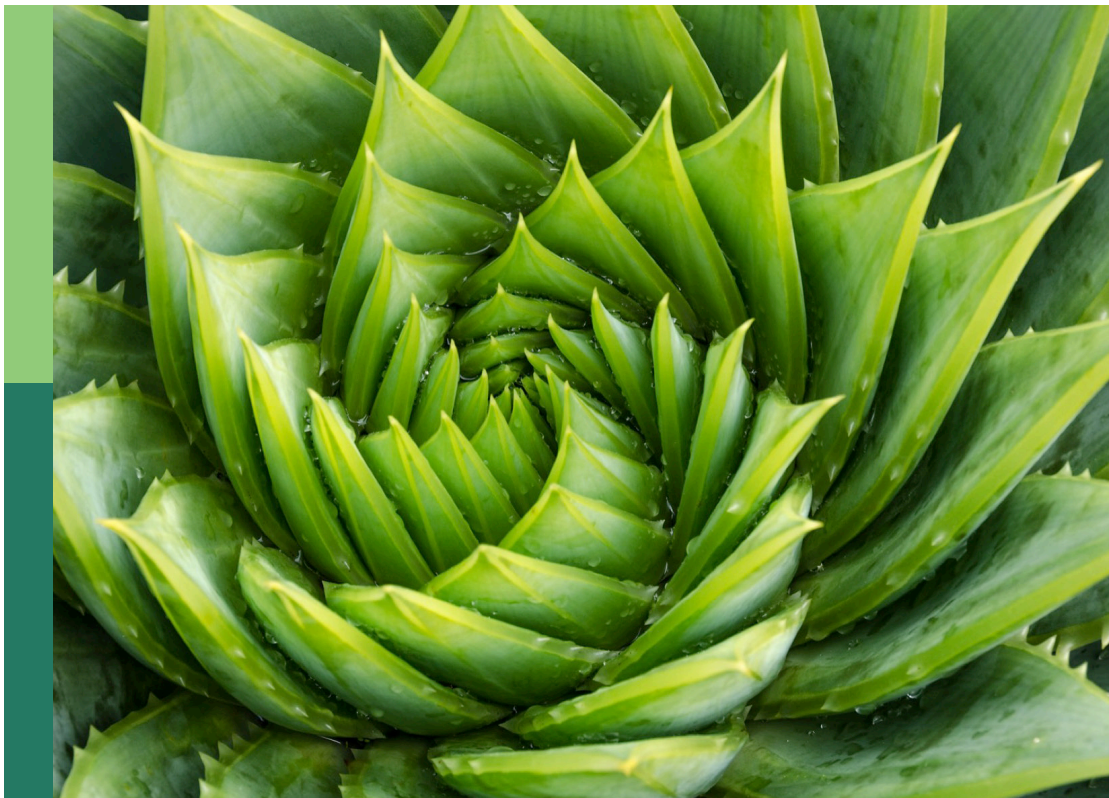
frontiers | Research Topics

Genomics in plant sciences: Understanding and development of stress-tolerant plants

Edited by
Yusuf Khan, Ayaz Dar, Hussain Touseef, Kashif Nawaz and Pranav Pankaj Sahu

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-2780-1
DOI 10.3389/978-2-8325-2780-1

About Frontiers

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

Frontiers journal series

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

Dedication to quality

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

What are Frontiers Research Topics?

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

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

Genomics in plant sciences: Understanding and development of stress-tolerant plants

Topic editors

Yusuf Khan — Oslo University Hospital, Norway

Ayaz Dar — University of Kashmir, India

Hussain Touseef — Matimate Agromart Pvt. Ltd. (Sevama AgriClinic Laboratory), India

Kashif Nawaz — Max Planck Institute for Plant Breeding Research, Germany

Pranav Pankaj Sahu — Global Change Research Institute, Czech Academy of Sciences, Czechia

Citation

Khan, Y., Dar, A., Touseef, H., Nawaz, K., Sahu, P. P., eds. (2023). *Genomics in plant sciences: Understanding and development of stress-tolerant plants*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2780-1

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 07 June 2023
DOI 10.3389/fpls.2023.1222818

OPEN ACCESS

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

REVIEWED BY
Rahul Singh Jasrotia,
Indian Council of Agricultural Research,
India

*CORRESPONDENCE
Pranav Pankaj Sahu
sahu.p@czechglobe.cz
Yusuf Khan
yusuf.zhc@gmail.com

RECEIVED 15 May 2023
ACCEPTED 26 May 2023
PUBLISHED 07 June 2023

CITATION
Dar AM, Touseef H, Nawaz K, Khan Y and Sahu PP (2023) Editorial: Genomics in plant sciences: understanding and development of stress-tolerant plants.
Front. Plant Sci. 14:1222818.
doi: 10.3389/fpls.2023.1222818

Editorial: Genomics in plant sciences: understanding and development of stress-tolerant plants

Ayaz Mahmood Dar[1], Hussain Touseef[2], Kashif Nawaz[3,4], Yusuf Khan[5]* and Pranav Pankaj Sahu[6]*

[1]Organic and Biomolecular Research, Department of Higher Education (GDC Sogam), University of Kashmir, Srinagar, Jammu and Kashmir, India, [2]Sevama AgriClinic and Laboratory, Matl Mate Agromart Pvt. Ltd., Bhavnagar, Gujarat, India, [3]Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, Köln, Germany, [4]Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, [5]Newborn Screening Department, University Hospital, Oslo, Norway, [6]Laboratory of Ecological Plant Physiology, Global Change Research Institute of the Czech Academy of Sciences, Brno, Czechia

KEYWORDS

genomics, crop genetics, breeding, genome-wide studies, multi-omics, abiotic, biotic, stress-tolerance

Editorial on the Research Topic
Genomics in plant sciences: understanding and development of stress-tolerant plants

Environmental stresses have a profound impact on the growth and productivity of plants (Boyer, 1982). The predicted growth of the human population and impending changes in climate could further threaten food availability around the World. Hence, it is of prime importance for researchers across the world to fast-track the development the high-yield stress-tolerant plants. The process requires a deeper understanding of genome structure, and biochemical, physiological, and molecular mechanisms underlying abiotic and biotic stress tolerance in plants. Plants have evolved complex molecular mechanisms to respond to both these stimuli (Bechtold and Field, 2018). It comprises the recognition of stress factors, activation of the stress signaling cascade, and synthesis of stress-related genes, proteins, or other molecules (Zhang et al., 2020). Moreover, genomics approaches like genetic diversity analysis, allelic variations identification, phylogenies, and understanding the adaptive evolution of genomes could also serve as vital tools to develop plant varieties mitigating food security issues (Cushman and Bohnert, 2000). Hence, an extensive effort is needed to utilize the plant genomics resources generated and to identify the key genes and regulators, which can withstand the diverse agroecological conditions/climates. This Research Topic represented the collection of nine original research articles and two reviews signifying the role of genomics, functional genomics, and multi-omics approaches in the development of stress-tolerant plants.

The identification of hotspots in the genomic architecture of plant species by chromosome-scale genome assembly, comparative genomics, and multi-omics approach could assist in mapping stress-associated genes and quantitative trait loci (QTLs). The

captured QTLs and SNPs markers in the genome not only provide information about the genomic architecture but could impart stress resilience in plants by marker-assisted selection and breeding. A study done by Sarkar et al. on maize successfully mapped the QTLs for the growth- and yield-related traits under water availability scenarios. The genes underlying these QTLs have specific functions in maintaining the growth and senescence, and hormone signaling under water deficit conditions. The nucleotide polymorphism in the promoter elements in rice has been shown to modulate the salt-responsive attributes by Haque et al. The comparative sequence analysis of K^+ salt-responsive transporters (*OsTPKa* and *OsHAK*) among two contrasting rice genotypes differed in salt-responsiveness revealing that both have unique TF-binding motifs. These modifications in the elements, played a key role in the coordinated expression of both transports, leading to an effective acclimation response in rice.

The candidate genes and proteins belonging to the flavonoid, phenylpropanoid, and α-linolenic acid metabolism were mapped on the tea (*Camellia sinensis*) genome by Wang et al. These potential targets in the tea genome could play an imperative role in the development of tea plants resistant to green leafhopper infestation. A similar approach was also applied to understand the genetic basis of cold adaptation in Asparagus beans by Liang et al. In the comparative genomics approach between contrasting bean accessions, a set of variable genes in two individual genomes was retrieved. A detailed analysis of these genes identified that in the genome of cold tolerant accession, Ningjiang 3 the presence of ABCC-type transports (a sub-gene family of ABC transporters) was linked with enhanced abiotic stress resilience.

A massive structural variation in the sesame pan-genome due to evolutionary adaptations was reported by Parakkunnel et al. These genetic novelties in APETALA2/Ethylene Responsive Factor (AP2/ERF) and WRKY transcription factors were due to alterations in the copy numbers, shifts in their location, and structural changes in the *cis*-element and exon-intron structure. These transcription factors (TFs) have a very specific role in the regulation of plants' defense against environmental stress, hence suggesting the adaptive selection pressure leading to structural variation in the sesame genome. Similarly, the comparative analysis of the chloroplast genome of black pepper (*Piper nigrum L.*) with related piper species was studied by Gaikwad et al. The study provided key insights into the chloroplast genome evolution, which will assist in the development of molecular markers and accelerate ongoing black pepper improvement programs.

Moreover, *in-silico* genome-wide analysis of key gene families and their comparative analysis could decipher the biotic and abiotic stress adaptations in plant species. A similar strategy was further exploited by Nie et al., to explore the role of leucine-rich repeat receptor-like kinases (LRR-RLKs) during *Bursaphelenchus xylophilus* infestation in *Pinus massoniana*. The candidate RLK genes retrieved from the transcriptome datasets were subjected to structural and molecular characterization. The study revealed that an LRR XII subfamily gene *PmRLKs32*, having ROS-producing properties played an important role in providing nematode resistance in pine trees. An extensive structural, functional, molecular, and evolutionary study on the Caffeoyl-CoA O-methyltransferase (CCoAOMT) gene family was carried out in two jute species by Akhter et al. Despite both species showing high syntenic conservation among *CCoAOMTs*, the gene expression level was modulated among them under various abiotic stresses and hormonal treatment. Hence, the key gene from the CCoAOMT gene family could be used to genetically engineered jute plants that can withstand a wide range of environmental stresses. Though, it is also important to functionally characterize the candidate genes for the required traits. For example, Zhang et al., showed how an MYC TF plays a major role in regulating growth and fruit quality in tomato. Through molecular study, they reported that the interplay between MYC2 and jasmonic acid (JA) and TOR signaling pathways regulates the growth and development of tomato seedlings.

This Research Topic contains two important review articles focused on genetic (structural changes) and epigenetic (reversible changes) modifications in the genome leading to stress adaptations in plants. The retrotransposons are varied in copy number and activated by various biotic and abiotic stresses due to retrotransposition bursts. The review by Papolu et al., focused on the dynamics of LTR retrotransposons and the associated mechanism of genome expression and adaptive evolution during high-temperature stress. The epigenetic changes are transgenerational and important for the stress memories in plants (Sahu et al., 2013). The review by Ramakrishnan et al., summarized the regulatory mechanism of memory establishment in plants under cold and heat stress, through histone modification and DNA methylation process.

Overall, the present research studies highlight the importance of genetic and genomic tools in the identification and development of stress-tolerant plants. The progress made towards understanding how structural changes in plant genomes could shape the unique response to environmental stresses will be critical for the improvement of crop plants.

Author contributions

All authors have contributed in writing, editing, reading and agreed to the final version of the manuscript.

Funding

PPS thanks the Czech Science Foundation, for providing the grant support (Grant no. 20-25845Y).

Conflict of interest

Author HT was employed by MatI Mate Agromart Pvt. Ltd.

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

Publisher's note

References

Bechtold, U., and Field, B. (2018). Molecular mechanisms controlling plant growth during abiotic stress. *J. Exp. Bot.* 69, 2753–2758. doi: 10.1093/jxb/ery157

Boyer, J. S. (1982). Plant productivity and environment. *Science* 218, 443–448. doi: 10.1126/science.218.4571.443

Cushman, J. C., and Bohnert, H. J. (2000). Genomic approaches to plant stress tolerance. *Curr. Opin. Plant Biol.* 3, 117–124. doi: 10.1016/S1369-5266(99)00052-7

Sahu, P. P., Pandey, G., Sharma, N., Puranik, S., Muthamilarasan, M., and Prasad, M. (2013). Epigenetic mechanisms of plant stress responses and adaptation. *Plant Cell Rep.* 32, 1151–1159. doi: 10.1007/s00299-013-1462-x

Zhang, H., Zhao, Y., and Zhu, J.-K. (2020). Thriving under stress: how plants balance growth and the stress response. *Dev. Cell* 55, 529–543. doi: 10.1016/j.devcel.2020.10.012

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 23 September 2022
DOI 10.3389/fpls.2022.1004387

OPEN ACCESS

EDITED BY
Yusuf Khan,
Inland Norway University of Applied Sciences, Norway

REVIEWED BY
Mingzhuo Li,
North Carolina State University, United States
Gagandeep Singh,
The University of Queensland, Australia

*CORRESPONDENCE
Fen Wang
fenmin521@163.com
Baoxing Song
baoxing.song@pku-iaas.edu.cn
Litang Lu
ltlv@gzu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 27 July 2022
ACCEPTED 25 August 2022
PUBLISHED 23 September 2022

CITATION
Wang F, Zhang B, Wen D, Liu R, Yao X, Chen Z, Mu R, Pei H, Liu M, Song B and Lu L (2022) Chromosome-scale genome assembly of *Camellia sinensis* combined with multi-omics provides insights into its responses to infestation with green leafhoppers.
Front. Plant Sci. 13:1004387.
doi: 10.3389/fpls.2022.1004387

Chromosome-scale genome assembly of *Camellia sinensis* combined with multi-omics provides insights into its responses to infestation with green leafhoppers

Fen Wang[1,2*†], Baohui Zhang[2,3†], Di Wen[1†], Rong Liu[1], Xinzhuan Yao[2,4], Zhi Chen[1], Ren Mu[1], Huimin Pei[1], Min Liu[5], Baoxing Song[1,6*] and Litang Lu[2,4*]

[1]The Department of Life Science and Agriculture, Qiannan Normal College for Nationalities, Duyun, China, [2]The Key Laboratory of Plant Resources Conservation and Germplasm Innovation in Mountainous Region (Ministry of Education), Guiyang, China, [3]Horticulture Institute (Guizhou Horticultural Engineering Technology Research Center), Guizhou Academy of Agricultural Sciences, Guiyang, China, [4]College of Tea Science, Guizhou University, Guiyang, China, [5]Biomarker Technologies Corporation, Beijing, China, [6]Peking University Institute of Advanced Agricultural Sciences, Weifang, China

The tea plant (*Camellia sinensis*) is an important economic crop, which is becoming increasingly popular worldwide, and is now planted in more than 50 countries. Tea green leafhopper is one of the major pests in tea plantations, which can significantly reduce the yield and quality of tea during the growth of plant. In this study, we report a genome assembly for DuyunMaojian tea plants using a combination of Oxford Nanopore Technology PromethION™ with high-throughput chromosome conformation capture technology and used multi-omics to study how the tea plant responds to infestation with tea green leafhoppers. The final genome was 3.08Gb. A total of 2.97Gb of the genome was mapped to 15 pseudo-chromosomes, and 2.79Gb of them could confirm the order and direction. The contig N50, scaffold N50 and GC content were 723.7kb, 207.72Mb and 38.54%, respectively. There were 2.67Gb (86.77%) repetitive sequences, 34,896 protein-coding genes, 104 miRNAs, 261 rRNA, 669 tRNA, and 6,502 pseudogenes. A comparative genomics analysis showed that DuyunMaojian was the most closely related to Shuchazao and Yunkang 10, followed by DASZ and tea-oil tree. The multi-omics results indicated that phenylpropanoid biosynthesis, α-linolenic acid metabolism, flavonoid biosynthesis and 50 differentially expressed genes, particularly peroxidase, played important roles in response to infestation with tea green leafhoppers (*Empoasca vitis* Göthe). This study on the tea tree is highly significant for its role in illustrating the evolution of its genome and discovering how the tea plant responds to infestation with tea green leafhoppers will contribute to a theoretical foundation to breed tea plants resistant to insects that will ultimately result in an increase in the yield and quality of tea.

KEYWORDS

DuyunMaojian, genome, chromatin interaction mapping, multi-omics, tea green leafhoppers, resistant to insects

Introduction

The tea tree, which has been grown in China for approximately 2,000 years (Kingdom-Ward, 1950), is a very important beverage crop with substantial economic value (Fang et al., 2022). The tea has many functions and effects, such as engendering liquid and allaying thirst, improving eyesight, preventing cancer and other healthy values (Ross and Kasum, 2002; Khan et al., 2008; Tounekti et al., 2013). China and India are the primary producers of tea (Bhattacharyya et al., 2022; Na Nagara et al., 2022; Nyhus Dhillon et al., 2022), and the tea industry in Sri Lanka is the second largest source of foreign exchange (Munasinghe et al., 2017). Tea is becoming increasingly popular worldwide and is now planted in more than 50 countries (Xia et al., 2020b; Yue et al., 2022) because of intriguing flavors (Han et al., 2016; Tan et al., 2019) and health benefits (Zhang et al., 2021b; Ahammed and Li, 2022). DuyunMaojian tea in Guizhou Province is a famous variety that won the Panama World's Fair Award in 1915 and the cultivation is thought to date back to 1368–1644 AD. The shape of DuyunMaojian tea is like a fishhook and is full of pekoe, which produces high-quality taste. Tea plantations in Guizhou Province occupy 470,000 ha, and ranked first in China by 2022. Duyun Maojian tea plants usually grow in a severe environment at 30° north latitude and an elevation of 800–1,200 m with little sunshine and much cloud and fog. It has a soft aroma, fresh and thick smell, and a sweet aftertaste. Thus, DuyunMaojian tea plant is a good material for breeding. However, tea green leafhoppers (*Empoasca vitis* Göthe) can significantly reduce the yield and quality of tea and cause losses in yield of 11–55% in the absence of effective means of prevention and treatment(Hazarika et al., 2009; Yang et al., 2013). The infestation of tea plants by these leafhoppers can cause physiological and biochemical defense responses, which primarily include the activation of salicylic acid, jasmonic acid, and other signaling pathways(Arimura et al., 2011; Zeng et al., 2019). Thus, we used the DuyunMaojian tea plant genome in combination with transcriptomic, proteomic and metabolomic studies to analyze the differentially expressed genes (DEGs), differentially expressed proteins (DEPs), and differential metabolites that play dominant effects in the response to tea green leafhoppers. We explored the important metabolic pathways in the DEGs, DEPs and differential metabolites that are involved in the molecular mechanism of tea plant responses to infestation with green leafhoppers, which provides a theoretical reference to develop insect-resistant cultivars.

Camellia sinensis genomes have been previously published from the cultivars Yukang 10 (Xia et al., 2017), Shuchazao (Wei et al., 2018; Chen et al., 2020; Xia et al., 2020a), Biyun (Zhang et al., 2020a), DASZ (Zhang et al., 2020b), Huangdan (Wang et al., 2021), Longjing 43 (Wang et al., 2020), and Tieguanyin (Zhang et al., 2021a). Two of the genome versions (Xia et al., 2017; Wei et al., 2018) are short-read–based assemblies that are highly fragmented, and the chromosome-scale genome assemblies of Shuchazao, DASZ, Longjing 43 were obtained using PacBio SMRT (Xia et al., 2020a; Wang et al., 2020) and high-throughput chromosome conformation capture (Hi-C) technology. Huangdan and Tieguanyin were assembled using PacBio HiFi (Wang et al., 2021; Zhang et al., 2021a) and Hi-C technology. In this study, we used Oxford Nanopore Technology (ONT) PromethION™ (Nicholls et al., 2019) combined with Hi-C technology (Feng et al., 2022; Lin et al., 2022) to assemble the genome of the elite tea cultivar DuyunMaojian. ONT can generate multi-kilobase-long RNA reads and has no bias for GC content and length. Thus, this study will present a theoretical reference for programs to breed tea tree.

Materials and methods

Sample collection, DNA preparation and library construction

An individual plant of DuyunMaojian tea was collected in December 2020 from Duyun City, Guizhou Province, China (N26°20′19″ E107°31′11″). The altitude is 786.1 m. Fresh and healthy leaves were frozen in liquid nitrogen after collection, and stored at −80°C before DNA extraction. DNA was extracted from the leaves using the CTAB method, which were examined by NanoDrop spectrophotometry (Thermo Fisher Scientific, Waltham, MA, United States), Qubit fluorimeter (Invitrogen, Carlsbad, CA, United States) and electrophoresis on a 0.35% agarose gel, and large segments were filtered using the BluePippin™ System (Sage Science, Inc., Beverly, MA, United States). We then prepared a library using the large segments of DNA, an ONT Template prep kit (SQK-LSK109; Oxford Nanopore Technologies, Oxford, United Kingdom) and an NEB Next FFPE DNA Repair Mix kit (New England Biolabs, Ipswich, MA, United States). The leaves for transcriptomic, proteomic, and metabolomics analyses were collected from plants same with the genome. Those trees were raised from cuttings planted in Duyun City, Guizhou Province, China (N26°20′19″ E107°31′11″) in November 2018. The leaves were collected from 12 individual trees with similar heights in June 2021. We made sure that 12 tea plants grown in the same environment and were not infected with pathogens. Six trees were infested by tea green leafhoppers for 24 h and another 6 trees were used as control. The infested and control group were placed in two gauze cages separately (Figure 1). There are three replicates in in each group. An Ilumina NovaSeq platform was used to sequence the transcriptome of tea tree infested or not by tea green leafhoppers. A DP441 QIAGEN RNAprep Pure Plant Plus Kit was used to extract all of the RNAs. NEBNext Ultra TM RNA Library Prep Kit for Illumina (New England Biolabs, United States) was used to generate sequecing libraries, and we purified the mRNA. We then fragmented it in NEBNext First Strand Synthesis Reaction Buffer. First strand and second strand cDNA were synthesized using random hexamer primers, M-MuLV Reverse Transcriptase, and DNA Polymerase I and RNase H, respectively. Finally, the library preparations and pair-end reads (PE150) were produced. Tandem

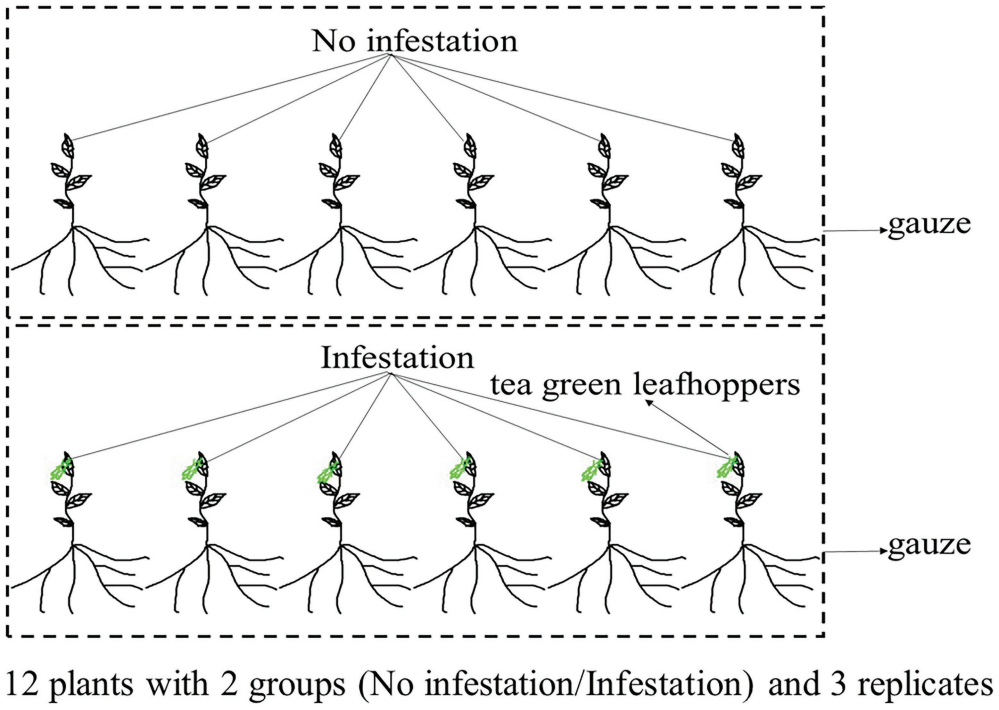


FIGURE 1
Experimental design. One group of the tea plants were not infested by tea green leafhoppers, and the other were infested by tea green leafhoppers.

mass tag (TMT) and liquid chromatography–tandem mass spectrometry (LC–MS/MS) were used to quantitatively analyze proteome. The tea tree leaves were frozen in liquid nitrogen and dissolved in lysis buffer. The quality of extracted proteins was detected by SDS-PAGE. We added 10 mM dithiothreitol to 100 μg of proteins, which was incubated for 1 h at 37°C and followed by alkylation with 40 mM iodoacetamide for 45 min in the dark at room temperature. Trypsin (150 enzyme to protein) was added to the samples, which were then incubated overnight at 37°C. The reaction was terminated by the addition of 0.1% formic acid (FA). Next, samples were desalted using a C_{18} chromatographic column, washed with 70% acetonitrile (ACN), and vacuum freeze-dried according to the manufacturer's instructions for the TMT kit. The LC–MS/MS spectra were then determined using XBridge Peptide BEH 5um C_{18}. The separated peptides were parsed using an Orbitrap Exploris™ 480 mass spectrometer. Ultra-performance liquid chromatography-quadrupole time-of-flight mass spectrometry (UPLC-QTOF-MS) was used to analyze the metabolome. The LC/MS system for metabolomics analysis was composed of a Waters Acquity I-Class PLUS ultra-high-performance liquid tandem Xevo G2-XS QT of high-resolution mass spectrometry. A Waters Xevo G2-XS QTOF high-resolution mass spectrometer can collect primary and secondary mass spectrometry data in the MSe mode under the control of the acquisition software (MassLynx V4.2, Waters, United States). Dual-channel data acquisition can be performed simultaneously on both low collision energy and high collision energy in each data acquisition cycle. The low collision energy is 2 V, the high collision energy range is 10 ~ 40 V, and the scanning frequency is 0.2 s to obtain a mass spectrum. The settings of the ESI ion source were as follows: capillary voltage: 2,000 V; cone voltage: 30 V; ion source temperature: 150°C; dissolvent gas temperature 500°C; backflush gas flow rate: 50 l/h; and desolventizing gas flow rate: 800 l/h.

Genome sequencing and assembly

Nanopore sequencing of 2 μg of gDNA was prepared using an NEB Next FFPE DNA Repair Mix kit (New England Biolabs) and subsequently processed using the ONT Template prep kit (Oxford Nanopore Technologies). The large segment library was premixed with loading beads and then pipetted onto a previously used and washed R9 flow cell. The library was sequenced on the ONT PromethION platform with the Corresponding R9 cell and ONT sequencing reagents kit (EXP-FLP001.PRO.6; Oxford Nanopore Technologies). The genome was assembled based on three ways: initial wtdbg2 (Ruan and Li, 2020) assembly and then SMARTdenovo assembly (Liu et al., 2021), followed by error correction using racon (Vaser et al., 2017) software and adjustment by Pilon (Walker et al., 2014) software. The assembly results were evaluated by the ratio of sequencing reads, Core Eukaryotic Genes Mapping Approach (CEGMA; Parra et al., 2007), and Benchmarking Universal Single-Copy Orthologs (BUSCO; Simão et al., 2015).

Hi-C sequencing

We constructed Hi-C fragment libraries that ranged from insert sizes of 300–700 bp as illustrated in Rao et al. (2014) and sequenced them using the Illumina Novaseq 6,000 System (San Diego, CA, United States). The low-quality reads were removed and the clean data truncated, and the trimmed reads were then aligned to the assembly genome using a Burrows-Wheeler Aligner (BWA; Li and Durbin, 2009). Only unique paired-end reads that could be aligned and had a mapping quality > 20 were conserved. The valid interaction pairs were employed to correct the scaffolds and then ordered by LACHESIS (Burton et al., 2013). Finally, the vast majority of the sequences were located on the chromosomes.

Repeat annotation

We integrated *de novo* and homology-based methods to recognize repetitive sequences in the genome of DuyunMaojian tea plants. We used LTR_FINDER (Xu and Wang, 2007) to look for homology. *De novo* predictions were conducted using RepeatScout (Price et al., 2005), and then PASTEClassifier (Hoede et al., 2014) and RepeatMasker were used to predict the repetitive sequences.

Gene prediction and functional annotation

We used *ab initio*, homology, and RNA-Seq methods to predict the protein-coding genes of the tea plant. We used Genscan (Burge and Karlin, 1997), Augustus (Stanke and Waack, 2003), GlimmerHMM (Majoros et al., 2004), GeneID, and SNAP (Korf, 2004) to perform *ab initio* predictions. We compared the protein-coding genes from *Arabidopsis thaliana*, black cottonwood (*Populus trichocarpa*), coffee (*Coffea canephora*), rice (*Oryza sativa*), and kiwifruit (*Actinidia chinensis*) to the DuyunMaojian tea plant genome. The homology methods were carried out using GeMoMa (Keilwagen et al., 2016, 2018) software. We used TransDecoder, GeneMarkS-T (Tang et al., 2015), and PASA (Campbell et al., 2006) to predict the transcriptome. We then integrated the three predictions described using EVM software. We annotated the tRNA genes using tRNAscan-SE (Lowe and Eddy, 1997). We recognized the microRNA and rRNA based on the Rfam (Griffiths-Jones et al., 2005) database. We predicted the pseudogenes using BLAST (Kent, 2002) and GeneWise. We functionally annotated the protein-coding genes of the tea plant genome by performing BLAST searches against the NR, gene ontology (GO), EuKaryotic Orthologous Groups (KOG), Kyoto Encyclopedia of Genes and Genomes (KEGG), and TrEMBL databases.

Comparative genomic analysis

We applied Orthofinder v2.5.1 software to recognize gene families of the tea plant and 14 other plant species, including Shuchazao, Yunkang10, DASZ, *Camellia oleifera Abel*, kiwifruit (*Actinidia chinensis*), coffee (*C. canephora*), *Arabidopsis thaliana*, cacao (*Theobroma cacao*), blueberry (*Vaccinium corymbosum*), *Rhododendron delavayi*, rice (*Oryza sativa*), apple (*Malus domestica*), grape (*Vitis vinifera*), *Citrus clementina*. We generated high-quality single-copy genes that were used to build the phylogenetic tree among the 15 species described above. The expansion or contraction events of gene families were computationally identified by CAFE v4.2 (De Bie et al., 2006). The synonymous substitution rates (*Ks*) of genes were calculated using wgdv1.1.132. We searched for LTR-RT sequences using LTR_FINDERv1.07 and LTRharvestv1.5.9 software. Collinearity analyses were conducted using VGSC35 software.

The DEGs, DEPs and differential metabolites of the transcriptomic, proteomic and metabolomics analyses

The level of expression of the transcriptome genes was calculated using fragments per kilobase of transcript per million mapped reads (FPKM). The DEGs were investigated through DESeq with fold change ≥ 2 and FDR < 0.01 defined as DEGs. TMT and liquid chromatography–tandem mass spectrometry (LC–MS/MS) were used to quantitatively analyze the proteins of tea leaves infested with tea green leafhoppers. Proteins with fold change ≥ 2 and FDR < 0.01 were assigned as DEPs. UPLC-QTOF-MS techniques were used to analyze the metabolites qualitatively and quantitatively. We used principal component analysis (PCA), orthogonal projections to latent structures-discrimination analysis (OPLS-DA), and fold change (FC) to filter out the differential metabolites. The protein–protein interaction network (PPIN) of the DEPs were constructed based on two method: (1) the Interolog (Matthews et al., 2001) method and BLAST were conducted between the DEPs and STRING databases (Szklarczyk et al., 2021); and (2) The existing DEP PPINs were extracted from STRING databases and TeaGPIN (Singh et al., 2021). We obtained the DEP PPINs by integrating the three DEPs PPINs.

Results

Chromosome-scale assembly and annotation of the DuyunMaojian tea tree genome

The genome of the tea plant designated DuyunMaojian was sequenced and assembled using a combination of ONT and Hi-C. ONT produced 289.67 Gb of clean data, and the sequencing

TABLE 1 Summary statistics of the DuyunMaojian tea plant assembly genome compared with those of other cultivars and species.

Genomic feature	DuyunMaojian using HIC	Shuchazao	Yunkang 10	DASZ	*Camellia oleifera*
Genome size (Gb)	2.97	2.94	3.02	3.11	2.89
Contig number	15,771	7,031	258,790	5,453	4,075
Contig N50 (kb)	723.697	600.46	19.96	2589.771	1,002
Contig max (kb)	6,456.183	2,885.68	257.648	16,831.527	/
Scaffold N50 (Mb)	207.721	218.1	0.45	204.21	185.364
Scaffold max (Mb)	250.703	/	3.51	336.927	/
GC content (%)	38.54	38.25	39.62	38.98	37.51
Gene number	34,896	50,525	36,951	33,021	42,426
Average gene length (bp)	6,961.34	4,906	3,549	8,050	3,955
Average exon length (bp)	232.25	245	237	211.7	317
Average intron length (bp)	1,117.79	973	640	1330.34	776
Reference	This study	Xia et al. (2020a)	Xia et al., (2017)	Zhang et al. (2020a,b)	Lin et al. (2022)

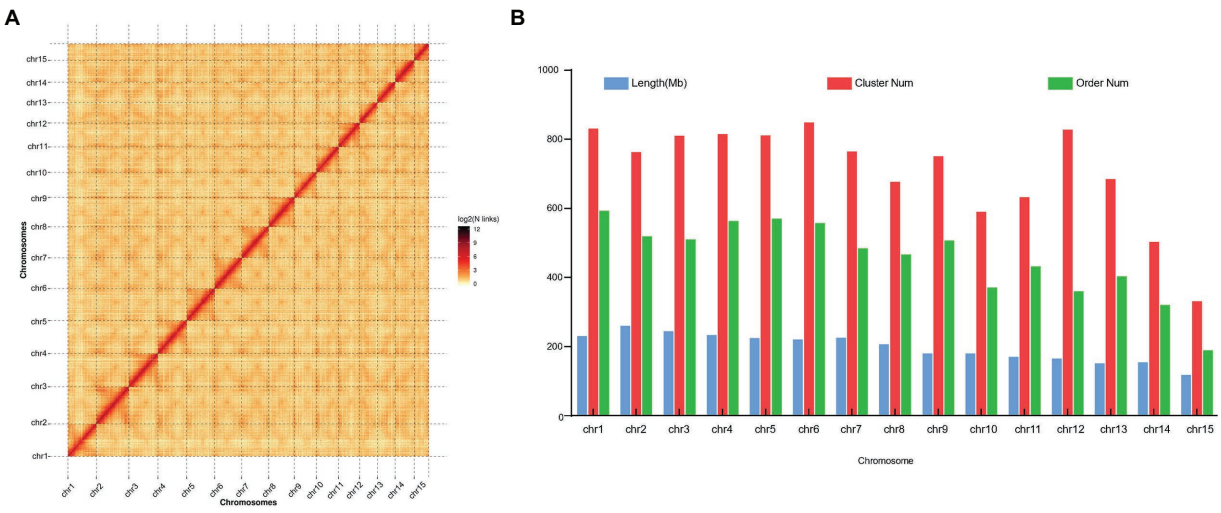


FIGURE 2
Hi-C heatmap based on the chromosome-scale assembly of the DuyunMaojian tea plant genome. **(A)** The heatmap represents the contact matrices generated by aligning the Hi-C data to the chromosome-scale assembly of the DuyunMaojian tea plant genome. **(B)** The assembly data statistics of each chromosome of the DuyunMaojian tea plant genome resulting from Hi-C.

depth was 85.56X (Supplementary Tables S1–S3). We got 3.35 Gb genome assembly and contig N50 was 778.61 kb. The GC content was 38.56% (Table 1). To validate the genome assembly quality, we first mapped all the high-throughput clean reads (916,216,962) to the assembled genome, the mapping rate was 91.66% and a proper mapping rate was 85.47% (Supplementary Table S4). Secondly, a CEGMAv2.5 result demonstrated that 90.61% of the core genes were found in the DuyunMaojian tea tree genome (Supplementary Table S5). Third, BUSCO v2.0 data demonstrated that 87.78% of the key genes were located in the assembly (Supplementary Table S6). A Hi-C analysis (Supplementary Tables S7–S10) was introduced to enhance the quality of tea plant genome and build a chromosome-scale assembly. After Hi-C assembly and manual adjustment, 2.97 Gb of the genome was mapped to 15 pseudo-chromosomes that anchored 96.35% of the assembled sequences and 2.79 Gb can confirm the order and direction (Figure 2). The redundant sequences from the heterozygous genome were removed by the Hi-C heat map, and the final genome was 3.08 Gb. After error correction, the contig N50, scaffold N50 and GC contents were 723.7 kb, 207.72 Mb and 38.54%, respectively (Table 1). The Hi-C heatmap perfectly showed 15 pseudo-chromosomes, which were designated chr1 to chr15, and the longest and shortest were chr2 and chr15 at 260,782,366 bp and 117,955,952 bp, respectively.

We identified 34,896 protein-coding genes and 2.67 Gb (86.77%) of repetitive sequences in the genome. Of the 34,896 genes, 33,481(95.95%) were annotated with the GO, KEGG, KOG, Pfam, TrEMBL and NR databases and we identified 16,674 SSRs. The prediction of noncoding RNA genes produced 104 miRNA, 261 rRNA, and 669 tRNA (Supplementary Table S11). In addition, 6,502 pseudogenes were annotated in our genome. Finally,

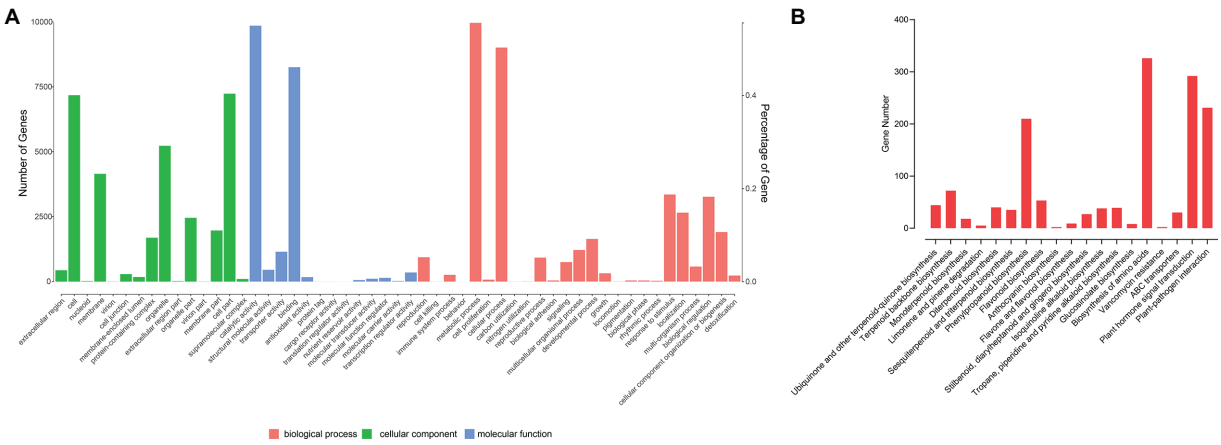


FIGURE 3
Gene annotation. **(A)** GO functional classification. **(B)** KEGG functional classification.

we obtained 2,997 motifs and 37,323 domains. Retrotransposons elements include LTR and non-LTR. The LTR-RT include *Copia* and *Gypsy*, while the non-LTR include LINE and SINE. *Copia* and *Gypsy* were the predominant LTR-RTs and accounted for 12.5 and 43.68% of the DuyunMaojian tea plant genome, respectively (Supplementary Table S12). LINE is an autonomously active transposable element in the tea plant genome and accounts for about 1.68% of the genome. There were 27,830 SINEs, which accounted for 0.17% of the DuyunMaojian tea plant genome. They are small retrotransposable elements, which do not encode any transposable genes that are active. *Copia*, *Gypsy*, LINE, and SINE contribute to the expansion of the DuyunMaojian tea plant genome. Thus, retrotransposon elements may perform important functions in the responses to external stimulus.

Functional annotation of tea tree genome

There were 17,919 genes annotated in the GO database (Supplementary Table S13). In cellular components, there were 15 secondary categories in which 7,177 genes, 7,236 genes and 5,229 genes were annotated to cell, cell part and organelle, respectively. In molecular function, 9,856, 8,257, and 1,149 genes were annotated to catalytic activity, binding, and transporter activity, respectively. In biological process, 9,963, 9,013, 3,354, 1,641, 318, 258, and 233 genes were annotated to metabolic process, cellular process, response to stimulus, developmental process, growth, immune system process, and detoxification, respectively (Figure 3A). A total of 11,407 genes were annotated to 128 pathways. A total of 44 genes were participated in ubiquinone and other terpenoid-quinone biosynthesis; 72 genes were took part in terpenoid backbone biosynthesis; 18 genes participated in monoterpenoid biosynthesis; 210 genes were took part in phenylpropanoid biosynthesis; 53 genes were involved with flavonoid biosynthesis; 70 genes were involved in α-linolenic acid metabolism; 292 genes participate in plant hormone signal transduction, and 231 genes participate in plant-pathogen interaction (Figure 3B). There were 17,700 genes annotated to the KOG database in all; 1,839 genes were classified to signal transduction mechanisms; 442 genes were classified into cell cycle, cell division, chromosome partitioning, and 213 genes were annotated to defense mechanism.

Comparative genomic analysis

Gene duplication within a species, evolution among species, and classification of species-specific genes (Figure 4A) were analyzed by comparing the genomes of DuyunMaojian tea plant with those of cacao (*T. cacao*), coffee (*C. canephora*), apple (*M. domestica*), *C. clementina*, kiwifruit (*A. chinensis*), blueberry (*V. corymbosum*), rice (*O. sativa*), *R. delavayi*, grape (*V. vinifera*), *A. thaliana*, Yunkang 10, Shuchazao, DASZ, and *C. oleifera*. The protein sequences of the DuyunMaojian tea plant and the other 14 species were classified using Orthofinderv2.5.1 (Emms, 2019). Thirty thousand and eight hundred and fifty three genes were grouped into 16,548 gene families (Table 2; Figure 4B). In addition, 215 gene families were specific to DuyunMaojian, and could be involved in specific biological processes, such as the large amount of synthesis of tea polyphenolics. IQ-TREEv1.6.11 (Nguyen and Schmidt, 2014), MAFFTv7.205 (Katoh et al., 2009), Gblocksv0.91b (Talavera and Castresana, 2007), and ModelFinder (Kalyaanamoorthy et al., 2017) software were used to study the evolutionary relationships between Duyun Maojian and the other 14 species using single-copy protein sequences. The results showed that DuyunMaojian had the closest relationship with already available tea tree genome of Shuchazao and Yunkang 10, followed by DASZ, *C. oleifera*, *V. corymbosum*, *R. delavayi*, *A. chinensis*, *C. canephora*, *V. vinifera*, *T. cacao*, *A. thaliana*, *C. clementina*, *M. domestica*, and *O. sativa* (Figure 4C). Shuchazao is evaluated to have diverged from Yunkang 10 about 15 (15–27)

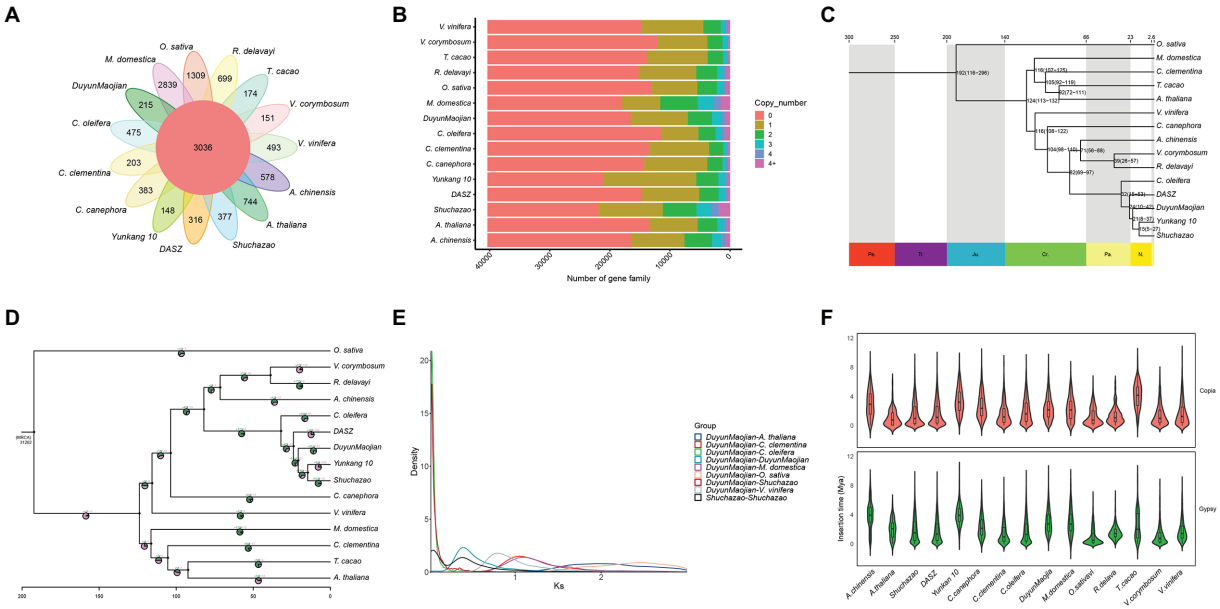


FIGURE 4
Comparative genomic analysis. **(A)**. A Venn diagram shows the shared and unique gene families among the DuyunMaojian and 14 other species. **(B)**. The copy number distribution of the DuyunMaojian tea plant and other 14 species. **(C)**. Phylogenetic tree of the DuyunMaojian tea plant and other 14 species. **(D)**. Expansion and contraction of gene families among the 15 plant species. **(E)**. Whole genome duplication events detected in the DuyunMaojian tea plant. **(F)**. Insertion times of copia and gypsy.

TABLE 2 Classification of gene family.

Name	Total gene	Cluster gene number	Total family	Unifamily
Vaccinium corymbosum	20,771	18,538	12,109	151
Rhododendron delavayi	32,410	26,819	15,208	699
Shuchazao	50,525	48,637	21,900	377
Theobroma cacao	21,254	20,614	13,712	174
DuyunMaojian	34,896	30,853	16,548	215
Oryza sativa	28,317	24,773	13,093	1,309
Arabidopsis thaliana	27,533	24,544	13,398	744
Malus domestica	55,414	43,649	17,974	2,839
Vitis vinifera	29,421	23,785	14,808	493
Coffea canephora	25,100	22,141	14,306	383
Citrus clementine	22,207	20,342	13,495	203
Actinidia chinensis	39,444	30,001	16,420	578
Yunkang 10	36,951	31,246	21,130	148
DASZ	28,981	24,339	14,630	316
Camellia oleifera	33,173	31,856	11,504	475

million years ago (MYA). DuyunMaojian is estimated to have diverged from Shuchazao and Yunkang 10 approximately 21 (8–37) MYA and split from DASZ approximately 24 (10–42) MYA. The cultivar DuyunMaojian is also evaluated to have diverged from *C. oleifera* approximately 32 (15–53) MYA and split from blueberry, *R. delavayi* and kiwifruit about 82 (69–97) MYA and from grape about 116 (108–122) MYA. We found that 201 gene families had expanded and 72 contracted in the DuyunMaojian tea plant genome using CAFÉv4.2 (De Bie et al., 2006; Figure 4D). The *Ks* of genes were calculated using wgdv1.1.1 (Zwaenepoel, 2019) and revealed that a recent whole-genome duplication (WGD) event (*Ks* =0.39) was occurred in DuyunMaojian tea plant genome (Figure 4E). We searched for LTR-RT sequences using LTR_FINDERv1.07 (Xu and Wang, 2007) and LTRharvestv1.5.9 (Ellinghaus et al., 2008) software and the insertion time of *Copia* and *Gypsy* among the 15 species (Figure 4F). Collinearity analyses were conducted using VGSC (Xu et al., 2016) software, and there were 42,543 collinear genes

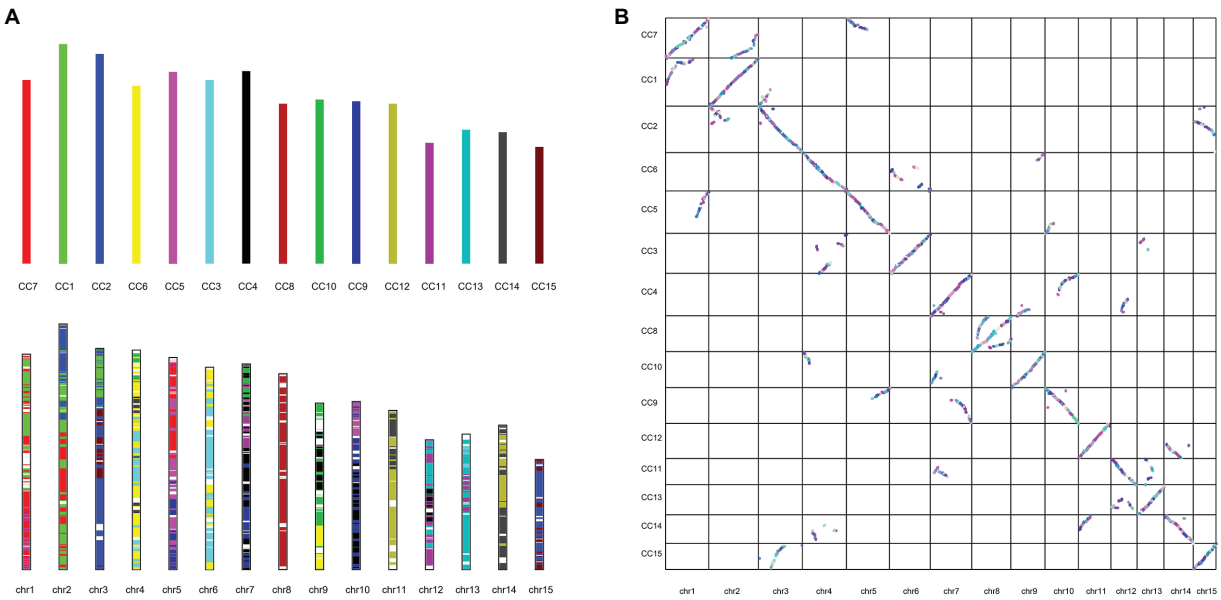


FIGURE 5
Collinearity analysis between DuyunMaojian and Shuchazao **(A)**. Evolutionary patterns of the chromosome-scale assembly of DuyunMaojian and Shuchazao. chr1-chr15 represent the 15 pseudo-chromosomes of DuyunMaojian. CC1-CC15 represent the 15 pseudo-chromosomes of Shuchazao. **(B)**. Comparison of chromosome-scale assembly of DuyunMaojian tea plant genome and Shuchazao.

between DuyunMaojian and Shuchazao (Figure 5). All these findings will greatly enhance our understanding of the diversification history of the DuyunMaojian tea plant genome.

Multi-omics analysis of the response of tea plants to infestation with tea green leafhoppers

Tea green leafhopper (Liu et al., 2017; Yang, 2017) can significantly reduce the yield and quality of tea during the growth of plant. These insects can cause a loss of 11–55% in tea yields in the absence of effective means of prevention and treatment. Thus, we also used the tea plant transcriptome, proteome and metabolome in combination with its genome to analyze the DEGs, DEPs and differential metabolites after leafhopper infestation. We paid more attention on tea plants responds to infestation with tea green leafhoppers in a short time rather than a long-term mechanism. At the same time, we are trying to screen for the key DEGs, DEPs, and differential metabolites between the two groups. Therefore, we only selected 24 h for comparison of two groups after infestation by tea green leafhoppers. There were 1,575 DEGs, 871 DEPs, and 41 differential metabolites, respectively. A total of 281 DEGs, 353 DEPs, and eight differential metabolites identified through these processes were annotated to the KEGG database, respectively (Supplementary Tables S14–S16). We found that phenylpropanoid biosynthesis pathway, flavonoid biosynthesis pathway, and α-linolenic acid metabolism were simultaneously identified in the transcriptome, proteome, metabolome and genome. Therefore, we hypothesized that proteins and metabolites involved in the three pathways play important roles in the resistance to tea green leafhopper. The results indicated that phenylpropanoid biosynthesis, α-linolenic acid metabolism, and flavonoid biosynthesis contained 50 DEPs that were filtered out and found in the tea plant transcriptome, proteome, metabolome and genome (Supplementary Table S17). Thus, we hypothesized that the three pathways and 50 DEPs play important roles in response to tea green leafhoppers, which will provide a theoretical foundation to breed tea plants resistant to insects and ultimately result in increases in tea yield and quality.

Protein–protein interaction network analysis of the response of tea plants to infestation with tea green leafhoppers

The predicted tea tree protein–protein interaction network (PPIN) was available in STRING database, TeaGPIN (Singh et al., 2021) and TeaLIPIN (Singh et al., 2019). There were 820 proteins and 1,067 protein–protein interactions (PPIs) in the predicted network based on the existing PPIN in the STRING database and the interolog method. We also extracted DEPs PPIN from TeaGPIN, in which there were 389 proteins and 583 PPIs (Supplementary Table S18). There were 362 shared PPIs in both network. We obtained the DEPs PPIN by integrating the two DEPs PPINs. In the merged DEPs PPIN (Supplementary Table S19), there were 1,047 nodes and 1,430 PPIs and a degree distribution of $P(k) \approx 53.574x^{-1.008}$, $R^2 = 0.684$. The topological properties (Figure 6) of the PPIN showed that the degree of distribution of the proteins obeyed a power-law distribution (Figure 6A). There

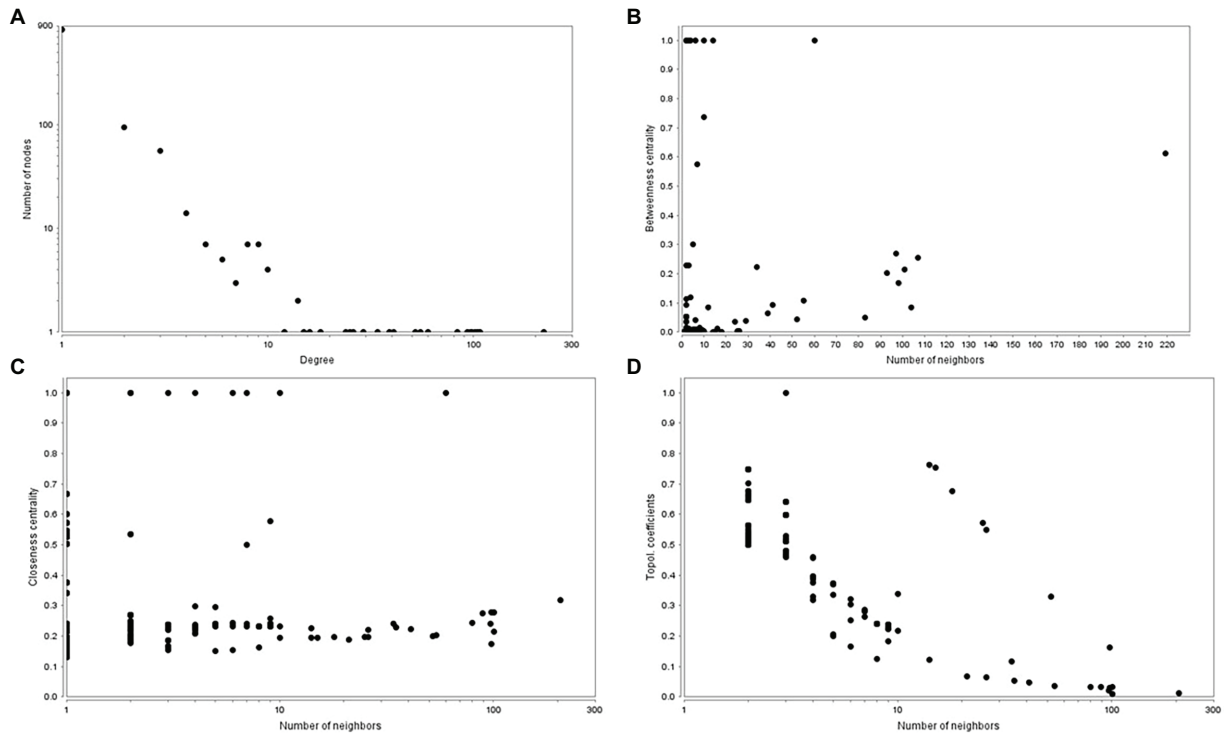


FIGURE 6
Topological properties of the PPINs. **(A)**. Degree of distribution. **(B)**. Betweenness centrality. **(C)**. Closeness centrality. **(D)**. Topological coefficient.

were a few proteins that have a high degree, which is a scale-free phenomenon. Highly connected proteins with central roles in the PPIN are hubs, and the degree of the hub proteins was calculated (Supplementary Table S20). These are the 21 proteins with the largest number of interactions, suggesting that they are also the most important proteins (hubs) in the PPIN. The betweenness centrality (Figure 6B) was a measure of a node's centrality in a network equal to the number of shortest paths from all vertices to all the others that pass through that node. Its value was between 0 and 1. $C_b(i) = \sum_{m \neq i \neq n}(\sigma_{mn}(i)/\sigma_{mn})$, where m and n were the nodes in the network that differed from i, and σ_{mn} denoted the number of the shortest paths from m to n. $\sigma_{mn}(i)$ was the number of shortest paths from m to n that i laid on. The closeness centrality (Figure 6C) of each node was also between 0 and 1 and was used to identify important positions within the network. $C_c(i) = 1/\text{avg.}(S[u,v])$, where $S[u,v]$ was the length of shortest path between the nodes u and v. $C_c(i)$ of node i is the reciprocal of the average shortest path length. The topological coefficient (Figure 6D) of protein i is $T_i = \text{avg.}(J(i,j))/n_i$, and $J(i,j)$ that indicate that all nodes j share at least one neighbor with i. The topological coefficient decreases with the increasing number of neighbors, revealing that the nodes with many neighbors are not artificially clustered together. Thus, the merged DEPs PPINs of the 50 DEPs is reliably based on the topological property analysis described above. We hypothesized that the 21 hubs with the largest degrees, particularly peroxidase, played important roles in the response to tea green leafhopper infestation.

Discussion

We reported the assembly of DuyunMaojian tea plant genome using a combination of the ONT PromethION™ combined with Hi-C technology, and the final genome was 3.08 Gb. A total of 2.97 Gb of the genome was mapped to 15 pseudo-chromosomes, among which 2.79 Gb can confirm the order and direction. The contig N50 was 723.7 kb, which was longer than that of Yunkang 10 (Xia et al., 2017) and shorter than that of Shuchazao (Xia et al., 2020a). The GC contents were 38.54%, which was consistent with those of the Yunkang 10 and Shuchazao genomes (Xia et al., 2020a). The ratio of *Copia* was higher than those in the CSSV1.1 genome, and the ratio of *Gypsy* was lower than those in the CSSV1.1 genome. In addition, we identified 16,674 SSRs, which was far less than those of the CSSV1.2 and CSA genomes.

A comparative genomics analysis showed that DuyunMaojian was the most closely related to Shuchazao and Yunkang 10, followed by DASZ, and *C. oleifera*. DuyunMaojian was evaluated to have diverged from *C. oleifera* approximately 32 (15–53) MYA, split from blueberry, *R. delavayi*, and kiwifruit about 82 (69–97) MYA and from grape about 116 (108–122) MYA. These estimates are similar to those of earlier studies (Shi et al., 2010; Fiz-Palacios et al., 2011; Wu, 2019). The *Ks* of genes were calculated using wgdv1.1.1 (Zwaenepoel, 2019) and revealed that a recent whole-genome duplication (WGD) event ($Ks = 0.39$) occurred in the DuyunMaojian tea plant genome, which was consistent with the findings of previous studies (Xia et al., 2017, 2020a; Wang et al.,

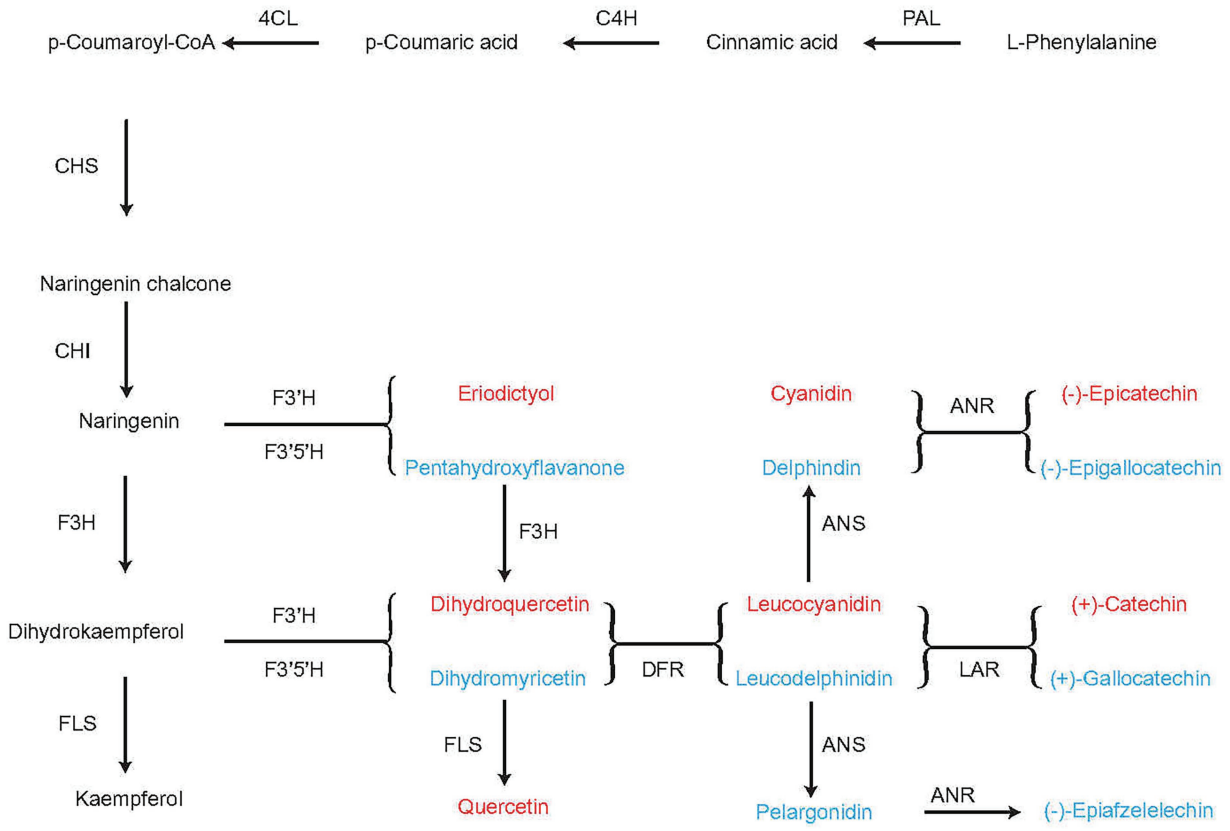


FIGURE 7
Flavonoid biosynthesis pathway in the tea plant. PAL, phenylalanine ammonia lyase; C4H: cinnamate 4-hydroxylase; 4CL: 4-coumarate-CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; DFR, dihydroflavonol 4-reductase; ANR, anthocyanidin reductase; F3′5′H, flavonoid-3′5′-hydroxylase; F3′H, flavonoid 3′-hydroxylase; FLS, flavonol synthase; LAR, leucoanthocyanidin reductase.

2020). These data provide theoretical references for the cultivation of excellent tea varieties.

There were 33,481 annotated genes in the assembly genome, which were lower than those in the CSSV1.2 genome (Xia et al., 2020a) and higher than those in the CSSV1.1 genome (Wei et al., 2018). Tea plant leaves contain an extraordinarily high level of flavonoids that contribute to its health benefits and significantly affect its flavor, taste, and mouthfeel. Flavonoids, particularly flavonol glycosides and catechins, are the major source of the bitterness and astringency of tea. There were 53 genes annotated to flavonoid biosynthesis, 210 genes annotated to phenylpropanoid biosynthesis, and 70 genes annotated to α-linolenic acid metabolism in the tea plant genome, which contained 50 DEGs filtered out that were identified in the transcriptome, proteome and metabolome. Thus, we hypothesized that the three pathways (Zhao et al., 2020) and 50 DEGs played important roles in the response to tea green leafhopper infestation.

When the tea plants were infested with tea green leafhoppers, they were influenced by DEPs, such as peroxidase. The putative flavonoid biosynthesis *via* the phenylpropanoid pathway are converted from cinnamoyl-CoA to pinobanksin 3-acetate by the enzymes chalcone synthase (CHS), chalcone isomerase (CHI), and flavanone-3-hydroxylase (F3H). L-Phenylalanine can be converted to *p*-coumaroyl-CoA through the subsequent action of phenylalanine ammonia lyase (PAL), cinnamate-4-hydroxylase (C4H), and 4-coumarate-CoA-ligase (4CL). *p*-Coumaroyl-CoA can be converted to naringenin through the action of CHS and CHI. Naringenin can be converted to either (−)-epicatechin through the subsequent actions of flavonoid 3′-hydroxylase (F3′H), flavonoid 3′, 5′-hydroxylase (F3′5′H), and anthocyanidin reductase (ANR) or to (−)-epigallocatechin through the subsequent action of 3′-hydroxylase (F3′H), flavonoid 3′, 5′-hydroxylase (F3′5′H), and anthocyanidin reductase (ANR). F3H catalyzes naringenin to dihydrokaempferol, which can be converted to either kaempferol through the action of flavonol synthase (FLS), or to dihydroquercetin, leucocyanidin, and finally (+)-catechin through the subsequent actions of F3′H, F3′5′H, dihydroflavonol 4-reductase (DFR), and leucoanthocyanidin reductase (LAR) or to (+)-gallocatechin through the subsequent actions of F3′H, F3′5′H, DFR and LAR. Naringenin can be converted either eriodictyol or pentahydroxyflavanone by the enzymes F3′H and F3′5′H, respectively. Eriodictyol and pentahydroxyflavanone can subsequently be converted to dihydroquercetin and dihydromyricetin by F3H, respectively, and finally converted to quercetin by FLS. DFR catalyzes the transformation of dihydromyricetin to leucodelphinidin, which can be converted to (−)-epiafzelechin by ANS and ANR (Figure 7). Thus, in the flavonoid biosynthesis pathway, we obtained

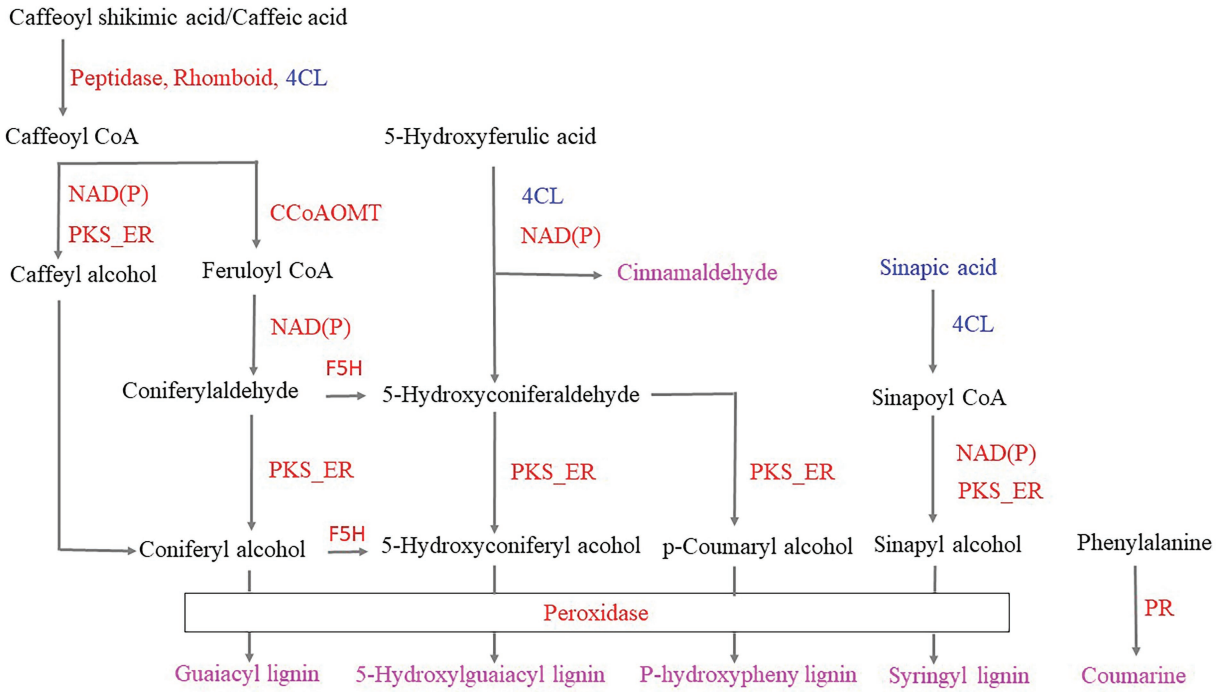


FIGURE 8
Responses of the phenylpropanoid biosynthetic pathway to infestation with tea green leafhoppers. Note: red proteins represent upregulation; blue proteins represent downregulation; pink metabolites were obtained in response to tea green leafhopper. PR: protein kinase; 4CL: 4-coumarate-CoA ligase; NAD(P): NAD(P)-bd_dom domain-containing protein; PKS_ER: PKS_ER domain-containing protein; Peptidase: peptidase_M16 domain-containing protein; Rhomboid: rhomboid domain-containing protein; CCoAOMT: caffeoyl-CoA o-methyltransferase.

(−)-epicatechin, (−)-epigallocatechin, (+)-catechin, (+)-gallocatechin, (−)-epiafzelelechin, kaempferol, and quercetin, which play important roles in improving the quality of tea and enabling it to respond to infestation by tea green leafhoppers. This result was consistent with those of previous studies (Treutter, 2006; Wang et al., 2016; Mei et al., 2017; Zhao et al., 2020).

Infestation of the tea plants with tea green leafhoppers resulted in the following changes in the phenylpropanoid biosynthesis pathway: (1) Phenylalanine can be converted to coumarin by upregulating protein kinase and other enzymes. (2) 5-Hydroxy ferulic acid can be converted to cinnamaldehyde based on the downregulation of 4CL and the upregulation of NAD(P). (3) 5-Hydroxyferulic acid can be converted to *p*-coumaryl alcohol by the subsequent action of the downregulated 4CL, upregulated NAD(P), and upregulated PKS_ER. Peroxidase then catalyzes coumaryl alcohol to *p*-hydroxyphenyl lignin. (4) Caffeoylshikimic acid/caffeic acid can be converted to caffeoyl CoA through the action of upregulated peptidase, Rhomboid/downregulated 4CL. Caffeoyl CoA can be converted to either caffeoyl alcohol and coniferyl alcohol through the subsequent action of upregulated NAD(P) and PKS_ER or to feruloyl-CoA through the upregulated CCoAOMT and converted to coniferyl aldehyde through the action of upregulated NAD(P). Upregulated PKS_ER then catalyzes coniferyl aldehyde to coniferyl alcohol. Finally, coniferyl alcohol was converted to guaiacyl lignin through the action of upregulated peroxidase. (5) Coniferyl aldehyde and coniferyl alcohol can be converted to 5-hydroxyconiferaldehyde and 5-hydroxyconiferyl alcohol, respectively, through the action of upregulated F5H. Moreover, on the basis of downregulated 4CL, upregulated NAD(P) and PKS_ER, 5-hydroxyconiferaldehyde and 5-hydroxyconiferyl alcohol were obtained, which were catalyzed by the upregulated peroxidase, resulting in 5-hydroxyguaiacyl lignin. (6) Sinapic acid was downregulated and could be converted to sinapylCoA through the action of downregulated 4CL, and sinapylCoA can be converted to sinapyl alcohol by the subsequent catalysis of upregulated NAD(P) and PKS_ER. Finally, syringyl lignin was obtained by the catalysis of upregulated peroxidase. Thus, we hypothesized that when the tea plant was infested, peroxidase, coumarin, cinnamaldehyde, *p*-hydroxyphenyl lignin, guaiacyl lignin, 5-hydroxyguaiacyl lignin, syringyl lignin, and the phenylpropanoid biosynthetic pathway were used to respond to the tea green leafhopper (Figure 8), which was in accordance with the previous studies (Zhou et al., 2019; Natukunda, 2021; Souza et al., 2021; Zhang et al., 2022). These findings suggest that the resistance of tea plants to tea leafhopper could be obtained by inducing the synthesis of lignin, which could have potential applications in preventing insect infestations on tea.

The α-linolenic acid metabolic pathway played important roles when the tea plants were infested with tea green leafhoppers. Under the catalysis of upregulated germin-like protein, phosphatidylcholine was converted to α-linolenic acid and then converted to stearidonic acid, which was used to respond to tea green leafhoppers. Moreover, α-linolenic acid can be either converted to 9(S)-hydroper

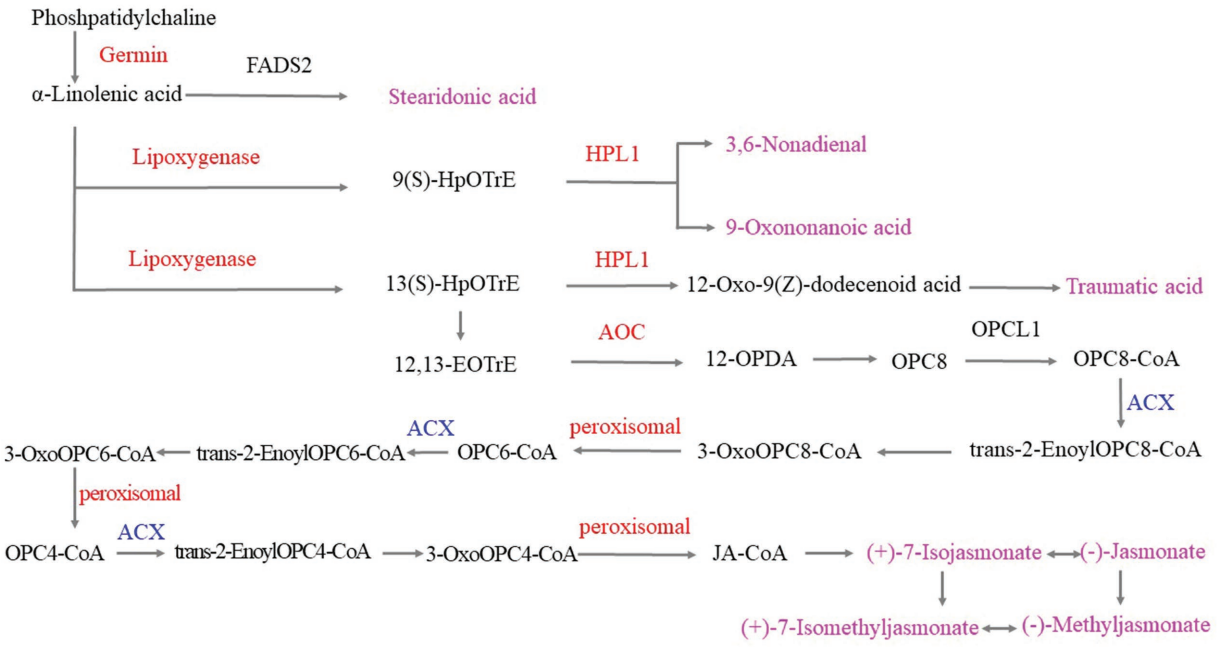


FIGURE 9
α-Linolenic acid metabolism in response to tea green leafhopper infestation in tea plants. Note: the red genes represent upregulation; the blue genes represent downregulation, and the pink metabolites were obtained in response to tea green leafhopper infestation.

oxyoctadecatrienoic acid (HpOTrE) by the upregulation of lipoxygenase or to 3,6-nonadienal and 9-oxononanoid acid by the upregulation of hydroperoxide lyase 1 (HPL1). Next, α-linolenic acid was catalyzed by lipoxygenase, HPL1 and other enzymes to produce traumatic acid. Moreover, responses to the infestation of tea green leafhoppers included the catalysis of α-linolenic acid by upregulated lipoxygenase to obtain 13(S)HpOTrE, which was catalyzed by the related enzyme and then catalyzed the production of 12,13-EOTrE. Upregulated allene-oxide cyclase (AOC) catalyzed the transformation of 12,13-EOTrE to 12-OPDA, and 12-OPDA was then converted to (+)-7-isojasmonate, (−)-jasmonate, (+)-7-isomethyljasmonate, and (−)-methyl jasmonate by the downregulated ACX and upregulated peroxisomal, which can interconvert these compounds. Thus, we hypothesized that stearidonic acid, 9(S)-HOTrE, 3,6-nonadienal, 9-oxononanoid acid, 10-OPDA, traumatic acid, (+)-7-isojasmonate, (−)-jasmonate, (+)-7-isomethyl jasmonate, and (−)-methyl jasmonate played important roles in the response to tea green leafhopper infestation, which was consistent with the findings of previous studies (Nafie and Hathout, 2011; Xin et al., 2016; Zhao et al., 2020; Figure 9).

In the PPIN, we hypothesized that the hubs DEPs (Supplementary Table S20) and peroxidase in particular, may play important roles in the response to infestation by tea green leafhoppers.

In brief, the chromosome-scale genome of tea plant is highly promising to help understand the evolution of tea genome and then discover how the tea plant responds to tea green leafhopper. Our results suggest that (1) Duyun Maojian had the closest relationship with Shuchazao and Yunkang 10, followed by DASZ, and *C. oleifera*. (2) A recent whole-genome duplication (WGD) event (*Ks* = 0.39) occurred in the DuyunMaojian tea plant genome (3) The tea plant could become resistant to tea leafhoppers by the biosynthesis of lignin, which could have potential application values to prevent insect infestations on tea plants. (4) (−)-Epicatechin, (−)-epigallocatechin, (+)-catechin, (+)-gallocatechin, (−)-epiafzelechin, kaempferol, and quercetin played important roles in the response to tea green leafhopper infestation. (5) Stearidonic acid, 9(S)-HOTrE, 3,6-nonadienal, 9-oxononanoid acid, 10-OPDA, traumatic acid, (+)-7-isojasmonate, (−)-jasmonate, (+)-7-isomethyljasmonate, and (−)-methyl jasmonate played important roles in response to tea green leafhopper infestation. (6) The hubs DEPs and peroxidase, in particular, may play important roles in the response to the tea green leafhopper infestation.

Data availability statement

The datasets presented in this study can be found in the NCBI Sequence Read Archive under the BioProject accession number PRJNA841059. The DuyunMaojian tea plant gene and functional annotations are available in the figshare database (DOI: doi.org/10.6084/m9.figshare.19972808.v1).

Author contributions

LL designed the study. DW and RL contributed to the sample preparation and genome sequencing. FW, BZ, and ML conducted the genome assembly, performed genome annotation and interpreted comparative genomic analysis. FW, BZ, XY, HP, RM, and ZC performed multi-omics analysis. FW wrote the

manuscript. LL, BS, and FW modified the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (grant number 32260080 and 31900486); Guizhou Provincial Science and Technology Foundation [(2019)1298]; Guizhou Provincial Education Department [ZDXK(2016)23, (2016)020, QNYSKYPT2018007, (2020)071, and (2015)68)]; the Scientific Research Project of Qiannan Normal University for Nationalities (QNYSKYTD2018011, Qnsyk201605, QNSY2018BS018, 2018xjg0520, QNYSXXK 2018005, QNSY2018ZJ006, and 2019xjg0303).

Conflict of interest

ML was employed by Biomarker Technologies Corporation.

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

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1004387/full#supplementary-material

References

Ahammed, G. J., and Li, X. (2022). Hormonal regulation of health-promoting compounds in tea (Camellia sinensis L.). *Plant Physiol. Biochem.* 185, 390–400. doi: 10.1016/j.plaphy.2022.06.021

Arimura, G.-I., Ozawa, R., and Maffei, M. E. (2011). Recent advances in plant early signaling in response to Herbivory. *Int. J. Mol. Sci.* 12, 3723–3739. doi: 10.3390/ijms12063723

Bhattacharyya, C., Imchen, M., Mukherjee, T., Haldar, S., Mondal, S., Mukherji, S., et al. (2022). Rhizosphere impacts bacterial community structure in the tea (*Camellia sinensis* (L.) *O. Kuntze.*) estates of Darjeeling. *India. Environ. Microbiol.* 24, 2716–2731. doi: 10.1111/1462-2920.15874

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268, 78–94. doi: 10.1006/jmbi.1997.0951

Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R., Kitzman, J. O., and Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* 31, 1119–1125. doi: 10.1038/nbt.2727

Campbell, M. A., Haas, B. J., Hamilton, J. P., Mount, S. M., and Buell, C. R. (2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC Genomics* 7, 1–17. doi: 10.1186/1471-2164-7-327

Chen, J.-D., Zheng, C., Ma, J.-Q., Jiang, C.-K., Ercisli, S., Yao, M.-Z., et al. (2020). The chromosome-scale genome reveals the evolution and diversification after the recent tetraploidization event in tea plant. *Hortic. Res.* 7:63. doi: 10.1038/s41438-020-0288-2

De Bie, T., Cristianini, N., Demuth, J. P., and Hahn, M. W. (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22, 1269–1271. doi: 10.1093/bioinformatics/btl097

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9, 1–14. doi: 10.1186/1471-2105-9-18

Emms, D. M. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238. doi: 10.1186/s13059-019-1832-y

Fang, Z., Jin, L., Ye, Y., He, W., Shu, Z., Shao, J., et al. (2022). Effects of different shading treatments on the biomass and Transcriptome profiles of tea leaves (*Camellia sinensis* L.) and the regulatory effect on phytohormone biosynthesis. Pdf. *Front. Plant Sci.* 13, 1–11. doi: 10.3389/fpls.2022.909765

Feng, F., Yao, Y., Wang, X. Q. D., Zhang, X., and Liu, J. (2022). Connecting high-resolution 3D chromatin organization with epigenomics. *Nat. Commun.* 13:2054. doi: 10.1038/s41467-022-29695-6

Fiz-Palacios, O., Schneider, H., Heinrichs, J., and Savolainen, V. (2011). Diversification of land plants: insights from a family-level phylogenetic analysis. *BMC Evol. Biol.* 11:341. doi: 10.1186/1471-2148-11-341

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R., and Bateman, A. (2005). Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* 33, D121–D124. doi: 10.1093/nar/gki081

Han, Z.-X., Rana, M. M., Liu, G.-F., Gao, M.-J., Li, D.-X., Wu, F.-G., et al. (2016). Green tea flavour determinants and their changes over manufacturing processes. *Food Chem.* 212, 739–748. doi: 10.1016/j.foodchem.2016.06.049

Hazarika, L. K., Bhuyan, M., and Hazarika, B. N. (2009). Insect pests of tea and their management. *Annu. Rev. Entomol.* 54, 267–284. doi: 10.1146/annurev.ento.53.103106.093359

Hoede, C., Arnoux, S., Moisset, M., Chaumier, T., Inizan, O., Jamilloux, V., et al. (2014). PASTEC: an automatic transposable element classification tool. *PLoS One* 9:e91929. doi: 10.1371/journal.pone.0091929

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Katoh, K., Asimenos, G., and Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* 537, 39–64. doi: 10.1007/978-1-59745-251-9_3

Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O., and Grau, J. (2018). Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinf.* 19, 1–12.

Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., and Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* 44:e89. doi: 10.1093/nar/gkw092

Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* 12, 656–664.

Khan, N., Adhami, V. M., and Mukhtar, H. (2008). Apoptosis by dietary agents for prevention and treatment of cancer. *Biochem. Pharmacol.* 76, 1333–1339. doi: 10.1016/j.bcp.2008.07.015

Kingdom-Ward, F. (1950). Does wild tea exist? *Nature* 165, 297–299. doi: 10.1038/165297a0

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* 5:59. doi: 10.1186/1471-2105-5-59

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Lin, P., Wang, K., Wang, Y., Hu, Z., Yan, C., Huang, H., et al. (2022). The genome of oil-camellia and population genomics analysis provide insights into seed oil domestication. *Genome Biol.* 23:14. doi: 10.1186/s13059-021-02599-2

Liu, J.-H., Sun, C.-Y., Long, J., and Guo, J.-J. (2017). Complete mitogenome of tea green leafhopper, Empoasca onukii (Hemiptera: Cicadellidae) from Anshun,

Guizhou Province in China. *Taylor Francis* 2, 808–809. doi: 10.1080/23802359.2017.1398616

Liu, H., Wu, S., Li, A., and Ruan, J. (2021). SMARTdenovo: a *de novo* assembler using long noisy reads. *Gigabyte* 2021, 1–9. doi: 10.46471/gigabyte.15

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.955

Majoros, W. H., Pertea, M., and Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20, 2878–2879. doi: 10.1093/bioinformatics/bth315

Matthews, L. R., Vaglio, P., Reboul, J., Ge, H., Davis, B. P., Garrels, J., et al. (2001). Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "Interologs." *Genome Res.* 11, 2120–2126. doi: 10.1101/gr.205301

Mei, X., Liu, X., Zhou, Y., Wang, X., Zeng, L., Fu, X., et al. (2017). Formation and emission of linalool in tea (*Camellia sinensis*) leaves infested by tea green leafhopper (*Empoasca* (Matsumurasca) *onukii* Matsuda). *Food Chem.* 237, 356–363. doi: 10.1016/j.foodchem.2017.05.124

Munasinghe, M., Deraniyagala, Y., Dassanayake, N., and Karunarathna, H. (2017). Economic, social and environmental impacts and overall sustainability of the tea sector in Sri Lanka. *Sustain. Prod. Consum.* 12, 155–169. doi: 10.1016/j.spc.2017.07.003

Na Nagara, V., Sarkar, D., Luo, Q., Biswas, J. K., and Datta, R. (2022). Health risk assessment of exposure to trace elements from drinking black and green tea marketed in three countries. *Biol. Trace Elem. Res.* 200, 2970–2982. doi: 10.1007/s12011-021-02863-3

Nafie, E., and Hathout, T. (2011). Jasmonic acid elicits oxidative defense and detoxification systems in Cucumis melo L cells. *Braz. J. Plant Physiol.* 23, 161–174.

Natukunda, M. I. (2021). Interaction between rag genes results in a unique synergistic transcriptional response that enhances soybean resistance to soybean aphids. *BMC Genomics* 22. doi: 10.1186/s12864-021-08147-3

Nguyen, L.-T., and Schmidt, H. A. (2014). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Nicholls, S. M., Quick, J. C., Tang, S., and Loman, N. J. (2019). Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *GigaScience* 8, 1–9. doi: 10.1093/gigascience/giz043

Nyhus Dhillon, C., Vossenaar, M., Weiligmann, B., Sanwal, N., Djimeu, E. W., Kneepkens, M., et al. (2022). A nutrition behavior change program moderately improves minimum diet diversity and Handwashing behaviors among tea workers in Assam and Tamil Nadu, India. *Food Nutr. Bull.* 43, 159–170. doi: 10.1177/03795721211070706

Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23, 1061–1067. doi: 10.1093/bioinformatics/btm071

Price, A. L., Jones, N. C., and Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. *Bioinformatics* 21, i351–i358. doi: 10.1093/bioinformatics/bti1018

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D map of the human genome at Kilobase resolution reveals principles of chromatin looping. *Cells* 159, 1665–1680. doi: 10.1016/j.cell.2014.11.021

Ross, J. A., and Kasum, C. M. (2002). Dietary flavonoids bioavailability, metabolic effects, and safety. *Annu. Rev. Nutr.* 22, 19–34. doi: 10.1146/annurev.nutr.22.111401.144957

Ruan, J., and Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* 17, 155–158. doi: 10.1038/s41592-019-0669-3

Shi, T., Huang, H., and Barker, M. S. (2010). Ancient genome duplications during the evolution of kiwifruit (Actinidia) and related Ericales. *Ann. Bot.* 106, 497–504. doi: 10.1093/aob/mcq129

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Singh, G., Singh, V., and Singh, V. (2019). Construction and analysis of an interologous protein–protein interaction network of *Camellia sinensis* leaf (TeaLIPIN) from RNA–Seq data sets. *Plant Cell Rep.* 38, 1249–1262. doi: 10.1007/s00299-019-02440-y

Singh, G., Singh, V., and Singh, V. (2021). Genome-wide interologous interactome map (TeaGPIN) of *Camellia sinensis*. *Genomics* 113, 553–564. doi: 10.1016/j.ygeno.2020.09.048

Souza, C. S. F., Souza, B. H. S., Parrella, R. A. C., Simeone, M. L. F., Nascimento, P. T., França, J. C. O., et al. (2021). Resistance of bmr energy sorghum hybrids to sugarcane borer and fall armyworm. *Braz. J. Biol.* 84, 1–9. doi: 10.1590/1519-6984.251883

Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19, ii215–ii225. doi: 10.1093/bioinformatics/btg1080

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49, D605–D612. doi: 10.1093/nar/gkaa1074

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564–577. doi: 10.1080/10635150701472164

Tan, H. R., Lau, H., Liu, S. Q., Tan, L. P., Sakumoto, S., Lassabliere, B., et al. (2019). Characterisation of key odourants in Japanese green tea using gas chromatography-olfactometry and gas chromatography-mass spectrometry. *LWT* 108, 221–232. doi: 10.1016/j.lwt.2019.03.054

Tang, S., Lomsadze, A., and Borodovsky, M. (2015). Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res.* 43:e78. doi: 10.1093/nar/gkv227

Tounekti, T., Joubert, E., Hernández, I., and Munné-Bosch, S. (2013). Improving the polyphenol content of tea. *Crit. Rev. Plant Sci.* 32, 192–215. doi: 10.1080/07352689.2012.747384

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

Vaser, R., Sovic, I., and Nagarajan, N. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746. doi: 10.1101/gr.214270.116

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:15. doi: 10.1371/journal.pone.0112963

Wang, X., Feng, H., Chang, Y., Ma, C., Wang, L., Hao, X., et al. (2020). Population sequencing enhances understanding of tea plant evolution. *Nat. Commun.* 11:10. doi: 10.1038/s41467-020-18228-8

Wang, Y.-N., Tang, L., Hou, Y., Wang, P., Yang, H., and Wei, C.-L. (2016). Differential transcriptome analysis of leaves of tea plant (Camellia sinensis) provides comprehensive insights into the defense responses to Ectropis oblique attack using RNA-Seq. *Funct. Integr. Genomics* 16, 383–398. doi: 10.1007/s10142-016-0491-2

Wang, P., Yu, J., Jin, S., Chen, S., Yue, C., Wang, W., et al. (2021). Genetic basis of high aroma and stress tolerance in the oolong tea cultivar genome. *Hortic. Res.* 8:107. doi: 10.1038/s41438-021-00542-x

Wei, C., Yang, H., Wang, S., Zhao, J., Liu, C., Gao, L., et al. (2018). Draft genome sequence of *Camellia sinensis* var. *sinensis* provides insights into the evolution of the tea genome and tea quality. *Proc. Natl. Acad. Sci.* 115, E4151–E4158. doi: 10.1073/pnas.1719622115

Wu, H. (2019). A high-quality *Actinidia chinensis* (kiwifruit) genome. *Hortic. Res.* 6, 1–9. doi: 10.1038/s41438-019-0202-y

Xia, E., Tong, W., Hou, Y., An, Y., Chen, L., Wu, Q., et al. (2020a). The reference genome of tea plant and Resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation. *Mol. Plant* 13, 1013–1026. doi: 10.1016/j.molp.2020.04.010

Xia, E.-H., Tong, W., Wu, Q., Wei, S., Zhao, J., Zhang, Z.-Z., et al. (2020b). Tea plant genomics: achievements, challenges and perspectives. *Hortic. Res.* 7:7. doi: 10.1038/s41438-019-0225-4

Xia, E.-H., Zhang, H.-B., Sheng, J., Li, K., Zhang, Q.-J., Kim, C., et al. (2017). The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. *Mol. Plant* 10, 866–877. doi: 10.1016/j.molp.2017.04.002

Xin, Z., Li, X., Li, J., Chen, Z., and Sun, X. (2016). Application of chemical elicitor (Z)-3-hexenol enhances direct and indirect plant defenses against tea geometrid *Ectropis obliqua*. *BioControl* 61, 1–12. doi: 10.1007/s10526-015-9692-1

Xu, Y., Bi, C., Wu, G., Wei, S., Dai, X., Yin, T., et al. (2016). VGSC: a web-based vector graph toolkit of genome Synteny and collinearity. *Biomed. Res. Int.* 2016:7823429. doi: 10.1155/2016/7823429

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35, W265–W268. doi: 10.1093/nar/gkm286

Yang, T. (2017). Molecular identification of spiders preying on *Empoasca vitis* in a tea plantation. *Sci. Rep.* 7, 1–10. doi: 10.1038/s41598-017-07668-w

Yang, Z., Baldermann, S., and Watanabe, N. (2013). Recent studies of the volatile compounds in tea. *Food Res. Int.* 53, 585–599. doi: 10.1016/j.foodres.2013.02.011

Yue, C., Chen, Q., Hu, J., Li, C., Luo, L., and Zeng, L. (2022). Genome-wide identification and characterization of GARP transcription factor gene family members reveal their diverse functions in tea plant (*Camellia sinensis*).Pdf. *Front. Plant Sci.* 13, 1–21. doi: 10.3389/fpls.2022.947072

Zeng, L., Watanabe, N., and Yang, Z. (2019). Understanding the biosyntheses and stress response mechanisms of aroma compounds in tea (*Camellia sinensis*) to safely and effectively improve tea aroma. *Crit. Rev. Food Sci. Nutr.* 59, 2321–2334. doi: 10.1080/10408398.2018.1506907

Zhang, X., Chen, S., Shi, L., Gong, D., Zhang, S., Zhao, Q., et al. (2021a). Haplotype-resolved genome assembly provides insights into evolutionary history of the tea plant *Camellia sinensis*. *Nat. Genet.* 53, 1250–1259. doi: 10.1038/s41588-021-00895-y

Zhang, Q., Li, T., Gao, M., Ye, M., Lin, M., Wu, D., et al. (2022). Transcriptome and Metabolome profiling reveal the resistance mechanisms of Rice against Brown Planthopper. *Int. J. Mol. Sci.* 23:4083. doi: 10.3390/ijms23084083

Zhang, Q.-J., Li, W., Li, K., Nan, H., Shi, C., Zhang, Y., et al. (2020a). The chromosome-level reference genome of tea tree unveils recent bursts of non-autonomous LTR Retrotransposons in driving genome size evolution. *Mol. Plant* 13, 935–938. doi: 10.1016/j.molp.2020.04.009

Zhang, Z., Zhang, X., Bi, K., He, Y., Yan, W., Yang, C. S., et al. (2021b). Potential protective mechanisms of green tea polyphenol EGCG against COVID-19. *Trends Food Sci. Technol.* 114, 11–24. doi: 10.1016/j.tifs.2021.05.023

Zhang, W., Zhang, Y., Qiu, H., Guo, Y., Wan, H., Zhang, X., et al. (2020b). Genome assembly of wild tea tree DASZ reveals pedigree and selection history of tea varieties. *Nat. Commun.* 11:3719. doi: 10.1038/s41467-020-17498-6

Zhao, X., Chen, S., Wang, S., Shan, W., Wang, X., Lin, Y., et al. (2020). Defensive responses of tea plants (*Camellia sinensis*) against tea green leafhopper attack: a multi-Omics study. *Front. Plant Sci.* 10:1705. doi: 10.3389/fpls.2019.01705

Zhou, Y., Liu, X., and Yang, Z. (2019). Characterization of Terpene synthase from tea green leafhopper being involved in formation of Geraniol in tea (*Camellia sinensis*) leaves and potential effect of Geraniol on insect-derived Endobacteria. *Biomol. Ther.* 9:808. doi: 10.3390/biom9120808

Zwaenepoel, A. (2019). Wgd—simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* 35, 2153–2155. doi: 10.1093/bioinformatics/bty915

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 27 October 2022
DOI 10.3389/fpls.2022.1013445

OPEN ACCESS

EDITED BY
Yusuf Khan,
Oslo University Hospital, Norway

REVIEWED BY
Arif Rashid,
Anhui Agricultural University, China
Yan Liang,
Northwest A&F University, China

*CORRESPONDENCE
Yang Yu
yuyang@syau.edu.cn
Na Cui
cuina@syau.edu.cn

†These authors share first authorship

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 07 August 2022
ACCEPTED 06 October 2022
PUBLISHED 27 October 2022

CITATION
Zhang Y, Xing H, Wang H, Yu L, Yang Z, Meng X, Hu P, Fan H, Yu Y and Cui N (2022) SlMYC2 interacted with the *SlTOR* promoter and mediated JA signaling to regulate growth and fruit quality in tomato.
Front. Plant Sci. 13:1013445.
doi: 10.3389/fpls.2022.1013445

SlMYC2 interacted with the *SlTOR* promoter and mediated JA signaling to regulate growth and fruit quality in tomato

Yujiao Zhang[1†], Hongyun Xing[1†], Haoran Wang[2], Lan Yu[1], Zhi Yang[1], Xiangnan Meng[1], Pengpeng Hu[3], Haiyan Fan[1,4], Yang Yu[1*] and Na Cui[1,4*]

[1]College of Bioscience and Biotechnology, Shenyang Agricultural University, Shenyang, China, [2]School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou, China, [3]Department of Foreign Language Teaching, Shenyang Agricultural University, Shenyang, China, [4]Key Laboratory of Protected Horticulture of Ministry of Education, Shenyang Agricultural University, Shenyang, China

Tomato (*Solanum lycopersicum*) is a major vegetable crop cultivated worldwide. The regulation of tomato growth and fruit quality has long been a popular research topic. MYC2 is a key regulator of the interaction between jasmonic acid (JA) signaling and other signaling pathways, and MYC2 can integrate the interaction between JA signaling and other hormone signals to regulate plant growth and development. TOR signaling is also an essential regulator of plant growth and development. However, it is unclear whether MYC2 can integrate JA signaling and TOR signaling during growth and development in tomato. Here, MeJA treatment and *SlMYC2* overexpression inhibited the growth and development of tomato seedlings and photosynthesis, but increased the sugar–acid ratio and the contents of lycopene, carotenoid, soluble sugar, total phenol and flavonoids, indicating that JA signaling inhibited the growth of tomato seedlings and altered fruit quality. When TOR signaling was inhibited by RAP, the JA content increased, and the growth and photosynthesis of tomato seedlings decreased, indicating that TOR signaling positively regulated the growth and development of tomato seedlings. Further yeast one-hybrid assays showed that SlMYC2 could bind directly to the *SlTOR* promoter. Based on GUS staining analysis, SlMYC2 regulated the transcription of *SlTOR*, indicating that SlMYC2 mediated the interaction between JA and TOR signaling by acting on the promoter of *SlTOR*. This study provides a new strategy and some theoretical basis for tomato breeding.

KEYWORDS

***Solanum lycopersicum*, JA signaling, *SlMYC2*, TOR signaling, growth and fruit quality**

Introduction

Tomato (*Solanum lycopersicum*) is a major vegetable crops, and the regulation of its growth and development, as well as yield and quality, have been popular research topics. JA is not only an important plant growth regulator but also part of an important hormone signaling pathway, playing a key role in plant growth and development (Huang et al., 2017). The exogenous application of JA inhibits various aspects of seedling growth, including primary root growth, leaf expansion, and hypocotyl elongation (Wasternack and Hause, 2013; Song et al., 2014; Kim et al., 2015). Moreover, JA treatment inhibits the expansion of true leaves and cotyledons (Zhang and Turner, 2008; Chehab et al., 2012; Aleman et al., 2016).

Important components of the JA signaling pathway include the protein COI1, complex SCF^{COI1}, protein JAZ, and transcription factor MYC2 (Chini et al., 2016; Li et al., 2021a; Song et al., 2022). MYC2 is the activating component of JA signaling and belongs to the transcription factor MYC basic helix loop helix (bHLH) IIIe subfamily. The MYC IIIe family in *Arabidopsis* consists of AtMYC2, AtMYC3, AtMYC4 and AtMYC5. However, in tomato, the MYC IIIe family consists of SlMYC1 and SlMYC2 (Heim et al., 2003; Chini et al., 2016; Goossens et al., 2016), and MYC2 is the main regulator of JA signaling activation (Yan et al., 2013; Xu et al., 2018; Liu et al., 2019; Min et al., 2020).

MYC2 is widely present in animals and plants and has a variety of regulatory functions (Dombrecht et al., 2007; Altmann et al., 2020). In *Arabidopsis thaliana*, MYC2 directly binds to the promoters of *PLT1* and *PLT2*, inhibiting their expression and thereby inhibiting the growth of the main roots (Chen et al., 2011). In rice, OsMYC2 regulates the expression of *OsMADS1*, a gene that regulates the development of floral organoids and activates the development of rice spikelets (Cai et al., 2014). In apple, MdMYC2 promotes the transcription of the ethylene synthesis genes *MdACS1* and *MdACO1*, which in turn promotes ethylene synthesis in fruit and promotes fruit ripening (Li et al., 2017). In addition, MYC2 is also involved in complex metabolism. MYC2 plays an important role in regulating of artemisinin biosynthesis (Shen et al., 2016), and MYC2 regulates SGA biosynthesis in tomato and potato (Cárdenas et al., 2016; Thagun et al., 2016). The MYC2 protein contains a JID domain and an acidic AD domain at its N-terminus, and it contains a bHLH-zip domain and an ACT domain at its C-terminus (Thines et al., 2007). Elevated levels of JA-Ile promote the degradation of JAZ by the SCF^{COI1} protease complex, followed by the transcriptional expression of *MYC2*, which initiates the expression of JA-responsive genes (Zhang et al., 2015; Breeze, 2019). This finding indicates that MYC2 plays an important role in JA-mediated plant metabolism and is a high-level transcriptional regulatory element in the JA signaling pathway.

MYC2 positively regulates the inhibition of hypocotyl elongation by red or far-red light and negatively regulates the inhibition of hypocotyl elongation by blue light (Biswas et al., 2003; Riemann et al., 2008; Yan et al., 2012; Yang et al., 2012; Riemann et al., 2013), suggesting that MYC2 has different functions under different conditions. Moreover, MYC2 can integrate the interaction between JA signaling and other hormone signaling to regulate plant growth and development (Li et al., 2021b). One such example is for the ABA receptor PYL6, which interacts with the JA master regulator MYC2 to regulate the transcriptional activity of *MYC2* (Aleman et al., 2016).

JA interacts with various signaling pathways to mediate plant growth inhibition. JAs promote the degradation of JAZ to activate MYC2 and release DELLA protein (Hou et al., 2010; Yang et al., 2012). EIN3 is activated by the JA-mediated breakdown of JAZ proteins and ethylene (ET)-mediated stabilization. It binds to and represses MYC2. EIN3 also transcriptionally activates *ORA59*, and ORA59 represses *MYC2* transcription. MYC2 can enhance its own transcription in the short term but repress it in the long term. EDS1 can repress MYC2 during AvrRps4-induced ETI. In addition, SA can promote JAZs degradation during ETI *via* NPR3 and NPR4. Generally, SA is an inhibitor of *MYC2* transcription. Abscisic acid (ABA) directly activates the transcription of *MYC2* and enhances the binding of the ABA receptor PYL6 to MYC2, which regulates *MYC2* transcriptional activity, and MYC2 has different effects on the *JAZ6* and *JAZ8* promoters (Aerts et al., 2021). The above findings suggest that MYC2 mediates the interaction of JA signaling with other signaling pathways (Aerts et al., 2021).

In both tomato and *Arabidopsis*, the active hormone JA-Ile promotes COI1-dependent degradation of the JAZ repressors and thereby activates the master TF MYC2. However, studies have found that the target genes downstream of MYC2 are different in tomato and *Arabidopsis*. In tomato, MYC2 positively and directly regulates the transcription of its downstream *MTFs*, which in turn regulates the expression of late wounding-responsive genes or pathogen-responsive genes. In contrast, in *Arabidopsis*, MYC2 positively regulates wounding-responsive genes while negatively regulating pathogen-responsive genes (Du et al., 2017).

Target of rapamycin (TOR) is also an important signaling pathway in regulating plant growth and development and plays a central role in integrating metabolic energy and hormone signaling (Dobrenel et al., 2016). TOR is a highly conserved serine/threonine protein kinase in eukaryotes (Boutouja et al., 2019). Recent studies have found that the plant TOR complex can coordinate energy, growth, hormones and other signals (Zhuo et al., 2020; Fu et al., 2021), enabling it to regulate plant growth and development, as well as nutrition and energy processes, through the integration of its downstream effector proteins E2Fa and SPS (Xiong et al., 2017; De Vleesschauwer et al., 2018; Brunkard et al., 2020; Fu et al., 2020; O'Leary et al., 2020; Wang et al., 2020). Studies have shown that TOR signaling

can regulate the growth and development of cotton fiber through the interaction of JAZ with JA signaling (Song et al., 2017). However, MYC2 is the core factor of signaling interactions, and it is unclear whether MYC2 can integrate JA signaling and TOR signaling pathways in response to the growth and development of tomato. Our study found that the presence of a cis-acting element of SlMYC2 on the *SlTOR* promoter, so it was speculated that SlMYC2 bound to the *SlTOR* promoter to regulate downstream response genes and thereby regulate the growth and development of tomato.

In this research, we analyzed the roles of JA signaling, SlMYC2 and TOR signaling in tomato growth and development. We demonstrated that JA signaling inhibited the growth and development of tomato plants and altered fruit quality, and TOR signaling positively regulated the growth and development of tomato seedlings. In addition, yeast one-hybrid and GUS assays indicated that SlMYC2 could directly bind to the *SlTOR* promoter and positively regulate its transcription. These results suggested that MYC2 could integrate JA signaling and TOR signaling in respond to growth and development of tomato, providing a new strategy and some theoretical basis for tomato breeding.

Materials and methods

Determination of seed germination and tomato growth

Tomato seeds were heated in water at 55°C for 5 min, then soaked in 70% alcohol for 1 min in an ultra-clean bench, then soaked in 5% sodium hypochlorite for 5 min, washed with ddH_2O, and then transferred into MS medium with sterile ophthalmic forceps. The substrate containing the seeds was placed in a light incubator and incubated for 3 d at 25/18°C protected from light; and then incubated under a light intensity of 20000 Lux, photoperiod of 16/8 h, and 25/18°C while germination. The germination rate, root length and hypocotyl length were observed and recorded.

Using 0.1% DMSO as a control, the seedlings were treated with RAP (10 μM), MHY1485 (5 μM), DIECA (200 μM) and MeJA (100 μM), respectively.

DNA construction and plant transformation

The *SlMYC2*-OE and *SlMYC2*-RNAi lines were kindly provided by Dr. Chuanyou Li's research group (Institute of Genetics and Developmental Biology Chinese Academy of Sciences). Transgenic lines of *SlMYC2*-OE and *SlMYC2*-RNAi were constructed using Gateway (Invitrogen) technology. A fragment of the *SlMYC2* open reading frame (1–400 bp) was selected. The sequence was then cloned and inserted into pCAMBIA-1301 under the control of the *CaMV35S* promoter to generate the construct pCAMBIA-1301-*SlMYC2*-RNAi. For *SlMYC2*-OE tomato plants, the full-length coding sequence of *SlMYC2* was amplified by PCR and cloned into the pGWB5 vector to generate the Pro35S:*SlMYC2*-GFP construct. The constructs were introduced into tomato cv M82 by *Agrobacterium tumefaciens*-mediated transformation. Transformants were selected based on their resistance to hygromycin. The *A. tumefaciens*-transformed T0 generation seeds were collected, and the T1 generation was screened by hygromycin and conformed to the ratio of 3:1, indicating that there were 1/3 homozygotes, and qRT–PCR was identified as homozygous and single copy. Then, the T1 generation was planted. After identification, 10 homozygous lines were selected, and the T2 generation of seeds from a single plant was collected. After hygromycin screening, no separation occurred, showing they were all homozygotes. The T3 seeds from the T2 generation were also collected, and the T4 seeds identified as homozygotes were collected through the same identification method as above. Homozygous seeds from T3 or T4 generation were all homozygotes after identification and screening. The homozygotes of T3 or T4 transgenic seedlings were used for phenotypic and molecular characterization (Du et al., 2017).

Construction and transformation of tomato *SlTOR* gene silencing by virus-induced gene silencing (VIGS) method. A pTRV-based VIGS was performed to knock out the *SlTOR* in tomato cotyledons. A 402 bp fragment within the 3'-region of the *SlTOR* cDNA was cloned into the pTRV2 vector (TRV : *SlTOR*). The gene-specific primers were listed in Table S1. 10 mL of each *A. tumefaciens* strain used was incubated overnight at 28° C in YEP medium supplemented with 100 $mg{\cdot}L^{-1}$ rifampicin and 50 $mg{\cdot}L^{-1}$ kanamycin. Then, 200 μL of each overnight culture was inoculated into a 20 mL portion of YEP medium containing the above antibiotics and incubated at 28°C until the culture reached the selected optical density of OD_{600} = 0.8-1.0. Induced *A. tumefaciens* strain EHA105 carrying different pTRV2-derived vectors (pTRV2 and pTRV2-*SlTOR*) was mixed with the pTRV1 *A. tumefaciens* strain EHA105 at a ratio of 1:1. The samples were spiked with 10 mM MES, 10 mM $MgCl_2$, and 200 μM acetosyringone (AS) and then the germinated tomato seeds (root length = 0.2-0.6 cm) were immersed in the bacteria and vacuumed for 4 min. The seeds were sown and cultured in a growth chamber at 22°C with a 16 h light and 8 h dark photoperiod (Yu et al., 2019).

Determination of tomato fruit quality

Fruit hardness was measured by a TMS-Pilot Precision texture analyzer (Food Technology Corporation, Virginia, USA), and the fruit color difference was measured by a color

difference meter from CHROMA METER (KONICA MINOLTA SENSING INC. Japan).

Lycopene was extracted with reference to Salvia-Trujillo and McClements (Salvia-Trujillo and McClements, 2016), total phenolic compounds were extracted with reference to Toor and Savage (Toor and Savage, 2005), and total flavonoids according to Jia et al. (Jia et al., 1999), while carotenoids were extracted by the acetone method and then all were determined spectrophotometrically. The contents of pectin, vitamin C, soluble protein, soluble sugar and acidity were also determined by spectrophotometrically.

Quantification of JA

High performance liquid chromatography (HPLC) was used to analyze the total JA content in the leaves of tomato plants.

Website and tools for promoter analysis

The sequences of the *SlTOR* promoter were obtained from the NCBI website (http://www.ncbi.nlm.nih.gov/). The promoter sequences of *SlTOR* were analyzed by the promoter prediction website PlantCARE.

RNA extraction and qRT–PCR

Total RNA from plants was extracted using TRIzol reagent (Gold Hi Plasmid Mini kit, CWO581M). BioDrop was used to measure the RNA concentration, and cDNA was synthesized by the FastKing cDNA First-Strand Synthesis Kit (catalog number: KR116) from Tiangen Biochemical Technology Co., Ltd. qRT–PCR was performed using a SYBR Green PCR Master Mix kit (TIANGEN, Beijing, China) on the CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Primer sequences were listed in Table S1. The reaction system (10 μL) contained 4.5 μL of 2× SuperReal PreMix Plus, 1 μL of cDNA, 0.5 μL of forward primer, 0.5 μL of reverse primer, and finally ddH_2O was added to 10 μL. The reaction program included 95°C for 15 min and 40 cycles of 95°C for 10 s and 60°C for 32 s. Dissolution curve program included 65°C for 5 s and 95°C for 5 min.

GUS staining and enzyme activity assay

The CDS of *SlMYC2* and the 2000 bp promoter region of *SlTOR* were cloned using 2×Super Pfx MasterMix from Comway Century Biotechnology Co., Ltd. The *SlMYC2* was then cloned into the pRI-101-GFP vector driven by the 35S promoter to obtain the *SlMYC2* overexpression vector, and then the empty vector pRI-101-GFP and the recombinant vector *SlMYC2*-GFP were transformed into *A. tumefaciens* EHA105, respectively. The *SlTOR* promoter was cloned into the pRI-101-GUS vector to obtain a reporter gene driving GUS expression, and this vector and the recombinant vector were transformed into *A. tumefaciens*. The bacterial broth was inoculated into YEP liquid medium and incubated at 28°C with shaking at 200 rpm until OD_{600} = 0.8-1.0. Took 20 mL of different bacterial solutions into a centrifuge tube and centrifuged at 5000 rpm for 10 min, and discarded the supernatant. Resuspended and washed the bacteria with 2 mL of immersion solution (containing 10 mM MES and 10 mM $MgCl_2{\cdot}6H_2O$) and centrifuged at 5000 rpm for 10 min, and discarded the supernatant. Resuspended cells in the soaking solution (containing 10 mM MES, 10 mM $MgCl_2{\cdot}6H_2O$, and 200 mM AS) to make different combinations of cells with $OD_{600} \approx 1$ and incubated at 28°C in the dark for 3 h. The reporter and effector were then mixed together in a 1:1 volume ratio to transform 5-week-old *Nicotiana benthamiana*. The empty pRI-101-GUS and pRI-101-GFP vectors served as controls. Primer sequences were listed in Table S1. The different combinations of bacterial solutions were injected into tobacco leaves using a needleless syringe and incubated overnight in the dark, then removed and incubated normally for 1 d. The injected tobacco leaves (taken for 4 d consecutive) were immersed in 20 mL of GUS staining solution, vacuum-treated in the dark for 10 min, and stained in the dark at 37°C for 24 h. The tobacco leaves were decolorized to transparency and scanned for analysis by a scanner.

Meanwhile, 0.5 g of injected tobacco leaves was placed in a 5 mL precooled centrifuge tube and set aside in liquid nitrogen. The leaves were ground to a powder using a precooled grinder, added to 2 mL of protein extract, mixed thoroughly and refrigerated at 4°C. The mixture was incubated for 1 h to fully react and then centrifuged at 2000 × g for 20 min at 4°C, removed the supernatant to a new centrifuge tube and centrifuged at 4°C and 2000 × g for another 10 min. Then 100 μL of the total protein supernatant was collected into a 1.5 mL centrifuge tube, and 900 μL of 37°C pretreated hot GUS extraction buffer was added. The tube was placed in a water bath at 37°C for 30 min before, added 200 μL to 800 μL of 0.2 M Na_2CO_3. The fluorescence intensity was measured using a fluorescence spectrophotometer at EM WL of 365 nm, EX WL of 455 nm and a slit of 3 nm.

Yeast one-hybrid assay

The pGADT7 vector was digested with *EcoR* I and *BamH* I, and pAbAi was digested by *Kpn* I and *Sal* I. The carrier fragment was then recovered. Finally, the enzymatically cleaved vector fragment and the target fragment were ligated by one-step ligation (see Table S1 for the primer sequence list).

The motifs of *SlTOR* promoter were analyzed using PlantCARE (http://bioinformatics.psb.ugent.be/webtools/

plantcare/html). Based on the analysis results, the *SlTOR* promoter with a fragment size of 1856 bp was divided into three overlapping fragments, and each fragment overlapped by about 50 bp to avoid destroying the binding sites. The *SlTOR* promoter with a fragment size of 400-900 bp was cloned using 2×Super Pfx MasterMix (CWBIO, Jiangsu, China), and then promoters of *SlTOR* were inserted into the pAbAi vector to construct the baits (pAbAi-$P_{SlTOR1/2/3}$). The pAbAi-P_{SlTOR} was introduced into Y1H Gold, encapsulated on SD/-Ura medium plates, and incubated for 2-4 d at 30°C in inverted position.

The plaque was identified by PCR. The positive plaques were recoated on a medium plate of SD/-Ura containing an AbA gradient (50-500 ng·mL^{-1}). The pAbAi vector with promoter was screened for AbA resistance, i.e., the lowest concentration of AbA without phage spots was screened for subsequent yeast one-hybrid assays.

AD-*SlMYC2* was transformed into a yeast strain containing the pAbAi-P_{SlTOR} linear vector and then encapsulated on SD/-Leu medium containing with AbA at the resistance concentration, and incubated upside down at 30°C to observe plaque growth.

ChIP–qPCR assay

SlMYC2-OE tomato seedlings grown to 15 d of age were used for ChIP–qPCR analysis. ChIP–qPCR assays were performed as described by Du (2017) (Du et al., 2017). The ChIP signal was analyzed using qPCR. Each ChIP value was normalized to its respective input DNA value, and the enrichment of DNA was displayed as a percentage of the input. Primer sequences were listed in Table S1.

Data processing

All experiments in this study were performed at least three repetitions. Experimental data were processed using Microsoft Excel 2010 and GraphPad Prism 6 for graphing. The significance of differences was determined by ANOVA or Student's *t* test using IBM SPSS 20 software ($P< 0.05$). The data in the graphs were the mean ± SD (standard deviation).

Results

JA signaling contributed to the growth and photosynthesis of tomato seedlings

JA is not only an important growth regulator but also an important signaling molecule for tomato growth and development. To investigate the regulatory effect of JA on growth and development, tomato seedlings were treated with MeJA (100 μM) and DIECA (200 μM. DIECA, JA biosynthesis inhibitor sodium diethyldithiocarbamate) (Ding et al., 2021). The results showed that MeJA inhibited plant height, stem thickness and dry weight (Figures 1A–E).

Photosynthesis is closely related to the dry matter accumulation of plants. The products of photosynthesis are the basis for plant growth and development, the formation of yield and quality. Therefore, the photosynthetic indexes of 5-week-old tomato seedlings were examined using a photosynthesis meter Li-6400XT. MeJA treatment significantly reduced the net photosynthetic rate (Pn) and chlorophyll a and b contents of tomato seedlings (Figures 1F–H). The root–shoot ratio increased slightly after MeJA treatment, but the difference was not significant compared with the control (Figure 1I). In addition, for all parameters, there were no significant differences between the DIECA treated and control groups (Figure 1).

The transcription factor SlMYC2 contributed to the growth and photosynthesis of tomato seedlings

SlMYC2 is a key transcription factor in the JA signaling pathway. To confirm the effect of SlMYC2 on the growth and development of tomato seedlings, we first examined the expression of *SlMYC2* after MeJA and DIECA treatment. *SlMYC2* expression was significantly enhanced after MeJA treatment. The expression of *SlMYC2* decreased slightly in DIECA treatment, but it was not significant compared with the control, indicating that JA signaling could activate the expression of *SlMYC2*(Figure S1).

Then, *SlMYC2* overexpression and silenced lines were used as experimental materials (Figure S2). The results indicated that the plant height, stem thickness and dry weight of the *SlMYC2* overexpression lines were significantly lower than those of WT, but there were no significant differences between *SlMYC2*-silenced lines and WT (Figures 2A–E).

In the *SlMYC2* overexpression lines, the Pn and chlorophyll a and b contents were lower than those in the WT. Among them, Pn was remarkably lower in the overexpression lines than in WT, but the chlorophyll a and b contents were not significant. The Pn and chlorophyll a and b contents in the *SlMYC2*-silenced lines showed an increasing trends; however, the chlorophyll b content was observably higher than that in the WT, while the Pn and chlorophyll a content were not significant (Figures 2F–H).

Nevertheless, the root–shoot ratio significantly increased in the *SlMYC2* overexpression lines, but the effect was not significant in the *SlMYC2*-silenced lines, suggesting that *SlMYC2* overexpression could enhance the number of healthy tomato seedlings (Figure 2I).

The above results indicated that JA and SlMYC2, key transcription factor in the JA signaling pathway, had a regulatory role in the growth and development of tomato seedlings and the net photosynthetic rate.

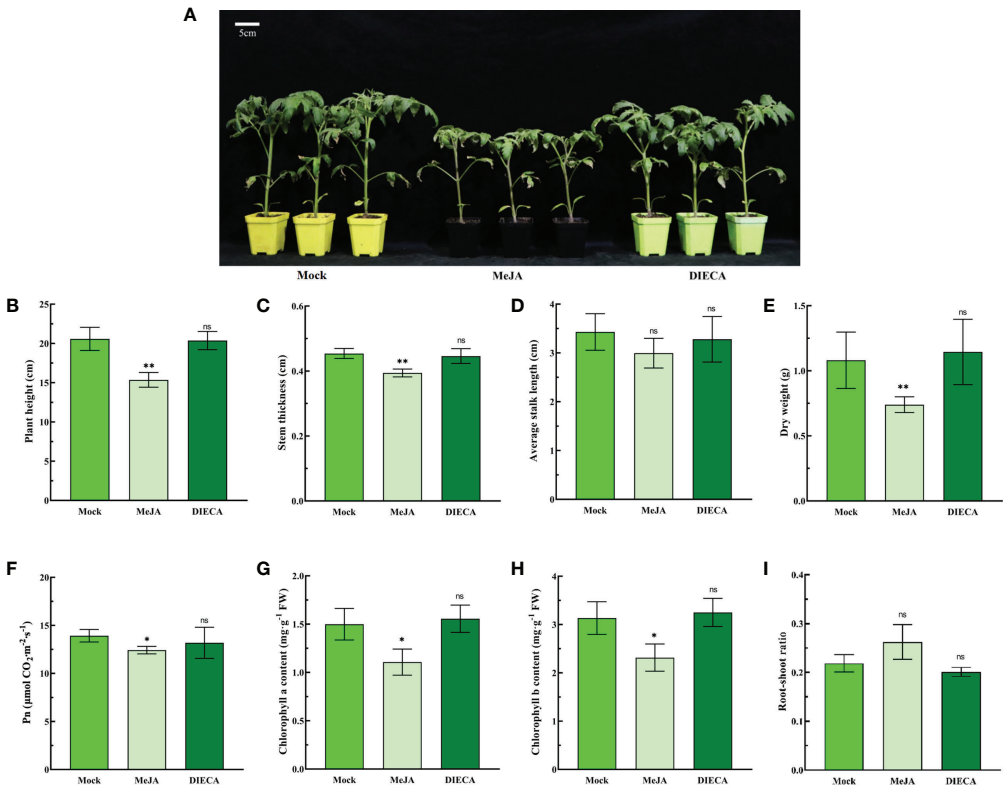


FIGURE 1
Effects of MeJA and DIECA on the growth and photosynthesis of tomato seedlings. **(A)** Phenotypic changes in 5-week-old seedlings after 1 week of treatment with 100 μM MeJA or 200 μM DIECA. **(B)** Plant height. **(C)** Stem thickness. **(D)** Average stalk length. **(E)** Dry weight. **(F)** Net photosynthetic rate (Pn). **(G)** Chlorophyll a content. **(H)** Chlorophyll b content. **(I)** Root–shoot ratio. Errors are standard deviations for three biological replicates (n=3). Differences between means were analyzed for significance using Student's *t* test, $*P < 0.05$, $**P < 0.01$. ns, no significant difference.

Regulation of the tomato fruit quality by the transcription factor SlMYC2

Tomato fruit quality is an important indicator for its commercialization. Therefore, the relevant quality indicators of mature tomato fruits were determined after MeJA and DIECA treatment (Figure S3). MeJA and DIECA treatment had no significant effect on tomato fruit firmness, indicating that their effects on fruit firmness were not obvious (Figure S3A). Although MeJA treatment significantly reduced the tomato fruit color index and total pectin (Figures S3B, C), it significantly increased the contents of lycopene, carotenoid, fructose, glucose, soluble sugar, starch, soluble protein, total phenol and flavonoids (Figures S3D–H, K, L, N, O), while the sugar–acid ratio also increased significantly (Figure S3J). MeJA treatment had no significant effects on total fruit acid or vitamin C content (Figures S3I, M). DIECA treatment only significantly improved the content of total pectin, but significantly decreased the content of flavonoids, however, while not significantly affecting other indicators (Figure S3). This result suggested that increasing JA could improve the quality of tomato fruits.

SlMYC2 is a key transcription factor in the JA signaling pathway; therefore, *SlMYC2* overexpression and *SlMYC2*-silenced lines were also used to determine tomato fruit quality (Figure 3). Tomato fruit firmness was significantly enhanced in the *SlMYC2* overexpression lines but significantly reduced in the silenced line (Figure 3A). Tomato fruit color index was significantly lower in the *SlMYC2* overexpression lines, but there was no significant difference between *SlMYC2*-silenced lines and WT (Figure 3B).

The color change of fruit pulp is mainly determined by the accumulation of carotenoids and lycopene. Therefore, the lycopene and carotenoids contents in tomato fruits of each group were further determined. The lycopene and carotenoid contents in mature fruits were significantly higher in the *SlMYC2* overexpression lines compared with those in the WT, but there was no significant difference in the *SlMYC2*-silenced lines (Figures 3D, E). The above results showed that *SlMYC2*-OE

10.3389/fpls.2022.1013445

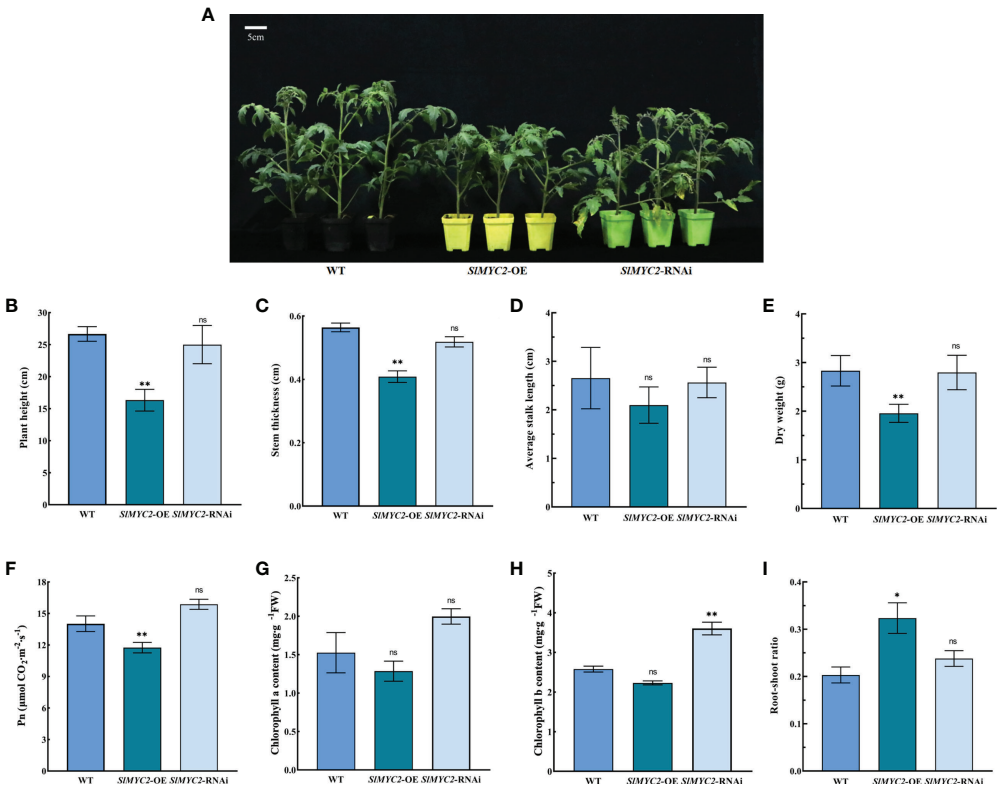


FIGURE 2
Effects of *SlMYC2* overexpression and silencing on the growth and photosynthesis of tomato seedlings. **(A)** Phenotypic of tomato seedlings (5-week-old). **(B)** Plant height. **(C)** Stem thickness. **(D)** Average stalk length. **(E)** Dry weight. **(F)** Net photosynthetic rate (Pn). **(G)** Chlorophyll a content. **(H)** Chlorophyll b content. **(I)** Root-shoot ratio. *SlMYC2*-OE, *SlMYC2* overexpression; *SlMYC2*-RNAi, *SlMYC2* silenced. Errors are standard deviations for three biological replicates (n=3). Differences between means were analyzed for significance using Student's *t* test, $*P < 0.05$, $**P < 0.01$. ns, no significant difference.

could increase the accumulation of lycopene and carotenoids and regulate fruit pulp color in mature tomato fruits.

The effects of SlMYC2 on tomato fruit were analyzed by measuring the nutrient composition of ripe fruit from the *SlMYC2* overexpression lines and *SlMYC2*-silenced lines. Total pectin content in ripe fruits from the *SlMYC2* overexpression lines was decreased significantly, whereas it increased significantly in the ripe fruits from *SlMYC2*-silenced lines (Figure 3C). *SlMYC2* overexpression significantly increased the contents of fructose, glucose, and soluble sugar but decreased the fruit acid content, whereas *SlMYC2* silencing had no significant effect (Figures 3F–I). However, *SlMYC2* overexpression increased the sugar–acid ratio in mature tomato fruits but decreased sugar–acid ratio in mature fruits of *SlMYC2*-silenced lines (Figure 3J). The contents of starch and total phenol in mature fruits of the *SlMYC2* overexpression lines were significantly higher than those of the control (Figures 3K, N), whereas the content of total vitamin C was increased with *SlMYC2* silencing (Figure 3M). The contents of soluble protein and flavonoids were significantly higher in mature fruits of *SlMYC2* overexpression lines, and the content of flavonoids was significantly lower in the *SlMYC2*-silenced line than in the control group, but the difference in soluble protein content was not significantly different (Figures 3L, O). These results revealed that JA signaling and the key transcription factor SlMYC2 were involved in the regulation of tomato fruit quality.

Interaction effect between TOR and JA signaling on the regulation of tomato seedling growth

TOR is conserved in eukaryotes and is one of the most important and highly conserved regulators of growth and development (Xiong et al., 2013). Three concentrations of RAP (SlTOR inhibitor) (1 μM, 5 μM and 10 μM) were employed to measure the germination rate and seedling height under different treatments. The results showed that the germination rate of tomato seeds and plant height were reduced by RAP treatment (Figures S4A, B). The 10 μM RAP

10.3389/fpls.2022.1013445

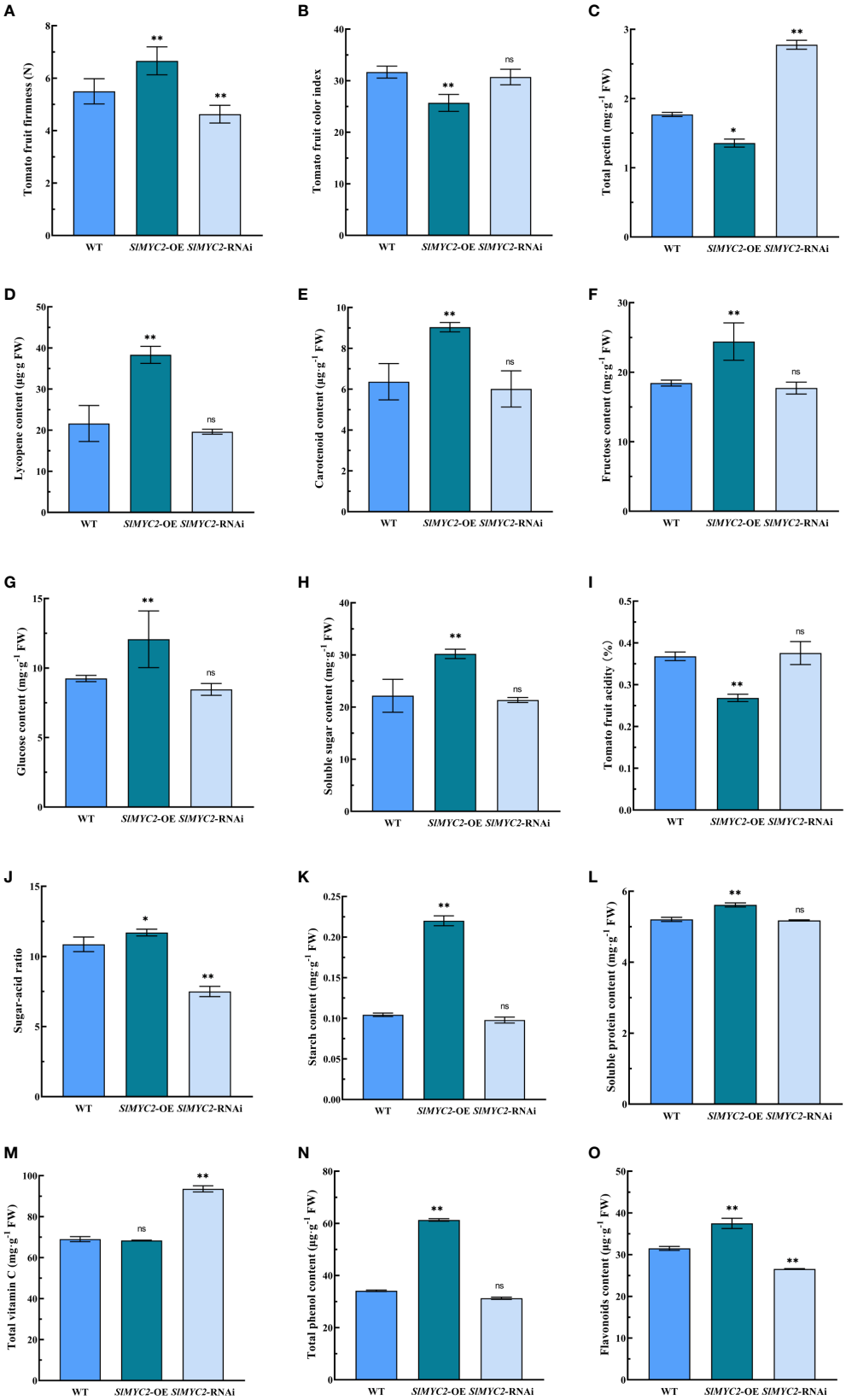


FIGURE 3 (Continued)

FIGURE 3 (Continued)
Effects of SlMYC2 overexpression and silencing on the quality of tomato fruit. **(A)** Tomato fruit firmness. **(B)** Tomato fruit color index. **(C)** Total pectin content. **(D)** Lycopene content. **(E)** Carotenoid content. **(F)** Fructose content. **(G)** Glucose content. **(H)** Soluble sugar content. **(I)** Tomato fruit acidity. **(J)** Sugar-acid ratio. **(K)** Starch content. **(L)** Soluble protein content. **(M)** Total vitamin C content. **(N)** Total phenol content. **(O)** Flavonoids conteSnltM. YC2-OE, SlMYC2 overexpression; SlMYC2-RNAi, SlMYC2 silenced. Errors are standard deviations for three biological replicates (n=3). Differences between means were analyzed for significance using Student's t test, *P < 0.05, **P < 0.01. ns, no significant difference.

concentration was superior to other concentrations, so 10 μM RAP was selected for further testing. Three MHY1485 (SlTOR activator) concentrations (1 μM, 5 μM and 10 μM) were tested. The results indicated that all concentrations could promote seed germination and seedling growth (Figures S4C, D, E) and accelerate relative hypocotyl length and root length (Figures S4F, G). However, overall, the 5 μM MHY1485 concentration was superior to other concentrations. Therefore, subsequent experiments were carried out using 5 μM MHY1485.

The treatments with RAP (10 μM) and MHY1485 (5 μM) were used for further experiments. The results demonstrated that RAP inhibited the growth of tomato seedlings, but promoted by MHY1485 (Figure 4A). Moreover, the seed germination, relative hypocotyl elongation and root elongation of tomato were significantly inhibited by MeJA treatment but promoted by DIECA (Figures 4B–G).

The photosynthetic changes of tomato seedlings were measured by a photosynthesis-testing instrument. When tomato seedlings were treated with RAP (10 μM), chlorophyll a and total chlorophyll (Figures 5A, C), the net photosynthetic rate (Pn) (Figure 5D) and the transpiration rate (Ts) (Figure 5E) of leaves were significantly reduced, while the changes in chlorophyll b (Figure 5B) and stomatal conductance (Gs) (Figure 5F) were not significant; however, the intercellular carbon dioxide concentration (Ci) was prominently increased (Figure 5G). The results suggested that the photosynthesis of tomato seedlings was inhibited after that RAP inhibited the TOR signaling pathway, but this change was caused by non-stomatal factors.

Under the synergistic treatment of MeJA and RAP, the TOR inhibitor RAP and jasmonic acid MeJA synergistically inhibited the growth of tomato seedlings, and the hypocotyl length and root length were significantly reduced (Table 1, Figure S5).

Further study showed that TOR affected the synthesis of endogenous JA. The content of JA in tomato seedlings was determined after treatment with the TOR inhibitor RAP and activator MHY1485. The results showed that RAP treatment increased the content of endogenous JA, but MHY1485 treatment significantly reduced the amount of JA (Figure 6A), indicating that TOR negatively regulated the synthesis of endogenous JA.

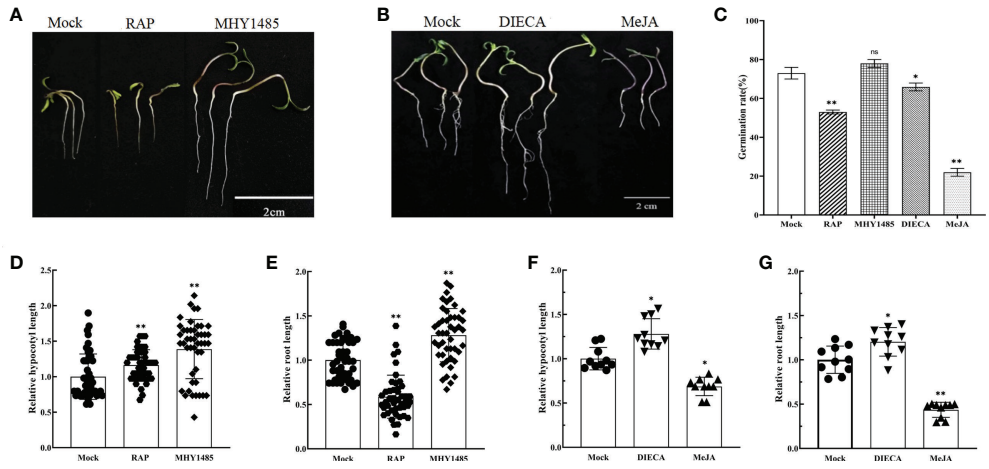


FIGURE 4
Effects of external application of RAP, MHY1485, DIECA and MeJA on tomato germination, relative hypocotyl length and root length. **(A)** Effects of external application of RAP and MHY1485 on tomato seedlings. **(B)** Effects of external application of DIECA and MeJA on tomato seedlings. **(C)** Effects of external application of RAP, MHY1485, DIECA and MeJA on tomato germination rate. **(D)** Effects of external application of RAP and MHY1485 on tomato relative hypocotyl length. **(E)** Effects of external application of RAP and MHY1485 on tomato relative root length. **(F)** Effects of external application of DIECA and MeJA on tomato relative hypocotyl length. **(G)** Effects of external application of DIECA and MeJA on tomato relative root length. Errors are standard deviations for three biological replicates (n=3). Differences between means were analyzed for significance using Student's *t* test, **P* < 0.05, ***P* < 0.01. ns, no significant difference.

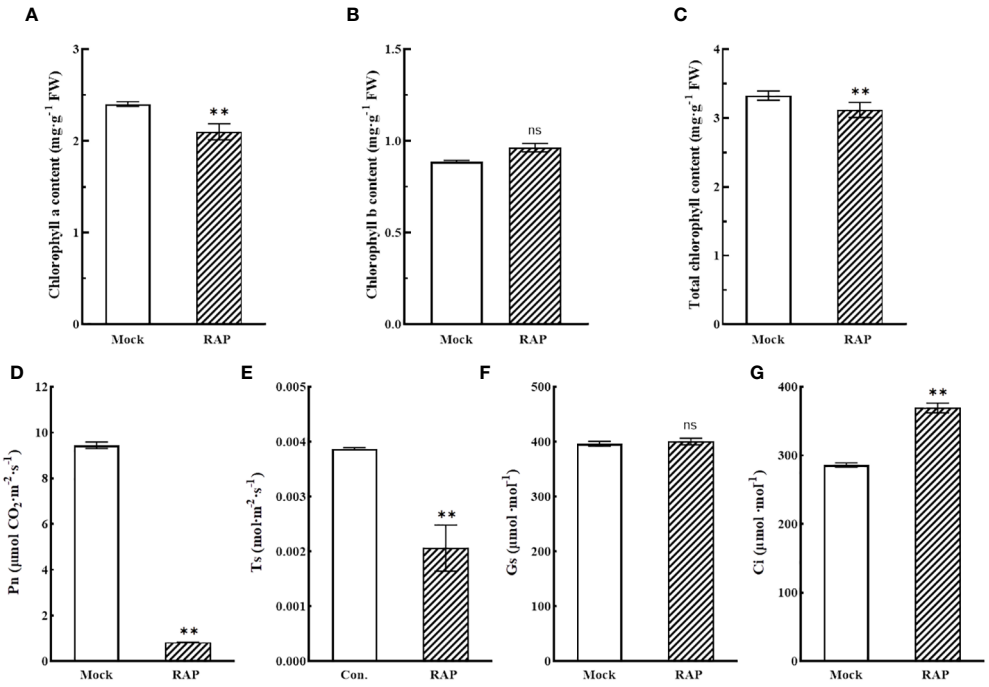


FIGURE 5
Changes in chlorophyll contents and photosynthetic indexes in leaves of tomato seedlings treated with RAP. **(A)** Chlorophyll a content in tomato leaves after RAP treatment. **(B)** Chlorophyll b content in tomato leaves after RAP treatment. **(C)** Total chlorophyll content in tomato leaves after RAP treatment. **(D)** Net photosynthetic rate in tomato leaves after RAP treatment. **(E)** Transpiration rate in tomato leaves after RAP treatment. **(F)** Stomatal conductance in tomato leaves after RAP treatment. **(G)** Intercellular carbon dioxide concentration in tomato leaves after RAP treatment. Errors are standard deviations for three biological replicates (n=3). Differences between means were analyzed for significance using Student's *t* test, $^{**}P < 0.01$. ns, no significant difference.

Then, RAP and MeJA treatment (Figure 6B) and TRV : *SlTOR* lines (Figure S6) were used to identify the expression of *SlMYC2* gene and downstream response genes in the JA signaling pathway. The results showed that *SlMYC2*, *TomLoxD* (JA synthesis gene) and *SlJA2L* (*SlMYC2* downstream target gene) were significantly upregulated after TOR signaling was suppressed by RAP and SlTOR silencing (Figures 6B, C). They were also upregulated by increasing MeJA, but *SlTOR* expression was downregulated by RAP and MeJA treatment, suggesting an interaction between TOR and JA signaling, which jointly regulated the growth of tomato seedlings.

TABLE 1 Effect of treatment with MeJA and RAP on tomato seedling growth.

Treatment	Hypocotyl length (cm)	Root length (cm)
Mock	5.558 ± 0.235a	8.057 ± 0.625a
MeJA	3.258 ± 0.228b	3.582 ± 0.637c
RAP	3.510 ± 0.243b	4.330 ± 0.437b
MeJA+RAP	2.327 ± 0.211c	3.118 ± 0.565c

Mock, Control group; MeJA, MeJA treatment; RAP, RAP treatment; MeJA+RAP, Treatment with both MeJA and RAP. Different letters indicate significant differences among treatments. Multiple comparisons of means were performed using Tukey's test at the 0.05 significance level.

SlMYC2 activated *SlTOR* expression by directly binding to the *SlTOR* promoter

Further bioinformatic analysis revealed that there were several predicted MYC2 binding elements existed in the *SlTOR* promoter (Figure 7A and Table S2), suggested that SlMYC2 might bind to the promoter of *SlTOR* and activate its expression *in vivo*. The binding ability of SlMYC2 to *SlTOR* promoter was firstly verified by yeast one-hybrid (Y1H) assay. In Y1H assay, according to the positions of the predicted elements, we divided the *SlTOR* promoter into three fragments: P_{SlTOR1}, P_{SlTOR2} and P_{SlTOR3} (Figure 7A). The results showed that yeast cells could growth under 550 ng·μL^{-1} AbA, indicating a potential interaction between SlMYC2 and P_{SlTOR3}. However, pAbAi-$P_{SlTOR1}$and pAbAi-P_{SlTOR2}were self-activated under the same AbA concentration (Figure 7B and Figure S7). To further confirm SlMYC2 bound to the *SlTOR* promoter *in vivo*, we performed chromatin immunoprecipitation (ChIP) experiments using *SlMYC2*-OE seedlings. We divided the *SlTOR* promoter into 7 segments for ChIP experiments (Figure 7A), and the details of the segments were shown in the Figures 7A and Table S2. The ChIP results showed that each fragment of the *SlTOR* promoter was enriched in some degree, but highest enrichment

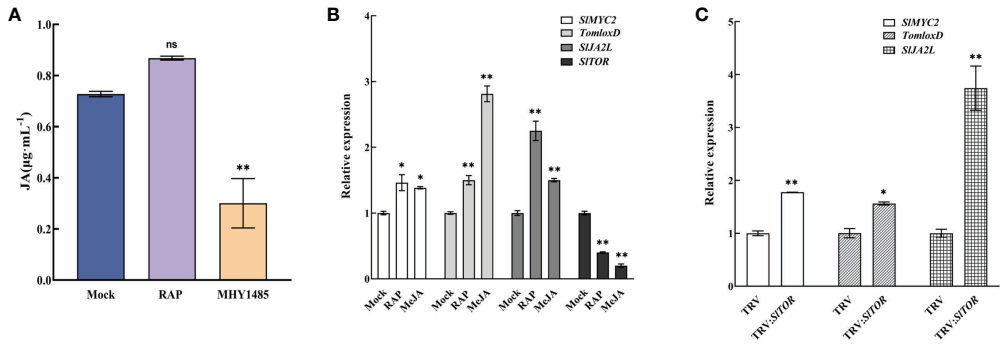


FIGURE 6
Crosstalk between JA signaling and TOR signaling pathways. **(A)** Endogenous JA content in tomato leaves after treatment with RAP and MHY1485. Mock, Control group. RAP, RAP treatment. MHY1485, MHY1485 treatment. **(B)** Expression of *SlMYC2*, *TomloxD*, *SlJA2L* and *SlTOR* after RAP and MeJA treatment. Mock, Control group. RAP, RAP treatment. MeJA, MeJA treatment. **(C)** Expression of *SlMYC2*, *TomloxD* and *SlJA2L* in TRV : *SlTOR*. TRV, only infiltrated with empty vectors. TRV : *SlTOR*, *SlTOR*-silenced lines. Errors are standard deviations for three biological replicates (n=3). Differences between means were analyzed for significance using Student's *t* test, *$P < 0.05$, **$P < 0.01$. ns, no significant difference.

occured at P5 segment which located in P_{SlTOR3} (Figure 7C), suggesting that potential binding site of SlMYC2 to *SlTOR* promoter might locate in -686 to -826 of the ATG upstream. Both Y1H and ChIP assays suggested that SlMYC2 bound to *SlTOR* promoter directly. To assay the activation effect of SlMYC2 on *SlTOR* transcription *in vivo*, transient transfections with different reporter and effector vectors were performed respectively. In the GUS histochemical staining assay, transformations with P_{SlTOR}::GUS and 35S::GFP were used as controls (Figures 7D, E). The staining results showed that GUS signal could not been observed when transformed with P_{SlTOR}:: GUS and 35S::GFP; however, it was appeared when co-transformed with both P_{SlTOR}::GUS and SlMYC2::GFP (Figure 7E). GUS enzyme activity analysis also supported the staining result. The GUS activity was significantly higher than those of the controls (Figure 7E), suggesting that SlMYC2 regulated the transcription of *SlTOR*. Moreover, we also found that the expression of *SlTOR* was obviously upregulated in the *SlMYC2* overexpression line (Figure 7F). Taken together, SlMYC2 might mediate the crosstalk between the JA and TOR signaling pathways through directly binding to the *SlTOR* promoter and activating its expression.

Discussion

JA signaling and the transcription factor SlMYC2 contributed to growth and fruit quality in tomato

The growth and development of seedlings are of great importance for obtaining vigorous seedlings and improving yield, quality and resistance. The growth and development of plants are regulated by many plant hormones. Recent studies have shown that JA plays an important role in the regulation of plant growth and development. The phytohormone JA regulates a wide range of biological processes (Chini et al., 2016; Huang et al., 2017; Li et al., 2021a; Song et al., 2022).

Jasmonates (JAs), which consist of jasmonic acid and its derivatives, such as methyl jasmonate (MeJA) and jasmonoyl-isoleucine, play multiple roles in growth and development as well as biotic and abiotic stress responses. Numerous studies have shown that JA is involved in root hair formation, stamen development, flowering, leaf senescence, anthocyanin biosynthesis and photosynthetic carbon fixation (He et al., 2002; Li et al., 2014; Qi et al., 2015; Ding et al., 2018; Wang et al., 2019; Han et al., 2020).

JA is both a hormone and a signaling molecule. MYC2 is a master regulator in the JA signaling pathway that is widely present in plants and animals (Zhang et al., 2015; Breeze, 2019) and has multiple regulatory functions (Chini et al., 2007; Sheard et al., 2010; Li et al., 2021b). MYC2 is an essential helix–loop–helix transcription factor that plays a critical role in the JA signaling pathway and is involved in a variety of plant growth and stress resistance processes. However, the effects of activation and inhibition of JA signaling, and enhancement and silencing of *SlMYC2* on the growth and photosynthesis of tomato seedlings, especially the regulation of tomato fruit quality, have not been systematically investigated.

The formation of fruit quality is inseparable from the growth of tomato seedlings. Only when the plants maintain continuous and healthily growth, can they provide sufficient nutrients for normal fruit development and maturation. To investigate the role of JA and its key transcription factor MYC2 in the growth of tomato seedlings, MeJA and DIECA were used as treatments, and *SlMYC2*-OE and *SlMYC2*-RANi lines were used to measure the

10.3389/fpls.2022.1013445

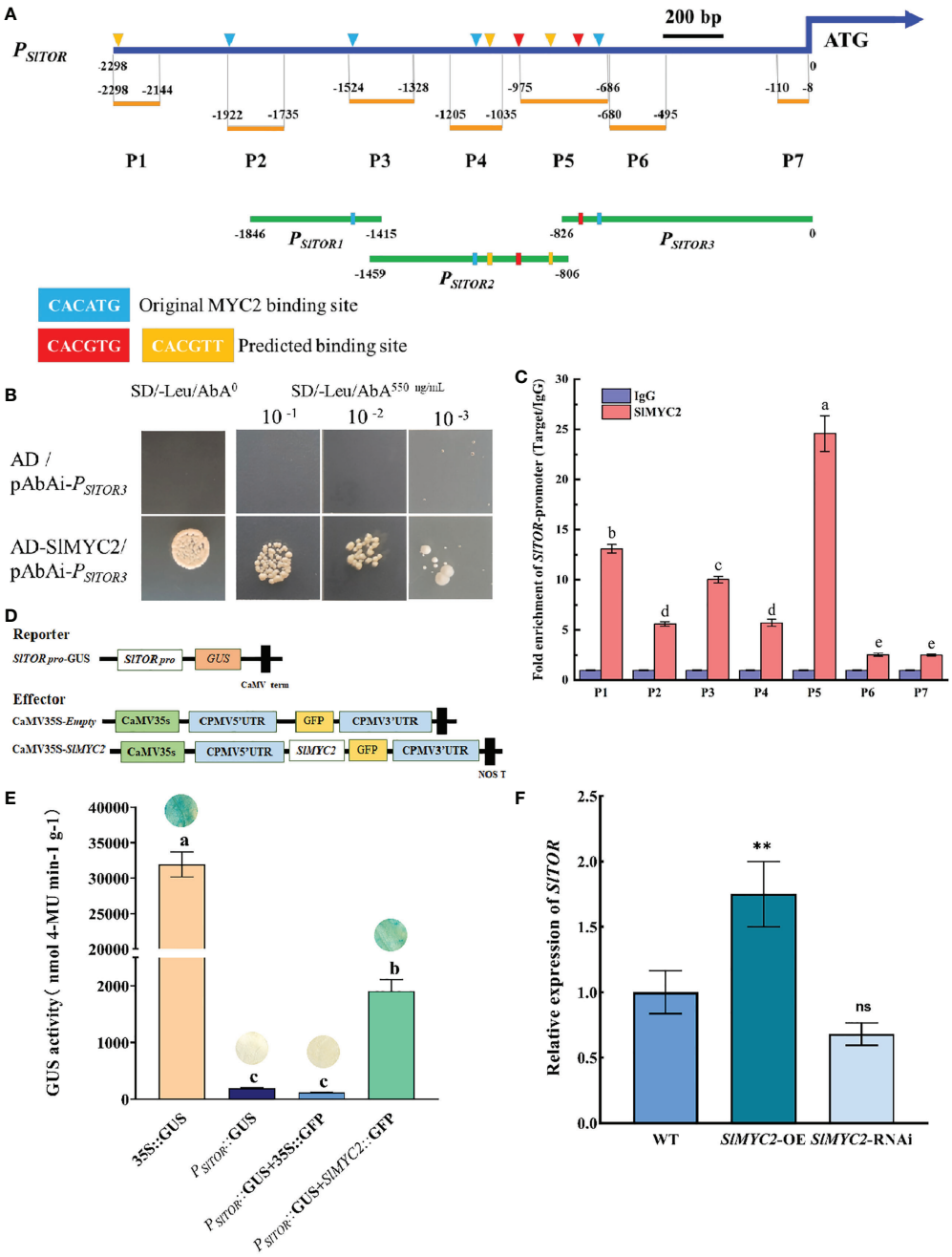


FIGURE 7
Interaction between SlMYC2 and the *SlTOR* promoter. **(A)** Schematic diagram of segments of *SlTOR* promoter in yeast one-hybrid and ChIP assays. P_{SlTOR1}, P_{SlTOR2} and P_{SlTOR3} are segments of the *SlTOR* promoter in yeast one-hybrid assays, P1-P7 are segments of the *SlTOR* promoter in ChIP assays. CACATG is the original MYC2 binding site, CACGTG and CACGTT are the predicted binding sites which are predicted by PlantCARE. **(B)** Yeast one-hybrid assay. Take 4 μL of bacterial solutions of different concentrations and spread it on a 550 ng·μL^{-1} AbA SD/-Leu plate to observe the yeast growth. **(C)** Schematic representation of the *SlTOR* promoter and chromatin immunoprecipitation (ChIP)–qPCR assays showing the binding of *SlMYC2*-OE to the *SlTOR* promoter upstream of the ATG start codon. P1-P7 represent partitions of the *SlTOR* promoter. Each ChIP value is normalized to its respective input DNA value, and enrichment is shown as the percentage of input. Error bars represent the SEM from three biological replicates. **(D, E)** Analysis of the *SlTOR* promoter activation by SlMYC2 using transient GUS activity assays. A 2-kb *SlTOR* upstream DNA fragment driving GUS was used as the reporter. The effectors were indicated in **(D)**. Measurement of transient GUS activity assay is shown in **(E)**. **(F)** Expression of *SlTOR* after SlMYC2 overexpression and silencing. Errors are standard deviations for three biological replicates (n=3). Differences between means were analyzed for significance using Student's *t* test, $*P < 0.05$, $**P < 0.01$. ns, no significant difference. Different letters indicate significant differences among treatments. Multiple comparisons of means were performed using Tukey's test at the 0.05 significance level.

growth and developmental indexes of tomato seedlings. The results showed that MeJA treatment and the *SlMYC2* overexpression inhibited the growth, development and photosynthesis of tomato seedlings but increased the root–shoot ratio. Increasing MeJA or the overexpression of *SlMYC2* inhibited the growth of tomato seedlings to some extent, but resulted in vigorous seedlings, indicating that JA signaling played a role in the growth and development of tomato seedlings.

Some studies showed that application of MeJA during preharvest and postharvest stages could enhance fruit antioxidant capacity and phenolic content, thereby extending the shelf life and improving fruit quality (Tzortzakis, 2007; Liu et al., 2018). The insignificant variation in lycopene content in fruits of lines with different jasmonic acid content indicates that the regulation of lycopene accumulation by jasmonic acid may be dependent on specific concentrations (Liu et al., 2012). Some experiments also show that the content of lycopene in the fruit of the JA synthetic mutant *spr2* decreases, suggesting a positive correlation between endogenous JA content and lycopene content in fruits. Exogenous application of MeJA can restore the lycopene content in mutant fruits (Min et al., 2018). Using the *SlMYC2*-silenced tomato fruit obtained by VIGS technology as the material, the results also show that SlMYC2 plays a critical role in MeJA-induced fruit chilling resistance, including the components related to fruit quality (Min et al., 2018). The above evidence suggests that JA and MYC2 play important roles in the regulation of tomato fruit quality. However, systematic studies on the changes of JA and SlMYC2 on tomato fruit quality are lacking.

In this research, we systematically studied the effects of JA and MYC2 on the quality of mature tomato fruits. The results showed that the contents of lycopene and carotenoid increased significantly after MeJA (100 μM) treatment and *SlMYC2* overexpression. Moreover, the contents of soluble sugar, protein, total phenol and flavonoids were also increased. MeJA processing and *SlMYC2* overexpression increased the sugar–acid ratio to improve the quality of fruits, and *SlMYC2* overexpression enhanced fruit firmness. Increasing MeJA or the overexpression of *SlMYC2* improved the quality of tomato fruits, and overexpression of *SlMYC2* increased fruit firmness and prolonged shelf life, indicating that JA signaling played an important role in tomato fruit quality.

Effects of SlMYC2 mediation on tomato growth and development through crosstalk between the JA and TOR signaling pathways

MYC2 is the key transcription factor in the JA signaling transduction pathway and is a critical component of JA signaling that interacts with other signaling pathways (Dombrecht et al., 2007; Altmann et al., 2020). MYC2 can integrate the interaction between JA signaling and other signals to regulate plant growth and development (Breeze, 2019). Target of rapamycin (TOR) is also an important signaling pathway that regulates plant growth and development, playing a central role in integrating metabolic energy and hormone signals (Dobrenel et al., 2016; Li et al., 2022). Through bioinformatic analysis, the binding elements of MYC2 were identified in the promoter of *SlTOR*. Therefore, it was hypothesized that SlMYC2 might be the node that mediated the crosstalk between the JA and TOR signaling pathways.

Studies have shown that JA signaling and TOR signaling are the core signals that regulate plant growth and development, such as seed germination, rhizome elongation, tricarboxylic acid cycle, starch storage and fruit quality (Xiong et al., 2013; Okello et al., 2015; Song et al., 2017). In addition, there may be an interaction between JA signaling and TOR signaling (Göbel and Feussner, 2009; Mosblech et al., 2009; Salem and Giavalisco, 2019). To clarify the role and relationship between JA and TOR signaling in tomato growth and development, tomato seedlings were treated with MHY1485 (an activator of TOR) and RAP (an inhibitor of TOR). When TOR signaling was inhibited by RAP, it increased the level of JA and inhibited the growth and photosynthesis of tomato seedlings, indicating that TOR signaling positively regulated the growth and development of tomato seedlings. Inhibition of TOR signaling and activation of JA signaling synergistically inhibited the tomato germination and seedling growth; that is, JA signaling and TOR signaling might be involved in crosstalk to regulate the growth and development of tomato seedlings. This result was consistent with previous findings (Song et al., 2017). Moreover, endogenous JA content was significantly reduced after activation of TOR signaling by MHY1485. The endogenous JA content increased after RAP inhibition of TOR, suggesting that TOR signaling inhibited JA regulation during plant growth and development by suppressing JA synthesis, and that SlMYC2 might mediate the crosstalk between JA signaling and TOR signaling pathways.

To show the effect of SlMYC2 on *SlTOR* transcription, the transcription of *SlTOR* in *SlMYC2*-silenced and overexpression lines was examined. The results showed that SlMYC2 promoted the accumulation of *SlTOR* transcription. Therefore, we tentatively determined that SlMYC2 could affect SlTOR signaling by regulating *SlTOR* transcription. To explore the mechanism by which SlMYC2 regulates *SlTOR* transcription, GUS staining, GUS enzyme activity and yeast one-hybrid assays were employed. In this study, yeast one-hybrid and ChIP assays obtained the same results that SlMYC2 could directly interact with *SlTOR*. Further analysis found that the highest enrichment occurred at P5 segment which located in P_{SlTOR3}, and the P_{SlTOR3} contained the CACGTG and CACATG motifs in P5 segment, suggesting that potential binding sites of SlMYC2 in *SlTOR* promoter might be located in -686 to -826 of the ATG upstream. It was determined that SlMYC2 could activate the transcriptional accumulation of *SlTOR* by binding to the

SlTOR promoter. This interaction might be one of the reasons why SlTOR regulated JA signaling by regulating SlMYC2.

The transcription factor MYC2 is a key factor in JA accumulation and signaling activation, and its regulation necessarily induces the response to JA signaling pathway (Heim et al., 2003; Chini et al., 2016; Goossens et al., 2016). SlTOR may regulate the response to the corresponding genes of JA by regulating the level of *SlMYC2*. Therefore, TRV-*SlTOR* lines were constructed to detect the expression of JA synthesis genes and SlMYC2 target genes (*TomLoxD* and *SlJA2L*). The results showed that the expression of *SlMYC2* and its target genes was significantly upregulated. The activation of *SlMYC2* and its target genes were significantly enhanced when SlTOR signaling pathway was inhibited, which uncovered that SlMYC2 mediated the interaction between JA and TOR signaling by acting on the promoter of *SlTOR*.

Studies in *Drosophila* and mammals have shown that TOR can activate PP2A to dephosphorylate MYC, leading to a decrease in MYC stability and protein content. When mTOR is inhibited, MYC2 synthesis is blocked and *MYC2* transcription is enhanced (Lee et al., 2017; Shen et al., 2018). However, more plant evidence is needed. In our unpublished data, the accumulation of SlMYC2 target genes and their protein content were significantly activated by the inhibition of TOR signaling. Therefore, subsequent studies will further reveal whether SlTOR regulates JA signaling through SlMYC2.

In conclusion, *SlMYC2* overexpression could significantly inhibit the growth of tomato seedlings but could improve the quality of tomato fruits. Biochemical experiments showed that inhibition of TOR signaling promoted JA synthesis in tomato and increased the amount of JA in the plant, releasing SlMYC2. SlMYC2 combined with the *SlTOR* promoter, thus activating the *SlTOR* gene, which might feedback on the inhibition of SlMYC2 (Figure S8). However, the feedback regulation of SlMYC2 by TOR signaling and the effect on JA signaling still need to be further explored. JA is an anti-stress hormone. The level of JA is low in plants under normal conditions, but under stress, JA is elevated to initiate SlMYC2 and activates *SlTOR* in combination with the *SlTOR* promoter; thus, feedback inhibits JA synthesis and *SlMYC2* expression, leading to a balance of growth and stress resistance. Therefore, the present study can lay the foundation for further revealing the mechanism of JA and TOR signaling interactions in regulating plant growth and development as well as stress resistance balance under stress.

Data availability statement

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

Author contributions

YZ and HX contributed to the experimental design, tomato planting and sampling, tobacco planting and experimentation, data processing and result analysis and writing. HW, LY, ZY, and XM contributed to the analysis of the data in the experiment. PH, HF, YY, and NC revised the paper. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the National Key Research and Development Program Projects of China (2019YFD1000300) and the Key Project of Science and Technology Research of Liaoning Provincial Education Department (LJKZ0631).

Acknowledgments

We are particularly grateful to Dr. Chuanyou Li (Institute of Genetics and Developmental Biology Chinese Academy of Sciences) for kindly providing the *SlMYC2* mutants. We would like to thank American Journal Experts (www.aje.com) for English language editing.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1013445/full#supplementary-material

References

Aerts, N., Mendes, M. P., 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

Aleman, F., Yazaki, J., Lee, M., Takahashi, Y., Kim, A. Y., Li, Z., et al. (2016). An ABA-increased interaction of the PYL6 ABA receptor with MYC2 transcription factor: a putative link of ABA and JA signaling. *Sci. Rep.* 6, 28941. doi: 10.1038/srep28941

Altmann, M., Altmann, S., Rodriguez, P. A., Weller, B., Elorduy Vergara, L., Palme, J., et al. (2020). Extensive signal integration by the phytohormone protein network. *Nature* 583 (7815), 271–276. doi: 10.1038/s41586-020-2460-0

Biswas, K. K., Neumann, R., Haga, K., Yatoh, O., and Iino, M. (2003). Photomorphogenesis of rice seedlings: A mutant impaired in phytochromemediated inhibition of coleoptile growth. *Plant Cell Physiol.* 44, 242–254. doi: 10.1093/pcp/pcg040

Boutouja, F., Stiehm, C. M., and Platta, H. W. (2019). mTOR: A cellular regulator interface in health and disease. *Cells* 8 (1), 18. doi: 10.3390/cells8010018

Breeze, E. (2019). Master MYCs: MYC2, the jasmonate signaling "master switch". *Plant Cell* 31 (1), 9–10. doi: 10.1105/tpc.19.00004

Brunkard, J. O., Xu, M., Scarpin, M. R., Chatterjee, S., Shemyakina, E. A., Goodman, H. M., et al. (2020). TOR dynamically regulates plant cell transport. *Proc. Natl. Acad. Sci. U. S. A.* 117 (9), 5049–5058. doi: 10.1073/pnas.1919196117

Cai, Q., Yuan, Z., Chen, M., Yin, C., Luo, Z., Zhao, X., et al. (2014). Jasmonic acid regulates spikelet development in rice. *Nat. Commun.* 5, 3476. doi: 10.1038/ncomms4476

Cárdenas, P. D., Sonawane, P. D., Pollier, J., Vanden Bossche, R., Dewangan, V., Weithorn, E., et al. (2016). GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the plant mevalonate pathway. *Nat. Commun.* 7, 1–16. doi: 10.1038/ncomms10654

Chehab, E. W., Yao, C., Henderson, Z., Kim, S., and Braam, J. (2012). *Arabidopsis* touch-induced morphogenesis is jasmonate mediated and protects against pests. *Curr. Biol.* 22, 701–706. doi: 10.1016/j.cub.2012.02.061

Chen, Q., Sun, J., Zhai, Q., Zhou, W., Qi, L., Xu, L. S., et al. (2011). The basic helix-loop-helix transcription factor MYC2 directly represses *PLETHORA* expression during jasmonate mediated modulation of the root stem cell niche in *Arabidopsis*. *Plant Cell* 23, 3335–3352. doi: 10.1105/tpc.111.089870

Chini, A., Fonseca, S., Fernandez, G., Adie, B., Chico, J. M., Lorenzo, O., et al. (2007). The JAZ family of repressors is the missing link in jasmonate signalling. *Nature* 448, 666–671. doi: 10.1038/nature06006

Chini, A., Gimenez-Ibanez, S., Goossens, A., and Solano, R. (2016). Redundancy and specificity in jasmonate signalling. *Curr. Opin. Plant Biol.* 33, 147–156. doi: 10.1016/j.pbi.2016.07.005

De Vleesschauwer, D., Filipe, O., Hoffman, G., Seifi, H. S., Haeck, A., Canlas, P., et al. (2018). Target of rapamycin signaling orchestrates growth-defense trade-offs in plants. *New Phytol.* 217, 305–319. doi: 10.1111/nph.14785

Ding, F., Wang, C., Xu, N., Wang, M., and Zhang, S. (2021). Jasmonic acid-regulated putrescine biosynthesis attenuates cold-induced oxidative stress in tomato plants. *Sci. Hortic.* 288, 110373. doi: 10.1016/j.scienta.2021.110373

Ding, F., Wang, M., and Zhang, S. (2018). Sedoheptulose-1,7-bisphosphatase is involved in methyl jasmonate- and dark-induced leaf senescence in tomato plants. *Int. J. Mol. Sci.* 19, 3673. doi: 10.3390/ijms19113673

Dobrenel, T., Caldana, C., Hanson, J., Robaglia, C., Vincentz, M., Veit, B., et al. (2016). TOR signaling and nutrient sensing. *Annu. Rev. Plant Biol.* 67, 261–285. doi: 10.1146/annurev-arplant-043014-114648

Dombrecht, B., Xue, G. P., Sprague, S. J., Kirkegaard, J. A., Ross, J. J., Reid, J. B., et al. (2007). MYC2 differentially modulates diverse jasmonate-dependent functions in *Arabidopsis*. *Plant Cell* 19, 2225–2245. doi: 10.1105/tpc.106.048017

Du, M., Zhao, J., Tzeng, D. T. W., Liu, Y., Deng, L., Yang, T., et al. (2017). MYC2 orchestrates a hierarchical transcriptional cascade that regulates jasmonate-mediated plant immunity in tomato. *Plant Cell* 29, 1883–1906. doi: 10.1105/tpc.16.00953

Fu, L., Liu, Y., Qin, G., Wu, P., Zi, H., Xu, Z., et al. (2021). The TOR-EIN2 axis mediates nuclear signalling to modulate plant growth. *Nature* 591, 288–292. doi: 10.1038/s41586-021-03310-y

Fu, L., Wang, P., and Xiong, Y. (2020). Target of rapamycin signaling in plant stress responses. *Plant Physiol.* 182, 1613–1623. doi: 10.1104/pp.19.01214

Göbel, C., and Feussner, I. (2009). Methods for the analysis of oxylipins in plants. *Phytochem* 70, 1485–1503. doi: 10.1016/j.phytochem.2009.07.040

Goossens, J., Fernández-Calvo, P., Schweizer, F., and Goossens, A. (2016). Jasmonates: signal transduction components and their roles in environmental stress responses. *Plant Mol. Biol.* 91, 673–689. doi: 10.1007/s11103-016-0480-9

Han, X., Zhang, M., Yang, M., and Hu, Y. (2020). *Arabidopsis* JAZ proteins interact with and suppress RHD6 transcription factor to regulate jasmonate-stimulated root hair development. *Plant Cell* 32, 1049–1062. doi: 10.1105/tcp.19.00617

He, Y., Fukushige, H., Hildebrand, D., and Gan, S. (2002). Evidence supporting a role of jasmonic acid in *Arabidopsis* leaf senescence. *Plant Physiol.* 128, 876–884. doi: 10.1104/pp.010843

Heim, M. A., Jakoby, M., Werber, M., Martin, C., Weisshaar, B., and Bailey, P. C. (2003). The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Mol. Biol. Evol.* 20, 735–747. doi: 10.1093/molbev/msg088

Hou, X., Lee, L. Y., Xia, K., Yan, Y., and Yu, H. (2010). DELLAs modulate jasmonate signaling *via* competitive binding to JAZs. *Dev. Cell.* 19, 884–894. doi: 10.1016/j.devcel.2010.10.024

Huang, H., Liu, B., Liu, L., and Song, S. (2017). Jasmonate action in plant growth and development. *J. Exp. Bot.* 68, 1349–1359. doi: 10.1093/jxb/erw495

Jia, Z., Tang, M., and Wu, J. (1999). The determination of flavonoid contents in mulberry and their scavenging effects on superoxide radicals. *Food Chem.* 64, 555–559. doi: 10.1016/S0308-8146(98)00102-2

Kim, J., Chang, C., and Tucker, M. L. (2015). To grow old: regulatory role of ethylene and jasmonic acid in senescence. *Front. Plant Sci.* 6, 20. doi: 10.13016/M2ZK55Q2J

Lee, J. E., Rayyan, M., Liao, A., Edery, I., and Pletcher, S. D. (2017). Acute dietary restriction acts *via* TOR, PP2A, and myc signaling to boost innate immunity in drosophila. *Cell Rep.* 20, 479–490. doi: 10.1016/j.celrep.2017.06.052

Li, D., Ding, Y., Cheng, L., Zhang, X., Cheng, S., Ye, Y., et al. (2022). Target of rapamycin (TOR) regulates the response to low nitrogen stress *via* autophagy and hormone pathways in. *Malus Hupehensis Horticulture Research* 9. doi: 10.1093/hr/uhac143

Li, T., Jia, K. P., Lian, H. L., Yang, X., Li, L., and Yang, H. Q. (2014). Jasmonic acid enhancement of anthocyanin accumulation is dependent on phytochrome a signaling pathway under far-red light in *Arabidopsis*. *Biochem. Biophys. Res. Commun.* 454, 78–83. doi: 10.1016/j.bbrc.2014.10.059

Li, Z., Luo, X., Ou, Y., Jiao, H., Peng, L. , Fu, X., et al. (2021a). JASMONATE-ZIM DOMAIN proteins engage polycomb chromatin modifiers to modulate jasmonate signaling in *Arabidopsis*. *Mol. Plant* 14, 732–747. doi: 10.1016/j.molp.2021.03.001

Li, Z., Min, D., Fu, X., Zhao, X., Wang, J. , Zhang, X., et al. (2021b). The roles of SlMYC2 in regulating ascorbate-glutathione cycle mediated by methyl jasmonate in postharvest tomato fruits under cold stress. *Sci. Hortic-Amsterdam.* 288, 110406. doi: 10.1016/j.scienta.2021.110406

Liu, Y., Du, M., Lei, D., Shen, J., and Zhai, Q. (2019). MYC2 regulates the termination of jasmonate signaling *via* an autoregulatory negative feedback loop. *Plant Cell.* 31, 106–127. doi: 10.1105/tpc.18.00405

Liu, H., Meng, F., Miao, H., Chen, S., Yin, T., Hu, S., et al. (2018). Effects of postharvest methyl jasmonate treatment on main health-promoting components and volatile organic compounds in cherry tomato fruits. *Food Chem.* 263, 194–200. doi: 10.1016/j.foodchem.2018.04.124

Liu, L., Wei, J., Zhang, M., Zhang, L., Li, C., and Wang, Q. (2012). Ethylene independent induction of lycopene biosynthesis in tomato fruits by jasmonates. *J. Exp. Bot.* 63, 5751–5761. doi: 10.1093/jxb/ers224

Li, T., Xu, Y., Zhang, L., Ji, Y., Tan, D., Yuan, H., et al. (2017). The jasmonate-activated transcription factor MdMYC2 regulates ETHYLENE RESPONSE FACTOR and ethylene biosynthetic genes to promote ethylene biosynthesis during apple fruit ripening. *Plant Cell.* 29, 1316–1334. doi: 10.1105/tpc.17.00349

Min, D., Li, F., Cui, X., Zhou, J., and Li, J. (2020). SlMYC2 are required for methyl jasmonate-induced tomato fruit resistance to *Botrytis cinerea*. *Food Chem.* 310, 125901. doi: 10.1016/j.foodchem.2019.125901

Min, D., Li, F., Zhang, X., Cui, X., Shu, P., Dong, L., et al. (2018). *SlMYC2* involved in methyl jasmonate-induced tomato fruit chilling tolerance. *J. Agric. Food Chem.* 66, 3110–3117. doi: 10.1021/acs.jafc.8b00299

Mosblech, A., Feussner, I., and Heilmann, I. (2009). Oxylipins: structurally diverse metabolites from fatty acid oxidation. *Plant Physiol. Biochem.* 47, 511–517. doi: 10.1016/j.plaphy.2008.12.011

Okello, R. C. O., Heuvelink, E., de Visser, P. H. B., Struik, P. C., and Marcelis, L. F. M. (2015). What drives fruit growth? *Funct. Plant Biol.* 42, 817–827. doi: 10.1071/FP15060

O'Leary, B. M., Oh, G. G. K., Lee, C. P., and Millar, A. H. (2020). Metabolite regulatory interactions control plant respiratory metabolism *via* target of rapamycin (TOR) kinase activation. *Plant Cell.* 32, 666–682. doi: 10.1105/tpc.19.00157

Qi, T., Huang, H., Song, S., and Xie, D. (2015). Regulation of jasmonate-mediated stamen development andseed production by a bHLH-MYB complex in *Arabidopsis*. *Plant Cell.* 27, 1620–1633. doi: 10.1105/tpc.15.00116

Riemann, M., Haga, K., Shimizu, T., Okada, K., Ando, S., Mochizuki, S., et al. (2013). Identification of rice *Allene oxide cyclase* mutants and the function of jasmonate for defence against *Magnaporthe oryzae*. *Plant J.* 74, 226–238. doi: 10.1111/tpj.12115

Riemann, M., Riemann, M., and Takano, M. (2008). Rice JASMONATE RESISTANT 1 is involved in phytochrome and jasmonate signalling. *Plant Cell Environ.* 31, 783–792. doi: 10.1111/j.1365-3040.2008.01790.x

Salem, M. A., and Giavalisco, P. (2019). Mutation in the *Arabidopsis* regulatory-associated protein TOR 1B (RAPTOR1B) leads to decreased jasmonates levels in leaf tissue. *Plant Signaling Behav.* 14, e1649567. doi: 10.1080/15592324.2019.1649567

Salvia-Trujillo, L., and McClements, D. J. (2016). Enhancement of lycopene bioaccessibility from tomato juice using excipient emulsions: Influence of lipid droplet size. *Food Chem.* 210, 295–304. doi: 10.1016/j.foodchem.2016.04.125

Sheard, L. B., Tan, X., Mao, H., Withers, J., Ben-Nissan, G., Hinds, T. R., et al. (2010). Jasmonate perception by inositol-phosphate-potentiated COI1-JAZ co-receptor. *Nature* 468, 400–405. doi: 10.1038/nature09430

Shen, Q., Lu, X., Yan, T., Fu, X., Lv, Z., Zhang, F., et al. (2016). The jasmonate-responsive AaMYC2 transcription factor positively regulates artemisinin biosynthesis in *Artemisia annua*. *New Phytol.* 210, 1269–1281. doi: 10.1111/nph.13874

Shen, P., Reineke, L. C., Knutsen, E., Chen, M., Pichler, M., Ling, H., et al. (2018). Metformin blocks MYC protein synthesis in colorectal cancer *via* m TOR-4 EBP-e IF 4E and MNK1-e IF 4G-e IF 4E signaling. *Mol. Oncol.* 12, 1856–1870. doi: 10.1002/1878-0261.12384

Song, H., Duan, Z., Wang, Z., Li, Y., Wang, Y., Li, C., et al. (2022). Genome-wide identification, expression pattern and subcellular localization analysis of the JAZ gene family in *Toona ciliata*. *Ind. Crop Prod.* 178, 114582. doi: 10.1016/j.indcrop.2022.114582

Song, S., Qi, T., Wasternack, C., and Xie, D. (2014). Jasmonate signaling and crosstalk with gibberellin and ethylene. *Curr. Opin. Plant Biol.* 21, 112–119. doi: 10.1016/j.pbi.2014.07.005

Song, Y., Zhao, G., Zhang, X., Li, L., Xiong, F., Zhuo, F., et al. (2017). The crosstalk between target of rapamycin (TOR) and jasmonic acid (JA) signaling existing in *Arabidopsis* and cotton. *Sci. Rep.* 7, 45830. doi: 10.1038/srep45830

Thagun, C., Imanishi, S., Kudo, T., Nakabayashi, R., Ohyama, K., Mori, T., et al. (2016). Jasmonate responsive ERF transcription factors regulate steroidal glycoalkaloid biosynthesis in tomato. *Plant Cell Physiol.* 57, 961–975. doi: 10.1093/pcp/pcw067

Thines, B., Katsir, L., Melotto, M., Niu, Y., Mandaokar, A., Liu, G., et al. (2007). JAZ repressor proteins are targets of the SCF COI1 complex during jasmonate signalling. *Nature* 448 (7154), 661–665. doi: 10.1038/nature05960

Toor, R. K., and Savage, G. P. (2005). Antioxidant activity in different fractions of tomatoes. *Food Res. Int.* 38, 487–494. doi: 10.1016/j.foodres.2004.10.016

Tzortzakis, N. G. (2007). Methyl jasmonate-induced suppression of anthracnose rot in tomato fruit. *Crop Prot.* 26, 1507–1513. doi: 10.1016/j.cropro.2006.12.014

Wang, J., Song, L., Gong, X., Xu, J., and Li, M. (2020). Functions of jasmonic acid in plant regulation and response to abiotic stress. *Int. J. Mol. Sci.* 21, 1446. doi: 10.3390/ijms21041446

Wang, M., Zhang, T., and Ding, F. (2019). Exogenous melatonin delays methyl jasmonate-triggered senescence in tomato leaves. *Agronomy* 9, 795. doi: 10.3390/agronomy9120795

Wasternack, C., and Hause, B. (2013). Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. *Ann. Bot.* 111, 1021–1058. doi: 10.1093/aob/mct067

Xiong, Y., McCormack, M., Li, L., Hall, Q., Xiang, C., and Sheen, J. (2013). Glucose-TOR signalling reprograms the transcriptome and activates meristems. *Nature* 496, 181–186. doi: 10.1038/nature12030

Xiong, F., Zhang, R., Meng, Z., Deng, K., Que, Y., Zhuo, F., et al. (2017). Brassinosteriod insensitive 2 (BIN2) acts as a downstream effector of the target of rapamycin (TOR) signaling pathway to regulate photoautotrophic growth in *Arabidopsis*. *New Phytol.* 213, 233–249. doi: 10.1111/nph.14118

Xu, J., van Herwijnen, Z. O., Dräger, D. B., Sui, C., Haring, M. A., and Schuurink, R. C. (2018). SlMYC1 regulates type VI glandular trichome formation and terpene biosynthesis in tomato glandular cells. *Plant Cell* 30, 2988–3005. doi: 10.1105/tpc.18.00571

Yan, Y., Christensen, S., Isakeit, T., Engelberth, J., Meeley, R., Hayward, A., et al. (2012). Disruption of OPR7 and OPR8 reveals the versatile functions of jasmonic acid in maize development and defense. *Plant Cell.* 24, 1420–1436. doi: 10.1105/tcp.111.094151

Yang, D. L., Yao, J., Mei, C. S., Tong, X. H., Zeng, L. J., Li, Q., et al. (2012). Plant hormone jasmonate prioritizes defense over growth by interfering with gibberellin signaling cascade. *Proc. Natl. Acad. Sci. U. S. A.* 109, E1192–E1200. doi: 10.1073/pnas.1201616109

Yan, L. H., Zhai, Q. Z., Wei, J. N., Wang, S. Y., Huang, T. T., Sun, M. M., et al. (2013). Role of tomato lipoxygenase d in wound-induced jasmonate biosynthesis and plant immunity to insect herbivores. *PloS Genet.* 9, e1003964. doi: 10.1371/journal.pgen.1003964

Yu, G., Wang, X., Chen, Q., Cui, N., Yu, Y., and Fan, H. (2019). Cucumber mildew resistance locus O interacts with calmodulin and regulates plant cell death associated with plant immunity. *Int. J. Mol. Sci.* 20, 2995. doi: 10.3390/ijms20122995

Zhang, Y., and Turner, J. G. (2008). Wound-induced endogenous jasmonates stunt plant growth by inhibiting mitosis. *PloS One* 3, e3699. doi: 10.1371/journal.pone.0003699

Zhang, F., Yao, J., Ke, J., Zhang, L., Lam, V. Q., Xin, X. F., et al. (2015). Structural basis of JAZ repression of MYC transcription factors in jasmonate signalling. *Nature* 525, 269–273. doi: 10.1038/nature14661

Zhuo, F., Xiong, F., Deng, K., Li, Z., and Ren, M. (2020). Target of rapamycin (TOR) negatively regulates ethylene signals in *Arabidopsis*. *Int. J. Mol. Sci.* 21, 2680. doi: 10.3390/ijms21082680

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 08 December 2022
DOI 10.3389/fpls.2022.1075279

OPEN ACCESS

EDITED BY
Hussain Touseef,
Matimate Agromart Pvt. Ltd. (Sevama AgriClinic Laboratory), India

REVIEWED BY
Mehanathan Muthamilarasan,
University of Hyderabad, India
Om Prakash Narayan,
Tufts University, United States

*CORRESPONDENCE
Qiang Wei
weiqiang@njfu.edu.cn
Zhijun Zhang
zjzhang@zafu.edu.cn

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 20 October 2022
ACCEPTED 23 November 2022
PUBLISHED 08 December 2022

CITATION
Ramakrishnan M, Zhang Z, Mullasseri S, Kalendar R, Ahmad Z, Sharma A, Liu G, Zhou M and Wei Q (2022) Epigenetic stress memory: A new approach to study cold and heat stress responses in plants. *Front. Plant Sci.* 13:1075279. doi: 10.3389/fpls.2022.1075279

Epigenetic stress memory: A new approach to study cold and heat stress responses in plants

Muthusamy Ramakrishnan[1], Zhijun Zhang[2,3]*, Sileesh Mullasseri[4], Ruslan Kalendar[5,6], Zishan Ahmad[1], Anket Sharma[7], Guohua Liu[1], Mingbing Zhou[7,8] and Qiang Wei[1]*

[1]Co-Innovation Center for Sustainable Forestry in Southern China, Bamboo Research Institute, Key Laboratory of National Forestry and Grassland Administration on Subtropical Forest Biodiversity Conservation, College of Biology and the Environment, Nanjing Forestry University, Nanjing, Jiangsu, China, [2]Bamboo Industry Institute, Zhejiang A&F University, Hangzhou, Zhejiang, China, [3]School of Forestry and Biotechnology, Zhejiang A&F University, Hangzhou, Zhejiang, China, [4]Department of Zoology, St. Albert's College (Autonomous), Kochi, Kerala, India, [5]Helsinki Institute of Life Science HiLIFE, Biocenter 3, University of Helsinki, Helsinki, Finland, [6]National Laboratory Astana, Nazarbayev University, Astana, Kazakhstan, [7]State Key Laboratory of Subtropical Silviculture, Bamboo Industry Institute, Zhejiang A&F University, Hangzhou, Zhejiang, China, [8]Zhejiang Provincial Collaborative Innovation Center for Bamboo Resources and High-Efficiency Utilization, Zhejiang A&F University, Hangzhou, Zhejiang, China

Understanding plant stress memory under extreme temperatures such as cold and heat could contribute to plant development. Plants employ different types of stress memories, such as somatic, intergenerational and transgenerational, regulated by epigenetic changes such as DNA and histone modifications and microRNAs (miRNA), playing a key role in gene regulation from early development to maturity. In most cases, cold and heat stresses result in short-term epigenetic modifications that can return to baseline modification levels after stress cessation. Nevertheless, some of the modifications may be stable and passed on as stress memory, potentially allowing them to be inherited across generations, whereas some of the modifications are reactivated during sexual reproduction or embryogenesis. Several stress-related genes are involved in stress memory inheritance by turning on and off transcription profiles and epigenetic changes. Vernalization is the best example of somatic stress memory. Changes in the chromatin structure of the *Flowering Locus C* (*FLC*) gene, a MADS-box transcription factor (TF), maintain cold stress memory during mitosis. *FLC* expression suppresses flowering at high levels during winter; and during vernalization, B3 TFs, cold memory *cis*-acting element and polycomb repressive complex 1 and 2 (PRC1 and 2) silence *FLC* activation. In contrast, the repression of *SQUAMOSA promoter-binding protein-like* (*SPL*) TF and the activation of *Heat Shock* TF (*HSFA2*) are required for heat stress memory. However, it is still unclear how stress memory is inherited by offspring, and the integrated view of the

regulatory mechanisms of stress memory and mitotic and meiotic heritable changes in plants is still scarce. Thus, in this review, we focus on the epigenetic regulation of stress memory and discuss the application of new technologies in developing epigenetic modifications to improve stress memory.

KEYWORDS

epigenetics, DNA methylation, chromatin remodelling, histone modifications, stress memory, somatic memory, transgenerational memory, intergenerational memory

Introduction

Extreme temperatures, such as cold and heat in the wake of climate change, significantly impact plant productivity by affecting gene expression. Therefore, plants memorize a stressful experience for a certain period of time and use their stress memory, such as transcriptional, somatic, intergenerational and transgenerational stress memory to increase their survival, when they are exposed to such conditions a second time (Xie et al., 2021; Brenya et al., 2022; Gallusci et al., 2022). Since plants do not have a brain, plant stress memory is completely different from human and animal memory. It is regulated by epigenetic changes and plays an important role in gene regulation by sensing and responding to cold and heat stress. This leads to the ability to develop appropriate physiological and molecular changes. In addition, epigenetically regulated stress memories are inherited through the next generations (Friedrich et al., 2019; Nguyen et al., 2022; Saeed et al., 2022; Sharma et al., 2022).

Salicylic acid and the downstream signalling protein are involved in such a memory process (Wu et al., 2012). During this process, the transcription of salicylic acid-responsive genes is activated. Elevated concentrations of salicylic acid result in changes in chromatin modification for these target genes. For example, salicylic acid-induced H3 acetylation, H4 acetylation and H3K4 methylation were increased at the promoter regions of stress-related genes (Butterbrodt et al., 2006), but chromatin modifications are dynamic when they are exposed to stress (Bhadouriya et al., 2021). Thus, changes in histone modifications either positively or negatively influence the expression of stress memory genes (Liu et al., 2022a). Constitutive photomorphogenesis 5A (CSN5A) regulates heat stress memory by increasing H3K4me3 in memory genes, *APX2* and *HSP22*, after heat stress (Singh et al., 2021). H3K4me3 also regulates transcriptional memory of stress-related genes such as *Fa-heat shock protein 17.8 Class II (FaHSP17.8-CII)* (Bi et al., 2021). Chromatin protein BRUSHY1 (BRU1) is required to maintain the transcriptional induction of heat stress memory genes (Brzezinka et al., 2019). In contrast, the changes of H3K27me3 of *Flowering Locus C* (*FLC*) gene maintain cold stress memory (Xie et al., 2021). This suggests that chromatin structure and accessibility play an important role in gene expression and are often associated with epigenetic regulations (epigenetic code) such as histone variants, histone post-translational modifications (PTMs), DNA methylation and some non-coding RNAs (Duan et al., 2018; Chang et al., 2020; Joseph and Young, 2022). In addition, histone modification and replacement of histones by histone variants are responsible for the inheritance of stress memory (Borg et al., 2021).

Nevertheless, plants have their limitations in adapting to stress conditions (Mahawar and Shekhawat, 2018; Khator et al., 2020), and it is still unclear to what extent these modifications contribute to the stability of systemic acquired stress resistance in terms of improved stress memory. Therefore, identifying epigenetic codes of plant stress memory responsible for stress responses is of great importance to developing stress-tolerant crops. However, the integrated view of epigenetic changes and their associations with stress memory and subsequent gene expression in plants is still scarce, except for a few reviews (Jacques et al., 2021; Xie et al., 2021; Gallusci et al., 2022; Liu et al., 2022a; Sharma et al., 2022). These reviews provided a general overview of stress memory during abiotic stress and raised many questions, such as how epigenetic stress memory allows plants to integrate information from previous stress conditions, how epimutations modulate stress memory, how memory types are maintained, etc. However, none focused exclusively on memory during cold and heat stress. In this review, we, therefore, discuss the histone modification and DNA methylation regulating chromatin structure and cold and heat stress memories. In addition, we focus on the application of emerging technologies in epigenetic modification and stress memory for a better future.

Histone and DNA modifications

Modified by various PTMs such as acetylation, methylation, phosphorylation, etc., histone proteins regulate gene expression, DNA damage repair, DNA replication, and recombination

(Han et al., 2022; Millán-Zambrano et al., 2022). Transcriptional activation or repression depends on which histone residues are modified and the type of PTMs (Sirko et al., 2022). For example, the methylation of lysine 4 and lysine 36 on histone H3 is associated with gene activation, while the methylation of lysine 9 and lysine 27 is associated with gene silencing. However, the histone code hypothesis states that different histone modifications in one or more tails sequentially or in combination become a histone code that can be read by other proteins to make different changes in gene expression (Millán-Zambrano et al., 2022). The histone code is responsible for storing plant epigenetic memory (Borg et al., 2020). Histone codes with specific combinations of histone variants and PTMs of the N-tail of histone proteins provide a large number of possibilities for nucleosome composition (Brewis et al., 2021), thereby playing an important role in regulating the expression of the genetic code. Similar to histone modification, DNA methylation, a conserved heritable epigenetic mark, influences nuclear gene expression and genome stability, thus regulating many biological processes. Moreover, changes in DNA methylation led to abnormalities in plant development (Lang et al., 2017). In this review, we discuss the epigenetic signatures that regulate plant stress memory under cold and heat stress conditions.

Regulation of cold stress in plants

Chilling (0 to 15°C) and freezing (<0°C) are two distinct types of cold stresses, which decrease plant growth by inducing photosynthesis-associated damages, chlorosis, unregulated apoptosis, loss of cell membrane and protein fluidity etc. (Banerjee et al., 2017). Several plants increase freezing tolerance (cold acclimation) when they are exposed to low non-freezing temperatures (Adhikari et al., 2022). However, several important crops like rice, maize, soybean, cotton and tomato are sensitive to cold stress, and they are incapable of acclimatising to cold stress, when ice forms in their tissue. In contrast, many plant species require exposure to prolonged cold periods so that they can flower during the next spring under favourable conditions. This is called vernalization (Antoniou-Kourounioti et al., 2021). Plant cold acclimatisation gradually decreases after vernalization, one of the best examples of plant epigenetic regulation to adapt to environmental cues (Luo and He, 2020; Kang et al., 2022). Genetic and molecular analyses have shown that epigenetic marks play a major role in vernalization, whose response is based on an epigenetic memory of plants (Sharma et al., 2022). Histone modifications and DNA methylation are the major epigenetic marks of plants that can be identified during cold stress (Banerjee et al., 2017), but it is still unknown whether the activation of epigenetic marks and stress memories differ in response to chilling and freezing.

Histone modifications during cold stress

Studies have revealed that histone modifications, such as methylation and acetylation, change under cold stress and target different molecular mechanisms, including cold-responsive genes that can be genetically manipulated (Bhadouriya et al., 2021; Kidokoro et al., 2022). Histone demethylases, *Jumonji C domain-containing* genes (*JMJC5*) regulate demethylation at the lysine or arginine residues of the histone and respond to circadian clock. *JMJC5* in *Medicago truncatula* was observed with the cold-dependent alternative splicing, which was reversed when plants returned to normal conditions, suggesting that the cold-dependent alternative splicing is associated with the epigenetic regulation of cold stress (Shen et al., 2016). Genomic regions containing active *cis*-acting elements, identified as DNase I hypersensitive sites (DHSs) in potatoes using chromatin analysis, play significant roles in regulating the cold stress response. Cold-stress induced DHSs are enriched in genic regions and regulate the expression of cold-responsive genes associated with bivalent histone modifications, H3K4me3 and H3K27me3 (Zeng et al., 2019).

In the regions of stress memory genes, both H3K4me3 and H3K27me3 were enriched, suggesting that cold-responsive genes increase chromatin accessibility under cold stress (Liu et al., 2014). In *Arabidopsis*, H3K27me3 decreases in two cold-responsive genes *COR15A* and *ATGOLS3* and gradually promotes the expression of these genes under increasing cold stress. When cold-exposed plants returned to normal conditions, the plants retained decreased H3K27me3 in *COR15A* and *ATGOLS3* and or histone replacement. Nevertheless, the expression of these two genes was not increased, suggesting that H3K27me3 does not inhibit gene activation but that H3K27me3 was inherited through cell division to serve as a memory marker for cold stress (Kwon et al., 2009). Polycomb repressive complex 2 (PRC2) increases H3K27me3 levels to suppress *FLC* activation during vernalization and maintain cold stress memory in chromatin regions in *Arabidopsis*, intimating that different stages of polycomb silencing are required to maintain epigenetic cold memory in *Arabidopsis* (Yang et al., 2017).

Histone acetylation normally promotes gene expression. Under cold stress, histone H3 acetylation promotes the activation of *COR* genes, and histone acetyltransferase HAC1 is necessary for a stress memory and enhances cold memory (Friedrich et al., 2019). Cold exposure also promotes histone deacetylation. *Zea mays* was

observed with the upregulation of histone deacetylases (HDACs) and led to deacetylation at the lysine residues on the histone subunits H3 and H4 during cold acclimatisation. Thus, HDACs' repression reduces the expression of cold-responsive genes *ZmDREB1* and *ZmCOR413* in *Zea mays* (Hu et al., 2011). In HDACs, three major families have been identified such as reduced potassium dependency protein 3 superfamily (RPD3/HDA1), silent information regulator protein 2 family (SIR2) and histone deacetylase 2 families (HD2) (Pandey et al., 2002).

In rice, 19 HDAC genes have been identified (Hu et al., 2009), and the expression of *OsHDA714*, *OsSRT701*, and *OsSRT702* is modulated by cold stress (Fu et al., 2007). *AtHD2D* was induced in *Arabidopsis* subjected to cold stress, resulting in slowed seed germination and flowering but a slow increase in malondialdehyde content; thus, transgenic plants are capable of greater cold tolerance (Han et al., 2016). In *Arabidopsis*, the upregulation of HDA6 promotes freezing tolerance, and *hda6* mutants reduced cold tolerance by affecting hundreds of cold-responsive genes, divulging a key function of HDA6 in cold response (To et al., 2011). *High Expression of Osmotically Responsive gene 15* (*HOS15*) involved in histone deacetylation is also necessary for cold response. The *hos15* mutant increased the acetylation of histone H4, and HOS15 interacts with HD2C and *C-REPEAT-BINDING FACTOR* (*CBF*) transcription factors (Figure 1). The degradation *HOS15* and HD2C (a partner of *HOS15*), mediated by the CULLIN4 (CUL4)-based E3 ubiquitin ligase complex, triggers histone acetylation and *CORs* genes, *COR15A* and *COR47*, revealing the importance of HDACs and *HOS15*-mediated histone modifications in cold stress memory (Zhu et al., 2008; Park et al., 2018). However, further focused research on crop plants using advanced high-throughput sequencing technologies such as chromatin immunoprecipitation-sequencing (ChIP-seq) targeting histone modifications could lead to plant development with enhanced stress memory capable of tolerating cold stress.

DNA methylation during cold stress in plants

Several studies reveal that DNA methylations, such as 5-methylcytosine (5mC) and N^6-methyladenine (6mA), regulate transcriptional activities in response to cold stress (Zhang et al., 2018; Sicilia et al., 2020; Kidokoro et al., 2022; Verma et al., 2022). Compared with 6mA, a new epigenetic marker, 5mC is the most-studied modification under cold stress. Although 6mA is widely distributed across *Arabidopsis* and rice genomes and associated with gene expression and stress responses (Zhang et al., 2018; Liu and He, 2020), 6mA functions in stress memory remain elusive. Therefore, we limit this review to the discussion of 5mC. Plants are able to transmit cold-stress-induced DNA methylation changes within the generation or as transgenerational epigenetic memories, allowing plants to respond to cold stress a second time (Sharma et al., 2022).

DNA methylation variations in response to cold acclimation were detected in *Chorispora bungeana* (Song et al., 2015), *Populus simonii* (Song et al., 2016) and *Phyllostachys edulis* (Ding et al., 2022). Cold acclimation induces both DNA methylation and demethylation; and demethylation regulates freezing tolerance in *Arabidopsis*. In cucumbers, cold stress impacts RNA-directed DNA methylation (RdDM), resulting in demethylation. However, the lowered activity of *MSH1* and *RNA-DIRECTED DNA METHYLATION 4* (*RDM4*), a key player in RdDM pathway, reduced cold response in *Arabidopsis*. *RDM4* modulates the expression of *CBF* TFs, which regulates cold stress memory (Liu and He, 2020; Xie et al., 2021). Cold stress reduced 5mC in vernalized plants and increased the expression *CASEIN KINASE II A-SUBUNIT* (*CKA2*) and *B-SUBUNIT* (*CKB4*), which elevated CK2 activity, resulting in the reduction of the clock gene *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*), an important player in photoperiod perception (Duan et al., 2017). Vernalization-induced DNA demethylation is not conserved in plants, and vernalization also influences non-CG methylation that can be maintained *via* mitosis (Liu and He, 2020).

5mC modulates ICE1-CBF-COR pathway associated with cold tolerance. In *Hevea brasiliensis*, the expression of *HbICE1* and *HbCBF2* was associated with DNA demethylation and cold tolerance (Tang et al., 2018). In *Arabidopsis*, the 5mC variation in *INTERACTOR OF LITTLE ELONGATION COMPLEX ELL SUBUNIT 1* (*ICE1*) and *ALLANTOINASE (ALN)* regulated freezing tolerance and seed dormancy and stimulated non-canonical RDR6 and ARGONAUTE 6 (AGO6)-dependent RdDM pathway. DNA hypermethylation of *AtICE1* resulted in the suppression of the genes involved in CBF pathway (Iwasaki et al., 2019; Xie et al., 2019). In *Brassica rapa*, cold-induced DNA demethylation promoted the expression of *MITOCHONDRIAL MALATE DEHYDROGENASE* (*BramMDH1*), whose overexpression increased heat tolerance in *Arabidopsis* but not in *Brassica rapa.* This indicates that DNA demethylation alone is not sufficient to increase heat tolerance in *Brassica rapa* but is sufficient to increase cold tolerance (Liu et al., 2017a). In *Rosa hybrida*, cold-induced DNA methylation (CHH context) reduced the expression of *RhAG*, whose improved suppression increased cold-induced petal number (Ma et al., 2015). In addition, reduced DNA and histone methylations (H3K9me2) and increased histone acetylation (H3K9ac) in heterochromatic regions regulate cold tolerance, suggesting that changes in both DNA and histones are necessary for cold acclimation (Kim et al., 2015). Like 5mC, the 6mA level is also reduced in response to cold stress (Zhang et al., 2018), but how both 5mC and 6mA functions connected in freezing tolerance remain largely obscure. Therefore, the characterization of cold stress-responsive epigenetic modifications is crucial for developing plants with enhanced cold stress memory.

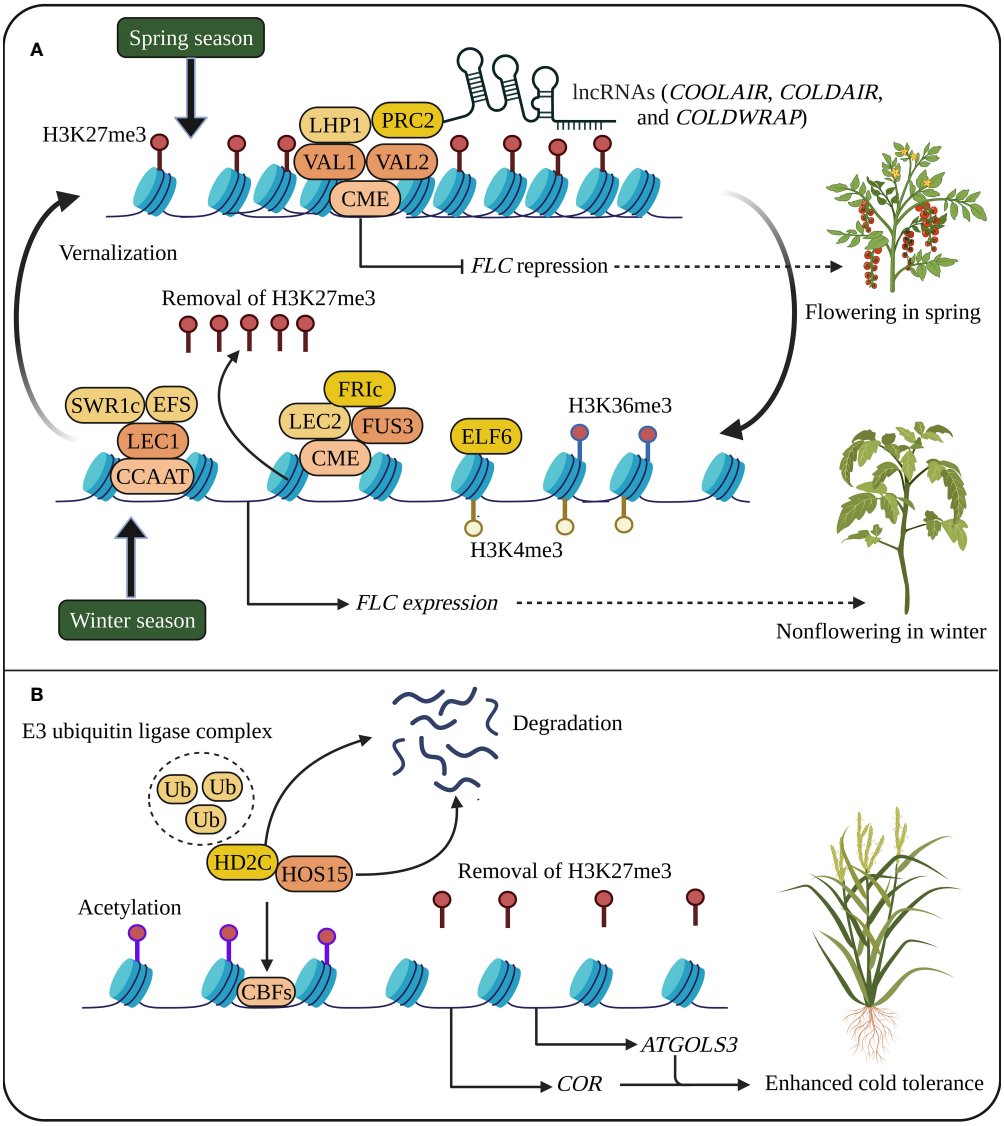


FIGURE 1
Cold stress memory in plants. **(A)** *Flowering Locus C* (*FLC*) expression suppresses flowering at high levels during winter. During vernalization, cold memory *cis*-acting element (CME), B3 transcription factors (TFs) such as *Viviparous1* (*VAL1* and *VAL2*), LIKE HETEROCHROMATIN PROTEIN 1 (LHP1), Polycomb repressive complex 2 (PRC2), and long noncoding RNAs (lncRNAs) silence *FLC* activation, and the spread of H3K27me3 throughout the *FLC* locus during cell division and DNA replication leads to *FLC* repression. In the next generation, epigenetic cold memory is reactivated during winter by the reactivation of *FLC* through TFs such as *Leafy Cotyledon 1* (*LEC1*), *LEC2* and *FUSCA3* (*FUS3*), *Early Flowering 6* (*ELF6*), *FRIGIDA* (*FRI*), *Early Flowering in short days* (*EFS*) and chromatin remodeling complex/PIE1 complex (SWR1c). **(B)** Cold stress memory regulates the expression of *Cold-regulated genes* (*CORs*) and *C-REPEAT-BINDING FACTORs* (*CBFs*) through histone methylation and acetylation. Under cold conditions, the CULLIN4 (CUL4)-based E3 ubiquitin ligase complex (Ub) degrades HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE 15 (HOS15) and HISTONE DEACETYLASE 2C (HD2C), resulting in the activation of histone acetylation and *CORs* genes (Park et al., 2018). The reduced H3K27me3 also promotes the expression of *CORs* and *ATGOLS3*. The schematic representation was adapted from (Xie et al., 2021) by adding additional information and created using BioRender.com.

Genes involved in cold acclimation in plants

Low temperatures cause membrane stiffening leading to calcium influx into cells and triggering signalling cascades regulated by multiple genes that lead to cold acclimation, suggesting that cold stress is sensed by plant cells through changes in the fluid state of the plasma membrane (Ma et al., 2015; Ding et al., 2019). Cold stress induces the expression of many transcription factors, such as *CBF*. The cold-responsive CBF signalling pathway is one of the known characterised mechanisms for transcriptional cold memory (Hwarari et al.,

2022). In *Arabidopsis*, CBF proteins bind to the *cis*-acting element DRE/CRT present in the promoter regions of the cold-responsive genes such as *COR* genes and activate their expression (Figure 1) (Adhikari et al., 2022). Cold stress rapidly induces *CBF1, CBF2* and *CBF3* genes along with *COR* gene expression. The overexpression of *CBFs* genes in plants increased cold tolerance and plant growth and development (late flowering) (Guo et al., 2018).

Depending on cold-induced epigenetic changes, the *CBF* genes are either positively or negatively regulated by other TFs or genes. For example, the mutant *cbf2* in *Arabidopsis* increased freezing tolerance and cold acclimation, but negatively regulated *CBF1* and *CBF3* (Novillo et al., 2004). COLD-RESPONSIVE PROTEIN KINASE 1 (CRPK1) is a cold-activated plasma membrane-localized protein and a negative regulator of CBF pathway. CRPK1 phosphorylates 14-3-3 proteins which interact with and destabilize CBF proteins in the nucleus, thus playing a negative role in cold acclimation (Liu et al., 2017c). Whereas the expression of *CBFs* is positively regulated by *ICE1* and *ICE2* (bHLH TFs), thereby maintaining the activation of *CBF/COR* genes (Guo et al., 2018). To degrade *ICE1* and to suppress the *ICE1*-mediated expression of *CBF/COR* genes, *ICE1* is ubiquitinated by *HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENES1* (*HOS1*) under cold stress. However, SIZ1 and SIZ2 modulate *ICE1* SUMOylation to suppress *ICE1* degradation, thus activating *CBF/COR* genes to increase cold memory and tolerance (Guo et al., 2018). The HOS15-HD2C complex promotes hypoacetylation of *COR* chromatin to inhibit *COR* genes, but the CUL4-based E3 ubiquitin ligase complex degrades HD2C and induces hyperacetylation of *COR* genes, which enhances the binding activity of CBFs to the promoters (Park et al., 2018). Therefore, the *CBF*-dependent gene expression through epigenetic regulation is a significant component of cold memory and acclimation in plants. However, in *Arabidopsis*, the transcriptome analysis showed that only 12% of the cold-responsive genes were regulated by *CBF* (CBF regulon/CBF-targeted genes), and no less than, 28% of the cold-responsive genes were not regulated by *CBF*, indicating that cold-responsive genes and cold memory genes are regulated by different TFs (Fowler and Thomashow, 2002).

NON-EXPRESSER OF PATHOGENESIS-RELATED GENES 1 (*NPR1*) gene is known to be the master regulator of plant pathogen response. *NPR1* is epigenetically regulated; indeed, the histone methyltransferase SET DOMAIN GROUP 8 (SDG8) binds it and induces upon H3K36 methylation (H3K36me3) its transcription upon pathogen infection (Zhang et al., 2020). However, recent genetic and molecular evidence has shown that *NPR1* plays an essential role in cold acclimation by independently regulating cold-induced gene expression of salicylic acid and TGA TFs (Olate et al., 2018). Also, the fact that *NPR1* might serve as a central hub integrating cold and pathogen signalling for a better adaptation is very interesting. Therefore, it seems important to insist that stress responses (biotic/abiotic) are intimately connected. In addition, understanding different signalling pathways in cold stress memory is required to improve cold stress response.

Regulation of heat stress in plants

Heat stress can cause cellular damage and cell death and produces an excessive amount of reactive oxygen species (ROS) and oxidative stress, resulting in morphological, physiological and biochemical changes, or in leaf shedding, flower and fruit abortion or even plant death (Hasanuzzaman et al., 2013). Heat-induced regulated cell death (RCD) such as apoptosis, necrosis and ferroptosis pathways selectively removes certain cells in specific tissues to maintain homeostasis under heat stress (Distéfano et al., 2021). Plant response to heat stress is a complex process that differs among species depending on the temperature range (28-48°C) and is regulated by different mechanisms such as hormone signalling, transcriptional process and epigenetic changes. The histone proteins H1, H2A, H2B and H3 with the exception of H4 have specialized histone variants that reprogram the gene expression responsible for heat acclimation and heat stress memory. Several key genes and networks regulating the heat response have been identified, and heat stress memory is associated with epigenetic changes (Li et al., 2021; Malik and Zhao, 2022). In contrast to hormonal regulation underlying heat stress at the transcriptional level, epigenetic regulation under stress in plants remains elusive, suggesting that identifying key mechanisms is important. The next sections will briefly discuss the epigenetic mechanisms and genes involved in heat stress response and memory.

Histone modifications during heat stress

Heat acclimation depends on a specific pattern of gene expression regulated by histone modifications, dynamic in response to heat stress and associated with gene transcription (Pecinka et al., 2010; Bhadouriya et al., 2021). Some histone modifications change rapidly under stress conditions, while others adapt gradually to the response (Kim et al., 2015). Histone H2A.Z, a variant of H2A comprising H2A, H2A.Bbd, H2A.X and H2A.Z variants, is involved in temperature sensing globally through their nucleosome occupancy. In *Arabidopsis*, the insertion of the histone variant H2A.Z into nucleosomes is regulated by a subunit of the chromatin remodeling complex

(SWR1), replacing H2A with H2A.Z in an ATP-dependent manner. *ACTIN-RELATED PROTEIN6* (*ARP6*) gene encodes SWR1, suggesting that *ARP6* is required for histone modification and mediates temperature response genes (Kumar and Wigge, 2010). In *Brachypodium*, the *H2A.Z* mutant decreased heat acclimation (Boden et al., 2013). Histone acetylation is necessary for H2A.Z deposition. This suggests that other histone variants and H2A.Z-containing nucleosomes modulate transcription in a temperature-dependent manner (Kim et al., 2015). Heat stress increases histone acetylation (H3K9ac and H4K5ac) in different tissues of maize and either does not alter or decreases histone methylation, divulging a key role of histone acetylation and methylation in chromatin decondensation (Wang et al., 2015). In *Arabidopsis*, the chromatin protein BRUSHY1 (BRU1) involved in chromatin organization facilitates the epigenetic inheritance of chromatin states and maintains the transcriptional induction of heat stress memory genes (Brzezinka et al., 2019).

The down-regulation of histone acetyltransferase GCN5 reduces heat tolerance. Heat tolerance of *Arabidopsis* depends on the expression of stress-related genes mediated by GCN5. Induction or overexpression of stress-related genes in the *gcn5* mutant increased stress memory and restored stress tolerance (Hu et al., 2019). HDACs such as HD2C, HDA6, HDA9, HDA15, and HDA19 regulate stress responses either negatively or positively. For example, in *Arabidopsis*, the interaction of HDA9 with POWERDRESS (PWR), a SANT-domain containing protein, increases heat tolerance (Tasset et al., 2018). In contrast, HDA15 decreases heat tolerance while interacting with TF, long Hypocotyl in Far Red1 (Shen et al., 2019).

Histone methylations regulate heat stress responses and memory. In *Arabidopsis*, H3K4me3 correlates with active gene expression, whereas the correlation of H3K4me2 with active transcription is negative in rice. A decrease in H3K4me2/H3K4me3 levels misregulated many genes related to plant development. Expression of several memory genes such as *HSP18.2*, *HSP21*, and *HSP22.0* correlates with H3K4me3 and H3K4me2 at their respective target loci, and higher accumulation of these modifications is associated with hyper induction of memory genes under continuous heat stress (Lämke et al., 2016; Bäurle, 2018; Liu et al., 2019). Like the regulation of cold acclimation by reduced H3K27me3, heat tolerance is also regulated by reduced H3K27me3 at *HSP17.6C* and *HSP22* genes mediated by JMJC histone demethylases, thus increasing heat stress memory in *Arabidopsis*. In addition, the removal of H3K27me3 by JMJC (Figure 2) is to maintain the accumulation of H3K4me3 at *HSP21* gene during heat response, suggesting that the balance between H3K27me3 and H3K4me3 levels is required for stress memory (Yamaguchi and Ito, 2021). Reduced H3K9me2 attenuates heat responses and is an important component in the regulation of *Fertilization-Independent Endosperm1* (*OsFIE1*), a member of PRC2, in *Arabidopsis* seed development under moderate heat stress (Folsom et al., 2014).

Histone SUMOylation contributes to the transcriptional regulation mediated by histone modifier HDACs and has conserved functions in heat acclimation. In *Arabidopsis*, heat stress decreased histone SUMOylation (H2B) and increased GCN5, and a variety of histone modifiers are involved in the regulation of SUMOylation (Kim et al., 2015). However, the modifications and their association with stress memory genes remain elusive and require considerable attention to investigate their functions in stress memory.

DNA methylation during heat stress in plants

Although studies show that DNA methylation (either hypomethylation or hypermethylation) regulates transcriptional activities in response to heat stress (Ding et al., 2022; Perrella et al., 2022; Sun et al., 2022), the relationship between DNA methylation and heat stress memory is still largely unclear. DNA methylation mediated by the RdDM pathway is known to regulate several heat stress-responsive genes, and this pathway is required for basal heat tolerance and the transition of transgenerational memory in *Arabidopsis*. Furthermore, *Arabidopsis* mutants with epigenetic defects showed that the transcriptional heat response depends on the RdDM pathway and HDA6 (Popova et al., 2013; Yang et al., 2020). In *Brassica napus*, DNA methylation changes between heat-tolerant and heat-sensitive genotypes, and the heat-sensitive genotypes displayed hypermethylation, suggesting that the level of DNA methylation is dynamic in response to heat stress (Gao et al., 2014). Heat stress induced active demethylation in *Arabidopsis*, especially after returning to normal temperature, and differentially methylated genes are associated with heat response (Korotko et al., 2021). Demethylation facilitates intergenerational stress memory and differentially methylated regions (DMRs) enriched with transposons were linked with the stress memory dependent on DNA methylation (Wibowo et al., 2016).

Heat-accelerated DNA methylation also regulates transposon silencing or activation to ameliorate stress memory. For instance, in *Arabidopsis* under heat stress, CHROMOMETHYLASE3 (CMT3), a DNA methyltransferase, activated retrotransposon *ONSEN* that is capable of binding heat shock (HS) TFs involved in transcriptional memory and is transcriptionally activated and increased their copies, which are involved in transgenerational stress memory. The *cmt3* mutants increased methylation at *ONSEN*, but CMT2-bound *ONSEN* chromatin in *cmt3* mutants accumulated H3K9me2, suggesting a collective role of transposon functions connected with DNA and histone modifications under stress conditions (Ito et al., 2011; Cavrak et al., 2014). In addition, retrotransposons maintained by RdDM suppress chromatin modifications and perform a role in stress memory. The transpositions of retrotransposon occur more frequently in the

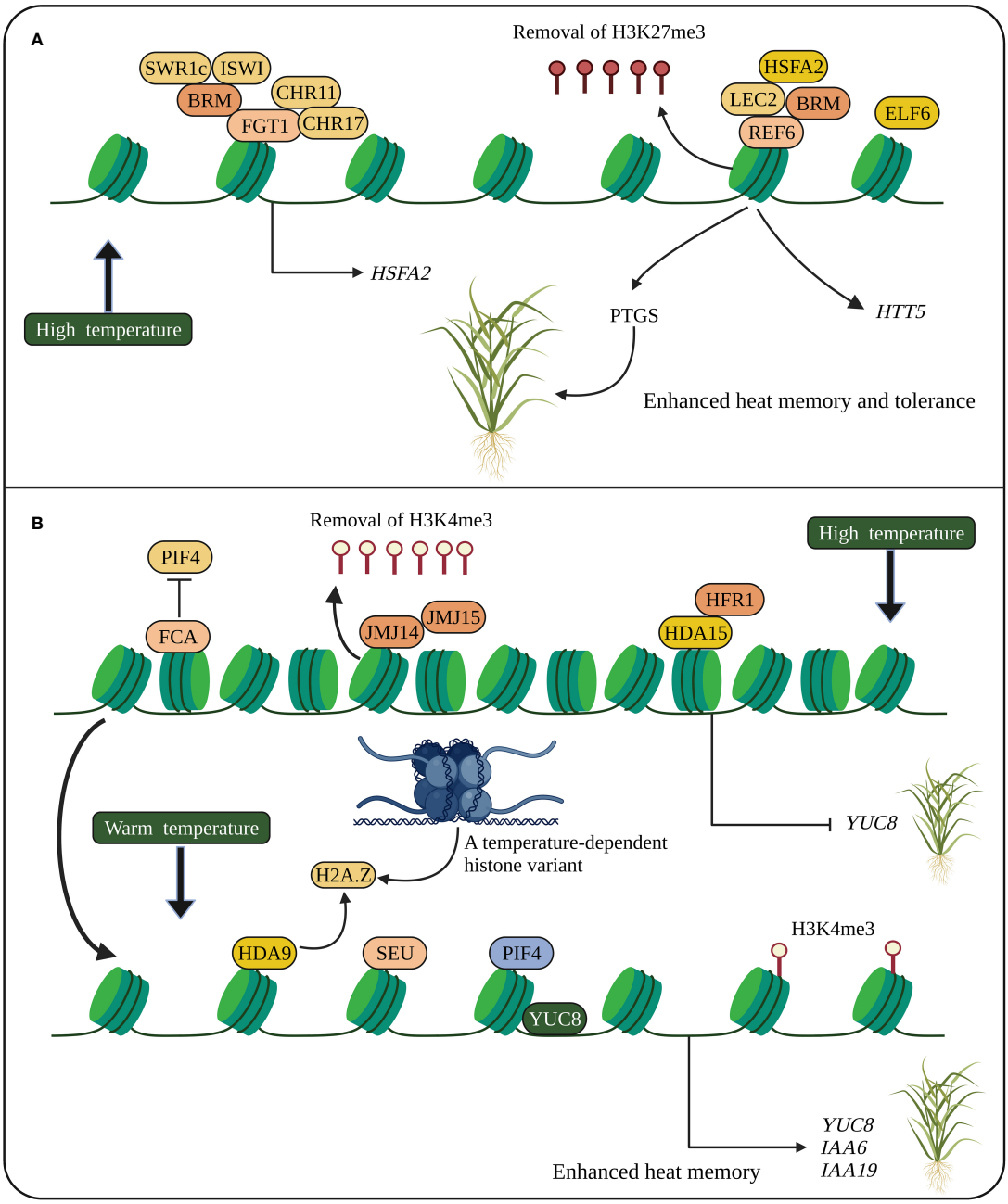


FIGURE 2
Heat stress memory in plants. **(A)** Heat memory is regulated by HSFA2-REF6-PTGS and FGT1. HSFA2 and REF6 are essential for heat memory and trigger the SGS3-interacting protein, which degrades SGS3 and inhibits PTGS. REF6 removes H3K27me3, and FGT1 and the chromatin remodelers such as BRM, CHR11 and CHR17 modulate nucleosome occupancy at the loci of heat memory genes. **(B)** Stress memory (thermomorphogenesis) is also regulated by PIF4. Epigenetic modifications and chromatin remodelers activate PIF4 to promote stress memory genes. A temperature-dependent histone variant (H2A.Z) modulating gene network is associated with YUC8 to promote PIF4 activation. HSFA2, heat shock transcription factor; *REF6*, *RELATIVE OF EARLY FLOWERING 6*; PTGS, posttranscriptional gene silencing; *SGS3*, *SUPPRESSOR OF GENE SILENCING 3*; *FGT1*, *FORGETTER1*; BRM, BRAHMA; CHR11 and CHR17, CHROMATIN-REMODELING PROTEIN 11 and 17, respectively; JMJC14 and JMJC15, Jumonji C domain-containing protein 14 and 15, respectively; *PIF4*, *PHYTOCHROME INTERACTING FACTOR 4*; *FCA*, *FLOWERING CONTROL LOCUS A*; *SEU*, *SEUSS*; *YUC8*, *YUCCC8*; *IAA*, *INDOLE-3-ACETIC ACID INDUCIBLE*. The schematic representation was adapted from (Xie et al., 2021) by adding additional information and created using BioRender.com.

heat-stressed progeny of RdDM mutants, implying that the RdDM machinery plays an important role in retrotransposon silencing and prevents the transgenerational memory of retrotransposons in *Arabidopsis* (Matsunaga et al., 2012; Sharma et al., 2022). However, the HS *cis*-acting element in the retrotransposon's promoter region can bind HS TFs and enable the transcription under heat stress conditions to enhance stress memory (Cavrak et al., 2014).

In maize, heat-induced demethylation is associated with the spliceosome pathway, suggesting that DNA demethylation regulates heat responses *via* RNA splicing (Qian et al., 2019). The interaction of spliceosome complex proteins with nuclear cyclophilins (CYPs) associated with spliceosomal components is required for heat acclimation in *Arabidopsis* (Jo et al., 2022). In addition, heat-induced genome-wide hypomethylation disrupted the metabolic pathways related to sugars and reactive oxygen species (ROS) (Liu and He, 2020). In addition, heat stress inhibited cell division in the tobacco cell line (BY-2 cells) and arrested cells in the pre-mitotic phase (somatic cells) by increasing the expression of *CycD3-1* and decreasing the transcripts of *NtCycA13*, *NtCyc29* and *CDKB1-1*. This was due to the altered expression of *CycD3-1* and *Nt-EXPA5* in conjunction with the methylation status of their promoters, suggesting that DNA methylation regulates cell cycle progression to control mitosis involved in somatic stress memory (Centomani et al., 2015).

Heat-induced DNA methylation in *Arabidopsis* affected two genes *At3g50770* (calmodulin-like 41(CML41) and *At5g43260* (Chaperon-like protein) containing transposon insertions in their promoters, and *At3g50770* expression is correlated with its promoter methylation. Heat-induced *NUCLEAR RNA POLYMERASE D1A* (NRPD1A) and *NRPD1B* upregulate *CML41* under heat stress. This suggests that the RNA polymerases PolIV and PolV (NRPD1 and NRPE1 respectively), main players in DNA methylation, may regulate transcripts not only through RdDM but also through other mechanisms (Naydenov et al., 2015). *Arabidopsis* mutants lacking DNA methyltransferase, such as *domains rearranged methylase1* (*drm1*), *drm2* and *cmt3*, showed less distinct-heat response; however, the prolonged-heat stress downregulated *METHYLASE1* (*MET1*) and *CMT3*, confirming that DNA methylation is dynamic under heat stress (Naydenov et al., 2015). However, the role of DNA methylation and its interaction with stress memory is poorly understood and requires considerable attention to investigate the role of DNA methylation in heat stress epigenetic memory.

Genes involved in heat tolerance

Several candidate genes and TFs associated with heat tolerance and stress memory have been found in many plant species. As molecular chaperones, heat shock proteins (HSPs) encoded by HSP family genes protect cells from heat stress through various strategies, including targeted protein localization. HSPs are classified into different families such as HSP10, HSP20, HSP40, HSP60, HSP70, HSP90 and HSP100 (Hasanuzzaman et al., 2013). HSPs are regulated by heat shock TFs (HSFs) interacting with heat-induced TFs and enzymes (Samakovli et al., 2020; Bourgine and Guihur, 2021). The heat shock response (HSR) and heat-responsive cyclic nucleotide-gated ion channels (CNGCs) trigger heat-induced gene expression to accumulate HSPs (Guihur et al., 2022).

Genes such as *HSP17.6C*, *HSP21* and *HSP22*, members of the small HSPs (sHSPs), are key factors of heat stress memory and associated with histone methylation (H3K4me3 and H3K27me3). *JMJ* genes balance H3K4me3 and H3K27me3 at *HSP21* during heat acclimation and activate *HSP* genes when sensing heat stress, whereas *jmj* mutants do not maintain heat memory (Yamaguchi and Ito, 2021; Yamaguchi et al., 2021). In *Arabidopsis*, the enzyme Filamentation Temperature-Sensitive H6 (FtsH6) degrades or resets HSP21 abundance during heat stress recovery. *HSFA2*, a positive regulator of stress memory and a member of the family of 21 heat stress TFs, activates *FTSH6* and improves stress memory in *ftsh6* mutants compared with wild-type plants (Sedaghatmehr et al., 2022).

The overexpression of *HSFA2* in *Arabidopsis* activates downstream targets *via* heat stress elements (HSEs) and induces the expression of *HSP21*, *HSP22*, *HSP18.2*, and *ASCORBATE PEROXIDASE 2* in response to heat stimuli (Yamaguchi, 2022). *JMJ*-and *HSFA2*-mediated histone modifications occur at the same *HSP22* locus. The mutants such as *hsp21, hsp22, hsp17.6c*, *heat-stress-associated 32* (*hsa32*) and *hsfa2* decreased heat acclimation, whereas the overexpression of these genes improved heat acclimation and memory (Yamaguchi, 2022), intimating that heat memory genes are *JMJ* and *HSFA2* dependent. Unlike the activation of *HSFA2*, the repression of *SQUAMOSA promoter-binding protein-like* (*SPL*) TF triggers heat memory genes (Yamaguchi, 2022). *HSFA1* and *HSFA2* TFs circuitously regulate *AGAMOUS LIKE 16* (*AGL16*), encoding a MADS-box TF and acting as a post-transcriptional memory factor (Xie et al., 2021). These results elucidate how heat memory ensures heat acclimation.

HEAT-INDUCED TAS1 TARGET (*HTT*) genes confer heat tolerance through the action of HSPs such as Hsp70-14, Hsp40, NUCLEAR FACTOR Y and SUBUNIT C2. *HSFA1a* TF binds directly to the HSE in the *HTT* promoter and can induce heat tolerance in *Arabidopsis* (Li et al., 2014). *FORGETTER1* (*FGT1*), encoding a plant homeodomain (PHD) finger protein, physically interacts with SWI2/SNF2 chromatin remodelers (BRAHMA (BRM) and CHROMATIN-REMODELING PROTEIN 11 (CHR11) and CHR17) to enhance heat stress memory by promoting heat-responsive genes. *FGT1* regulates nucleosome dynamics at the gene loci of *HSP22*, *HSP18.2*, *HSP21* and *HSA32* after initial heat stress and mediates stress memory (Brzezinka et al., 2016). The *fgt1* mutant reduced *HSA32* expression, thereby

impairing heat acclimation. The mutants *brm and chr11 chr17* also showed heat memory defects (Yamaguchi, 2022). *FGT2*, encoding a TYPE-2C PROTEIN PHOSPHATASE (PP2C), interacts with PHOSPHOLIPASE D α2 (PLDα2), which activates lipid composition and stress memory. Like the *fgt1* mutant, the *fgt2* mutant also exhibited an increased heat memory defect, indicating that *FGT1* and *FGT2* are essential for heat memory (Xie et al., 2021).

Heat-induced post-transcriptional gene silencing (PTGS) shows transgenerational stress memory, and the *SUPPRESSOR OF GENE SILENCING 3* (*SGS3*) is required for PTGS. *HSFA2* and *RELATIVE OF EARLY FLOWERING 6* (*REF6*) regulate heat memory transmission. HSFA2-REF6 triggers SGS3-interacting protein, which inhibits PTGS by degrading SGS3 (Figure 2) (Xie et al., 2021). Small RNA-regulated *AGO1* involved in PTGS is also required for heat stress-induced transgenerational inheritance in *Brassica rapa* (Bilichak et al., 2015). Stress memory (thermomorphogenesis) is also regulated by the bHLH TF, *PHYTOCHROME INTERACTING FACTOR 4* (*PIF4*), induced at warm temperatures. The chromatin remodeler and epigenetically activated PIF4 bind directly to the *YUCCC8* (*YUC8*) promoter to promote stress memory at warm temperatures. The RNA-binding protein *FLOWERING CONTROL LOCUS A* (*FCA*) removes the occupancy of PIF4 from *YUC8* by interacting with a histone demethylase to mediate stress memory (Figure 2) (Xie et al., 2021). Altogether suggests that *HSPs*, *HSFs* and *SGS3* are key regulators of heat stress memory. However, how HSFs and JMJs interact with HSPs and regulate histone modifications remains to be elucidated.

Stress-induced epigenetic memory and its heritability

Several studies show that plants not only store their stress experiences but also retain them over multiple generations (Latzel et al., 2016; Liu et al., 2022b). In general, stress-induced epigenetic modifications return to baseline levels after the stress ends. However, some modifications remain stable and can be inherited as epigenetic stress memory over multiple generations through mitotic and meiotic cell divisions, while other modifications are reactivated during sexual reproduction or embryogenesis (D'urso and Brickner, 2014; Tricker, 2015). Epigenetic memory, without alteration of nucleotide sequences, requires distinct chromatin changes and impacts gene activation and related mechanisms on different time scales and under different environmental stresses. Cellular, transcriptional, and transgenerational memory are the three types of epigenetic memory that are dynamic and mitotically or meiotically heritable and triggered by an antecedent stress exposure. Cellular memory (somatic) is a mitotically heritable transcriptional state established during primary development with tissue-specific TFs and transcriptional programs. Transcriptional memory is also mitotically heritable changes for several generations in response to previous stress experiences. In contrast, transgenerational memory is meiotically heritable changes and can impact the offspring's behaviour (D'urso and Brickner, 2014).

Plants use transcriptional, somatic, intergenerational and transgenerational stress memory (Figure 3), and several epigenetic mechanisms regulating those memories have been identified (Bäurle and Trindade, 2020; Sharma et al., 2022). Somatic stress memory lasts only for a specific period of time within a generation (a few days or weeks). For example, vernalization memory is maintained for a few weeks by H3K27me3 of the *FLC* gene (Yang et al., 2017). PRC1 and PRC2 are required to keep somatic memory genes active by mediating histone methylation (H3K27me), which is a repressive bookmark during mitosis and is obligatory for the maintenance and inheritance of active chromatin states. This suggests that somatic memory necessitates both chromatin repressive and modifying complexes (D'urso and Brickner, 2014).

Transgenerational memory continues into the next stress-free generation, where epigenetic changes are independent (Shanker et al., 2020). The genome-wide distribution of DNA methylation and regulatory mechanisms involving chromatin modifications play an important role in the transgenerational inheritance of stress memory (Ou et al., 2012; Tricker, 2015; Cong et al., 2019). Intergenerational memory persists for at least two stress-free generations (Lämke and Bäurle, 2017; Oberkofler et al., 2021). Nevertheless, intergenerational memory is reversible after a stress-free generation, suggesting that this memory is stress-dependent. This memory is mainly maternally inherited. The DEMETER (DME) (DNA demethylase) suppresses paternal inheritance, but paternal inheritance is restored in *dme* mutants. The RdDM-mediated demethylation regulates intergenerational memory (Wibowo et al., 2016). However, how intergenerational and transgenerational memories are inherited remains unclear.

Histone H3K4me2 and H3K4me3 regulate acquired transcriptional memory in primed plants (Jaskiewicz et al., 2011), and the hypermethylation of these marks maintains transcriptional memory during heat stress (Lämke et al., 2016). Besides this, transcriptional memory also requires histone variant H2A.Z, and the mutants lacking this variant did not retain memory (D'urso and Brickner, 2014). Not all genes are stress-induced memory genes. Thus, memory genes have been classified into two types: 1, genes that show persistent activation or suppression after exposure to stress, and 2, genes that show altered response to recurrent stress compared to the response of naive plants (Bäurle, 2018). However, because plants use both male and female specific germline mechanisms (Ono and Kinoshita, 2021), stress-induced epigenetic memory is still unclear, although there are several examples of priming of plant defences and evidence of induced genetic rearrangements. Therefore, further studies are needed to clarify the involvement and persistence of epigenetic memory.

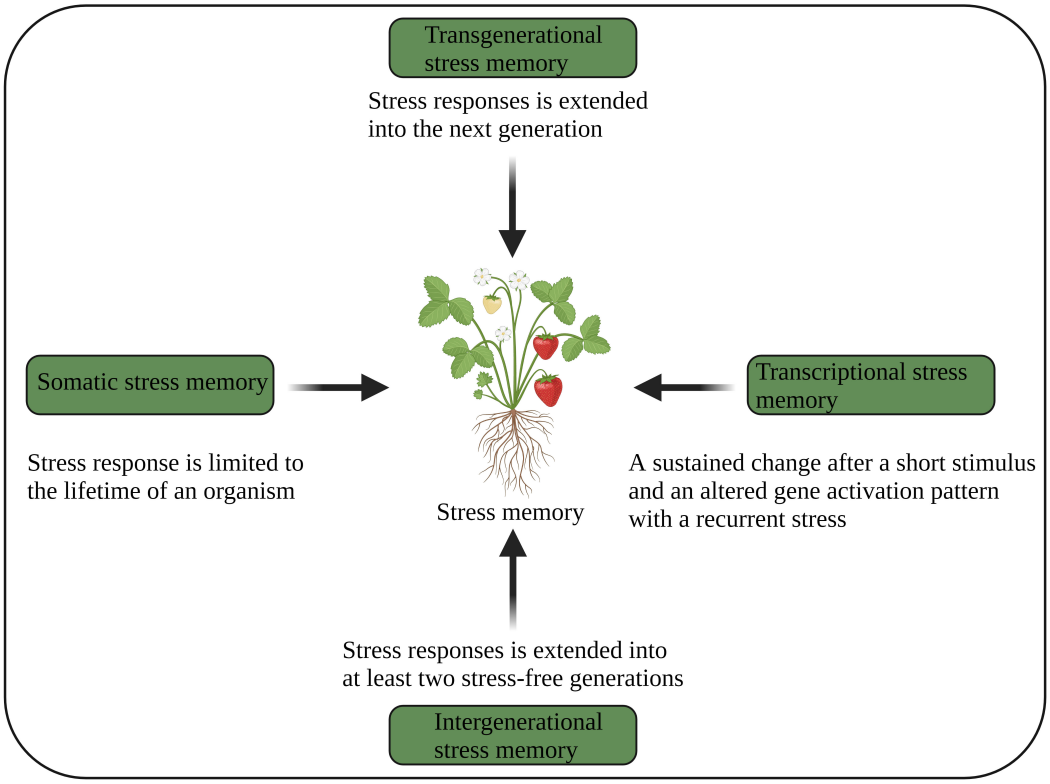


FIGURE 3
Different aspects of plant stress memory (Oberkofler et al., 2021). The image was created using BioRender.com.

Emerging technologies to analyse stress-induced epigenetic memory

Although high-throughput sequencing has facilitated genome assembly and gene annotation, the accurate annotation and assembly remain complex due to pseudogenes and duplicate copies of transposons. This prevents the application of stress memory in crop improvement. By using artificial intelligence to find predictive patterns in data and perform specific tasks based on a given dataset, machine learning could improve the annotations of transcriptomics, epigenomics and proteomics and accurately classify differentially expressed genes, proteins, pseudogenes and transposon copies (Libbrecht and Noble, 2015; Sartor et al., 2019). Machine learning can also highly accurately detect tissue-specific methylation profiles and gene expression in crop plants (N'diaye et al., 2020). Deep learning, a subset of machine learning, is also an emerging tool for genomics and its functional annotations and can uncover stress-induced interactions between chromosomes and genes, including lncRNA-miRNA interactions between different regulatory elements (Xu et al., 2022). In addition, deep Learning contributes to a better understanding of the topology of the genetic network and phenotype changes, as well as to the automation of genome prediction for various epigenetic changes in plants (Kautsar et al., 2017; Gazestani and Lewis, 2019; Wang et al., 2021a).

Oxford Nanopore sequencing, a third-generation approach, can sequence DNA and RNA without chemical labelling and PCR amplification of the sample and supports profiling of genome, epigenome, transcriptome and epitranscriptome with a single assay with 98.3% accuracy (Wang et al., 2021b; Wan et al., 2022). Machine learning can detect DNA modifications by analysing ion current signals from direct DNA sequencing with nanopores (Wan et al., 2022). In addition, single-cell sequencing (Tang et al., 2019) and the assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Ji et al., 2020; Grandi et al., 2022) await applications to investigate how DNA and histone modifications interact with gene expression in individual cells during stress memory stimulation. High-throughput phenotyping has emerged as a new perspective for non-destructive field-based phenotyping (Jangra et al., 2021). These technologies offer new opportunities to study stress-induced epigenetic memory, but these methods are insufficient for memory development because epigenetic changes have low specificity. Further improvements in machine learning and

Oxford Nanopore sequencing will lead to a better understanding of epigenetic regulation and stress memory. A trained deep and machine learning algorithm with high-throughput sequencing and phenotyping can be used in plant phenology. The method could also be used to identify and classify plant stress and quantify stress memory expression for future plant stress and further development.

Future perspective to increase epigenetic stress memory

1. A global database of chromatin changes (e.g., plant chromatin state database (PCSD, http://systemsbiology.cau.edu.cn/chromstates) (Liu et al., 2017b)) in plants grown at low and high temperatures is essential for a better understanding of epigenetic stress memory and may serve as an advanced research option to improve stress memory.

2. Molecular characterization of epigenetic stress memory is essential to link plant stress responses to phenotypic traits at all levels. The epigenetic regulation of certain stress-induced metabolites, such as volatile carbons, needs to be elucidated because they often function as interspecies communication molecules.

3. Some plant species are known to possess certain unique properties for stress tolerance and are more important than any other organism. Therefore, future research should focus on the functions of HSPs that are not yet elucidated in these plants, and extensive efforts are needed to identify specific epigenetic determinants that protect plants from the deleterious effects of cold and heat stress.

4. Practical implementation of epigenetic modifications is expensive compared to conventional plant breeding, which takes 5-10 years to develop new varieties. Therefore, a global initiative to optimise the cost-benefit ratio is needed to utilise epigenetic modification and its memory in plant breeding.

5. The use of epimutations or epialleles is a highly recommended technique to increase epigenetic stress memory, and technological advances in high-throughput sequencing and phenotyping could significantly alter this effective strategy.

6. Since epigenetic changes can be inherited by offspring as stress memory, targeted studies should be conducted to decipher the epigenetic codes of plant responses under stress conditions to further improve stress memory. Further development of high-throughput chromatin profiling could lead to significant progress in deciphering epigenetic changes and stress memory.

7. Determining the precise relationship between epigenetic modifications such as DNA methylation and histone modifications associated with epigenetic traits could lead to the development of cold and heat stress memory in crops. However, we are still trying to understand the interplay between specific epigenetic changes and epigenetic traits. Therefore, further studies could relate DNA methylation and histone modification to epigenetic traits and memory.

8. The heritability of stress-induced histone and DNA modifications needs further investigation, as these modifications are inherently dynamic and require epigenetic protein complexes to function properly.

9. The advancement of deep learning and machine learning with high-throughput sequencing could advance epigenetic research and thus the understanding of stress memory. High-throughput epigenetic studies on the mechanisms involved in the transmission of epigenetic memory are also needed to understand epigenetic memory better.

10. Creating big-data databases (e.g., quantitative PTMs, (http://qptmplants.omicsbio.info), a valuable resource for plant PTMs (Xue et al., 2022)) for epigenomics and the interpretation of multidimensional data are important to achieve a higher level of genetic gain. In addition, exploring epigenetic regulation using machine learning and big data tools could lead to a centralised infrastructure and data consortium for a better understanding of epigenetic memory that will greatly improve epigenetic stress memory.

11. Plant priming, i.e., exposure to non-lethal stress maintained over a period of time, is an adaptive method to enhance epigenetic memory, enabling them to respond robustly after second stress exposure (Ling et al., 2018). Therefore, artificially induced independent epigenetic changes through plant priming could lead to transgenerational inheritance (Bilichak and Kovalchuk, 2016).

Conclusion and future opportunities

Plants require optimal temperature for ideal growth, and the stress response in higher plants is a complicated process that differs from species to species. The plant stress response is regulated by thousands of genes in conjunction with epigenetic mechanisms that are dynamic depending on different stress conditions. The epigenetic mechanisms, inherited as epigenetic stress memory through mitotic and meiotic cell divisions, are well understood in *Arabidopsis*, but the mechanisms in crop plants are still unknown. Epigenetic stress memory fine-tunes the relationship between DNA and histone modifications and regulates the transition of epigenetic changes. Stress-induced H3K9me2 and H3K27me3, associated with DNA methylation and chromatin structures, regulate stress memory and play important roles in memory development. However, it is still unclear how epigenetic memories are inherited and how they are maintained. Enzymes such as writers, readers and erasers of epigenetic changes involved are still mysterious in stress memory; thus, it is essential to identify novel enzymes and their functions. Therefore, discovering the exact mechanism regulating the transmission of epigenetic memory will lead to an improvement in stress memory.

Recent epigenetic modifications (Figure 4) such as 6mA in DNA (Li et al., 2022), epitranscriptomics (RNA modification) such as N^6-methyladenosine (m^6A), 5-methylcytosine (m^5C), 5-hydroxymethylcytosine (hm^5C), N^7-methylguanosine (m^7G), Pseudouridine, 2′-*O*-methylation etc. (Hu et al., 2022; Murakami and Jaffrey, 2022; Ramakrishnan et al., 2022) play essential roles in plant development and stress response. Therefore, RNA modification is a new avenue for studying stress memory. Similarly, PTMs of proteins (epiproteome) (Figure 4) are key regulators of gene and protein activity (Song et al., 2020). Consequently, deciphering the epigenetic codes of stress memory requires future studies with advances in high-throughput sequencing and chromatin profiling technologies. The combined studies of epigenome, epitranscriptome and epiproteome of the specific cell and tissue on stress memory genes and their effects on epigenetic changes will greatly improve the understanding of epigenetic stress memory in specific cells and tissues. In addition, advances in machine learning, deep learning algorithms and big data analytics could help identify new epigenetic stress memories.

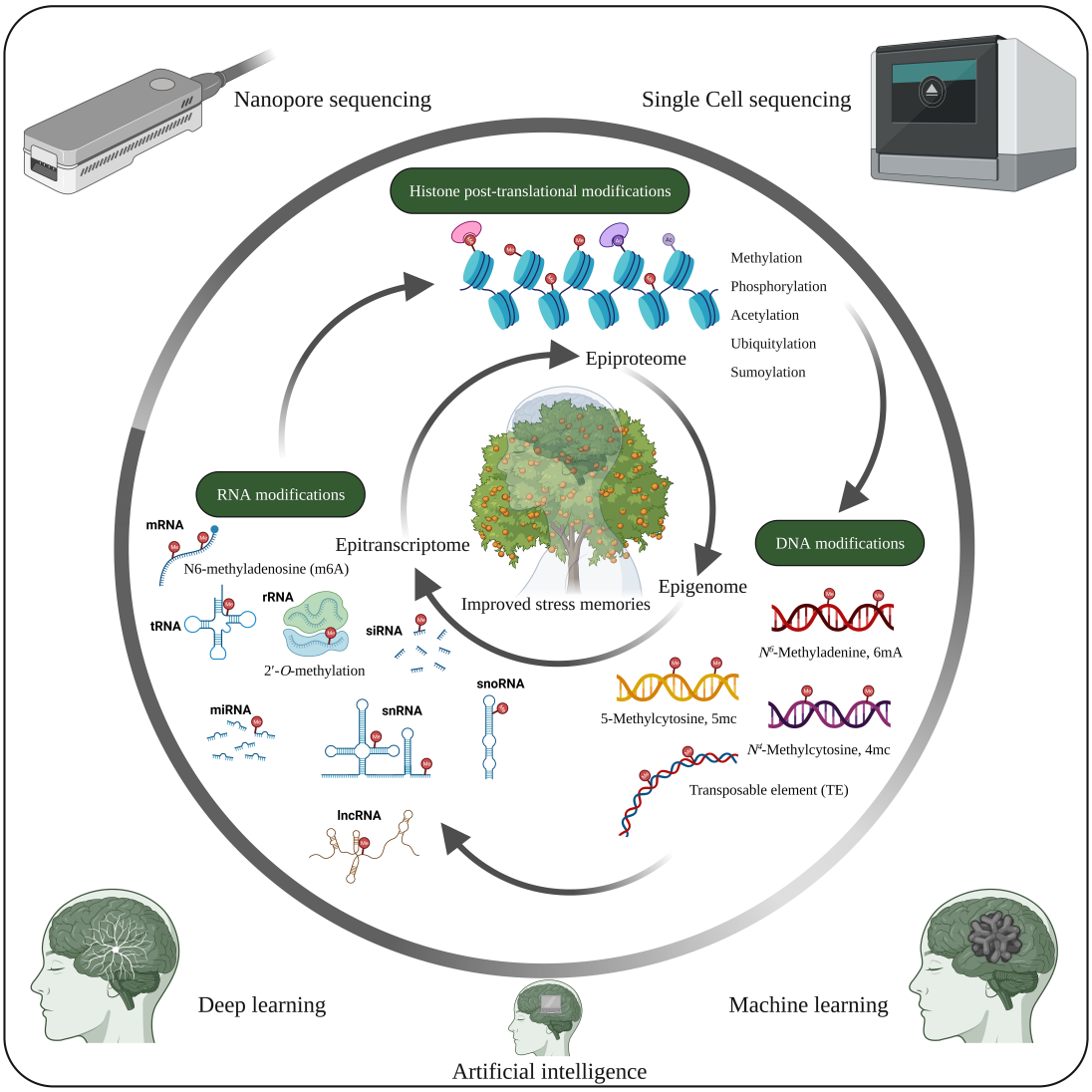


FIGURE 4
Future opportunities in plant stress memory. Cellular functions and stress tolerance depend on DNA, RNA and histone modifications, which are closely interconnected and lead to different phenotypic traits. However, the functions of the tissue-specific epigenome, epitranscriptome and epiproteome with respect to stress memory have not yet been extensively studied. Therefore, studies on the tissue-specific epigenome, epitranscriptome and epiproteome, along with recent developments such as artificial intelligence, machine learning, deep learning, nanopore sequencing and single-cell sequencing, could help identify potential mechanisms underlying stress memory. The image was created using BioRender.com.

Author contributions

MR, ZZ, and QW planned, designed and wrote the review. MR, SM, QW, ZZ, RK, and MZ outlined and edited the review. MR drew the images. MR, RK, ZA, MZ, SM, AS, GL, ZZ, and QW edited and revised the review. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by a grant from the National Natural Science Foundation of China (31770721), a grant from Jiangxi Shuangqian Program (S2019DQKJ2030), the Natural Science Foundation of Zhejiang Province (LY20C020001), the Natural Science Foundation for Distinguished Young Scholars of Nanjing Forestry University (JC2019004), the Qing Lan Project of Jiangsu Higher Education Institutions, and a project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions. The authors are also grateful for the support of Metasequoia Faculty Research Start-up Funding (163100028) at Bamboo Research Institute, Nanjing Forestry University for the first author MR.

Acknowledgments

The authors wish to thank Derek Ho (The University of Helsinki Language Centre, Finland) for outstanding editing and proofreading of the manuscript.

Conflict of interest statement

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

Adhikari, L., Baral, R., Paudel, D., Min, D., Makaju, S. O., Poudel, H. P., et al. (2022). Cold stress in plants: Strategies to improve cold tolerance in forage species. *Plant Stress* 4, 100081. doi: 10.1016/j.stress.2022.100081

Antoniou-Kourounioti, R. L., Zhao, Y., Dean, C., and Howard, M. (2021). Feeling every bit of winter – distributed temperature sensitivity in vernalization. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.628726

Banerjee, A., Wani, S. H., and Roychoudhury, A. (2017). Epigenetic control of plant cold responses. *Front. Plant Sci.* 8, 1643. doi: 10.3389/fpls.2017.01643

Bäurle, I. (2018). Can't remember to forget you: Chromatin-based priming of somatic stress responses. *Semin. Cell Dev. Biol.* 83, 133–139. doi: 10.1016/j.semcdb.2017.09.032

Bäurle, I., and Trindade, I. (2020). Chromatin regulation of somatic abiotic stress memory. *J. Exp. Bot.* 71, 5269–5279. doi: 10.1093/jxb/eraa098

Bhadouriya, S. L., Mehrotra, S., Basantani, M. K., Loake, G. J., and Mehrotra, R. (2021). Role of chromatin architecture in plant stress responses: An update. *Front. Plant Sci.* 11, 603380–603380. doi: 10.3389/fpls.2020.603380

Bilichak, A., Ilnytskyy, Y., Wóycicki, R., Kepeshchuk, N., Fogen, D., and Kovalchuk, I. (2015). The elucidation of stress memory inheritance in *Brassica rapa* plants. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00005

Bilichak, A., and Kovalchuk, I. (2016). Transgenerational response to stress in plants and its application for breeding. *J. Exp. Bot.* 67, 2081–2092. doi: 10.1093/jxb/erw066

Bi, A., Wang, T., Wang, G., Zhang, L., Wassie, M., Amee, M., et al. (2021). Stress memory gene FaHSP17.8-CII controls thermotolerance *via* remodeling PSII and ROS signaling in tall fescue. *Plant Physiol.* 187, 1163–1176. doi: 10.1093/plphys/kiab205

Boden, S. A., Kavanová, M., Finnegan, E. J., and Wigge, P. A. (2013). Thermal stress effects on grain yield in *Brachypodium distachyon* occur *via* H2A.Z-nucleosomes. *Genome Biol.* 14, R65. doi: 10.1186/gb-2013-14-6-r65

Borg, M., Jacob, Y., Susaki, D., Leblanc, C., Buendía, D., Axelsson, E., et al. (2020). Targeted reprogramming of H3K27me3 resets epigenetic memory in plant paternal chromatin. *Nat. Cell Biol.* 22, 621–629. doi: 10.1038/s41556-020-0515-y

Borg, M., Jiang, D., and Berger, F. (2021). Histone variants take center stage in shaping the epigenome. *Curr. Opin. Plant Biol.* 61, 101991. doi: 10.1016/j.pbi.2020.101991

Bourgine, B., and Guihur, A. (2021). Heat shock signaling in land plants: from plasma membrane sensing to the transcription of small heat shock proteins. *Front. Plant Sci.* 12, 710801. doi: 10.3389/fpls.2021.710801

Brenya, E., Pervin, M., Chen, Z.-H., Tissue, D. T., Johnson, S., Braam, J., et al. (2022). Mechanical stress acclimation in plants: Linking hormones and somatic memory to thigmomorphogenesis. *Plant Cell Environ.* 45, 989–1010. doi: 10.1111/pce.14252

Brewis, H. T., Wang, A. Y., Gaub, A., Lau, J. J., Stirling, P. C., and Kobor, M. S. (2021). What makes a histone variant a variant: Changing H2A to become H2A.Z. *PloS Genet.* 17, e1009950. doi: 10.1371/journal.pgen.1009950

Brzezinka, K., Altmann, S., and Bäurle, I. (2019). BRUSHY1/TONSOKU/MGOUN3 is required for heat stress memory. *Plant Cell Environ.* 42, 771–781. doi: 10.1111/pce.13365

Brzezinka, K., Altmann, S., Czesnick, H., Nicolas, P., Gorka, M., Benke, E., et al. (2016). *Arabidopsis FORGETTER1* mediates stress-induced chromatin memory through nucleosome remodeling. *eLife* 5, e17061. doi: 10.7554/eLife.17061.037

Butterbrodt, T., Thurow, C., and Gatz, C. (2006). Chromatin immunoprecipitation analysis of the tobacco PR-1a- and the truncated CaMV 35S promoter reveals differences in salicylic acid-dependent TGA factor binding and histone acetylation. *Plant Mol. Biol.* 61, 665–674. doi: 10.1007/s11103-006-0039-2

Cavrak, V. V., Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L. M., and Mittelsten Scheid, O. (2014). How a retrotransposon exploits the plant's heat stress response for its activation. *PloS Genet.* 10, e1004115. doi: 10.1371/journal.pgen.1004115

Centomani, I., Sgobba, A., D'addabbo, P., Dipierro, N., Paradiso, A., De Gara, L., et al. (2015). Involvement of DNA methylation in the control of cell growth during heat stress in tobacco BY-2 cells. *Protoplasma* 252, 1451–1459. doi: 10.1007/s00709-015-0772-y

Chang, Y. N., Zhu, C., Jiang, J., Zhang, H., Zhu, J. K., and Duan, C. G. (2020). Epigenetic regulation in plant abiotic stress responses. *J. Integr. Plant Biol.* 62, 563–580. doi: 10.1111/jipb.12901

Cong, W., Miao, Y., Xu, L., Zhang, Y., Yuan, C., Wang, J., et al. (2019). Transgenerational memory of gene expression changes induced by heavy metal stress in rice (*Oryza sativa* l.). *BMC Plant Biol.* 19, 282. doi: 10.1186/s12870-019-1887-7

Ding, Y., Shi, Y., and Yang, S. (2019). Advances and challenges in uncovering cold tolerance regulatory mechanisms in plants. *New Phytol.* 222, 1690–1704. doi: 10.1111/nph.15696

Ding, Y., Zou, L.-H., Wu, J., Ramakrishnan, M., Gao, Y., Zhao, L., et al. (2022). The pattern of DNA methylation alteration, and its association with the expression changes of non-coding RNAs and mRNAs in moso bamboo under abiotic stress. *Plant Sci.* 325, 111451. doi: 10.1016/j.plantsci.2022.111451

Distéfano, A. M., López, G. A., Setzes, N., Marchetti, F., Cainzos, M., Cascallares, M., et al. (2021). Ferroptosis in plants: triggers, proposed mechanisms, and the role of iron in modulating cell death. *J. Exp. Bot.* 72, 2125–2135. doi: 10.1093/jxb/eraa425

Duan, W., Zhang, H., Zhang, B., Wu, X., Shao, S., Li, Y., et al. (2017). Role of vernalization-mediated demethylation in the floral transition of *Brassica rapa*. *Planta* 245, 227–233. doi: 10.1007/s00425-016-2622-3

Duan, C. G., Zhu, J. K., and Cao, X. (2018). Retrospective and perspective of plant epigenetics in China. *J. Genet. Genomics* 45, 621–638. doi: 10.1016/j.jgg.2018.09.004

D'urso, A., and Brickner, J. H. (2014). Mechanisms of epigenetic memory. *Trends Genet.* 30, 230–236. doi: 10.1016/j.tig.2014.04.004

Folsom, J. J., Begcy, K., Hao, X., Wang, D., and Walia, H. (2014). Rice fertilization-independent Endosperm1 regulates seed size under heat stress by controlling early endosperm development. *Plant Physiol.* 165, 238–248. doi: 10.1104/pp.113.232413

Fowler, S., and Thomashow, M. F. (2002). *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *Plant Cell* 14, 1675–1690. doi: 10.1105/tpc.003483

Friedrich, T., Faivre, L., Bäurle, I., and Schubert, D. (2019). Chromatin-based mechanisms of temperature memory in plants. *Plant Cell Environ.* 42, 762–770. doi: 10.1111/pce.13373

Fu, W., Wu, K., and Duan, J. (2007). Sequence and expression analysis of histone deacetylases in rice. *Biochem. Biophys. Res. Commun.* 356, 843–850. doi: 10.1016/j.bbrc.2007.03.010

Gallusci, P., Agius, D. R., Moschou, P. N., Dobránszki, J., Kaiserli, E., and Martinelli, F. (2022). Deep inside the epigenetic memories of stressed plants. *Trends Plant Sci.*, 1–12. doi: 10.1016/j.tplants.2022.09.004

Gao, G., Li, J., Li, H., Li, F., Xu, K., Yan, G., et al. (2014). Comparison of the heat stress induced variations in DNA methylation between heat-tolerant and heat-sensitive rapeseed seedlings. *Breed Sci.* 64, 125–133. doi: 10.1270/jsbbs.64.125

Gazestani, V. H., and Lewis, N. E. (2019). From genotype to phenotype: Augmenting deep learning with networks and systems biology. *Curr. Opin. Syst. Biol.* 15, 68–73. doi: 10.1016/j.coisb.2019.04.001

Grandi, F. C., Modi, H., Kampman, L., and Corces, M. R. (2022). Chromatin accessibility profiling by ATAC-seq. *Nat. Protoc.* 17, 1518–1552. doi: 10.1038/s41596-022-00692-9

Guihur, A., Rebeaud, M. E., and Goloubinoff, P. (2022). How do plants feel the heat and survive? *Trends Biochem. Sci.* 47, 824–838. doi: 10.1016/j.tibs.2022.05.004

Guo, X., Liu, D., and Chong, K. (2018). Cold signaling in plants: Insights into mechanisms and regulation. *J. Integr. Plant Biol.* 60, 745–756. doi: 10.1111/jipb.12706

Han, D., Yu, Z., Lai, J., and Yang, C. (2022). Post-translational modification: a strategic response to high temperature in plants. *aBIOTECH* 3, 49–64. doi: 10.1007/s42994-021-00067-w

Han, Z., Yu, H., Zhao, Z., Hunter, D., Luo, X., Duan, J., et al. (2016). *AtHD2D* gene plays a role in plant growth, development, and response to abiotic stresses in *Arabidopsis thaliana*. *Front. Plant Sci.* 7, 310. doi: 10.3389/fpls.2016.00310

Hasanuzzaman, M., Nahar, K., Alam, M. M., Roychowdhury, R., and Fujita, M. (2013). Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. *Int. J. Mol. Sci.* 14, 9643–9684. doi: 10.3390/ijms14059643

Hu, J., Cai, J., Xu, T., and Kang, H. (2022). Epitranscriptomic mRNA modifications governing plant stress responses: underlying mechanism and potential application. *Plant Biotechnol. J.* 20, 2245–2257. doi: 10.1111/pbi.13913

Hu, Y., Lu, Y., Zhao, Y., and Zhou, D. X. (2019). Histone acetylation dynamics integrates metabolic activity to regulate plant response to stress. *Front. Plant Sci.* 10, 1236. doi: 10.3389/fpls.2019.01236

Hu, Y., Qin, F., Huang, L., Sun, Q., Li, C., Zhao, Y., et al. (2009). Rice histone deacetylase genes display specific expression patterns and developmental functions. *Biochem. Biophys. Res. Commun.* 388, 266–271. doi: 10.1016/j.bbrc.2009.07.162

Hu, Y., Zhang, L., Zhao, L., Li, J., He, S., Zhou, K., et al. (2011). Trichostatin a selectively suppresses the cold-induced transcription of the *ZmDREB1* gene in maize. *PloS One* 6, e22132. doi: 10.1371/journal.pone.0022132

Hwarari, D., Guan, Y., Ahmad, B., Movahedi, A., Min, T., Hao, Z., et al. (2022). ICE-CBF-COR signaling cascade and its regulation in plants responding to cold stress. *Int. J. Mol. Sci.* 23, 1549. doi: 10.3390/ijms23031549

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., and Paszkowski, J. (2011). An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* 472, 115–119. doi: 10.1038/nature09861

Iwasaki, M., Hyvärinen, L., Piskurewicz, U., and Lopez-Molina, L. (2019). Non-canonical RNA-directed DNA methylation participates in maternal and environmental control of seed dormancy. *eLife* 8, e37434. doi: 10.7554/eLife.37434.029

Jacques, C., Salon, C., Barnard, R. L., Vernoud, V., and Prudent, M. (2021). Drought stress memory at the plant cycle level: A review. *Plants (Basel)* 10, 1873. doi: 10.3390/plants10091873

Jangra, S., Chaudhary, V., Yadav, R. C., and Yadav, N. R. (2021). High-throughput phenotyping: A platform to accelerate crop improvement. *Phenomics* 1, 31–53. doi: 10.1007/s43657-020-00007-6

Jaskiewicz, M., Conrath, U., and Peterhänsel, C. (2011). Chromatin modification acts as a memory for systemic acquired resistance in the plant stress response. *EMBO Rep.* 12, 50–55. doi: 10.1038/embor.2010.186

Ji, Z., Zhou, W., Hou, W., and Ji, H. (2020). Single-cell ATAC-seq signal extraction and enhancement with SCATE. *Genome Biol.* 21, 161. doi: 10.1186/s13059-020-02075-3

Jo, S. H., Park, H. J., Lee, A., Jung, H., Park, J. M., Kwon, S. Y., et al. (2022). The *Arabidopsis* cyclophilin CYP18-1 facilitates PRP18 dephosphorylation and the splicing of introns retained under heat stress. *Plant Cell* 34, 2383–2403. doi: 10.1093/plcell/koac084

Joseph, F. M., and Young, N. L. (2022). Histone variant-specific post-translational modifications. *Semin. Cell Dev. Biol.* 135, 73–84. doi: 10.1016/j.semcdb.2022.02.012

Kang, H., Nugroho, A. B. D., Park, M., Kim, J. A., Lee, S. W., Moon, H., et al. (2022). Vernalization regulates flowering genes and modulates glucosinolates biosynthesis in chinese cabbage. *J. Plant Biol.* 65, 157–173. doi: 10.1007/s12374-021-09344-z

Kautsar, S. A., Suarez Duran, H. G., Blin, K., Osbourn, A., and Medema, M. H. (2017). plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res.* 45, W55–w63. doi: 10.1093/nar/gkx305

Khator, K., Mahawar, L., and Shekhawat, G. S. (2020). NaCl Induced oxidative stress in legume crops of Indian thar desert: an insight in the cytoprotective role of HO1, NO and antioxidants. *Physiol. Mol. Biol. Plants* 26, 51–62. doi: 10.1007/s12298-019-00728-7

Kidokoro, S., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2022). Transcriptional regulatory network of plant cold-stress responses. *Trends Plant Sci.* 27, 922–935. doi: 10.1016/j.tplants.2022.01.008

Kim, J.-M., Sasaki, T., Ueda, M., Sako, K., and Seki, M. (2015). Chromatin changes in response to drought, salinity, heat, and cold stresses in plants. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00114

Korotko, U., Chwiałkowska, K., Sańko-Sawczenko, I., and Kwasniewski, M. (2021). DNA Demethylation in response to heat stress in *Arabidopsis thaliana*. *Int. J. Mol. Sci.* 22, 1555. doi: 10.3390/ijms22041555

Kumar, S. V., and Wigge, P. A. (2010). H2A.Z-containing nucleosomes mediate the thermosensory response in *Arabidopsis*. *Cell* 140, 136–147. doi: 10.1016/j.cell.2009.11.006

Kwon, C. S., Lee, D., Choi, G., and Chung, W. I. (2009). Histone occupancy-dependent and -independent removal of H3K27 trimethylation at cold-responsive genes in *Arabidopsis*. *Plant J.* 60, 112–121. doi: 10.1111/j.1365-313X.2009.03938.x

Lämke, J., and Bäurle, I. (2017). Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. *Genome Biol.* 18, 124. doi: 10.1186/s13059-017-1263-6

Lämke, J., Brzezinka, K., Altmann, S., and Bäurle, I. (2016). A hit-and-run heat shock factor governs sustained histone methylation and transcriptional stress memory. *EMBO J.* 35, 162–175. doi: 10.15252/embj.201592593

Lang, Z., Wang, Y., Tang, K., Tang, D., Datsenka, T., Cheng, J., et al. (2017). Critical roles of DNA demethylation in the activation of ripening-induced genes and inhibition of ripening-repressed genes in tomato fruit. *Proc. Natl. Acad. Sci. U. S. A.* 114, E4511–e4519. doi: 10.1073/pnas.1705233114

Latzel, V., Rendina González, A. P., and Rosenthal, J. (2016). Epigenetic memory as a basis for intelligent behavior in clonal plants. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01354

Libbrecht, M. W., and Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* 16, 321–332. doi: 10.1038/nrg3920

Li, N., Euring, D., Cha, J. Y., Lin, Z., Lu, M., Huang, L.-J., et al. (2021). Plant hormone-mediated regulation of heat tolerance in response to global climate change. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.627969

Li, S., Liu, J., Liu, Z., Li, X., Wu, F., and He, Y. (2014). HEAT-INDUCED TAS1 TARGET1 mediates thermotolerance *via* HEAT STRESS TRANSCRIPTION FACTOR A1a-directed pathways in *Arabidopsis*. *Plant Cell* 26, 1764–1780. doi: 10.1105/tpc.114.124883

Ling, Y., Serrano, N., Gao, G., Atia, M., Mokhtar, M., Woo, Y. H., et al. (2018). Thermopriming triggers splicing memory in *Arabidopsis*. *J. Exp. Bot.* 69, 2659–2675. doi: 10.1093/jxb/ery062

Liu, H., Able, A. J., and Able, J. A. (2022a). Priming crops for the future: rewiring stress memory. *Trends Plant Sci.* 27, 699–716. doi: 10.1016/j.tplants.2021.11.015

Liu, N., Fromm, M., and Avramova, Z. (2014). H3K27me3 and H3K4me3 chromatin environment at super-induced dehydration stress memory genes of *Arabidopsis thaliana*. *Mol. Plant* 7, 502–513. doi: 10.1093/mp/ssu001

Liu, J., and He, Z. (2020). Small DNA methylation, big player in plant abiotic stress responses and memory. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.595603

Liu, Z., Jia, Y., Ding, Y., Shi, Y., Li, Z., Guo, Y., et al. (2017c). Plasma membrane CRPK1-mediated phosphorylation of 14-3-3 proteins induces their nuclear import to fine-tune CBF signaling during cold response. *Mol. Cell* 66, 117–128.e115. doi: 10.1016/j.molcel.2017.02.016

Liu, T., Li, Y., Duan, W., Huang, F., and Hou, X. (2017a). Cold acclimation alters DNA methylation patterns and confers tolerance to heat and increases growth rate in *Brassica rapa*. *J. Exp. Bot.* 68, 1213–1224. doi: 10.1093/jxb/erw496

Liu, Y., Liu, K., Yin, L., Yu, Y., Qi, J., Shen, W.-H., et al. (2019). H3K4me2 functions as a repressive epigenetic mark in plants. *Epigenet. Chromatin* 12, 40. doi: 10.1186/s13072-019-0285-6

Liu, X., Quan, W., and Bartels, D. (2022b). Stress memory responses and seed priming correlate with drought tolerance in plants: an overview. *Planta* 255, 45. doi: 10.1007/s00425-022-03828-z

Liu, Y., Tian, T., Zhang, K., You, Q., Yan, H., Zhao, N., et al. (2017b). PCSD: a plant chromatin state database. *Nucleic Acids Res.* 46, D1157–D1167. doi: 10.1093/nar/gkx919

Li, H., Zhang, N., Wang, Y., Xia, S., Zhu, Y., Xing, C., et al. (2022). DNA N6-methyladenine modification in eukaryotic genome. *Front. Genet.* 13, 914404. doi: 10.3389/fgene.2022.914404

Luo, X., and He, Y. (2020). Experiencing winter for spring flowering: A molecular epigenetic perspective on vernalization. *J. Integr. Plant Biol.* 62, 104–117. doi: 10.1111/jipb.12896

Ma, N., Chen, W., Fan, T., Tian, Y., Zhang, S., Zeng, D., et al. (2015). Low temperature-induced DNA hypermethylation attenuates expression of RhAG, an AGAMOUS homolog, and increases petal number in rose (*Rosa hybrida*). *BMC Plant Biol.* 15, 237. doi: 10.1186/s12870-015-0623-1

Mahawar, L., and Shekhawat, G. S. (2018). Haem oxygenase: A functionally diverse enzyme of photosynthetic organisms and its role in phytochrome chromophore biosynthesis, cellular signalling and defence mechanisms. *Plant Cell Environ.* 41, 483–500. doi: 10.1111/pce.13116

Malik, S., and Zhao, D. (2022). Epigenetic regulation of heat stress in plant male reproduction. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.826473

Matsunaga, W., Kobayashi, A., Kato, A., and Ito, H. (2012). The effects of heat induction and the siRNA biogenesis pathway on the transgenerational transposition of *ONSEN*, a *copia-like* retrotransposon in *Arabidopsis thaliana*. *Plant Cell Physiol.* 53, 824–833. doi: 10.1093/pcp/pcr179

Millán-Zambrano, G., Burton, A., Bannister, A. J., and Schneider, R. (2022). Histone post-translational modifications — cause and consequence of genome function. *Nat. Rev. Genet.* 23, 563–580. doi: 10.1038/s41576-022-00468-7

Murakami, S., and Jaffrey, S. R. (2022). Hidden codes in mRNA: Control of gene expression by m^6A. *Mol. Cell* 82, 2236–2251. doi: 10.1016/j.molcel.2022.05.029

Naydenov, M., Baev, V., Apostolova, E., Gospodinova, N., Sablok, G., Gozmanova, M., et al. (2015). High-temperature effect on genes engaged in DNA methylation and affected by DNA methylation in *Arabidopsis*. *Plant Physiol. Biochem.* 87, 102–108. doi: 10.1016/j.plaphy.2014.12.022

N'diaye, A., Byrns, B., Cory, A. T., Nilsen, K. T., Walkowiak, S., Sharpe, A., et al. (2020). Machine learning analyses of methylation profiles uncovers tissue-specific gene expression patterns in wheat. *Plant Genome* 13, e20027. doi: 10.1002/tpg2.20027

Nguyen, N. H., Vu, N. T., and Cheong, J. J. (2022). Transcriptional stress memory and transgenerational inheritance of drought tolerance in plants. *Int. J. Mol. Sci.* 23, 12918. doi: 10.3390/ijms232112918

Novillo, F., Alonso, J. M., Ecker, J. R., and Salinas, J. (2004). CBF2/DREB1C is a negative regulator of *CBF1/DREB1B* and *CBF3/DREB1A* expression and plays a central role in stress tolerance in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 3985–3990. doi: 10.1073/pnas.0303029101

Oberkofler, V., Pratx, L., and Bäurle, I. (2021). Epigenetic regulation of abiotic stress memory: maintaining the good things while they last. *Curr. Opin. Plant Biol.* 61, 102007. doi: 10.1016/j.pbi.2021.102007

Olate, E., Jiménez-Gómez, J. M., Holuigue, L., and Salinas, J. (2018). NPR1 mediates a novel regulatory pathway in cold acclimation by interacting with HSFA1 factors. *Nat. Plants* 4, 811–823. doi: 10.1038/s41477-018-0254-2

Ono, A., and Kinoshita, T. (2021). Epigenetics and plant reproduction: Multiple steps for responsibly handling succession. *Curr. Opin. Plant Biol.* 61, 102032. doi: 10.1016/j.pbi.2021.102032

Ou, X., Zhang, Y., Xu, C., Lin, X., Zang, Q., Zhuang, T., et al. (2012). Transgenerational inheritance of modified DNA methylation patterns and enhanced tolerance induced by heavy metal stress in rice (*Oryza sativa* l.). *PloS One* 7, e41143. doi: 10.1371/journal.pone.0041143

Pandey, R., Müller, A., Napoli, C. A., Selinger, D. A., Pikaard, C. S., Richards, E. J., et al. (2002). Analysis of histone acetyltransferase and histone deacetylase families of *Arabidopsis thaliana* suggests functional diversification of chromatin modification among multicellular eukaryotes. *Nucleic Acids Res.* 30, 5036–5055. doi: 10.1093/nar/gkf660

Park, J., Lim, C. J., Shen, M., Park, H. J., Cha, J.-Y., Iniesto, E., et al. (2018). Epigenetic switch from repressive to permissive chromatin in response to cold stress. *Proc. Natl. Acad. Sci. U. S. A.* 115, E5400–E5409. doi: 10.1073/pnas.1721241115

Pecinka, A., Dinh, H. Q., Baubec, T., Rosa, M., Lettner, N., and Scheid, O. M. (2010). Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in. *Arabidopsis Plant Cell* 22, 3118–3129. doi: 10.1105/tpc.110.078493

Perrella, G., Bäurle, I., and Van Zanten, M. (2022). Epigenetic regulation of thermomorphogenesis and heat stress tolerance. *New Phytol.* 234, 1144–1160. doi: 10.1111/nph.17970

Popova, O. V., Dinh, H. Q., Aufsatz, W., and Jonak, C. (2013). The RdDM pathway is required for basal heat tolerance in *Arabidopsis*. *Mol. Plant* 6, 396–410. doi: 10.1093/mp/sst023

Qian, Y., Hu, W., Liao, J., Zhang, J., and Ren, Q. (2019). The dynamics of DNA methylation in the maize (*Zea mays* l.) inbred line B73 response to heat stress at the seedling stage. *Biochem. Biophys. Res. Commun.* 512, 742–749. doi: 10.1016/j.bbrc.2019.03.150

Ramakrishnan, M., Rajan, K. S., Mullasseri, S., Palakkal, S., Kalpana, K., Sharma, A., et al. (2022). The plant epitranscriptome: revisiting pseudouridine and 2'-*O*-methyl RNA modifications. *Plant Biotechnol. J.* 20, 1241–1256. doi: 10.1111/pbi.13829

Saeed, F., Chaudhry, U. K., Bakhsh, A., Raza, A., Saeed, Y., Bohra, A., et al. (2022). Moving beyond DNA sequence to improve plant stress responses. *Front. Genet.* 13, 874648. doi: 10.3389/fgene.2022.874648

Samakovli, D., Tichá, T., Vavrdová, T., Ovečka, M., Luptovčiak, I., Zapletalová, V., et al. (2020). YODA-HSP90 module regulates phosphorylation-dependent inactivation of SPEECHLESS to control stomatal development under acute heat stress in *Arabidopsis*. *Mol. Plant* 13, 612–633. doi: 10.1016/j.molp.2020.01.001

Sartor, R. C., Noshay, J., Springer, N. M., and Briggs, S. P. (2019). Identification of the expressome by machine learning on omics data. *Proc. Natl. Acad. Sci. U. S. A.* 116, 18119–18125. doi: 10.1073/pnas.1813645116

Sedaghatmehr, M., Stüwe, B., Mueller-Roeber, B., and Balazadeh, S. (2022). Heat shock factor HSFA2 fine-tunes resetting of thermomemory *via* plastidic metalloprotease FtsH6. *J. Exp. Bot.* 73, 6394–6404. doi: 10.1093/jxb/erac257

Shanker, A. K., Bhanu, D., and Maheswari, M. (2020). Epigenetics and transgenerational memory in plants under heat stress. *Plant Physiol. Rep.* 25, 583–593. doi: 10.1007/s40502-020-00557-x

Sharma, M., Kumar, P., Verma, V., Sharma, R., Bhargava, B., and Irfan, M. (2022). Understanding plant stress memory response for abiotic stress resilience: Molecular insights and prospects. *Plant Physiol. Biochem.* 179, 10–24. doi: 10.1016/j.plaphy.2022.03.004

Shen, Y., Lei, T., Cui, X., Liu, X., Zhou, S., Zheng, Y., et al. (2019). *Arabidopsis* histone deacetylase HDA15 directly represses plant response to elevated ambient temperature. *Plant J.* 100, 991–1006. doi: 10.1111/tpj.14492

Shen, Y., Wu, X., Liu, D., Song, S., Liu, D., and Wang, H. (2016). Cold-dependent alternative splicing of a *Jumonji c domain-containing* gene *MtJMJC5* in *Medicago truncatula*. *Biochem. Biophys. Res. Commun.* 474, 271–276. doi: 10.1016/j.bbrc.2016.04.062

Sicilia, A., Scialò, E., Puglisi, I., and Lo Piero, A. R. (2020). Anthocyanin biosynthesis and DNA methylation dynamics in sweet orange fruit [*Citrus sinensis* l. (Osbeck)] under cold stress. *J. Agric. Food Chem.* 68, 7024–7031. doi: 10.1021/acs.jafc.0c02360

Singh, A. K., Dhanapal, S., Finkelshtein, A., and Chamovitz, D. A. (2021). CSN5A subunit of COP9 signalosome is required for resetting transcriptional stress memory after recurrent heat stress in *Arabidopsis*. *Biomolecules* 11, 668. doi: 10.3390/biom11050668

Sirko, A., Gotor, C., and Wawrzyńska, A. (2022). Editorial: Protein posttranslational modifications in plant responses to abiotic stress - women in plant science series. *Front. Plant Sci.* 13, 1049173. doi: 10.3389/fpls.2022.1049173

Song, Y., Ci, D., Tian, M., and Zhang, D. (2016). Stable methylation of a non-coding RNA gene regulates gene expression in response to abiotic stress in *Populus simonii*. *J. Exp. Bot.* 67, 1477–1492. doi: 10.1093/jxb/erv543

Song, H., Liu, D., Dong, S., Zeng, L., Wu, Z., Zhao, P., et al. (2020). Epitranscriptomics and epiproteomics in cancer drug resistance: therapeutic implications. *Sig. Transduction Targeting Ther.* 5, 193. doi: 10.1038/s41392-020-00300-w

Song, Y., Liu, L., Feng, Y., Wei, Y., Yue, X., He, W., et al. (2015). Chilling- and freezing-induced alterations in cytosine methylation and its association with the cold tolerance of an alpine subnival plant, *Chorispora bungeana*. *PloS One* 10, e0135485. doi: 10.1371/journal.pone.0135485

Sun, M., Yang, Z., Liu, L., and Duan, L. (2022). DNA Methylation in plant responses and adaption to abiotic stresses. *Int. J. Mol. Sci.* 23, 6910. doi: 10.3390/ijms23136910

Tang, X., Huang, Y., Lei, J., Luo, H., and Zhu, X. (2019). The single-cell sequencing: new developments and medical applications. *Cell Biosci.* 9, 53. doi: 10.1186/s13578-019-0314-y

Tang, X., Wang, Q., Yuan, H., and Huang, X. (2018). Chilling-induced DNA demethylation is associated with the cold tolerance of *Hevea brasiliensis*. *BMC Plant Biol.* 18, 70. doi: 10.1186/s12870-018-1276-7

Tasset, C., Singh Yadav, A., Sureshkumar, S., Singh, R., van der Woude, L., Nekrasov, M., et al. (2018). POWERDRESS-mediated histone deacetylation is essential for thermomorphogenesis in *Arabidopsis thaliana*. *PloS Genet.* 14, e1007280. doi: 10.1371/journal.pgen.1007280

To, T. K., Nakaminami, K., Kim, J. M., Morosawa, T., Ishida, J., Tanaka, M., et al. (2011). *Arabidopsis* HDA6 is required for freezing tolerance. *Biochem. Biophys. Res. Commun.* 406, 414–419. doi: 10.1016/j.bbrc.2011.02.058

Tricker, P. (2015). Transgenerational inheritance or resetting of stress-induced epigenetic modifications: two sides of the same coin. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00699

Verma, N., Giri, S. K., Singh, G., Gill, R., and Kumar, A. (2022). Epigenetic regulation of heat and cold stress responses in crop plants. *Plant Gene* 29, 100351. doi: 10.1016/j.plgene.2022.100351

Wang, Y., Zhang, P., Guo, W., Liu, H., Li, X., Zhang, Q., et al. (2021a). A deep learning approach to automate whole-genome prediction of diverse epigenomic modifications in plants. *New Phytol.* 232, 880–897. doi: 10.1111/nph.17630

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., and Au, K. F. (2021b). Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* 39, 1348–1365. doi: 10.1038/s41587-021-01108-x

Wang, P., Zhao, L., Hou, H., Zhang, H., Huang, Y., Wang, Y., et al. (2015). Epigenetic changes are associated with programmed cell death induced by heat stress in seedling leaves of *Zea mays*. *Plant Cell Physiol.* 56, 965–976. doi: 10.1093/pcp/pcv023

Wan, Y. K., Hendra, C., Pratanwanich, P. N., and Göke, J. (2022). Beyond sequencing: machine learning algorithms extract biology hidden in nanopore signal data. *Trends Genet.* 38, 246–257. doi: 10.1016/j.tig.2021.09.001

Wibowo, A., Becker, C., Marconi, G., Durr, J., Price, J., Hagmann, J., et al. (2016). Hyperosmotic stress memory in *Arabidopsis* is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA glycosylase activity. *eLife* 5, e13546. doi: 10.7554/eLife.13546.044

Wu, Y., Zhang, D., Chu, J. Y., Boyle, P., Wang, Y., Brindle, I. D., et al. (2012). The *Arabidopsis* NPR1 protein is a receptor for the plant defense hormone salicylic acid. *Cell Rep.* 1, 639–647. doi: 10.1016/j.celrep.2012.05.008

Xie, H., Sun, Y., Cheng, B., Xue, S., Cheng, D., Liu, L., et al. (2019). Variation in *ICE1* methylation primarily determines phenotypic variation in freezing tolerance in *Arabidopsis thaliana*. *Plant Cell Physiol.* 60, 152–165. doi: 10.1093/pcp/pcy197

Xie, W., Tang, Q., Yan, F., and Tao, Z. (2021). Transcriptional memory and response to adverse temperatures in plants. *J. Zhejiang Univ. Sci. B.* 22, 791–804. doi: 10.1631/jzus.B2100287

Xue, H., Zhang, Q., Wang, P., Cao, B., Jia, C., Cheng, B., et al. (2022). qPTMplants: an integrative database of quantitative post-translational modifications in plants. *Nucleic Acids Res.* 50, D1491–D1499. doi: 10.1093/nar/gkab945

Xu, D., Yuan, W., Fan, C., Liu, B., Lu, M.-Z., and Zhang, J. (2022). Opportunities and challenges of predictive approaches for the non-coding RNA in plants. *Front. Plant Sci.* 13, 890663. doi: 10.3389/fpls.2022.890663

Yamaguchi, N. (2022). Heat memory in plants: histone modifications, nucleosome positioning and miRNA accumulation alter heat memory gene expression. *Genes Genet. Syst.* 96, 229–235. doi: 10.1266/ggs.21-00040

Yamaguchi, N., and Ito, T. (2021). JMJ histone demethylases balance H3K27me3 and H3K4me3 levels at the HSP21 locus during heat acclimation *in Arabidopsis*. *Biomolecules* 11, 852. doi: 10.3390/biom11060852

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

Yang, H., Berry, S., Olsson, T. S. G., Hartley, M., Howard, M., and Dean, C. (2017). Distinct phases of polycomb silencing to hold epigenetic memory of cold in *Arabidopsis*. *Science* 357, 1142–1145. doi: 10.1126/science.aan1121

Yang, X., Sanchez, R., Kundariya, H., Maher, T., Dopp, I., Schwegel, R., et al. (2020). Segregation of an *MSH1* RNAi transgene produces heritable non-genetic memory in association with methylome reprogramming. *Nat. Commun.* 11, 2214. doi: 10.1038/s41467-020-16036-8

Zeng, Z., Zhang, W., Marand, A. P., Zhu, B., Buell, C. R., and Jiang, J. (2019). Cold stress induces enhanced chromatin accessibility and bivalent histone modifications H3K4me3 and H3K27me3 of active genes in potato. *Genome Biol.* 20, 123. doi: 10.1186/s13059-019-1731-2

Zhang, Q., Liang, Z., Cui, X., Ji, C., Li, Y., Zhang, P., et al. (2018). N6-methyladenine DNA methylation in *Japonica* and *Indica* rice genomes and its association with gene expression, plant development, and stress responses. *Mol. Plant* 11, 1492–1508. doi: 10.1016/j.molp.2018.11.005

Zhang, X., Ménard, R., Li, Y., Coruzzi, G. M., Heitz, T., Shen, W. H., et al. (2020). *Arabidopsis* SDG8 potentiates the sustainable transcriptional induction of the *Pathogenesis-related genes PR1* and *PR2* during plant defense response. *Front. Plant Sci.* 11, 277. doi: 10.3389/fpls.2020.00277

Zhu, J., Jeong, J. C., Zhu, Y., Sokolchik, I., Miyazaki, S., Zhu, J.-K., et al. (2008). Involvement of *Arabidopsis* HOS15 in histone deacetylation and cold tolerance. *Proc. Natl. Acad. Sci. U. S. A.* 105, 4945–4950. doi: 10.1073/pnas.0801029105

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 09 December 2022
DOI 10.3389/fpls.2022.1064847

OPEN ACCESS

EDITED BY
Hussain Touseef,
Matimate Agromart Pvt. Ltd. (Sevama AgriClinic Laboratory), India

REVIEWED BY
Karthikeyan Adhimoolam,
Jeju National University, South Korea
Sunil Kumar Sahu,
Beijing Genomics Institute (BGI), China

*CORRESPONDENCE
Muthusamy Ramakrishnan
ramky@njfu.edu.cn
Ping Yang
yangping@zafu.edu.cn
Mingbing Zhou
zhoumingbing@zafu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 08 October 2022
ACCEPTED 21 November 2022
PUBLISHED 09 December 2022

CITATION
Papolu PK, Ramakrishnan M, Mullasseri S, Kalendar R, Wei Q, Zou L-H, Ahmad Z, Vinod KK, Yang P and Zhou M (2022) Retrotransposons: How the continuous evolutionary front shapes plant genomes for response to heat stress. *Front. Plant Sci.* 13:1064847. doi: 10.3389/fpls.2022.1064847

Retrotransposons: How the continuous evolutionary front shapes plant genomes for response to heat stress

Pradeep K. Papolu[1†], Muthusamy Ramakrishnan[1,2*†], Sileesh Mullasseri[3], Ruslan Kalendar[4,5], Qiang Wei[2], Long–Hai Zou[1], Zishan Ahmad[2], Kunnummal Kurungara Vinod[6], Ping Yang[1,7*] and Mingbing Zhou[1,7*]

[1]State Key Laboratory of Subtropical Silviculture, Bamboo Industry Institute, Zhejiang A&F University, Hangzhou, Zhejiang, China, [2]Co-Innovation Center for Sustainable Forestry in Southern China, Bamboo Research Institute, Key Laboratory of National Forestry and Grassland Administration on Subtropical Forest Biodiversity Conservation, College of Biology and the Environment, Nanjing Forestry University, Nanjing, Jiangsu, China, [3]Department of Zoology, St. Albert's College (Autonomous), Kochi, Kerala, India, [4]Helsinki Institute of Life Science HiLIFE, Biocenter 3, University of Helsinki, Helsinki, Finland, [5]National Laboratory Astana, Nazarbayev University, Astana, Kazakhstan, [6]Division of Genetics, ICAR - Indian Agricultural Research Institute, New Delhi, India, [7]Zhejiang Provincial Collaborative Innovation Center for Bamboo Resources and High-Efficiency Utilization, Zhejiang A&F University, Hangzhou, Zhejiang, China

Long terminal repeat retrotransposons (LTR retrotransposons) are the most abundant group of mobile genetic elements in eukaryotic genomes and are essential in organizing genomic architecture and phenotypic variations. The diverse families of retrotransposons are related to retroviruses. As retrotransposable elements are dispersed and ubiquitous, their "copy-out and paste-in" life cycle of replicative transposition leads to new genome insertions without the excision of the original element. The overall structure of retrotransposons and the domains responsible for the various phases of their replication is highly conserved in all eukaryotes. The two major superfamilies of LTR retrotransposons, Ty1/*Copia* and Ty3/*Gypsy*, are distinguished and dispersed across the chromosomes of higher plants. Members of these superfamilies can increase in copy number and are often activated by various biotic and abiotic stresses due to retrotransposition bursts. LTR retrotransposons are important drivers of species diversity and exhibit great variety in structure, size, and mechanisms of transposition, making them important putative actors in genome evolution. Additionally, LTR retrotransposons influence the gene expression patterns of adjacent genes by modulating potential small interfering RNA (siRNA) and RNA-directed DNA methylation (RdDM) pathways. Furthermore, comparative and evolutionary analysis of the most important crop genome sequences and advanced technologies have elucidated the epigenetics and structural and functional modifications driven by LTR retrotransposon during speciation. However, mechanistic insights into LTR retrotransposons remain obscure in plant

development due to a lack of advancement in high throughput technologies. In this review, we focus on the key role of LTR retrotransposons response in plants during heat stress, the role of centromeric LTR retrotransposons, and the role of LTR retrotransposon markers in genome expression and evolution.

KEYWORDS

transposable element, retrotransposons, LTR, genetic diversity, siRNAs, RdDM pathways, Ty1/copia, Ty3/gypsy

Introduction

Eukaryotic genomes contain repetitive elements, such as transposable elements (TEs), that are present in multiple copies throughout the genome. TEs are tandemly arrayed, interspersed throughout the genome, and can be processed as pseudogenes. TEs are major components of eukaryotic genomes and can change their position within genomes (Lisch, 2013; Bourque et al., 2018). TEs were first described in maize by Barbara McClintock in the middle of the twentieth century and she named them jumping genes (Ravindran, 2012; Goodier, 2016). Although TEs are a source of spontaneous mutations, their expression and activity can also increase the stress response to different biotic and abiotic stresses (Ramakrishnan et al., 2021). Moreover, TE specificity has now been associated with the adaptation of plants to a range of these stresses. TEs have deep evolutionary origins and continuous diversification and come in a bewildering variety of forms and shapes (Bourque et al., 2018; Klein and O'Neill, 2018) in most eukaryotic genomes (Wicker et al., 2007; Muñoz-López and García-Pérez, 2010; Gorbunova et al., 2021). TEs are primarily classified into DNA transposons (Class II) and retrotransposons (Class I) based on their mechanism of transposition (Boeke et al., 1985). Both classes are further divided into subclasses based on the mechanism of chromosomal integration. Class I has two major classes, Long Terminal Repeats (LTR) retrotransposons (LTR retrotransposons) and non-LTR retrotransposons (Wessler et al., 1995). LTR retrotransposons and related elements are abundant in plant genomes and include functional genes encoding structural and enzymatic proteins (Galindo-González et al., 2017). LTR retrotransposon mobility is ensured through an RNA intermediate, allowing a copy-and-paste approach for their transposition. Their encoded RNA is reverse transcribed using their own (or not) encoded enzymes that reform a double-stranded DNA from the single-stranded RNA matrix at a new location. LTR retrotransposon integration occurs by cleavage and strand-transfer reaction catalyzed by an integration, similar to retroviruses (Brown et al., 1987). On the other hand, non-LTR retrotransposons include both long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs)s (Luan et al., 1993).

The two major superfamilies of LTR retrotransposons are Ty1/*Copia* and Ty3/*Gypsy*, which are classified based on the retroviral structural homology and domain order organization of the *pol* gene (Capy, 2005). These LTR retrotransposons exist universally in plant and animal genomes (Malik and Eickbush, 2001; Mangiavacchi et al., 2021). LTR-RTs are more active in plants and their functions are fine-tuned by epigenetic modifications. Although LTR retrotransposons in plants have attracted great attention in recent years, a more comprehensive understanding of the diverse functions of LTR retrotransposons can be gained from further studies. This review provides an overview of the processes associated with LTR retrotransposons involved in precise gene regulation in the plant genome. We also focus on the key role of LTR retrotransposon in plant heat response. Further, we discuss the LTR-derived small interfering RNA (siRNAs), LTR retrotransposon delivery system, centromeric LTR retrotransposons, the application of LTR-based molecular markers, and their contributions towards genome expression and evolution.

Genome-wide distribution and analysis of LTR families

LTR retrotransposons comprise about 10% to 90% of the total eukaryote genome in most plants. The distribution of LTRs differs among the major families of Ty3/*Gypsy* and Ty1/*Copia* elements in all plant genomes (Supplementary Table 1). Ty3/*Gypsy* elements are enriched in euchromatic sub-telomeric regions, whereas, Ty1/*Copia* elements are more frequent in heterochromatic pericentromeric regions (Jedlicka et al., 2019). Moreover, Ty3/*Gypsy* elements play crucial roles in host epigenetic response and are more heterogenous than Ty1/*Copia* elements. Although both families are found in a large number of copies in higher plants, these families were first identified in *Drosophila* (Sant et al., 2000). Members of these

superfamilies differ primarily in the arrangement of the gene coding for polymerase function within the polyprotein (POL) region. Ty1/*Copia* elements have a *pol* gene organized as the domains protease (PR), integrase (INT), reverse transcriptase (RT), and ribonuclease H (RNase H) (PR-INT-RT-RNase H). Ty3/*Gypsy* elements are organized as PR-RT-RNase H-INT domains (Sant et al., 2000).

LTR retrotransposons from genomes of about 300 plant species have been identified and are associated with diverse structural, functional annotation, and classification information (Zhou et al., 2021). Thus, this information may provide useful resources for investigating the evolutionary dynamics and functional implications of LTR retrotransposons in plant genomes (Kalendar et al., 2004; Moisy et al., 2014; Kalendar et al., 2020). Moreover, understanding the evolutionary forces governing TE polymorphism is crucial to understanding phenotypic variation in plants (Catlin and Josephs, 2022). Therefore, exploring the role of TEs leading to phenotypic variation and its regulation in plants has significant economic importance in the development of more efficient crops (Kalendar et al., 2008; Malaviya et al., 2021).

LTRs under heat stress

The impact of TEs on the structure, function and evolution of multiple plant genes have paved the way for epigenetic techniques that address diverse stresses in various crop species. TEs can be highly sensitive to different abiotic and biotic stresses, including salt, cold, heat, wounds, and infections (Mhiri et al., 1997; Ivashuta et al., 2002; Grandbastien et al., 2005; Buchmann et al., 2009; Naito et al., 2009; Ito et al., 2011; Lanciano and Mirouze, 2018). Several studies (Table 1) revealed that LTR retrotransposons become activated under certain epigenetic processes, such as siRNA regulation, DNA methylation, LTR retrotransposon integration, and chromatin modification (Grandbastien, 2015; Schorn et al., 2017). Moreover, LTR retrotransposons play a crucial role in the regulation of gene activity at the transcriptional and post-transcriptional level and in genome epigenetic regulation of stress resistance in a wide range of organisms (Galindo-González et al., 2017).

The transcriptional gene silencing of several LTR retrotransposons of *Arabidopsis* is accomplished by the loss of nucleosome and heterochromatin decondensation, which was restored upon recovery from heat stress (Pecinka et al., 2010). This indicates the role of environmental stress leading to epigenetic regulation. Moreover, heat-activated LTR retrotransposons play a crucial role in shaping a genome over an evolutionary period (Wessler, 1996; Masuta et al., 2018). Recently, we reported that the role of two LTR retrotransposons, *PHRE1* and *PHRE2* (Ty3/*Gypsy*), in Moso bamboo (*Phyllostachys edulis*) indicated that the 5' LTR acts as a promoter and can increase transposition activity during heat stress (Papolu et al., 2021).

A heat-responsive *ONSEN* retrotransposon is conserved among the *Brassica* species, and Adzuki bean exhibited upregulated transcript levels, and full-length extrachromosomal DNA accumulated in the stress-treated plants (Boonjing et al., 2020). The *ONSEN* family in most species of *Brassicaceae* showed integration into active chromatin, which was promoted by heat stress (Ito et al., 2013). Furthermore, there is a correlation between the heat-responsive elements (HREs) of *Copia* families and putative high-affinity heat shock factor binding HREs within the LTRs in seven *Brassicaceae* species. Moreover, the strong HRE of *ONSEN* is conserved over millions of years (Pietzenuk et al., 2016).

The active full-length Ty1/*Copia*, GBRE-1, showed increased expression under heat stress in *Gossypium hirsutum*, and its expression was similar to that of the *ONSEN* retrotransposon (Cao et al., 2015). The heat stress response and heat accumulation of Ty3/*Gypsy* retrotransposon in *Cryptomeria japonica* exhibited differential expression due to preheating treatment with heat shock factors, indicating the impact of LTR retrotransposons in the regulation of heat response systems in plants (Ujino-Ihara, 2020).

Several studies revealed the active role of *ONSEN* in regulating heat stress (Cavrak et al., 2014; Nozawa et al., 2021), including the regulatory role of siRNA. In *Arabidopsis*, *ONSEN* is activated by protracted exposure to heat stress (Ito et al., 2011; Matsunaga et al., 2012; Matsunaga et al., 2015; Ito et al., 2016). The genetic consequences of transposition bursts of the *Arabidopsis* LTR retrotransposon *Copia*78 family generated a novel progeny of chromosomally integrated LTRs consisting of a high frequency of intrafamily recombination and significant sequence diversity of LTR retrotransposons under heat stress (Sanchez et al., 2017). However, the role of LTRs, especially the Ty1/*Copia* and the Ty3/*Gypsy* superfamilies, requires further investigations to reveal their role in heat stress regulation. Such investigations will further the possibilities of developing crops to increase resistance to heat stresses due to global warming.

The function of LTR-derived siRNA biogenesis

Small noncoding RNAs (sRNAs) are the sequence-specific modulators of gene expression and precisely involved in the regulation of plant immunity (Borges and Martienssen, 2015). sRNAs interfere with the expression of particular genes with complementary nucleotide sequences by degrading mRNA after transcription, thus preventing translation (Laganà et al., 2015). Based on differences in biogenesis and function, sRNAs can be classified into several major classes, including: microRNAs (miRNAs), hairpin-derived siRNAs (hp-siRNAs), natural antisense siRNAs (natsiRNAs), heterochromatic siRNAs

TABLE 1 Summary of LTRs under heat stress and resulting phenotypes.

Target	LTR family	Host plant	Promoter	Findings	Temp point	Reference
*Copia*78	Ty1/*Copia*	*Arabidopsis*	35S promoter	Epigenetic regulation at ambient temperature was transcriptionally activated upon exposure of *Arabidopsis* plants to prolonged heat stress	37°C for 30 h	(Pecinka et al., 2010)
Copia-type retrotransposons	Ty1/*Copia*	*Arabidopsis*	35S promoter	Heterochromatin-associated silencing in *Arabidopsis* plants subjected to a particular temperature regime is released in a genome-wide manner	37°C for 15 h	(Tittel-Elmer et al., 2010)
ONSEN	Ty1/*Copia*	*Arabidopsis*	*ONSEN* promoter	*ONSEN* insertions confer heat responsiveness to nearby genes	37°C for 24 h	(Ito et al., 2011)
Retrotransposon-like sequences (*LEA*, *P5CS2*, *AbaH*)	Ty3/*Gypsy* and Ty1/*Copia*	*Pinus sylvestris*	LTR promoters	The transcriptional activation of different types of retrotransposon elements in the Scots pine genome was observed in response to heat-stress conditions	40°C for 16 h	(Voronova et al., 2011)
ONSEN	Ty1/*Copia*	*Arabidopsis*	*ONSEN* promoter	Under stress, high accumulation of the transcripts and amplified DNA copies of *ONSEN* were detected in callus	37°C for 24 h	(Matsunaga et al., 2012)
*FaRE*1	Ty1/*Copia*	*Fragaria ananassa*	*FaRE*1 promoter	The promoter of FaRE1 may act as different signal transduction pathways in response to stress	47°C for 32 h	(He et al., 2012)
ONSEN	Ty1/*Copia*	*Arabidopsis*	*ONSEN* promoter	Heat-induced transcriptional activation of *ONSEN* family in several species of Brassicaceae	37°C for 24 h	(Ito et al., 2013)
ONSEN	Ty1/*Copia*	*Arabidopsis*	*ONSEN* promoter	Plant heat shock transcription factor in periods of heat stress exploits the heat stress response to achieve transposon activation	37°C for 30 h	(Cavrak et al., 2014)
PtIGF7, PtGypsyX1, PtCopiaX1	Ty3/*Gypsy* and Ty1/*Copia*	*Pinus sylvestris*	LTR promoters	Stress conditions induced transcriptional activation of a wide range of retrotransposon sequences	40°C for 16 h	(Voronova et al., 2014)
GBRE-1	Ty1/Copia	*Gossypium barbadense* and *G. hirsutum*	GBRE-1 promoter	The expression level was increased under the heat-stress condition in *G. hirsutum*.	37°C for 24 h	(Cao et al., 2015)
ONSEN	Ty1/*Copia*	*Arabidopsis*	*ONSEN* promoter	Transcriptional activation of *ONSEN* was regulated by a small interfering RNA (siRNA)-related pathway, and the activation may also be induced by stress	37°C for 24 h	(Matsunaga et al., 2015)
ONSEN	Ty1/*Copia*	*Arabidopsis*	*ONSEN* promoter	Transposons activated by environmental stress may alter the genome in a potentially powerful manner	37°C for 24 h	(Ito et al., 2016)
ONSEN	Ty1/*Copia*	*Arabidopsis*	*ONSEN* promoter	Several *ONSEN* copies in Col-0 were activated by heat stress and maintained their transpositional activity in the progeny	37°C for 24 h	(Masuda et al., 2016)
ONSEN (*Copia*78)	Ty1/*Copia*	*Arabidopsis*	*ONSEN* promoter	*ONSEN* heat-responsive elements (HREs) accumulated mutations and lost heat-responsiveness	37°C for 24 h	(Pietzenuk et al., 2016)
ONSEN	Ty1/*Copia*	*Brassicaceae*	*ONSEN* promoter	Several new insertions were detected in a regenerated plant derived from heat-stressed tissues and its self-fertilized progeny	37°C for 24 h	(Masuta et al., 2017)
Copia78 or *ONSEN*	Ty1/*Copia*	*Arabidopsis*	*ONSEN* promoter	Chromosomally integrated LTR retrotransposons consisting of pairwise recombination products were produced in a process comparable to the sexual exchange of genetic information	37°C for 24 h	(Sanchez et al., 2017)
ONSEN	Ty1/*Copia*	*Arabidopsis*	*ONSEN* promoter	High inter-and intraplant variation in the number and chromosomal position of new insertions	37°C for 24 h	(Gaubert et al., 2017)
ONSEN	Ty1/*Copia*	*Vigna angularis*	*ONSEN* promoter	*ONSEN* element can be fully activated in the calli	40°C for 24 h	(Masuta et al., 2018)
HuTy1P4	Ty3/*Gypsy* and Ty1/*Copia*	*Hylocereus undatus*	*Pitaya* LTRs promoter	The Ty1/*Copia* and Ty3/*Gypsy* retrotransposons were usually silent but maybe expressed after exposure to abiotic stresses	45°C for 24 h	(Nie et al., 2019)
HUO	Ty1/*Copia*	*Oryza* genus	LTR promoters	Multiple HUO copies may trigger genomic instability through altering genome-wide DNA methylation and small RNA (sRNA) biogenesis and changing global gene expression, resulting in decreased disease resistance and yield	45°C for 10 h	(Peng et al., 2019)
LTRs (*CJHS018732* and *CJHS031206*)	Ty3/*Gypsy*	*Cryptomeria japonica*	LTR promoters	The expression of Ty3/*Gypsy* type retrotransposons was dramatically induced under stress	45°C for 120 min	(Ujino-Ihara, 2020)
ONSEN	Ty1/*Copia*	*Arabidopsis*	*ONSEN* promoter	Extrachromosomal DNA of *ONSEN* accumulated in heat-treated plants	40°C for 24 h and 28° C for 24	(Boonjing et al., 2020)

(Continued)

TABLE 1 Continued

Target	LTR family	Host plant	Promoter	Findings	Temp point	Reference
Heat-induced LTRs	Ty3/*Gypsy* and Ty1/*Copia*	*Arabidopsis*	LTR promoters	Heat activation of TEs exhibited a high correlation with the reduction of chromosomal interactions involving peri centromeres	37°C for 72 h	(Sun et al., 2020)
ONSEN	Ty1/*Copia*	*Arabidopsis*	*ONSEN* promoter	Under heat stress, loss-of-function of chromomethylase3 (CMT3) mutation led to increased CHH methylation at *ONSEN*	37°C for 24 h	(Nozawa et al., 2021)
LTR/*Copia* and LTR/*Gypsy*	Ty3/*Gypsy* and Ty1/*Copia*	*Arabidopsis*	*ONSEN* promoter	HistoneH1 repressed *Copia* elements by maintaining DNA methylation under heat	37°C for 36 h	(Liu et al., 2021)
MAGO1/2	Ty3/*Gypsy*	*Zea mays*	pCsVMV promoter	Argonaute-dependent, RNA-guided mechanism is critical in maize plants to sustain male fertility under stress conditions	38°C for 8 h	(Lee et al., 2021)
ONSEN	Ty1/*Copia*	*Arabidopsis*	*ONSEN* promoter	*ONSEN* transcript level was increased in the drd1 mutant relative to wild type under heat stress	37°C for 24 h	(Takehira et al., 2021)

(hetsiRNAs) and secondary siRNAs. miRNAs and siRNAs are the two major classes of plant sRNAs. The role of miRNAs in plant development, immunity, and intracellular immune receptors is well documented (Song et al., 2019; Wang S. et al., 2021; Dong et al., 2022). siRNAs are best known for their role in silencing viral RNAs, replication, and genome reprogramming (Kong et al., 2022).

siRNAs are specifically generated from double-stranded RNA (dsRNA) precursors derived from noncoding transcripts, inverted repeat sequences, sense and anti-sense transcripts, and exogenous RNAs (Xie et al., 2004). The dsRNAs are primarily processed into mature 21-24-nt siRNAs by various Dicer-like enzymes (DCL 1-4) and loaded into AGOs to form RISCs. DCL1 processes primary miRNAs into 21-nt-long mature miRNAs. DCL2 is involved in antiviral strategies and cleaves viral dsRNA into 21-22 nt long siRNAs, which target viral transcripts. DCL3 is involved in silencing processes targeting TEs and produces siRNAs approximately 24 nt in length. Finally, DCL4 generates 21-nt transacting siRNAs (tasiRNAs), which silence specific genes. siRNAs can be divided into two main classes: RDR6-dependent secondary siRNAs and RNA polymerase IV-dependent siRNAs (P4-siRNAs) (Kong et al., 2022). Secondary siRNAs are generated by transcripts from noncoding genes, e.g., tasiRNA loci, and protein-coding genes within large gene families, e.g., the nucleotide-binding leucine-rich repeats (NB -LLRs) (Sanan-Mishra et al., 2021). P4-siRNAs, especially 24-nt long, are mainly produced by heterochromatic regions, and TEs are linked to RdDM to induce transcriptional gene silencing (Ito, 2012; Lopez-Gomollon and Baulcombe, 2022).

siRNA pathways are significantly involved in retrotransposon silencing and may mediate different forms of epigenetic regulation in plants (Figure 1) (Lippman et al., 2003). In addition, siRNAs derived from TEs act as a trigger for host silencing mechanisms (Table 2). For example, siRNA silencing of a different class of LTR-retrotransposon mutants was shown to impact retrotransposon methylation, chromatin remodeling, and histone modification in *Arabidopsis* (Lippman et al., 2003). The mutagenic activity of LTR retrotransposons, especially in the pollen vegetative nucleus of *Arabidopsis*, is suppressed by siRNA silencing that may transmit the TEs to next-generation offspring (Slotkin et al., 2009). Remarkably, siRNAs suppress transposons by RNA-directed DNA methylation (RdDM), thus in turn leading to TEs becoming epigenetically silenced (Nosaka et al., 2012). In maize, loss of RNA-dependent RNA polymerase 2 (RDR2) function in the mediator of paramutation1 (mop1) results in the reactivation of transcriptionally silenced mutator retrotransposon and a substantial reduction in the accumulation of siRNAs. This suggests that the RDR2 pathway is an independent mechanism for silencing LTR retrotransposons in complex genomes like maize (Jia et al., 2019).

In *Arabidopsis*, siRNA targeted LTR retrotransposons are associated with reduced gene expression due to RdDM silencing. However, the effect of RdDM silencing was lower in *A. lyrata*, and thus showed differential transposon proliferation among species (Hollister et al., 2011). In addition, the transcriptionally active LTR retrotransposons in *Arabidopsis* produced RdDM-dependent siRNAs, indicating the function of RNA-dependent RNA Polymerase 6 (RDR6) and RNA Polymerase IV (Pol IV). These are independent in the silencing of TEs, in which Pol IV-RdDM functions to initiate TE silencing in an RNA Polymerase II expression-independent manner. In contrast, RDR6-RdDM functions to recognize active Polymerase II-derived TE mRNA transcripts to reestablish DNA methylation and TE silencing (Nuthikattu et al., 2013). Moreover, the targeting specificity of RDR6-RdDM function for full-length LTR retrotransposons in *Arabidopsis* have full-length transposon mRNA to be cleaved by primary 21-22-nt siRNAs and thus the RNA cleavage specificity drives the initiation of epigenetic transcriptional silencing targeted to LTR retrotransposons and transgenes (Panda et al., 2016). The function of DNA methylation to transcriptionally active LTR retrotransposons has demonstrated that mRNA-derived 21-22-nt siRNAs are directly incorporated into the ARGONAUTE 6 (AGO6) protein and in turn guide the AGO6 to its chromatin

10.3389/fpls.2022.1064847

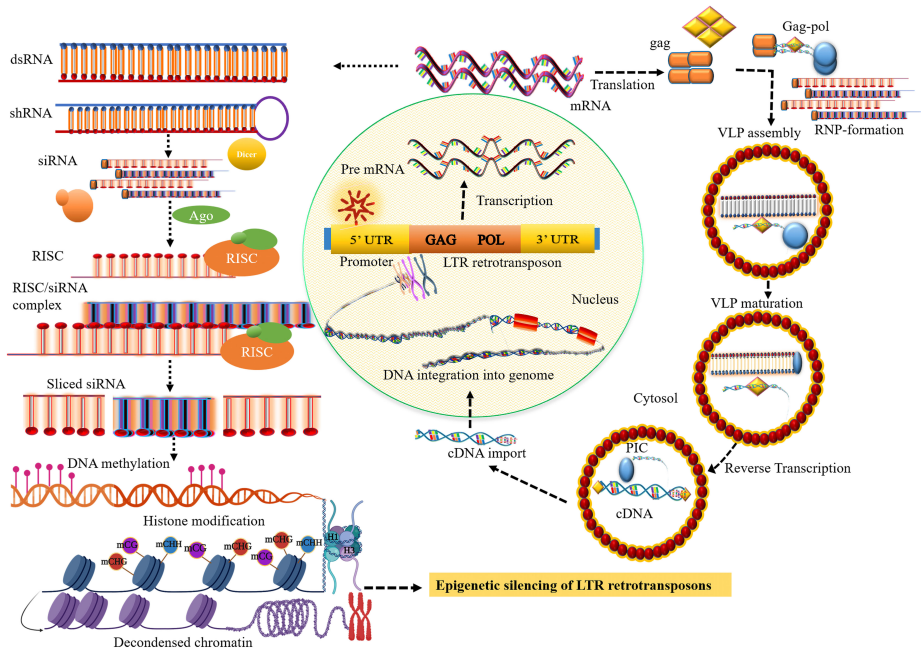


FIGURE 1
Small RNA biosynthesis and the transposition mechanism of long terminal repeat (LTR) retrotransposons (LTR retrotransposons). The transposition cycle initiates with the transcription of LTR retrotransposon that produces RNA template strand and major structural proteins for reverse transcription. After cleavage of the major structural gag-pol polyprotein by the viral protease (PR) activity, the gag protein-containing capsid and nucleic acid-binding domain involve the formation of cytoplasmic virus-like particles (VLPs). The polyprotein, pol comprising catalytic domain for replication encodes pepsin-like aspartate proteases, integrase, reverse transcriptase, and ribonuclease H proteins, which are crucial for reverse transcription and transposition of retrotransposons. Collectively with the RNA template, reverse transcription most likely takes place within VLPs that produce the cDNA which is then imported into the nucleus. The integrase that involves the formation of DNA nicks at the target sites is inserted into a new chromosomal locus to generate a new copy of retrotransposons and their insertion into the genome. The core siRNA silencing pathway: dsRNA or shRNA is processed into siRNA duplexes by Dicer RNase III. Subsequently, the siRNA or RNA-induced silencing complex (RISC) then binds to the complementary sequence of the target mRNA resulting in the degradation of the target transcript, establishing methylation of DNA through RdDM, and inducing histone modification, and heterochromatin formation.

targets to establish epigenetic transcriptional silencing of TEs in RdDM (McCue et al., 2015).

Recently, Nerd, a plant-specific GW repeat protein triggered by siRNA-dependent DNA methylation in *Arabidopsis*, was found to play a central role in integrating chromatin-based RNA silencing supported by binding both histone H3 and Ago2 proteins and to contribute to siRNA accumulation at a Nerd-targeted locus of LTR retrotransposons. This suggests that RdDM might preferentially target LTR retrotransposons and other repeat sequences (Pontier et al., 2012). The establishment of virus-induced gene silencing (VIGS) mediated RdDM function in *Arabidopsis* requires RNA Polymerase V (Pol V) and *de novo* methyltransferase 2 (DRM2). However, dicer-like-3 and Pol IV pathway components are not required for such functions. Perhaps the DNA methylation in VIGS is guided by virus-derived 21-22-nt siRNAs, thus suggesting VIGS-RdDM is a tool for retrotransposon silencing in *Arabidopsis* (Bond and Baulcombe, 2015). Later, the retrotransposon virus-like particles in *Arabidopsis* are activated by DDM1 mutations, giving rise to 21-22-nt siRNA through RNA-dependent RNA polymerase 6 (RDR6). This suggests that virus-like particle (VLP) DNA could also provide a powerful tool for identifying active LTR retrotransposons from the complex genome and their control at the transcriptional and post-transcriptional levels (Lee et al., 2020a). However, TE-derived siR815 drives RdDM of ST1 promoter and leads to transcriptional suppression of ST1, which abolished the WRKY45 transcription factor in rice resistance to *Xanthomonas oryzae* (Zhang et al., 2016).

The stress-induced full-length *Rider* LTR retrotransposons in the tomato genome indicate that RdDM controls *Rider* activity through siRNA production and DNA methylation, which may contribute to phenotypic variation through epigenetic alteration induced during environmental stress (Benoit et al., 2019). Furthermore, *Arabidopsis* mutations in the Argonaute9 protein (AGO9) indicated that AG09 can interact with 24-nt small RNAs (sRNA) corresponding to LTR retrotransposons expression in the ovule. AGO9 is also necessary for silencing repetitive genomic regions involved in heterochromatin formation. Thus, the AGO9-dependent pathway may be responsible for the epigenetic control of gametogenesis in

TABLE 2 List of small interfering RNA (siRNA), micro RNAs (miRNAs), and small RNAs (sRNAs) derived from LTR retrotransposons and their functions.

Plant species (siRNA/ miRNA, size)	Expression pattern	Response	Reference
Arabidopsis (siRNAs, 25)	Down	LTR-siRNAs tend to be susceptible to different forms of epigenetic regulation	(Lippman et al., 2003)
Arabidopsis (siRNAs, 24)	Up	siRNA produced from TEs activated in the pollen vegetative nucleus can target silencing in gametes	(Slotkin et al., 2009)
Maize (mop1) (siRNAs, 24)	Down	RDR2 pathway is an independent mechanism for silencing retrotransposons, genes, and siRNAs	(Jia et al., 2009)
Arabidopsis (siRNAs, 24)	Down	RNA-directed DNA methylation (RdDM) silencing is lower in *Arabidopsis*, which may lead to differential transposon proliferation among species	(Hollister et al., 2011)
MuDR element of maize (siRNAs, 24)	Up	RNA silencing pathway is associated with reduced expression of a regulator of trans-acting siRNA (tasiRNA) pathway and changes in epigenetic regulation of a maize transposon	(Li et al., 2010)
Arabidopsis MOM1 (siRNAs, 21-24)	Up	Functional cooperation of MOM1 and Pol-V regulates the degree of transcriptional gene silencing (TGS).	(Yokthongwattana et al., 2010)
Arabidopsis (siRNAs, 24)	Up	AGO9 preferentially binds to 24-nt sRNAs and may be a significant source of silencing LTRs in ovule	(Durán-Figueroa and Vielle-Calzada, 2010)
Arabidopsis ONSEN (siRNAs, 21-24)	Down	Plays a crucial role in the siRNA pathway in restricting a burst of retrotransposition that may generate novel, stress-responsive regulatory gene networks	(Ito et al., 2011)
Veju element of Wheat (siRNAs, 24)	Down	Intergeneric hybridization and allopolyploidization result in the deregulation of sRNAs and the associated reduction in transposon methylation	(Kenan-Eichler et al., 2011)
Rice (miR820, 24)	Down	The sRNAs silencing might act as a regulator of interactions between hosts and their parasitic elements	(Nosaka et al., 2012)
Arabidopsis (siRNA854, 21-22)	UP	Stress response mediated by siRNA854 incorporation into Argonaute1 protein regulates UB1b gene expression during cellular stress	(McCue et al., 2012)
Arabidopsis (siRNAs, 21-24)	Up	Nerd protein triggers chromatin-based RNA silencing pathway in plants	(Pontier et al., 2012)
Arabidopsis (siRNAs, 21-24)	Up	Distinct functions of Pol IV-RdDM and RDR6-RdDM collectively reestablish transposon methylation and epigenetic silencing	(Nuthikattu et al., 2013)
Arabidopsis Evade (EVD) (siRNAs, 21-24)	Up	Potent trans silencing by 24-nt LTR-derived siRNAs can establish functional *de novo* TE silencing at EVD-proximal genes	(Marí-Ordóñez et al., 2013)
Rice (*OsDCL3a*) (siRNAs, 24)	Down	*OsDCL3a*-dependent 24-nt siRNAs derived from transposons influence the expression of nearby genes and affect functional agricultural traits in rice	(Wei et al., 2014)
Arabidopsis (easiRNAs, 21)	Down	miRNA-directed 21-nt easiRNA biogenesis preferentially targets long-term heterochromatic silencing and host defense	(Creasey et al., 2014)
Arabidopsis (siRNAs, 21-22)	Up	21-22 nt siRNAs are directly incorporated into the AGO6 protein and guide AGO6 to its chromatin targets to establish TE-RdDM function	(McCue et al., 2015)
Arabidopsis virus-derived sRNAs (21-24)	Up	Virus-derived 24-nt sRNAs can reinforce VIGS-RdDM as a tool for epigenetic silencing	(Bond and Baulcombe, 2015)
Rice siR815 (21)	Down	TE-siR815-induced suppression of promoter elements of st1 results in WRKY45-mediated disease resistance by RdDM	(Zhang et al., 2016)
Arabidopsis sRNAs (23-24)	Up	RDR6-RdDM preferentially targets LTRs and suppressing mobilization in plants is epigenetically inherited in new generations	(Panda et al., 2016)
Strawberry (*fve–miR1511*, 24)	Up	miRNA targets LTR silencing and specifically contributes to genome stability, size, and architecture	(Šurbanovski et al., 2016)
Arabidopsis (siRNAs, 24)	Down	siRNAs independent of DCLs (sidRNAs) are associated with Ago4 and may drive heterochromatin DNA methylation	(Ye et al., 2016)
Arabidopsis (sRNAs, 24)	Down	Pol IV-dependent sRNAs (P4 RNAs) produced by Pol IV and RDRs may function as trigger RNAs to initiate DNA methylation by dicer-independent RdDM	(Yang et al., 2016)
Moso bamboo (siRNAs,21-24)	Down	Both 21-nt siRNA and-nt siRNAs derived from LTRs may be involved in the epigenetic regulation of host genes and may be responsible for diverse phenotypes	(Zhou et al., 2017b)
Arabidopsis (siRNA854, 24)	Up	Transposon-derived siRNA854 produced in the vegetative cell of pollen controls translation of UBP1b connected to triploid seed viability	(Wang et al., 2018)
Sweet pepper (miRNAs, 24 and siRNAs, 24)	Up and down	Differentially expressed 24-nt hetsiRNAs and 21-nt and 24-nt phasiRNAs may be employed to improve the quality and quantity of fruit	(Taller et al., 2018a)

(Continued)

TABLE 2 Continued

Plant species (siRNA/ miRNA, size)	Expression pattern	Response	Reference
Norway spruce pollen (sRNAs, 24)	Up	Tissue-specific transposon-derived 24-nt sRNAs may provide insights into the functional diversification of sRNAs in TE between gymnosperms and angiosperms	(Nakamura et al., 2019)
Tomato Rider (siRNAs, 24)	Up	Rider stress-induced retrotransposon may be a potential source of epigenetic variations involving siRNAs and RdDM pathway	(Benoit et al., 2019)
Arabidopsis embryonic (siRNAs, 24)	Up and down	Chromatin-mediated *de novo* production of sRNAs may provide cell-autonomous homeostasis to help reestablish euchromatic and heterochromatic states	(Papareddy et al., 2020)
Arabidopsis (sRNAs, 21-24)	Up	Pol IV switches to generating 21-22 nt siRNAs that are associated with AGO1 and may function in regulating gene expression	(Panda et al., 2020)
Arabidopsis (sRNAs, 21-22)	Down	The 21-22nt easiRNAs that depend on RDR6 may be responsible for LTR silencing at transcriptional gene silencing (TGS) and post-transcriptional gene silencing (PTGS) levels	(Lee et al., 2020b)

plants (Durán-Figueroa and Vielle-Calzada, 2010). In a recent report on pepper, pepper-specific heterochromatin-associated 24-nt siRNAs (hetsiRNAs) and 21-24-nt phased siRNAs (phasiRNAs) produced from transposons were preferentially expressed in seeds and placenta, indicating that pepper fruit quality and quantity is associated with changes in sRNA abundance (Taller et al., 2018b). The dynamics of TE-derived embryonic siRNAs in *Arabidopsis* could promote re-methylation of euchromatic and heterochromatic TEs in a new generation, therefore the decondensed chromatin-mediated 24-nt siRNA transcription may provide cell-autonomous silencing of transposons (Papareddy et al., 2020). The TE-siRNAs generated by plant-specific Pol IV can participate in RdDM, whereas other siRNAs and microRNAs (21-22-nt) are associated with *Argonaute1* (AGO1), suggesting that Pol I-dependent 21-22-nt siRNAs may participate in post-transcriptional regulation (Panda et al., 2016; Panda et al., 2020).

In maize, a link between the vegetative phase and the initiation of epigenetic silencing of MuDR retrotransposon is associated with a reduction of mutant expression during plant development. This is associated with an increase in trans-acting siRNA (*LRR*s) levels, which in turn is responsible for silencing epigenetic regulation of the MuDR element (Li et al., 2010). The regulatory interplay between *MOM1* mutants of LTR-retrotransposon in *Arabidopsis* and RNA polymerase-V may regulate the intensity and siRNA accumulation at the transgenic locus and the transcriptional gene silencing at the locus is accompanied by DNA methylation (Yokthongwattana et al., 2010). The heat-induced *ONSEN* retrotransposon in *Arabidopsis* showed its accumulation was stimulated in mutants deficient in the biogenesis of siRNAs, suggesting a considerable role of the siRNA pathway triggered by environmental stress during retrotransposition (Ito et al., 2011).

In wheat, high-throughput sRNA sequencing in parental, hybrid, and allopolyploid plants showed that miRNAs and the TE-derived siRNAs respond differently to changes at the ploidy level, and the siRNA pools were significantly reduced upon allopolyploidization. This, in turn, causes siRNA deregulation and the associated reduction in CpG methylation of LTR retrotransposons, which may contribute to genome instability at the initial stage of speciation (Kenan-Eichler et al., 2011). The Fatima family LTR retrotransposons of polyploid wheat are highly specific to B-genome and proliferated before allopolyploid wheat formation (Salina et al., 2011). Likewise, in hexaploidy wheat TriRe-1, LTR retrotransposons have a specific amplification history of B-genome progenitors, implying that genome-specific TriRe-1 may be utilized for the development of wheat molecular markers (Monden et al., 2014b).

In rice, a transposon produces microRNA820 (miR820) to suppress host silencing. The miR820 negatively regulates the expression of *de novo* DNA methyltransferase gene OsDRM2, indicating that transposon-derived siRNA silencing might act as a regulator of interactions between the host and their TEs (Nosaka et al., 2012). The Dicer-like 3 homolog OsDCL3a produces 24-nt siRNAs that target gibberellin (GA) and brassinosteroid (BR) homeostasis-related genes by association with TEs, which suppress the expression of nearby genes and may control important agricultural traits in rice (Wei et al., 2014). Whereas Dicer-like (DCL) proteins and 24-nt siRNAs are not required for DNA methylation at RdDM target loci, P4 sRNA transcripts generated by Pol IV and RNA-dependent RNA polymerases (RDRs) may function as RNA-triggered gene silencing of retrotransposons to initiate DNA methylation through the RdDM pathway (Yang et al., 2016). However, the biogenesis of TE heterochromatin-associated siRNAs in *Arabidopsis* is mechanically distinct from gene-regulating microRNAs (miRNA) or tasiRNAs. This suggests that the TE-derived siRNA854 regulates UBP1b mutant gene expression during the stress response, and the accumulation of siRNA854 is under the same trans-generational epigenetic regulation and inheritance pattern as the *Arabidopsis* LTR retrotransposons (McCue et al., 2012). Evd LTR-derived 24-nt siRNAs can silence transactive Evd copies in *Arabidopsis*. Reciprocal crossing between F11 and F14 plants resulted in the silencing of all F11-derived Evd copies. In addition, an Evd

RNA and 3′ gag-derived siRNAs of 21-22 nt were below detection in F1 plants, indicating effective trans silencing by LTR-triggered 24-nt siRNAs (Marí-Ordóñez et al., 2013).

In *Arabidopsis*, Post-transcriptional gene silencing (PTGS) mediated by miRNA-directed siRNA biogenesis specifically targets retrotransposon transcripts, whereas transcriptional gene silencing (TGS) of LTR retrotransposons is mediated by 24-nt heterochromatic (het) siRNA. Together, LTR retrotransposons give rise to the most abundant 21-nt epigenetically activated siRNAs (easiRNAs) in *ddm1* and methyltransferase1 (*met1*) mutants, and in the nucleus of pollen grains and callus cultures. Consequently, this supports an antagonistic relationship between PTGS and TGS in plants (Creasey et al., 2014).

In moso bamboo, both 21-nt siRNA and 24-nt siRNA have targets within LTR regions of retrotransposons. The high number of siRNAs derived from LTR retrotransposons may be responsible for diverse phenotypes of moso bamboo (Zhou et al., 2017a). The silencing mechanism of LTR retrotransposons is mediated by the most abundantly expressed miRNA, fve-miR1511. This fve-miR1511 is generated from a single locus that specifically targets LTR transcripts at the PBS site for methionyl initiator tRNA, which is essential for reverse transcription. This may contribute to features such as genome stability size and architecture in strawberries (Šurbanovski et al., 2016). The distinct class of 24-nt siRNAs independent of Dicer-like 3 (DCL3) is associated with effector AGO4 and is capable of driving DNA methylation and is subsequently subjected to 3'-5' exonucleolytic activity for maturation. Therefore, this class may be the initial trigger of *de novo* DNA methylation (Ye et al., 2016).

In addition, the transposon-associated sRNAs in pollen and cell culture of Norway spruce are responsible for tissue-specific and environmentally induced gene repression. This may provide insights into the diversification process of sRNA in transposon silencing between angiosperms and gymnosperms (Nakamura et al., 2019). The enhanced retrotransposon expression in *Botrytis cinerea* leads to the suppression of plant defense-related genes during infection. Retrotransposons are pathogenicity factors that manipulate host gene expression by encoding trans-species sRNAs (BcsRNAs) and therefore have a broad impact on host-microbe interactions and pathology (Porquier et al., 2021).

Previously, understanding of the sRNA activity in plants generally came from their prominent functions in plant development. Now, there is a greater understanding of the complex molecular mechanisms involved in sRNA biogenesis and function in plants. sRNAs play a significant role in the diversification and specialization of gene silencing. This is because there are several pathways for sRNA biogenesis and function, which are related to evolution. However, most sRNA classes contribute to biotic and abiotic stress and transgenerational inheritance, and the stability of acquired sRNA-based responses has not been characterized.

However, unless sRNAs in isolated cell types and single cells can be profiled, understanding of the specificities and interplay between the different gene-silencing mechanisms operating in plant cells will remain limited. Therefore, focused research on the aspects described above is necessary to manage stress-induced agricultural losses and the development of stress-resistant crops.

Nanoparticle-based LTR retrotransposon delivery system into plants

Nanomaterial-mediated delivery of biomolecules and therapeutics has been extensively studied in animals, but its potential for plant-based systems lags behind (Demirer et al., 2019a). Several previous studies have used nanoparticles to deliver plasmid DNAs (Cunningham et al., 2018; Wang J. W. et al., 2019; Lv et al., 2020), proteins (Wang J. W. et al., 2019; Wang J. W. et al., 2021), small interfering RNAs (Demirer et al., 2020), and intact plant cells (Serag et al., 2011; Demirer et al., 2019a). Carbon nanotubes have been used to perform stable genetic transformation in bacterial (Castillo et al., 2021; Weise et al., 2022) and mammalian (Golestanipour et al., 2018) cells. In our recent study, we used for the first time an efficient polyethylenimine (PEI)-walled carbon nanotube (SWNT) diffusion method to introduce the LTR retrotransposon plasmid DNA into Moso bamboo plants without transgene integration (Papolu et al., 2021) (Figure 2). We found that internalization of nanoparticles in the intact plant cells resulted in increased GFP expression in the leaves after 72 hours. The carbon nanotubes enable the transport of plasmids without integration of transgenes into crop plants (Kwak et al., 2019). GFP were expressed in various tissues such as roots, leaves, protoplasts, and immature tissues (Ali et al., 2022). Enhanced GFP expression in leaf protoplasts by the use of carbon nanomaterials has been demonstrated in arugula (*Eruca sativa*), *Gossypium hirsutum* (cotton), and wheat (*Triticum aestivum*) (Demirer et al., 2019b; Kwak et al., 2019). The use of nanoparticle-mediated transformation has also been demonstrated for siRNA gene silencing production (Demirer et al., 2020; Zhang et al., 2021). Further focused studies on LTR retrotransposon delivery system are required to explore the molecular mechanisms of LTR retrotransposons in the plant genome.

Roles of centromere-specific retrotransposable elements

LTR retrotransposons are greatly responsible for plant genome evolution and are enriched in the pericentromeric region of host genomes. Active retrotransposable elements are

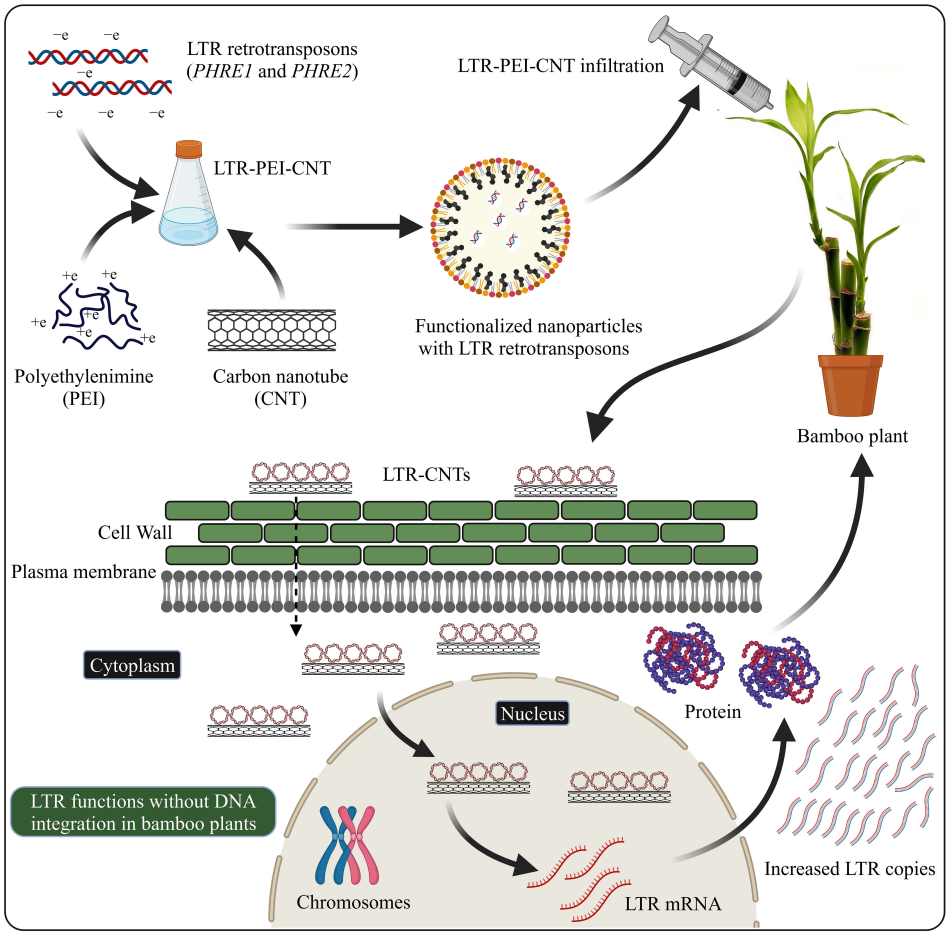


FIGURE 2
LTR retrotransposons with carboxylated carbon nanotubes (COOH-CNTs) and polyethylenimine (PEI) delivery into bamboo leaves. Covalently modified COOH-CNTs with PEI carrying a net positive charge are incubated with negatively charged LTR vectors. Bamboo leaves infiltrated with LTR–CNTs produce LTR transcripts and proteins and increased LTR copies, without the genome integration. The schematic representation is based on Demirer et al., (2019a) and Papolu et al., (2021) and created with BioRender.com.

also highly mutagenic and often target protein-coding genes for insertion. In addition, these elements cause chromosome breakage, illegitimate recombination, and genome rearr angement. Therefore, active retrotransposable elements are recognized to play a central role in maintaining chromatin structures, centromeric functions, and regulation of gene expression in their hosts (Shapiro, 2014). Moreover, they are largely responsible for plant genome size variation (Girard and Freeling, 1999; Slotkin and Martienssen, 2007). Centromeric sequences play a central role in chromosome distribution during the mitotic and meiotic cell lifecycle (de Castro Nunes et al., 2018). Centromeric retrotransposons (CR) were first discovered in the grass as centromere-specific sequences (Miller et al., 1998; Presting et al., 1998). Remarkably, in plants, they are usually surrounded and dispersed by LTR retrotransposon sequences (Neumann et al., 2011). The centromere-targeting retrotransposable elements can replace centromeric tandem repeats that bind centromere-specific proteins and may act as a substrate for the efficient repair of frequent double-stranded breaks (Presting, 2018). The centromere-specific histone H3 (CENH3)-associated sequences of centromeric retrotransposons and satellite DNAs are the important structural elements in epigenetic centromere function (Keçeli et al., 2020). Retrotransposable elements can be used to deduce centromere positions, as some elements target active centromeres during integration (Presting, 2018). However, the roles of retrotransposable elements in centromere functions remain unclear. Centromere-targeting elements may be able to replace centromeric tandem repeats. Therefore, centromeric retrotransposons of several plant species have been investigated during the last two decades.

In wheat, the FISH analysis revealed that the sequence of pHind258 was homologous to integrase and the LTRs of centromeric Ty3-*gypsy* retrotransposons of cereal species

(Ito et al., 2004). A 67-kb clone (R11H) containing Ty3/*gypsy* retrotransposon-related sequences was also identified, which showed strong hybridization signals on the centromeres (Fukui et al., 2001). The expansion of centromeric retrotransposon sequences on dicentric chromosomes to chromosome arms and the formation of multiple centromeres in wheat-rye hybrids may be responsible for chromosome breakage in the next-generation offspring and may be associated with chromosomal rearrangement, stability, and novel chromosome formations (Guo et al., 2016).

In cotton species, centromere-associated sequences are composed of A and D genomes, and the location of the functional centromere co-localizes with centromere retrotransposon hybridization on metaphase mitotic chromosomes. Additionally, FISH and dot-blot hybridization revealed that centromere retrotransposons are present only in D-genome diploid species, indicating that retrotransposons may have invaded the A-genome centromere from the D genome during allopolyploidization (Luo et al., 2012). In addition, LTRs generated from a sequenced bacterial artificial chromosome (BAC) were located in the D progenitor in *Gossypium raimondii* but not in the A progenitor *G. herbaceum*, indicating that the centromeric regions of triploid cotton may be derived from D progenitor (Zhang et al., 2014). Moreover, centromeric retrotransposable elements from the different progenitor genomes may become activated during genomic instability following allopolyploidization (Divashuk et al., 2016). Thus, allopolyploid offers an opportunity to understand the evolution of centromeric sequences from resident TEs (Hartley and O'Neill, 2019).

In maize, centromeric retrotransposons represent a transcriptionally active component of centromeres from a wide range of angiosperm species and play a central role in plant centromere evolution (Neumann et al., 2011). However, a recent study revealed that the centromeric retrotransposons can give rise to CRM1 and CRM4 tandem repeats in maize. Nevertheless, maize centromeres are fluid genomic regions whose borders are heavily influenced by the interplay of retrotransposons and epigenetic marks. Distinct CRM1TR sequence variation may lead to gene conversion, which is the main cause of sequence variation and may increase the size of the satellite repeat locus (Sharma et al., 2013). Furthermore, FISH and chromatin immunoprecipitation (ChIP) with anti-CENH3 antibodies in maize and soybean revealed that centromeres differ in size and contain a higher density of CENH3 chip reads, indicating that the tandem satellite repeats and interspersed centromeric retrotransposons may be shaped primarily by retrotransposons (Wolfgruber et al., 2009; Tek et al., 2010). Additionally, various repetitive elements in maize, including centromeric retrotransposon, CentC, and CentA, are found preferentially near the centromeres of the A chromosome hybridized to distinct sites from centromere on the B chromosome, revealing a high concentration of centromeric repeats at the major location on the B chromosome (Lamb et al., 2005). A comparative genomic analysis of centromeric retrotransposons in maize revealed that the maize B chromosome co-existed with the A chromosome during retrotransposition, suggesting that the B chromosome had its origins from A chromosome elements (Theuri et al., 2005). The cores of maize centromeres contain primarily CentC arrays and a cluster of centromere-specific retrotransposons of maize. The structural relationship between CentC, centromeric retrotransposons, and CENH3 was visualized by sequential detection procedure on stretched centromeres, demonstrating that the maize centromeres constantly incorporate oat CENH3 nucleosomes (Jin et al., 2005).

Tobacco cell lines have been identified with an expression of a HaloTag7-fused CENH3 centromeric-tandem repetitive DNA sequences located with CENH3 by a HaloTag7-based chromatin affinity purification system. Further, FISH and ChIP analysis indicated that repeats were chromosome-specific centromeric retrotransposons (Nagaki et al., 2012). Moreover, the centromeric retrotransposons derived from BAC clones act as centromeric DNA sequences in tobacco and the estimated amplification timings of centromeric retrotransposons were different in the two ancestral diploid species of tobacco, indicating that retrotransposons accumulate especially in CENH3-binding regions of tobacco species (Nagaki et al., 2011).

In *Brassica* species, centromere retrotransposons are the major repeats in centromeric and pericentromeric heterochromatin, and the distribution of the species in allotetraploid relatives indicates that repetitive elements are A-genome specific (Lim et al., 2007). In addition, ChIP and immunostaining analysis with anti-CENH3 antibodies showed that both centromere-specific retrotransposons and centromeric tandem repeats represent a dominant component of the diploid and allotetraploid *Brassica* species and are directly associated with CENH3 proteins (Wang et al., 2011). Recently, the centromeric-specific retrotransposon in *Brassica* species showed that the centromeric repeats spread and proliferated between the diploid species possessing A, B, or C genomes after polyploidization, implying that centromeric retrotransposons are particularly important in the evolution and polyploidization of the *Brassica* genome (Wang G.-X. et al., 2019). Furthermore, the repetitive elements in Brassica species that are conserved in pericentromeres, sub-telomeres, and telomeres rapidly diverged during the evolution of A/C and B genome lineages. Furthermore, these repeats may be associated with genomic stability and may provide insights into genome evolution during Brassica polyploidization (Koo et al., 2011). BACs derived from the rapid proliferation of nested LTR retrotransposons in *Brassica* species may play an evolutionarily important role in the formation of centromere regions (Wei et al., 2013).

In rice, the contribution of LTR retrotransposons to the evolution of gene structure and function indicates that Ty3/

Gypsy elements are more abundant than Ty1/*Copia* elements, and the intrachromosomal distribution of retrotransposons across chromosome 10 is non-random with the highest density being present in the pericentromeric region (Gao et al., 2004). Moreover, the structural features of LTR retrotransposons in rice indicated that centromeric retrotransposons and CentO satellite repeats are harbored in the core region of the rice chromosome 4-specific centromere, indicating the fragmental duplication of arrays of satellite repeats is mainly responsible for the amplification of centromere satellite DNA and rapid reshuffling of CentO satellites (Ma and Bennetzen, 2006). Although the centers of rice centromeres are occupied by a CentO satellite repeat and a centromere-specific retrotransposon, the CentO satellite is quantitatively variable among 12 rice chromosomes and is interrupted by centromeric retrotransposons, therefore suggesting that CentO satellite and centromere-specific retrotransposons may be the key DNA components for centromere function in rice (Cheng et al., 2002).

The position of CENH3 nucleosomes in rice centromeres is regularly spaced with 155-bp periodicity on CentO satellite repeats but not on non-CentO sequences, suggesting that centromeric repeats evolve for the stabilization of CENH3 nucleosomes (Zhang et al., 2013). Evidence also suggests that suppression of LTR retrotransposon proliferation through the formation of heterochromatin may be an advantage in large genomes in eukaryotes that have a high content of LTR retrotransposons (Cossu et al., 2017). The centromeric retrotransposons of rice are enriched with heterochromatin and its constitutive sequences are transcribed in all the tested rice organs. The centromeric transcripts are differentially processed into sRNAs, indicating a crucial role in the RNAi-mediated pathway for heterochromatin formation and centromere function (Neumann et al., 2007).

Recently, the phylogenetic relationships of centromeric retrotransposons in grasses show that horizontal transfer of centromeric retrotransposon between oryzoid (rice) and panicoid (maize, sorghum, *Setaria*, *Panicum*, and *Coix*) lineages and interelement recombination are important factors in the evolution of centromeric retrotransposons (Sharma and Presting, 2014).

In sugarcane, the characterization of centromere-associated DNA sequences indicated that centromeric retrotransposable elements and centromeric tandem repeats may directly interact with CENH3 in sugarcane centromeres (Nagaki and Murata, 2005). Moreover, the centromeric satellites had the formation and evolutionary stability for 7 million years and exhibited different ploidy levels and unusually longer monomeric repeats that lacked translation phasing on the CENH3 nucleosomes. This indicates that they originated from a retrotransposon and may form extrachromosomal circular DNAs (eccDNAs) (Huang et al., 2021).

In the grass family, centromere-specific retrotransposons discovered in BAC clones revealed that both centromere-specific and non-centromere-specific repeats are the primary DNA elements of maize centromeres and may play a significant role in grass family evolution (Nagaki et al., 2003). Similarly, a centromeric LTR retrotransposon of *Brachypodium distachyon* derived from centromeric BAC sequences was found in high copy number and is enriched in *B. distachyon* centromeric regions, indicating that *Brachypodium* centromeric retrotransposons are highly divergent among other grass species (Qi et al., 2013).

In the potato genome, retrotransposon-related sequences and satellite repeat-based centromeres can rapidly proliferate from neocentromeres by *de novo* amplification and can associate with the CENH3 nucleosome (Gong et al., 2012). The LTR retrotransposons identified using BAC inserts in Beta species have a chromodomain that is highly similar to centromeric retrotransposons in rice, maize, and barley. Based on sequence diversity, LTRs may have been transposed within the last 60 000 years, indicating that their large-scale genomic organization and transcriptional activity may play an important structural role in centromeres of chromosomes (Weber and Schmidt, 2009).

The annotations and comparison of the centromeric region of *Coffea*, which is rich in several centromeric retrotransposon family elements, showed that the role of LTR retrotransposons may be more diverse in plants and may extend beyond the chromodomains (de Castro Nunes et al., 2018). The centromeric region of tomato chromosome 12 is composed of nested repeat sequences, including LTR retrotransposons and chloroplast DNA insertions. A block of CAA trinucleotide microsatellite repeats was found in the centromere and pericentromeric region of chromosome 12, suggesting that microsatellite arrays like CAA blocks may be a component of tomato centromeres (Yang et al., 2005). A high copy number of tandem repeats in *Allium* species is located in all chromosomes and differs in sequence, structure, chromosome level, and genome organization. These repeats are transcribed and associated with the insertions of retrotransposons and organelle DNA, which can be used for future applications of its association with kinetochore protein CENH3 (Kirov et al., 2020). Likewise, the chromosomal organization of centromeric retrotransposons in the genomes of *Allium cepa* and *A. fistulosum* are localized in centromeric regions and the chromosomes of *A. fistulosum* are expressed less in centromeric regions and were abundant in other chromosomal regions (Kiseleva et al., 2014). Holocentromeres in *Rhynchospora pubera* is composed of centromeric units interspersing the gene containing chromatin. A cell-dependent shuffling of multiple centromeric units results in the formation of functional centromeres during mitosis; genome-wide analysis indicated that different types of holocentromeres may exist in different species, with and without repetitive elements among eukaryotes (Marques et al., 2015).

In *Arabidopsis*, the centromere-enriched retrotransposons are significantly diverged between two different species and can target their integration preferentially into the centromere

region on each of the different chromosomes in the karyotype (Birchler and Presting, 2012). Furthermore, the structure and organization of centromere-specific retrotransposons and CentO-F satellites in *Oryza brachyantha* indicate that CentO-F satellites are located within the chromosomal regions and are characterized by tandemly repeated satellite DNA flanked by centromeric retrotransposons. This may explain its potential impact on functional centromeres in *Oryza* species (Yi et al., 2013). FRetro3 centromeric retrotransposons are located in the functional domains of *O. brachyantha* centromeres and have replaced centromeric retrotransposons of rice as dominant centromeric retroelements in *Oryza* species (Gao et al., 2009). The retrotransposon of *A. lyrata* Tal1 was introduced into *Arabidopsis* by tissue culture-mediated transformation and showed that the highest retrotransposed copies were found in centromeric repeats of *Arabidopsis*, which suggests dynamic controls for the evolution of the retrotransposon-rich heterochromatin regions (Tsukahara et al., 2012). Furthermore, the structural heterozygosity and chromosomal rearrangements of tissue-specific retrotransposons and tandem repeat copy number in *Aegilops speltoides* indicate that the tissue-specific pattern of retrotransposons emerges during cell proliferation and this may reflect the reorganization of individual genomes under rapid environmental changes (Shams and Raskina, 2018). However, significant advancements in epigenetics and different types of plant centromeres may be essential to increase the number of sequenced genomes (Oliveira and Torres, 2018).

Over the last two decades, several varied approaches have been used to study the genomes of many plant species. Studies on agriculturally important plant species are particularly important. Following genome sequencing of crop plants, genome sequencing within the genus should be the next targeted research for genomic analysis. Further research should be conducted on genome organization and comparisons at the chromosome, sequence, functional, and evolutionary levels (Voronova et al., 2020).

Several studies demonstrated that LTR retrotransposons participate in centromere-specific transposition and may be a driving force in plant centromere evolution. However, there are many mechanisms involved in the organization of genome functions and in maintaining complex programs of genome organization. Therefore, studies resolving the questions above require novel technologies in molecular biological, cytogenetic, biochemical, and genetic methods. Such studies may provide a clearer understanding of the relationship between plant evolution and LTR retrotransposons.

Applications of LTR-retrotransposon as molecular marker system

Retrotransposable and related elements are highly abundant in eukaryotic genomes and insert into new genomic locations by a mechanism that involves the reverse transcription of an RNA intermediate. Changes in the copy number of repeat elements and internal rearrangements on both homologous chromosomes occur after the induction of recombinational processes during the meiotic prophase. The insertion of LTR retrotransposons is random and occurs in the transposition process in the continuous evolution of a species. This can provide a wealth of information for the study of evolution, species, and genome differentiation.

Retrotransposon-based DNA marker applications have become a key element of research on genetic variability and diversity (Vuorinen et al., 2018; Ghonaim et al., 2020; Kalendar et al., 2021b). The scope of their usage includes creating genetic maps and the identification of individuals or lines carrying certain genetic polymorphic variations (Khapilina et al., 2021a). LTR retrotransposon-derived molecular genetic marker systems have been employed in deciphering the genetic diversity of crop plants (Kalendar et al., 1999; Kalendar and Schulman, 2006; Kalendar et al., 2011; Kalendar et al., 2018; Kalendar et al., 2021a). The retrotransposon-based marker systems are highly effective in detecting the effects of environmental stress on retrotransposon activation (Kalendar et al., 2008; Belyayev et al., 2010). Moreover, the detection of TE expression, including polymorphisms and the diversity of the transposon transcriptional landscape, may provide new insight into host-TE interactions (Lanciano and Cristofari, 2020). In addition, LTR retrotransposons are associated with key genes involved in potential applications of genome assembly, genome variation, gene tagging, and functional analysis of genes, indicating their crucial role as markers in molecular breeding (Potter, 2005).

In pepper (*C. annuum*), LTR retrotransposons were inserted 6 million years ago and exhibit chromosomal insertional preferences, which may be a useful tool to design species-specific retrotransposon-based markers (Yañez-Santos et al., 2021). The combination of active LTR retrotransposons and Inter-Retrotransposon Amplified Polymorphism (IRAP) markers (Kalendar et al., 1999; Hosid et al., 2012) may be a suitable system for genetic fidelity assessment of tissue-culture-generated plants in sugarcane (Shingote et al., 2019) and better germplasm management in *Xanthosoma* and *Colocasia* (Doungous et al., 2015). The IRAP marker system in LTR retrotransposon insertions of flax genome appeared to be suitable for the identification of retrotransposon polymorphisms and showed a high level of plant adaptation in a radioactive environment (Smýkal et al., 2011; Lancíková and Žiarovská, 2020). The IRAP and REMAP markers of the cassava genome produced high polymorphism and may be suitable for the investigation of genetic diversity and relationships among cassava cultivars (Kuhn et al., 2016). A comparative analysis of two LTR retrotransposons, BARE-1 and Jeli, may provide a potential source of polymorphic Sequence-Specific Amplification Polymorphism (SSAP) markers for genetic diversity in diploid wheat (Konovalov et al., 2010). The LTR retrotransposon based SSAP markers in cashew and myrtle genomes exhibited a

significantly higher proportion of polymorphic markers than those of AFLP (Waugh et al., 1997; Syed et al., 2005; Woodrow et al., 2012).

The genetic maps generated with several retrotransposon-based markers such as iPBS (inter-Priming Binding Site) (Kalendar et al., 2010; Doungous et al., 2020; Khapilina et al., 2021b) and REMAP (REtrotransposon-Microsatellite Amplified Polymorphism) (Kalendar et al., 1999) exhibited regions of different marker densities, indicating that the distribution of retrotransposons in lentil is non-random and widespread throughout the lentil genome. This may be useful in lentil breeding by marker-assisted selection (Rey-Banos et al., 2017). The development of Retrotransposon-Based Insertion Polymorphism (RBIP) markers (Flavell et al., 1998) derived from sweet potatoes can determine intraspecific variability. These markers can also be used as core primer pairs for evaluating genetic diversity and constructing linkage maps of various plant species, guiding breeding and germplasm research (Meng et al., 2021). The RBIP marker was shown to be duplicated several times during the development of Asian pear cultivars and may provide a comprehensive picture of the complex relationship and evolution of *Pyrus* species (Jiang et al., 2015). Likewise, genome-wide analysis of RBIP markers in the *Melilotus* genome revealed considerable polymorphism information content (PIC), indicating that these markers are highly informative and may be used for implementing genetic improvement in the *Melilotus* genus (Ouyang et al., 2021). Furthermore, RBIP markers used for DNA profiling of Japanese, Chinese, and European pear cultivars revealed that retrotransposons have transposed during Asian pear evolution or reflect the genetic relationship between Asian and European pears. Thus, suitable combinations of retrotransposon insertions may be useful for cultivar-specific DNA markers (Kim et al., 2012). The polymorphism markers generated from several retrotransposon families and the effectiveness of the dominant (IRAP) and codominant (RBIP) marker systems for assessing the genetic diversity among different potato varieties were compared. Distinct DNA profiles for Ty1/*Copia* and Ty3/*Gypsy* retrotransposons are active in the genome and may contribute to potato genome organization (Sharma and Nandineni, 2014). High-throughput RBIP data analysis indicated that may strongly support the model of independent domestications for *Pisum sativum* species, which in turn provides a broad understanding of the diversity and evolution of *Pisum* (Jing et al., 2010). Likewise, a wide variety of LTR retrotransposon-based markers generated from peas, broad beans, and Norway spruce may be useful in revealing polymorphisms associated with the corresponding retrotransposons within the *Pisum* genus (Pearce et al., 1999). The non-random distribution of abundant LTR retrotransposons within the lentil genome indicates that defective and non-autonomous retrotransposons are highly frequent and maybe a suitable source of genetic markers for further genetic analysis (Rey-Banos et al., 2017). The novel Ty1/*Copia* and Ty3/*Gypsy* LTR retrotransposons derived from *Lilium* species indicate that they were non-autonomous retrotransposons. IRAP analysis using the LTR sequence of these retrotransposons may provide a new approach to analyzing the species relationship among *Lilium* species (Lee et al., 2015). In *Cleistogenes songorica* and strawberry genomes, various LTR retrotransposon-based molecular markers were developed and exhibited a high level of polymorphism frequency and high transferability of polymorphic primer pairs. This suggests that RBIP markers may be useful in future studies on genetic diversity, QTL mapping, population structure, and the evolution of germplasm accessions in *C. songorica* and related grasses (Monden et al., 2014a; Ma et al., 2022). Several LTR retrotransposon markers derived from chokecherry genome sequences indicated that retrotransposon markers in map construction and genetic mapping may facilitate genetic research in Rosaceous species (Liang et al., 2016).

Role of LTR retrotransposons in plant evolution

Evolution is primarily a change in physiological and genetic composition; therefore, variation is a significant process in evolution. Like in most eukaryotes, TEs are the most variable parts of the plant genome (Lisch, 2013). TEs can make dramatic differences in the overall architecture of the genomes of even closely related plant species. Moreover, TEs make up most of all plant DNA (Bennetzen et al., 2005). Gene inactivation is one of the most common TE-induced phenotypic changes. Therefore, the propensity of some TEs to insert into or near genes has been successfully utilized for generating new null mutations (Hsing et al., 2007; Settles et al., 2007; Candela and Hake, 2008), and this is also a major driver of genome size evolution (Hawkins et al., 2006; Neumann et al., 2006; Piegu et al., 2006) Therefore, the evolutionary potential of TEs, especially LTR retrotransposons, should be thoroughly explored to gain a better understanding on the evolutionary characteristics of plants. Retroviruses and LTR retrotransposons share similar gene architecture, but LTR retrotransposons lack the envelop gene and an extracellular stage in their lifecycle. It has been proposed that these retroviruses emerged from the LTR retrotransposon family Ty3/Gypsy by acquiring the envelope gene (Malik and Eickbush, 2001), but this evolutionary relationship is not confirmed.

Genome relationships and LTR retrotransposon diversity can be used to understand the genomic relationship among the members of a genus or family in plants. Recently, genome relationships and LTR retrotransposon diversity in three cultivated *Capsicum* strains were analyzed and a close

relationship among the species was revealed (de Assis et al., 2020). Moreover, genome-wide analysis of LTR retrotransposons and their impact on evolution has been explored in several plants (Roulin et al., 2009; Beulé et al., 2015; Giordani et al., 2016; Mascagni et al., 2017; Keidar et al., 2018; Liu et al., 2018; Akakpo et al., 2020; Mascagni et al., 2020; Ouyang et al., 2021). TE amplification is the main mechanism behind plant genome size increase and evolution (Gantuz et al., 2022). The proliferation of LTR retrotransposons is related to genome reorganization caused by hybridization or polyploidization (Vicient and Casacuberta, 2017). Moreover, allopolyploidization is associated with rapid structural and functional alterations of genomes (Leitch and Leitch, 2008) and this is recognized as the major mechanism behind adaptation and speciation in the plant kingdom (Ramsey and Schemske, 1998). In addition, polyploidy increases genome size and activates TE amplification, and the resultant genome rearrangement may alter their balance in epigenetic silencing (O'Neill et al., 1998; Ozkan et al., 2001; Madlung et al., 2005; Petit et al., 2007; Parisod et al., 2009). TEs are known to associate with recombination-driven sequence loss that leads to major structural changes (Parisod et al., 2010). In plants, TE abundance is correlated with the recombination rate of some TE families (Daron et al., 2014). In maize, LTR retrotransposons are enriched in regions of low recombination (Stitzer et al., 2021). Moreover, a negative correlation between LTR retrotransposons and recombination was also reported in many other plant species, such as soya bean, rice, and bread wheat (Tian et al., 2009; Tian et al., 2012; Daron et al., 2014). Angiosperm genomes are unstable at the level of chromosome number, genome size, and repetitive DNA content; most genes are found as single-gene groups surrounded by nested TEs (Sanmiguel and Bennetzen, 1998). Furthermore, in maize, any two alleles of the same gene diverged >2 million years ago (Wang and Dooner, 2006). Although gene content and organization are mostly similar, variation in copy number and gene order has been observed in grass plants (Bennetzen, 2007; Springer et al., 2009). However, variation in copy number and its influence on genome order and evolution should be explored to gain a better understanding of the influence of LTR retrotransposons in plant evolution.

In general, LTR retrotransposons are one of the key elements that drive evolution by mechanisms of recombination and gene duplications. Moreover, TEs affect the genome when the mobile elements are closer to the genome or even from a considerable distance. This is because TEs can move. Therefore, TEs, especially LTR retrotransposons, have a significant role in the evolution of the plant kingdom because of their wide occurrence. Further focused studies are required to explore the role and the exact process of LTR retrotransposons in plant evolution, which may provide further insight into the molecular mechanisms of evolution in the plant kingdom.

Future perspective

1. Retrotransposable elements represent up to 90% of the total genome in most plants. Several studies describe the role of LTR retrotransposons in epigenetic regulation. Exploring the role of retrotransposable elements leading to phenotypic variation and its regulation in plants may have significant economic importance in the field of plant breeding and agriculture.
2. Investigations on the role of LTR retrotransposons, especially the Ty1/*Copia* and the Ty3/*Gypsy* superfamilies, may reveal their roles in heat stress regulation, which will provide a further understanding of the possibilities of developing smart crops that are resistant to heat stresses due to global warming.
3. Genetic engineering methods and epigenetic modifications using LTR retrotransposons may have future scope in the field of smart agriculture by developing smarter crops.
4. Further research should focus on profiling sRNAs in isolated cell types and single cells. This may further understand the specificities and interplay between the different gene-silencing mechanisms in plant cells.
5. There is currently a limited understanding of most sRNA classes that contribute to biotic and abiotic stress and the transgenerational inheritance and stability of acquired sRNA-based responses. This should be a focus of further research in the development of stress-resilient crops and plant breeding in general.
6. LTR retrotransposons participate in centromere-specific transposition and may be a driving force in plant centromere evolution. Further studies should focus on genome organization and comparisons at the chromosome, sequence, functional, and evolutionary levels.
7. Genome-sequencing studies on agriculturally important plant species are important; genome sequencing within the genus should be targeted for subsequent research.
8. Significant advancements in epigenetics and different types of plant centromeres are required to increase the number of sequenced genomes. This increase should further understand the relationship between plant evolution and LTR retrotransposons.
9. Further investigations are necessary to gain a better understanding of the variation in copy number of LTR retrotransposons and its influence on evolution and genetic variation.

10. TEs, especially LTR retrotransposons, contribute significantly to intraspecific phenotypic variation in plants. Therefore, understanding the dynamics governing LTR retrotransposons is a crucial research focus for evolutionary biologists.

Conclusion

Retrotransposons are a class of mobile genetic elements that are universally distributed in plant genomes. Their distribution and transposition activities are significantly associated with plant evolution. Several studies of LTR retrotransposons have provided valuable insights into the mechanism of the genome evolution of plants. The genomes of most plant species exhibit dynamic variations in size and other structural features of LTRs. In chromatin modification, reduced DNA methylation often promotes the expression of retrotransposons. A wide variety of genetic factors are responsible for retrotransposon expressions, such as miRNAs, ncRNAs, piRNAs, RdRPs, risiRNAs, siRNAs, ta-siRNAs, ra-siRNAs, nat-siRNAs, dsRNAs, endo-siRNAs, viRNAs, heterochromatin, DNA methylation, histone post-translational modifications, and gene silencing pathways. Moreover, the potential biological functions of plant sRNAs to acquire information from different tissues and shift it across generations may improve future plant research. The development of RNA biogenesis mechanisms leads to the regulation of biological processes coupled with plant development and environmental responses. Retrotransposable elements, considered a kind of genetic pool, have tremendous potential in genome analysis, biodiversity research, gene mapping, gene cloning, and functional analysis.

A high proportion of LTR retrotransposons are involved in multiple epigenetic mechanisms, including stress tolerance, transpositional activity, regulation of gene expression, DNA methylation, histone modification, and chromatin remodeling and their interconnected networks in the plant genome. Increasing research interest in such epigenetic mechanisms may contribute to a greater understanding of their central role in genome organization and evolution. Therefore, an integrated TE database with epigenetic information will be a valuable resource for future research focused on assessing the possible contribution of LTR retrotransposons to develop single-molecule real-time sequencing and transcriptome variations resulting from advancements of genome annotation and investigations of plant genetic diversity. Moreover, advancement in the forthcoming reference genomes in association with novel sequence technologies may lead to the implementation of long-read sequencing. This will further enhance understanding of various aspects of genome disruption of LTR retrotransposons.

Environmental stresses affecting crops grown under field conditions are a major threat in the global warming era, and the activities of retrotransposons show a close relationship with such stresses. During environmental stress, LTR retrotransposons are more active and induce mutational and insertional polymorphisms. LTR retrotransposon-mediated molecular genetic markers are a highly polymorphic and efficient system. Moreover, as this does not influence genetic structure across species, DNA marker investigations will be a promising tool for exploring crop diversity and germplasm. Furthermore, several studies revealed that centromere-specific retrotransposons are conserved in pericentromeres, sub-telomeres, and telomeres and have rapidly diverged during the evolution of A, B, and C diploid genome lineages. Moreover, recent developments in genomics-based on whole-genome sequencing and 3D nuclear organization, allele-specific histone modification, and RNA Pol II binding profiles may facilitate the understanding of epigenetic regulation of differential gene expression between homologous chromosomes.

Another consideration is nanoparticle-based biomolecule delivery systems. In these systems, biomolecules such as DNA, RNA, and protein can be efficiently delivered and incorporated into the plant genome. This method can be utilized to make desirable epigenetic modifications in crop plants. In addition, high-throughput sequencing technology combined with artificial intelligence approaches for big data analysis may be beneficial in providing a more comprehensive picture of the interplay between LTR retrotransposon-induced epigenetic changes. Further collaborative studies are required to understand the complexity of LTR retrotransposons in evolutionary and organismal biology.

Author contributions

PKP and MR planned, designed, and wrote the review. PKP, MR, SM, QW, RK, PY, and MZ outlined and edited the review. PKP, MR, RK, QW, SM, LHZ, ZA, KKV, PY, and MZ edited and revised the review.

Funding

This work was funded by a grant from the National Natural Science Foundation of China (Grant Nos 31870656). This review was also supported by grants from the Zhejiang Provincial Natural Science Foundation of China (Grant No. LZ19C160001). The authors are grateful for the support of Metasequoia Faculty Research Start-up Funding (grant number 163100028) at the Bamboo Research Institute, Nanjing Forestry University for the co-first author MR.

Acknowledgments

The authors wish to thank the University of Helsinki Language Centre, Finland for the outstanding editing and proofreading of the 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.2022.1064847/full#supplementary-material

References

Akakpo, R., Carpentier, M. C., Ie Hsing, Y., and Panaud, O. (2020). The impact of transposable elements on the structure, evolution and function of the rice genome. *New Phytol.* 226, 44–49. doi: 10.1111/nph.16356

Ali, Z., Serag, M. F., Demirer, G. S., Torre, B., di Fabrizio, E., Landry, M. P., et al. (2022). DNA–Carbon nanotube binding mode determines the efficiency of carbon nanotube-mediated DNA delivery to intact plants. *ACS Appl. Nano Mater.* 5, 4663–4676. doi: 10.1021/acsanm.1c03482

Belyayev, A., Kalendar, R., Brodsky, L., Nevo, E., Schulman, A. H., and Raskina, O. (2010). Transposable elements in a marginal plant population: Temporal fluctuations provide new insights into genome evolution of wild diploid wheat. *Mobile DNA* 1, 1–16. doi: 10.1186/1759-8753-1-6

Bennetzen, J. L. (2007). Patterns in grass genome evolution. *Curr. Opin. Plant Biol.* 10, 176–181. doi: 10.1016/j.pbi.2007.01.010

Bennetzen, J. L., Ma, J., and Devos, K. M. (2005). Mechanisms of recent genome size variation in flowering plants. *Ann. Bot.* 95, 127–132. doi: 10.1093/aob/mci008

Benoit, M., Drost, H.-G., Catoni, M., Gouil, Q., Lopez-Gomollon, S., Baulcombe, D., et al. (2019). Environmental and epigenetic regulation of rider retrotransposons in tomato. *PloS Genet.* 15, e1008370. doi: 10.1371/journal.pgen.1008370

Beulé, T., Agbessi, M. D., Dussert, S., Jaligot, E., and Guyot, R. (2015). Genome-wide analysis of LTR-retrotransposons in oil palm. *BMC Genomics* 16, 1–14. doi: 10.1186/s12864-015-2023-1

Birchler, J. A., and Presting, G. G. (2012). Retrotransposon insertion targeting: A mechanism for homogenization of centromere sequences on nonhomologous chromosomes. *Genes Dev.* 26, 638–640. doi: 10.1101/gad.191049.112

Boeke, J. D., Garfinkel, D. J., Styles, C. A., and Fink, G. R. (1985). Ty Elements transpose through an RNA intermediate. *Cell* 40, 491–500. doi: 10.1016/0092-8674(85)90197-7

Bond, D. M., and Baulcombe, D. C. (2015). Epigenetic transitions leading to heritable, RNA-mediated *de novo* silencing in arabidopsis thaliana. *Proc. Natl. Acad. Sci.* 112, 917–922. doi: 10.1073/pnas.1413053112

Boonjing, P., Masuta, Y., Nozawa, K., Kato, A., and Ito, H. (2020). The effect of zebularine on the heat-activated retrotransposon ONSEN in arabidopsis thaliana and vigna angularis. *Genes Genet. Syst.* 95 (4), 165–172. doi: 10.1266/ggs.19-00046

Borges, F., and Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* 16, 727–741. doi: 10.1038/nrm4085

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., et al. (2018). Ten things you should know about transposable elements. *Genome Biol.* 19, 1–12. doi: 10.1186/s13059-018-1577-z

Brown, P. O., Bowerman, B., Varmus, H. E., and Bishop, J. M. (1987). Correct integration of retroviral DNA *in vitro*. *Cell* 49, 347–356. doi: 10.1016/0092-8674(87)90287-X

Buchmann, R. C., Asad, S., Wolf, J. N., Mohannath, G., and Bisaro, D. M. (2009). Geminivirus AL2 and L2 proteins suppress transcriptional gene silencing and cause genome-wide reductions in cytosine methylation. *J. Virol.* 83, 5005–5013. doi: 10.1128/JVI.01771-08

Candela, H., and Hake, S. (2008). The art and design of genetic screens: Maize. *Nat. Rev. Genet.* 9, 192–203. doi: 10.1038/nrg2291

Cao, Y., Jiang, Y., Ding, M., He, S., Zhang, H., Lin, L., et al. (2015). Molecular characterization of a transcriptionally active Ty1/copia-like retrotransposon in gossypium. *Plant Cell Rep.* 34, 1037–1047. doi: 10.1007/s00299-015-1763-3

Capy, P. (2005). Classification and nomenclature of retrotransposable elements. *Cytogenetic Genome Res.* 110, 457–461. doi: 10.1159/000084978

Castillo, A. E. D. R., De León-Rodriguez, A., Terrones, M., and de la Rosa, A. P. B. (2021). Multi-walled carbon nanotubes enhance the genetic transformation of bifidobacterium longum. *Carbon* 184, 902–909. doi: 10.1016/j.carbon.2021.08.052

Catlin, N. S., and Josephs, E. B. (2022). The important contribution of transposable elements to phenotypic variation and evolution. *Curr. Opin. Plant Biol.* 65, 102140. doi: 10.1016/j.pbi.2021.102140

Cavrak, V. V., Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L. M., and Mittelsten Scheid, O. (2014). How a retrotransposon exploits the plant's heat stress response for its activation. *PloS Genet.* 10, e1004115. doi: 10.1371/journal.pgen.1004115

Cheng, Z., Dong, F., Langdon, T., Ouyang, S., Buell, C. R., Gu, M., et al. (2002). Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell* 14, 1691–1704. doi: 10.1105/tpc.003079

Cossu, R. M., Casola, C., Giacomello, S., Vidalis, A., Scofield, D. G., and Zuccolo, A. (2017). LTR Retrotransposons show low levels of unequal recombination and high rates of intraelement gene conversion in large plant genomes. *Genome Biol. Evol.* 9, 3449–3462. doi: 10.1093/gbe/evx260

Creasey, K. M., Zhai, J., Borges, F., Van Ex, F., Regulski, M., Meyers, B. C., et al. (2014). miRNAs trigger widespread epigenetically activated siRNAs from transposons in arabidopsis. *Nature* 508, 411–415. doi: 10.1038/nature13069

Cunningham, F. J., Goh, N. S., Demirer, G. S., Matos, J. L., and Landry, M. P. (2018). Nanoparticle-mediated delivery towards advancing plant genetic engineering. *Trends Biotechnol.* 36, 882–897. doi: 10.1016/j.tibtech.2018.03.009

Daron, J., Glover, N., Pingault, L., Theil, S., Jamilloux, V., Paux, E., et al. (2014). Organization and evolution of transposable elements along the bread wheat chromosome 3B. *Genome Biol.* 15, 1–15. doi: 10.1186/s13059-014-0546-4

de Assis, R., Baba, V. Y., Cintra, L. A., Gonçalves, L. S. A., Rodrigues, R., and Vanzela, A. L. L. (2020). Genome relationships and LTR-retrotransposon diversity in three cultivated capsicum L.(Solanaceae) species. *BMC Genomics* 21, 1–14. doi: 10.1186/s12864-020-6618-9

de Castro Nunes, R., Orozco-Arias, S., Crouzillat, D., Mueller, L. A., Strickler, S. R., Descombes, P., et al. (2018). Structure and distribution of centromeric retrotransposons at diploid and allotetraploid coffea centromeric and pericentromeric regions. *Front. Plant Sci.* 9, 175. doi: 10.3389/fpls.2018.00175

Demirer, G. S., Zhang, H., Goh, N. S., González-Grandío, E., and Landry, M. P. (2019a). Carbon nanotube–mediated DNA delivery without transgene integration in intact plants. *Nat. Protoc.* 14, 2954–2971. doi: 10.1038/s41596-019-0208-9

Demirer, G. S., Zhang, H., Goh, N. S., Pinals, R. L., Chang, R., and Landry, M. P. (2020). Carbon nanocarriers deliver siRNA to intact plant cells for efficient gene knockdown. *Sci. Adv.* 6, eaaz0495. doi: 10.1126/sciadv.aaz0495

Demirer, G. S., Zhang, H., Matos, J. L., Goh, N. S., Cunningham, F. J., Sung, Y., et al. (2019b). High aspect ratio nanomaterials enable delivery of functional genetic material without DNA integration in mature plants. *Nat. Nanotechnol.* 14, 456–464. doi: 10.1038/s41565-019-0382-5

Divashuk, M. G., Khuat, T. M. L., Kroupin, P. Y., Kirov, I. V., Romanov, D. V., Kiseleva, A. V., et al. (2016). Variation in copy number of Ty3/Gypsy centromeric retrotransposons in the genomes of thinopyrum intermedium and its diploid progenitors. *PloS One* 11, e0154241. doi: 10.1371/journal.pone.0154241

Dong, Q., Hu, B., and Zhang, C. (2022). microRNAs and their roles in plant development. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.824240

Doungous, O., Kalendar, R., Adiobo, A., and Schulman, A. H. (2015). Retrotransposon molecular markers resolve cocoyam (*Xanthosoma sagittifolium*) and taro (*Colocasia esculenta*) by type and variety. *Euphytica* 206, 541–554. doi: 10.1007/s10681-015-1537-6

Doungous, O., Kalendar, R., Filippova, N., and Ngane, B. K. (2020). Utility of iPBS retrotransposons markers for molecular characterization of African gnetum species. *Plant Biosystems* 154 (5), 587–592. doi: 10.1080/11263504.2019.1651782

Durán-Figueroa, N., and Vielle-Calzada, J.-P. (2010). ARGONAUTE9-dependent silencing of transposable elements in pericentromeric regions of arabidopsis. *Plant Signaling Behav.* 5, 1476–1479. doi: 10.4161/psb.5.11.13548

Flavell, A. J., Knox, M. R., Pearce, S. R., and Ellis, T. N. (1998). Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. *Plant J.* 16, 643–650. doi: 10.1046/j.1365-313x.1998.00334.x

Fukui, K.-N., Suzuki, G., Lagudah, E. S., Rahman, S., Appels, R., Yamamoto, M., et al. (2001). Physical arrangement of retrotransposon-related repeats in centromeric regions of wheat. *Plant Cell Physiol.* 42, 189–196. doi: 10.1093/pcp/pce026

Galindo-González, L., Mhiri, C., Deyholos, M. K., and Grandbastien, M.-A. (2017). LTR-Retrotransposons in plants: Engines of evolution. *Gene* 626, 14–25. doi: 10.1016/j.gene.2017.04.051

Gantuz, M., Morales, A., Bertoldi, M. V., Ibañez, V. N., Duarte, P. F., Marfil, C. F., et al. (2022). Hybridization and polyploidization effects on LTR-retrotransposon activation in potato genome. *J. Plant Res.* 135, 81–92. doi: 10.1007/s10265-021-01354-9

Gao, D., Gill, N., Kim, H. R., Walling, J. G., Zhang, W., Fan, C., et al. (2009). A lineage-specific centromere retrotransposon in oryza brachyantha. *Plant J.* 60, 820–831. doi: 10.1111/j.1365-313X.2009.04005.x

Gao, L., McCarthy, E. M., Ganko, E. W., and McDonald, J. F. (2004). Evolutionary history of oryza sativa LTR retrotransposons: A preliminary survey of the rice genome sequences. *BMC Genomics* 5, 1–18. doi: 10.1186/1471-2164-5-18

Gaubert, H., Sanchez, D. H., Drost, H.-G., and Paszkowski, J. (2017). Developmental restriction of retrotransposition activated in arabidopsis by environmental stress. *Genetics* 207, 813–821. doi: 10.1534/genetics.117.300103

Ghonaim, M., Kalendar, R., Barakat, H., Elsherif, N., Ashry, N., and Schulman, A. H. (2020). High-throughput retrotransposon-based genetic diversity of maize germplasm assessment and analysis. *Mol. Biol. Rep.* 47, 1589–1603. doi: 10.1007/s11033-020-05246-4

Giordani, T., Cossu, R. M., Mascagni, F., Marroni, F., Morgante, M., Cavallini, A., et al. (2016). Genome-wide analysis of LTR-retrotransposon expression in leaves of populus× canadensis water-deprived plants. *Tree Genet. Genomes* 12, 1–14. doi: 10.1007/s11295-016-1036-5

Girard, L., and Freeling, M. (1999). Regulatory changes as a consequence of transposon insertion. *Dev. Genet.* 25, 291–296. doi: 10.1002/(SICI)1520-6408(1999)25:4<291::AID-DVG2>3.0.CO;2-5

Golestanipour, A., Nikkhah, M., Aalami, A., and Hosseinkhani, S. (2018). Gene delivery to tobacco root cells with single-walled carbon nanotubes and cell-penetrating fusogenic peptides. *Mol. Biotechnol.* 60, 863–878. doi: 10.1007/s12033-018-0120-5

Gong, Z., Wu, Y., Koblížková, A., Torres, G. A., Wang, K., Iovene, M., et al. (2012). Repeatless and repeat-based centromeres in potato: implications for centromere evolution. *Plant Cell* 24, 3559–3574. doi: 10.1105/tpc.112.100511

Goodier, J. L. (2016). Restricting retrotransposons: a review. *Mobile DNA* 7, 1–30. doi: 10.1186/s13100-016-0070-z

Gorbunova, V., Seluanov, A., Mita, P., McKerrow, W., Fenyö, D., Boeke, J. D., et al. (2021). The role of retrotransposable elements in ageing and age-associated diseases. *Nature* 596, 43–53. doi: 10.1038/s41586-021-03542-y

Grandbastien, M.-A. (2015). LTR Retrotransposons, handy hitchhikers of plant regulation and stress response. *Biochim. Biophys. Acta (BBA) Gene Regul. Mech.* 1849, 403–416. doi: 10.1016/j.bbagrm.2014.07.017

Grandbastien, M.-A., Audeon, C., Bonnivard, E., Casacuberta, J., Chalhoub, B., Costa, A.-P., et al. (2005). Stress activation and genomic impact of Tnt1 retrotransposons in solanaceae. *Cytogenetic Genome Res.* 110, 229–241. doi: 10.1159/000084957

Guo, X., Su, H., Shi, Q., Fu, S., Wang, J., Zhang, X., et al. (2016). *De novo* centromere formation and centromeric sequence expansion in wheat and its wide hybrids. *PloS Genet.* 12, e1005997. doi: 10.1371/journal.pgen.1005997

Hartley, G., and O'Neill, R. J. (2019). Centromere repeats: Hidden gems of the genome. *Genes* 10, 223. doi: 10.3390/genes10030223

Hawkins, J. S., Kim, H., Nason, J. D., Wing, R. A., and Wendel, J. F. (2006). Differential lineage-specific amplification of transposable elements is responsible for genome size variation in gossypium. *Genome Res.* 16, 1252–1261. doi: 10.1101/gr.5282906

He, P., Ma, Y., Dai, H., Li, L., Liu, Y., Li, H., et al. (2012). Characterization of the hormone and stress-induced expression of FaRE1 retrotransposon promoter in strawberry. *J. Plant Biol.* 55, 1–7. doi: 10.1007/s12374-011-9180-9

Hollister, J. D., Smith, L. M., Guo, Y.-L., Ott, F., Weigel, D., and Gaut, B. S. (2011). Transposable elements and small RNAs contribute to gene expression divergence between arabidopsis thaliana and arabidopsis lyrata. *Proc. Natl. Acad. Sci.* 108, 2322–2327. doi: 10.1073/pnas.1018222108

Hosid, E., Brodsky, L., Kalendar, R., Raskina, O., and Belyayev, A. (2012). Diversity of long terminal repeat retrotransposon genome distribution in natural populations of the wild diploid wheat *Aegilops speltoides*. *Genetics* 190, 263–274. doi: 10.1534/genetics.111.134643

Hsing, Y.-I., Chern, C.-G., Fan, M.-J., Lu, P.-C., Chen, K.-T., Lo, S.-F., et al. (2007). A rice gene activation/knockout mutant resource for high throughput functional genomics. *Plant Mol. Biol.* 63, 351–364. doi: 10.1007/s11103-006-9093-z

Huang, Y., Ding, W., Zhang, M., Han, J., Jing, Y., Yao, W., et al. (2021). The formation and evolution of centromeric satellite repeats in saccharum species. *Plant J.* 106, 616–629. doi: 10.1111/tpj.15186

Ito, H. (2012). Small RNAs and transposon silencing in plants. *Development Growth Differ.* 54, 100–107. doi: 10.1111/j.1440-169X.2011.01309.x

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., and Paszkowski, J. (2011). An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* 472, 115–119. doi: 10.1038/nature09861

Ito, H., Kim, J.-M., Matsunaga, W., Saze, H., Matsui, A., Endo, T. A., et al. (2016). A stress-activated transposon in arabidopsis induces transgenerational abscisic acid insensitivity. *Sci. Rep.* 6, 1–12. doi: 10.1038/srep23181

Ito, H., Nasuda, S., and Endo, T. R. (2004). Erratum: A direct repeat sequence associated with the centromeric retrotransposons in wheat. *Genome* 47, 998–998. doi: 10.1139/g04-901

Ito, H., Yoshida, T., Tsukahara, S., and Kawabe, A. (2013). Evolution of the ONSEN retrotransposon family activated upon heat stress in brassicaceae. *Gene* 518, 256–261. doi: 10.1016/j.gene.2013.01.034

Ivashuta, S., Naumkina, M., Gau, M., Uchiyama, K., Isobe, S., Mizukami, Y., et al. (2002). Genotype-dependent transcriptional activation of novel repetitive elements during cold acclimation of alfalfa (Medicago sativa). *Plant J.* 31, 615–627. doi: 10.1046/j.1365-313X.2002.01383.x

Jedlicka, P., Lexa, M., Vanat, I., Hobza, R., and Kejnovsky, E. (2019). Nested plant LTR retrotransposons target specific regions of other elements, while all LTR retrotransposons often target palindromes and nucleosome-occupied regions: In silico study. *Mobile DNA* 10, 1–14. doi: 10.1186/s13100-019-0186-z

Jia, H. M., Jia, H. J., Cai, Q. L., Wang, Y., Zhao, H. B., Yang, W. F., et al. (2019). The red bayberry genome and genetic basis of sex determination. *Plant Biotechnol. J.* 17, 397–409. doi: 10.1111/pbi.12985

Jia, Y., Lisch, D. R., Ohtsu, K., Scanlon, M. J., Nettleton, D., and Schnable, P. S. (2009). Loss of RNA–dependent RNA polymerase 2 (RDR2) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. *PloS Genet.* 5, e1000737. doi: 10.1371/journal.pgen.1000737

Jiang, S., Zong, Y., Yue, X., Postman, J., Teng, Y., and Cai, D. (2015). Prediction of retrotransposons and assessment of genetic variability based on developed retrotransposon-based insertion polymorphism (RBIP) markers in pyrus l. *Mol. Genet. Genomics* 290, 225–237. doi: 10.1007/s00438-014-0914-5

Jing, R., Vershinin, A., Grzebyta, J., Shaw, P., Smýkal, P., Marshall, D., et al. (2010). The genetic diversity and evolution of field pea (Pisum) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. *BMC Evol. Biol.* 10, 1–20. doi: 10.1186/1471-2148-10-44

Jin, W., Lamb, J. C., Vega, J. M., Dawe, R. K., Birchler, J. A., and Jiang, J. (2005). Molecular and functional dissection of the maize b chromosome centromere. *Plant Cell* 17, 1412–1423. doi: 10.1105/tpc.104.030643

Kalendar, R., Amenov, A., and Daniyarov, A. (2018). Use of retrotransposon-derived genetic markers to analyse genomic variability in plants. *Funct. Plant Biol.* 46, 15–29. doi: 10.1071/FP18098

Kalendar, R., Antonius, K., Smýkal, P., and Schulman, A. H. (2010). iPBS: a universal method for DNA fingerprinting and retrotransposon isolation. *Theor. Appl. Genet.* 121, 1419–1430. doi: 10.1007/s00122-010-1398-2

Kalendar, R., Flavell, A., Ellis, T., Sjakste, T., Moisy, C., and Schulman, A. H. (2011). Analysis of plant diversity with retrotransposon-based molecular markers. *Heredity* 106, 520–530. doi: 10.1038/hdy.2010.93

Kalendar, R., Grob, T., Regina, M., Suoniemi, A., and Schulman, A. (1999). IRAP and REMAP: Two new retrotransposon-based DNA fingerprinting techniques. *Theor. Appl. Genet.* 98, 704–711. doi: 10.1007/s001220051124

Kalendar, R., Raskina, O., Belyayev, A., and Schulman, A. H. (2020). Long tandem arrays of Cassandra retroelements and their role in genome dynamics in plants. *Int. J. Mol. Sci.* 21, 2931. doi: 10.3390/ijms21082931

Kalendar, R., Muterko, A., and Boronnikova, S. (2021a). Retrotransposable elements: DNA fingerprinting and the assessment of genetic diversity. *Methods in molecular biology* 2222, 263–286. doi: 10.1007/978-1-0716-0997-2_15

Kalendar, R., Sabot, F., Rodriguez, F., Karlov, G. I., Natali, L., and Alix, K. (2021b). Editorial: Mobile elements and plant genome evolution, comparative analyzes and computational tools. *Front. Plant Sci.* 12, 735134. doi: 10.3389/fpls.2021.735134

Kalendar, R., and Schulman, A. H. (2006). IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. *Nat. Protoc.* 1, 2478–2484. doi: 10.1038/nprot.2006.377

Kalendar, R., Tanskanen, J., Chang, W., Antonius, K., Sela, H., Peleg, O., et al. (2008). Cassandra Retrotransposons carry independently transcribed 5S RNA. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5833–5838. doi: 10.1073/pnas.0709698105

Kalendar, R., Vicient, C. M., Peleg, O., Anamthawat-Jonsson, K., Bolshoy, A., and Schulman, A. H. (2004). Large Retrotransposon derivatives: abundant, conserved but nonautonomous retroelements of barley and related genomes. *Genetics* 166, 1437–1450. doi: 10.1534/genetics.166.3.1437

Keçeli, B. N., Jin, C., Van Damme, D., and Geelen, D. (2020). Conservation of centromeric histone 3 interaction partners in plants. *J. Exp. Bot.* 71, 5237–5246. doi: 10.1093/jxb/eraa214

Keidar, D., Doron, C., and Kashkush, K. (2018). Genome-wide analysis of a recently active retrotransposon, au SINE, in wheat: content, distribution within subgenomes and chromosomes, and gene associations. *Plant Cell Rep.* 37, 193–208. doi: 10.1007/s00299-017-2213-1

Kenan-Eichler, M., Leshkowitz, D., Tal, L., Noor, E., Melamed-Bessudo, C., Feldman, M., et al. (2011). Wheat hybridization and polyploidization results in deregulation of small RNAs. *Genetics* 188, 263–272. doi: 10.1534/genetics.111.128348

Khapilina, O., Raiser, O., Danilova, A., Shevtsov, V., Turzhanova, A., and Kalendar, R. (2021a). DNA Profiling and assessment of genetic diversity of relict species allium altaicum pall. on the territory of Altai. *PeerJ* 9, e10674. doi: 10.7717/peerj.10674

Khapilina, O., Turzhanova, A., Danilova, A., Tumenbayeva, A., Shevtsov, V., Kotukhov, Y., et al. (2021b). Primer binding site (PBS) profiling of genetic diversity of natural populations of endemic species *Allium ledebourianum* schult. *BioTech* 10, 23. doi: 10.3390/biotech10040023

Kim, H., Terakami, S., Nishitani, C., Kurita, K., Kanamori, H., Katayose, Y., et al. (2012). Development of cultivar-specific DNA markers based on retrotransposon-based insertional polymorphism in Japanese pear. *Breed. Sci.* 62, 53–62. doi: 10.1270/jsbbs.62.53

Kirov, I., Odintsov, S., Omarov, M., Gvaramiya, S., Merkulov, P., Dudnikov, M., et al. (2020). Functional allium fistulosum centromeres comprise arrays of a long satellite repeat, insertions of retrotransposons and chloroplast DNA. *Front. Plant Sci.* 11, 1668. doi: 10.3389/fpls.2020.562001

Kiseleva, A., Kirov, I., and Khrustaleva, L. (2014). Chromosomal organization of centromeric Ty3/gypsy retrotransposons in allium cepa l. and allium fistulosum l. *Russian J. Genet.* 50, 586–592. doi: 10.1134/S102279541404005X

Klein, S. J., and O'Neill, R. J. (2018). Transposable elements: genome innovation, chromosome diversity, and centromere conflict. *Chromosome Res.* 26, 5–23. doi: 10.1007/s10577-017-9569-5

Kong, X., Yang, M., Le, B. H., He, W., and Hou, Y. (2022). The master role of siRNAs in plant immunity. *Mol. Plant Pathol.* 23, 1565–1574. doi: 10.1111/mpp.13250

Konovalov, F. A., Goncharov, N. P., Goryunova, S., Shaturova, A., Proshlyakova, T., and Kudryavtsev, A. (2010). Molecular markers based on LTR retrotransposons BARE-1 and jeli uncover different strata of evolutionary relationships in diploid wheats. *Mol. Genet. Genomics* 283, 551–563. doi: 10.1007/s00438-010-0539-2

Koo, D.-H., Hong, C. P., Batley, J., Chung, Y. S., Edwards, D., Bang, J.-W., et al. (2011). Rapid divergence of repetitive DNAs in brassica relatives. *Genomics* 97, 173–185. doi: 10.1016/j.ygeno.2010.12.002

Kuhn, B., Mangolin, C. A., Souto, E. R., Vicient, C. M., and Machado, M. F. (2016). Development of retrotransposon-based markers IRAP and REMAP for cassava (Manihot esculenta). *Gen. Mol. Res.* 15 (2), gmr.15027149. doi: 10.4238/gmr.15027149

Kwak, S.-Y., Lew, T. T. S., Sweeney, C. J., Koman, V. B., Wong, M. H., Bohmert-Tatarev, K., et al. (2019). Chloroplast-selective gene delivery and expression in planta using chitosan-complexed single-walled carbon nanotube carriers. *Nat. Nanotechnol.* 14, 447–455. doi: 10.1038/s41565-019-0375-4

Laganà, A., Veneziano, D., Russo, F., Pulvirenti, A., Giugno, R., Croce, C. M., et al. (2015). Computational design of artificial RNA molecules for gene regulation. *RNA Bioinformatics* 1269, 393–412. doi: 10.1007/978-1-4939-2291-8_25

Lamb, J. C., Kato, A., and Birchler, J. A. (2005). Sequences associated with a chromosome centromeres are present throughout the maize b chromosome. *Chromosoma* 113, 337–349. doi: 10.1007/s00412-004-0319-z

Lanciano, S., and Cristofari, G. (2020). Measuring and interpreting transposable element expression. *Nat. Rev. Genet.* 21, 721–736. doi: 10.1038/s41576-020-0251-y

Lanciano, S., and Mirouze, M. (2018). Transposable elements: All mobile, all different, some stress responsive, some adaptive? *Curr. Opin. Genet. Dev.* 49, 106–114. doi: 10.1016/j.gde.2018.04.002

Lancíková, V., and Žiarovská, J. (2020). Inter-retrotransposon amplified polymorphism markers revealed long terminal repeat retrotransposon insertion polymorphism in flax cultivated on the experimental fields around Chernobyl. *J. Environ. Sci. Health Part A* 55, 957–963. doi: 10.1080/10934529.2020.1760016

Lee, S. C., Ernst, E., Berube, B., Borges, F., Parent, J.-S., Ledon, P., et al. (2020a). Arabidopsis LTR retrotransposons and their regulation by epigenetically activated small RNA. *bioRxiv* 2020.01.24.919167. doi: 10.1101/2020.01.24.919167

Lee, S. C., Ernst, E., Berube, B., Borges, F., Parent, J.-S., Ledon, P., et al. (2020b). Arabidopsis retrotransposon virus-like particles and their regulation by epigenetically activated small RNA. *Genome Res.* 30, 576–588. doi: 10.1101/gr.259044.119

Lee, S.-I., Kim, J.-H., Park, K.-C., and Kim, N.-S. (2015). LTR-Retrotransposons and inter-retrotransposon amplified polymorphism (IRAP) analysis in *Lilium* species. *Genetica* 143, 343–352. doi: 10.1007/s10709-015-9833-6

Lee, Y.-S., Maple, R., Dürr, J., Dawson, A., Tamim, S., Del Genio, C., et al. (2021). A transposon surveillance mechanism that safeguards plant male fertility during stress. *Nat. Plants* 7, 34–41. doi: 10.1038/s41477-020-00818-5

Leitch, A., and Leitch, I. (2008). Genomic plasticity and the diversity of polyploid plants. *science* 320, 481–483. doi: 10.1126/science.1153585

Liang, Y., Lenz, R. R., and Dai, W. (2016). Development of retrotransposon-based molecular markers and their application in genetic mapping in chokecherry (Prunus virginiana l.). *Mol. Breed.* 36, 1–10. doi: 10.1007/s11032-016-0535-2

Li, H., Freeling, M., and Lisch, D. (2010). Epigenetic reprogramming during vegetative phase change in maize. *Proc. Natl. Acad. Sci.* 107, 22184–22189. doi: 10.1073/pnas.1016884108

Lim, K. B., Yang, T. J., Hwang, Y. J., Kim, J. S., Park, J. Y., Kwon, S. J., et al. (2007). Characterization of the centromere and peri-centromere retrotransposons in brassica rapa and their distribution in related brassica species. *Plant J.* 49, 173–183. doi: 10.1111/j.1365-313X.2006.02952.x

Lippman, Z., May, B., Yordan, C., Singer, T., Martienssen, R., and Becker, P. (2003). Distinct mechanisms determine transposon inheritance and methylation *via* small interfering RNA and histone modification. *PloS Biol.* 1, e67. doi: 10.1371/journal.pbio.0000067

Lisch, D. (2013). How important are transposons for plant evolution? *Nat. Rev. Genet.* 14, 49–61. doi: 10.1038/nrg3374

Liu, S., De Jonge, J., Trejo-Arellano, M. S., Santos-González, J., Köhler, C., and Hennig, L. (2021). Role of H1 and DNA methylation in selective regulation of transposable elements during heat stress. *New Phytol.* 229, 2238–2250. doi: 10.1111/nph.17018

Liu, Z., Liu, Y., Liu, F., Zhang, S., Wang, X., Lu, Q., et al. (2018). Genome-wide survey and comparative analysis of long terminal repeat (LTR) retrotransposon families in four gossypium species. *Sci. Rep.* 8, 1–10. doi: 10.1038/s41598-018-27589-6

Lopez-Gomollon, S., and Baulcombe, D. C. (2022). Roles of RNA silencing in viral and non-viral plant immunity and in the crosstalk between disease resistance systems. *Nat. Rev. Mol. Cell Biol.* 23, 645–662. doi: 10.1038/s41580-022-00496-5

Luan, D. D., Korman, M. H., Jakubczak, J. L., and Eickbush, T. H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72, 595–605. doi: 10.1016/0092-8674(93)90078-5

Luo, S., Mach, J., Abramson, B., Ramirez, R., Schurr, R., Barone, P., et al. (2012). The cotton centromere contains a Ty3-gypsy-like LTR retroelement. *PloS One* 7, e35261. doi: 10.1371/journal.pone.0035261

Lv, Z., Jiang, R., Chen, J., and Chen, W. (2020). Nanoparticle-mediated gene transformation strategies for plant genetic engineering. *Plant J.* 104, 880–891. doi: 10.1111/tpj.14973

Ma, J., and Bennetzen, J. L. (2006). Recombination, rearrangement, reshuffling, and divergence in a centromeric region of rice. *Proc. Natl. Acad. Sci.* 103, 383–388. doi: 10.1073/pnas.0509810102

Madlung, A., Tyagi, A. P., Watson, B., Jiang, H., Kagochi, T., Doerge, R. W., et al. (2005). Genomic changes in synthetic arabidopsis polyploids. *Plant J.* 41, 221–230. doi: 10.1111/j.1365-313X.2004.02297.x

Malaviya, D. R., Roy, A. K., Kaushal, P., Pathak, S., and Kalendar, R. (2021). Phenotype study of multifoliolate leaf formation in trifolium alexandrinum l. *PeerJ.* 9, e10874. doi: 10.7717/peerj.10874

Malik, H. S., and Eickbush, T. H. (2001). Phylogenetic analysis of ribonuclease h domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. *Genome Res.* 11, 1187–1197. doi: 10.1101/gr.185101

Mangiavacchi, A., Liu, P., Della Valle, F., and Orlando, V. (2021). New insights into the functional role of retrotransposon dynamics in mammalian somatic cells. *Cell. Mol. Life Sci.* 78, 5245–5256. doi: 10.1007/s00018-021-03851-5

Marí-Ordóñez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V., and Voinnet, O. (2013). Reconstructing *de novo* silencing of an active plant retrotransposon. *Nat. Genet.* 45, 1029–1039. doi: 10.1038/ng.2703

Marques, A., Ribeiro, T., Neumann, P., Macas, J., Novák, P., Schubert, V., et al. (2015). Holocentromeres in rhynchospora are associated with genome-wide centromere-specific repeat arrays interspersed among euchromatin. *Proc. Natl. Acad. Sci.* 112, 13633–13638. doi: 10.1073/pnas.1512255112

Mascagni, F., Giordani, T., Ceccarelli, M., Cavallini, A., and Natali, L. (2017). Genome-wide analysis of LTR-retrotransposon diversity and its impact on the evolution of the genus helianthus (L.). *BMC Genomics* 18, 1–16. doi: 10.1186/s12864-017-4050-6

Mascagni, F., Vangelisti, A., Usai, G., Giordani, T., Cavallini, A., and Natali, L. (2020). A computational genome-wide analysis of long terminal repeats retrotransposon expression in sunflower roots (Helianthus annuus l.). *Genetica* 148, 13–23. doi: 10.1007/s10709-020-00085-4

Masuda, S., Nozawa, K., Matsunaga, W., Masuta, Y., Kawabe, A., Kato, A., et al. (2016). Characterization of a heat-activated retrotransposon in natural accessions of arabidopsis thaliana. *Genes Genet. Syst.* 91 (6), 293–299. doi: 10.1266/ggs.16-00045

Masuta, Y., Kawabe, A., Nozawa, K., Naito, K., Kato, A., and Ito, H. (2018). Characterization of a heat-activated retrotransposon in vigna angularis. *Breed. Sci.* 68, 168–176. doi: 10.1270/jsbbs.17085

Masuta, Y., Nozawa, K., Takagi, H., Yaegashi, H., Tanaka, K., Ito, T., et al. (2017). Inducible transposition of a heat-activated retrotransposon in tissue culture. *Plant Cell Physiol.* 58, 375–384. doi: 10.1093/pcp/pcw202

Matsunaga, W., Kobayashi, A., Kato, A., and Ito, H. (2012). The effects of heat induction and the siRNA biogenesis pathway on the transgenerational transposition of ONSEN, a copia-like retrotransposon in arabidopsis thaliana. *Plant Cell Physiol.* 53, 824–833. doi: 10.1093/pcp/pcr179

Matsunaga, W., Ohama, N., Tanabe, N., Masuta, Y., Masuda, S., Mitani, N., et al. (2015). A small RNA mediated regulation of a stress-activated retrotransposon and the tissue specific transposition during the reproductive period in arabidopsis. *Front. Plant Sci.* 6, 48. doi: 10.3389/fpls.2015.00048

Ma, T., Wei, X., Zhang, Y., Li, J., Wu, F., Yan, Q., et al. (2022). Development of molecular markers based on LTR retrotransposon in the cleistogenes songorica genome. *J. Appl. Genet.* 63, 61–72. doi: 10.1007/s13353-021-00658-9

McCue, A. D., Nuthikattu, S., Reeder, S. H., and Slotkin, R. K. (2012). Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. *PloS Genet.* 8, e1002474. doi: 10.1371/journal.pgen.1002474

McCue, A. D., Panda, K., Nuthikattu, S., Choudury, S. G., Thomas, E. N., and Slotkin, R. K. (2015). ARGONAUTE 6 bridges transposable element m RNA-derived si RNA s to the establishment of DNA methylation. *EMBO J.* 34, 20–35. doi: 10.15252/embj.201489499

Meng, Y., Su, W., Ma, Y., Liu, L., Gu, X., Wu, D., et al. (2021). Assessment of genetic diversity and variety identification based on developed retrotransposon-based insertion polymorphism (RBIP) markers in sweet potato (Ipomoea batatas (L.) lam.). *Sci. Rep.* 11, 1–12. doi: 10.1038/s41598-021-95876-w

Mhiri, C., Morel, J.-B., Vernhettes, S., Casacuberta, J. M., Lucas, H., and Grandbastien, M.-A. (1997). The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress. *Plant Mol. Biol.* 33, 257–266. doi: 10.1023/A:1005727132202

Miller, J. T., Dong, F., Jackson, S. A., Song, J., and Jiang, J. (1998). Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. *Genetics* 150, 1615–1623. doi: 10.1093/genetics/150.4.1615

Moisy, C., Schulman, A. H., Kalendar, R., Buchmann, J. P., and Pelsy, F. (2014). The Tvv1 retrotransposon family is conserved between plant genomes separated by over 100 million years. *Theor. Appl. Genet.* 127, 1223–1235. doi: 10.1007/s00122-014-2293-z

Monden, Y., Fujii, N., Yamaguchi, K., Ikeo, K., Nakazawa, Y., Waki, T., et al. (2014a). Efficient screening of long terminal repeat retrotransposons that show high insertion polymorphism *via* high-throughput sequencing of the primer binding site. *Genome* 57, 245–252. doi: 10.1139/gen-2014-0031

Monden, Y., Takai, T., and Tahara, M. (2014b). *Characterization of a novel retrotransposon TriRe-1 using nullisomic-tetrasomic lines of hexaploid wheat* (Faculty of Agriculture, Okayama University) 103, 21–30.

Muñoz-López, M., and García-Pérez, J. L. (2010). DNA Transposons: Nature and applications in genomics. *Curr. Genomics* 11, 115–128. doi: 10.2174/138920210790886871

Nagaki, K., and Murata, M. (2005). Characterization of CENH3 and centromere-associated DNA sequences in sugarcane. *Chromosome Res.* 13, 195–203. doi: 10.1007/s10577-005-0847-2

Nagaki, K., Shibata, F., Kanatani, A., Kashihara, K., and Murata, M. (2012). Isolation of centromeric-tandem repetitive DNA sequences by chromatin affinity purification using a HaloTag7-fused centromere-specific histone H3 in tobacco. *Plant Cell Rep.* 31, 771–779. doi: 10.1007/s00299-011-1198-4

Nagaki, K., Shibata, F., and Murata, M. (2011). A mosaic structure of centromeric DNA in tobacco. *Genes Genet. Syst.* 86, 434–434. GENETICS SOC JAPAN NATIONAL INST GENETICS YATA 1111, MISHIMA, SHIZUOKA-KEN.

Nagaki, K., Song, J., Stupar, R. M., Parokonny, A. S., Yuan, Q., Ouyang, S., et al. (2003). Molecular and cytological analyses of large tracks of centromeric DNA reveal the structure and evolutionary dynamics of maize centromeres. *Genetics* 163, 759–770. doi: 10.1093/genetics/163.2.759

Naito, K., Zhang, F., Tsukiyama, T., Saito, H., Hancock, C. N., Richardson, A. O., et al. (2009). Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* 461, 1130–1134. doi: 10.1038/nature08479

Nakamura, M., Köhler, C., and Hennig, L. (2019). Tissue-specific transposon-associated small RNAs in the gymnosperm tree, Norway spruce. *BMC Genomics* 20, 1–10. doi: 10.1186/s12864-019-6385-7

Neumann, P., Koblízková, A., Navrátilová, A., and Macas, J. (2006). Significant expansion of vicia pannonica genome size mediated by amplification of a single type of giant retroelement. *Genetics* 173, 1047–1056. doi: 10.1534/genetics.106.056259

Neumann, P., Navrátilová, A., Koblížková, A., Kejnovský, E., Hřibová, E., Hobza, R., et al. (2011). Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mobile DNA* 2, 1–16. doi: 10.1186/1759-8753-2-4

Neumann, P., Yan, H., and Jiang, J. (2007). The centromeric retrotransposons of rice are transcribed and differentially processed by RNA interference. *Genetics* 176, 749–761. doi: 10.1534/genetics.107.071902

Nie, Q., Qiao, G., Peng, L., and Wen, X. (2019). Transcriptional activation of long terminal repeat retrotransposon sequences in the genome of pitaya under abiotic stress. *Plant Physiol. Biochem.* 135, 460–468. doi: 10.1016/j.plaphy.2018.11.014

Nosaka, M., Itoh, J.-I., Nagato, Y., Ono, A., Ishiwata, A., and Sato, Y. (2012). Role of transposon-derived small RNAs in the interplay between genomes and parasitic DNA in rice. *PLoS Genet.* 8 (9), e1002953. doi: 10.1371/journal.pgen.1002953

Nozawa, K., Chen, J., Jiang, J., Leichter, S. M., Yamada, M., Suzuki, T., et al. (2021). DNA Methyltransferase CHROMOMETHYLASE3 prevents ONSEN transposon silencing under heat stress. *PloS Genet.* 17, e1009710. doi: 10.1371/journal.pgen.1009710

Nuthikattu, S., McCue, A. D., Panda, K., Fultz, D., DeFraia, C., Thomas, E. N., et al. (2013). The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Physiol.* 162, 116–131. doi: 10.1104/pp.113.216481

O'Neill, R. J. W., O'Neill, M. J., and Graves, J. A. M. (1998). Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* 393, 68–72. doi: 10.1038/29985

Oliveira, L. C., and Torres, G. A. (2018). Plant centromeres: genetics, epigenetics and evolution. *Mol. Biol. Rep.* 45, 1491–1497. doi: 10.1007/s11033-018-4284-7

Ouyang, Z., Wang, Y., Ma, T., Kanzana, G., Wu, F., and Zhang, J. (2021). Genome-wide identification and development of LTR retrotransposon-based molecular markers for the melilotus genus. *Plants* 10, 890. doi: 10.3390/plants10050890

Ozkan, H., Levy, A. A., and Feldman, M. (2001). Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops–triticum*) group. *Plant Cell* 13, 1735–1747. doi: 10.1105/tpc.010082

Panda, K., Ji, L., Neumann, D. A., Daron, J., Schmitz, R. J., and Slotkin, R. K. (2016). Full-length autonomous transposable elements are preferentially targeted by expression-dependent forms of RNA-directed DNA methylation. *Genome Biol.* 17, 1–19. doi: 10.1186/s13059-016-1032-y

Panda, K., McCue, A. D., and Slotkin, R. K. (2020). Arabidopsis RNA polymerase IV generates 21–22 nucleotide small RNAs that can participate in RNA-directed DNA methylation and may regulate genes. *Philos. Trans. R. Soc. B* 375, 20190417. doi: 10.1098/rstb.2019.0417

Papareddy, R. K., Páldi, K., Paulraj, S., Kao, P., Lutzmayer, S., and Nodine, M. D. (2020). Chromatin regulates expression of small RNAs to help maintain transposon methylome homeostasis in arabidopsis. *Genome Biol.* 21, 1–24. doi: 10.1186/s13059-020-02163-4

Papolu, P. K., Ramakrishnan, M., Wei, Q., Vinod, K. K., Zou, L.-H., Yrjala, K., et al. (2021). Long terminal repeats (LTR) and transcription factors regulate PHRE1 and PHRE2 activity in moso bamboo under heat stress. *BMC Plant Biol.* 21, 1–19. doi: 10.1186/s12870-021-03339-1

Parisod, C., Alix, K., Just, J., Petit, M., Sarilar, V., Mhiri, C., et al. (2010). Impact of transposable elements on the organization and function of allopolyploid genomes. *New Phytol.* 186, 37–45. doi: 10.1111/j.1469-8137.2009.03096.x

Parisod, C., Salmon, A., Zerjal, T., Tenaillon, M., Grandbastien, M. A., and Ainouche, M. (2009). Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in spartina. *New Phytol.* 184, 1003–1015. doi: 10.1111/j.1469-8137.2009.03029.x

Pearce, S. R., Stuart-Rogers, C., Knox, M. R., Kumar, A., Ellis, T. N., and Flavell, A. J. (1999). Rapid isolation of plant Ty1-copia group retrotransposon LTR sequences for molecular marker studies. *Plant J.* 19, 711–717. doi: 10.1046/j.1365-313x.1999.00556.x

Pecinka, A., Dinh, H. Q., Baubec, T., Rosa, M., Lettner, N., and Scheid, O. M. (2010). Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in arabidopsis. *Plant Cell* 22, 3118–3129. doi: 10.1105/tpc.110.078493

Peng, Y., Zhang, Y., Gui, Y., An, D., Liu, J., Xu, X., et al. (2019). Elimination of a retrotransposon for quenching genome instability in modern rice. *Mol. Plant* 12, 1395–1407. doi: 10.1016/j.molp.2019.06.004

Petit, M., Lim, K. Y., Julio, E., Poncet, C., De Borne, F. D., Kovarik, A., et al. (2007). Differential impact of retrotransposon populations on the genome of allotetraploid tobacco (Nicotiana tabacum). *Mol. Genet. Genomics* 278, 1–15. doi: 10.1007/s00438-007-0226-0

Piegu, B., Guyot, R., Picault, N., Roulin, A., Saniyal, A., Kim, H., et al. (2006). Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in oryza australiensis, a wild relative of rice. *Genome Res.* 16, 1262–1269. doi: 10.1101/gr.5290206

Pietzenuk, B., Markus, C., Gaubert, H., Bagwan, N., Merotto, A., Bucher, E., et al. (2016). Recurrent evolution of heat-responsiveness in brassicaceae COPIA elements. *Genome Biol.* 17, 1–15. doi: 10.1186/s13059-016-1072-3

Pontier, D., Picart, C., Roudier, F., Garcia, D., Lahmy, S., Azevedo, J., et al. (2012). NERD, a plant-specific GW protein, defines an additional RNAi-dependent chromatin-based pathway in arabidopsis. *Mol. Cell* 48, 121–132. doi: 10.1016/j.molcel.2012.07.027

Porquier, A., Tisserant, C., Salinas, F., Glassl, C., Wange, L. E., Enard, W., et al. (2021). Retrotransposons as pathogenicity factors of the plant pathogenic fungus botrytis cinerea. *Genome Biol.* 22, 225. doi: 10.1186/s13059-021-02446-4

Potter, S. (2005). Long terminal repeat (Ltr) type retrotransposons in populus species: A uniquely abundant and informative class of molecular markers for forest biotechnology. *For. Genet.* 12 (1), 35–44.

Presting, G. G. (2018). Centromeric retrotransposons and centromere function. *Curr. Opin. Genet. Dev.* 49, 79–84. doi: 10.1016/j.gde.2018.03.004

Presting, G. G., Malysheva, L., Fuchs, J., and Schubert, I. (1998). A TY3/GYPSY retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. *Plant J.* 16, 721–728. doi: 10.1046/j.1365-313x.1998.00341.x

Qi, L., Wu, J., Friebe, B., Qian, C., Gu, Y., Fu, D., et al. (2013). Sequence organization and evolutionary dynamics of brachypodium-specific centromere retrotransposons. *Chromosome Res.* 21, 507–521. doi: 10.1007/s10577-013-9378-4

Ramakrishnan, M., Satish, L., Kalendar, R., Narayanan, M., Kandasamy, S., Sharma, A., et al. (2021). The dynamism of transposon methylation for plant development and stress adaptation. *Int. J. Mol. Sci.* 22, 11387. doi: 10.3390/ijms222111387

Ramsey, J., and Schemske, D. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* 29 (1), 467–501.

Ravindran, S. (2012). Barbara McClintock and the discovery of jumping genes. *Proc. Natl. Acad. Sci.* 109, 20198–20199. doi: 10.1073/pnas.1219372109

Rey-Banos, R., Saenz de Miera, L. E., García, P., and Pérez de la Vega, M. (2017). Obtaining retrotransposon sequences, analysis of their genomic distribution and use of retrotransposon-derived genetic markers in lentil (*Lens culinaris medik.*). *PloS One* 12, e0176728. doi: 10.1371/journal.pone.0176728

Roulin, A., Piegu, B., Fortune, P. M., Sabot, F., d'Hont, A., Manicacci, D., et al. (2009). Whole genome surveys of rice, maize and sorghum reveal multiple horizontal transfers of the LTR-retrotransposon Route66 in poaceae. *BMC Evol. Biol.* 9, 1–10. doi: 10.1186/1471-2148-9-58

Salina, E. A., Sergeeva, E. M., Adonina, I. G., Shcherban, A. B., Belcram, H., Huneau, C., et al. (2011). The impact of Ty3-gypsy group LTR retrotransposons Fatima on b-genome specificity of polyploid wheats. *BMC Plant Biol.* 11, 1–14. doi: 10.1186/1471-2229-11-99

Sanan-Mishra, N., Abdul Kader Jailani, A., Mandal, B., and Mukherjee, S. K. (2021). Secondary siRNAs in plants: Biosynthesis, various functions, and applications in virology. *Front. Plant Sci.* 12, 610283. doi: 10.3389/fpls.2021.610283

Sanchez, D. H., Gaubert, H., Drost, H.-G., Zabet, N. R., and Paszkowski, J. (2017). High-frequency recombination between members of an LTR retrotransposon family during transposition bursts. *Nat. Commun.* 8, 1–7. doi: 10.1038/s41467-017-01374-x

Sanmiguel, P., and Bennetzen, J. L. (1998). Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Ann. Bot.* 82, 37–44. doi: 10.1006/anbo.1998.0746

Sant, V., Sainani, M., Sami-Subbu, R., Ranjekar, P., and Gupta, V. (2000). Ty1-copia retrotransposon-like elements in chickpea genome: their identification, distribution and use for diversity analysis. *Gene* 257, 157–166. doi: 10.1016/S0378-1119(00)00405-4

Schorn, A. J., Gutbrod, M. J., LeBlanc, C., and Martienssen, R. (2017). LTR-Retrotransposon control by tRNA-derived small RNAs. *Cell* 170, 61–71. e11. doi: 10.1016/j.cell.2017.06.013

Serag, M. F., Kaji, N., Gaillard, C., Okamoto, Y., Terasaka, K., Jabasini, M., et al. (2011). Trafficking and subcellular localization of multiwalled carbon nanotubes in plant cells. *ACS Nano* 5, 493–499. doi: 10.1021/nn102344t

Settles, A. M., Holding, D. R., Tan, B. C., Latshaw, S. P., Liu, J., Suzuki, M., et al. (2007). Sequence-indexed mutations in maize using the UniformMu transposon-tagging population. *BMC Genomics* 8, 1–12. doi: 10.1186/1471-2164-8-116

Shams, I., and Raskina, O. (2018). Intraspecific and intraorganismal copy number dynamics of retrotransposons and tandem repeat in aegilops speltoides tausch (Poaceae, triticeae). *Protoplasma* 255, 1023–1038. doi: 10.1007/s00709-018-1212-6

Shapiro, J. A. (2014). Epigenetic control of mobile DNA as an interface between experience and genome change. *Front. Genet.* 5. doi: 10.3389/fgene.2014.00087

Sharma, V., and Nandineni, M. R. (2014). Assessment of genetic diversity among Indian potato (Solanum tuberosum l.) collection using microsatellite and retrotransposon based marker systems. *Mol. Phylogenet. Evol.* 73, 10–17. doi: 10.1016/j.ympev.2014.01.003

Sharma, A., and Presting, G. G. (2014). Evolution of centromeric retrotransposons in grasses. *Genome Biol. Evol.* 6, 1335–1352. doi: 10.1093/gbe/evu096

Sharma, A., Wolfgruber, T. K., and Presting, G. G. (2013). Tandem repeats derived from centromeric retrotransposons. *BMC Genomics* 14, 1–11. doi: 10.1186/1471-2164-14-142

Shingote, P. R., Mithra, S. A., Sharma, P., Devanna, N. B., Arora, K., Holkar, S. K., et al. (2019). LTR Retrotransposons and highly informative ISSRs in combination are potential markers for genetic fidelity testing of tissue culture-raised plants in sugarcane. *Mol. Breed.* 39, 25. doi: 10.1007/s11032-019-0931-5

Slotkin, R. K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* 8, 272–285. doi: 10.1038/nrg2072

Slotkin, R. K., Vaughn, M., Borges, F., Tanurdžić, M., Becker, J. D., Feijó, J. A., et al. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* 136, 461–472. doi: 10.1016/j.cell.2008.12.038

Smýkal, P., Bačová-Kerteszová, N., Kalendar, R., Corander, J., Schulman, A. H., and Pavelek, M. (2011). Genetic diversity of cultivated flax (*Linum usitatissimum* l.) germplasm assessed by retrotransposon-based markers. *Theor. Appl. Genet.* 122, 1385–1397. doi: 10.1007/s00122-011-1539-2

Song, X., Li, Y., Cao, X., and Qi, Y. (2019). MicroRNAs and their regulatory roles in plant–environment interactions. *Annu. Rev. Plant Biol.* 70, 489–525. doi: 10.1146/annurev-arplant-050718-100334

Springer, N. M., Ying, K., Fu, Y., Ji, T., Yeh, C.-T., Jia, Y., et al. (2009). Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PloS Genet.* 5, e1000734. doi: 10.1371/journal.pgen.1000734

Stitzer, M. C., Anderson, S. N., Springer, N. M., and Ross-Ibarra, J. (2021). The genomic ecosystem of transposable elements in maize. *PloS Genet.* 17, e1009768. doi: 10.1371/journal.pgen.1009768

Sun, L., Jing, Y., Liu, X., Li, Q., Xue, Z., Cheng, Z., et al. (2020). Heat stress-induced transposon activation correlates with 3D chromatin organization rearrangement in arabidopsis. *Nat. Commun.* 11, 1–13. doi: 10.1038/s41467-020-15809-5

Šurbanovski, N., Brilli, M., Moser, M., and Si-Ammour, A. (2016). A highly specific micro RNA-mediated mechanism silences LTR retrotransposons of strawberry. *Plant J.* 85, 70–82. doi: 10.1111/tpj.13090

Syed, N., Sureshsundar, S., Wilkinson, M., Bhau, B., Cavalcanti, J., and Flavell, A. (2005). Ty1-copia retrotransposon-based SSAP marker development in cashew (Anacardium occidentale l.). *Theor. Appl. Genet.* 110, 1195–1202. doi: 10.1007/s00122-005-1948-1

Takehira, K., Hayashi, Y., Nozawa, K., Chen, L., Suzuki, T., Masuta, Y., et al. (2021). DRD1, a SWI/SNF-like chromatin remodeling protein, regulates a heat-

activated transposon in arabidopsis thaliana. *Genes Genet. Syst.* 96 (3), 151–158. doi: 10.1266/ggs.21-00005

Taller, D., Balint, J., Gyula, P., Nagy, T., Barta, E., Baksa, I., et al. (2018a). Expansion of capsicum annum fruit is linked to dynamic tissue-specific differential expression of miRNA and siRNA profiles. *PloS One* 13, e0200207. doi: 10.1371/journal.pone.0200207

Taller, D., Bálint, J., Gyula, P., Nagy, T., Barta, E., Baksa, I., et al. (2018b). Correction: Expansion of capsicum annum fruit is linked to dynamic tissue-specific differential expression of miRNA and siRNA profiles. *PloS One* 13, e0203582. doi: 10.1371/journal.pone.0203582

Tek, A. L., Kashihara, K., Murata, M., and Nagaki, K. (2010). Functional centromeres in soybean include two distinct tandem repeats and a retrotransposon. *Chromosome Res.* 18, 337–347. doi: 10.1007/s10577-010-9119-x

Theuri, J., Phelps-Durr, T., Mathews, S., and Birchler, J. (2005). A comparative study of retrotransposons in the centromeric regions of a and b chromosomes of maize. *Cytogenetic Genome Res.* 110, 203–208. doi: 10.1159/000084953

Tian, Z., Rizzon, C., Du, J., Zhu, L., Bennetzen, J. L., Jackson, S. A., et al. (2009). Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons? *Genome Res.* 19, 2221–2230. doi: 10.1101/gr.083899.108

Tian, Z., Zhao, M., She, M., Du, J., Cannon, S. B., Liu, X., et al. (2012). Genome-wide characterization of nonreference transposons reveals evolutionary propensities of transposons in soybean. *Plant Cell* 24, 4422–4436. doi: 10.1105/tpc.112.103630

Tittel-Elmer, M., Bucher, E., Broger, L., Mathieu, O., Paszkowski, J., and Vaillant, I. (2010). Stress-induced activation of heterochromatic transcription. *PloS Genet.* 6, e1001175. doi: 10.1371/journal.pgen.1001175

Tsukahara, S., Kawabe, A., Kobayashi, A., Ito, T., Aizu, T., Shin-i, T., et al. (2012). Centromere-targeted *de novo* integrations of an LTR retrotransposon of arabidopsis lyrata. *Genes Dev.* 26, 705–713. doi: 10.1101/gad.183871.111

Ujino-Ihara, T. (2020). Transcriptome analysis of heat stressed seedlings with or without pre-heat treatment in cryptomeria japonica. *Mol. Genet. Genomics* 295, 1163–1172. doi: 10.1007/s00438-020-01689-3

Vicient, C. M., and Casacuberta, J. M. (2017). Impact of transposable elements on polyploid plant genomes. *Ann. Bot.* 20 (2), 195–207. doi: 10.1093/aob/mcx078

Voronova, A., Belevich, V., Jansons, A., and Rungis, D. (2014). Stress-induced transcriptional activation of retrotransposon-like sequences in the scots pine (*Pinus sylvestris* l.) genome. *Tree Genet. Genomes* 10, 937–951. doi: 10.1007/s11295-014-0733-1

Voronova, A., Jansons, Ā, and Ruņģis, D. (2011). Expression of retrotransposon-like sequences in scots pine (*Pinus sylvestris*) in response to heat stress. *Environ. Exp. Biol.* 9, 121–127.

Voronova, A., Rendón-Anaya, M., Ingvarsson, P., Kalendar, R., and Ruņģis, D. (2020). Comparative study of pine reference genomes reveals transposable element interconnected gene networks. *Genes* 11, 1216. doi: 10.3390/genes11101216

Vuorinen, A. L., Kalendar, R., Fahima, T., Korpelainen, H., Nevo, E., and Schulman, A. H. (2018). Retrotransposon-based genetic diversity assessment in wild emmer wheat (*Triticum turgidum ssp. dicoccoides*). *Agronomy* 8, 107. doi: 10.3390/agronomy8070107

Wang, J. W., Cunningham, F. J., Goh, N. S., Boozarpour, N. N., Pham, M., and Landry, M. P. (2021). Nanoparticles for protein delivery in planta. *Curr. Opin. Plant Biol.* 60, 102052. doi: 10.1016/j.pbi.2021.102052

Wang, Q., and Dooner, H. K. (2006). Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus. *Proc. Natl. Acad. Sci.* 103, 17644–17649. doi: 10.1073/pnas.0603080103

Wang, J. W., Grandio, E. G., Newkirk, G. M., Demirer, G. S., Butrus, S., Giraldo, J. P., et al. (2019). Nanoparticle-mediated genetic engineering of plants. *Mol. Plant* 12, 1037–1040. doi: 10.1016/j.molp.2019.06.010

Wang, G., He, Q., Liu, F., Cheng, Z., Talbert, P. B., and Jin, W. (2011). Characterization of CENH3 proteins and centromere-associated DNA sequences in diploid and allotetraploid brassica species. *Chromosoma* 120, 353–365. doi: 10.1007/s00412-011-0315-z

Wang, G.-X., He, Q.-Y., Zhao, H., Cai, Z.-X., Guo, N., Zong, M., et al. (2019). ChIP-cloning analysis uncovers centromere-specific retrotransposons in brassica nigra and reveals their rapid diversification in brassica allotetraploids. *Chromosoma* 128, 119–131. doi: 10.1007/s00412-019-00701-z

Wang, Y., Liang, W., and Tang, T. (2018). Constant conflict between gypsy LTR retrotransposons and CHH methylation within a stress-adapted mangrove genome. *New Phytol.* 220, 922–935. doi: 10.1111/nph.15209

Wang, S., Liang, H., Xu, Y., Li, L., Wang, H., Sahu, D. N., et al. (2021). Genome-wide analyses across viridiplantae reveal the origin and diversification of small RNA pathway-related genes. *Commun. Biol.* 4, 412. doi: 10.1038/s42003-021-01933-5

Waugh, R., McLean, K., Flavell, A., Pearce, S., Kumar, A., Thomas, B., et al. (1997). Genetic distribution of bare–1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Mol. Gen. Genet. MGG* 253, 687–694. doi: 10.1007/s004380050372

Weber, B., and Schmidt, T. (2009). Nested Ty3-gypsy retrotransposons of a single beta procumbens centromere contain a putative chromodomain. *Chromosome Res.* 17, 379–396. doi: 10.1007/s10577-009-9029-y

Wei, L., Gu, L., Song, X., Cui, X., Lu, Z., Zhou, M., et al. (2014). Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. *Proc. Natl. Acad. Sci.* 111, 3877–3882. doi: 10.1073/pnas.1318131111

Weise, K., Winter, L., Fischer, E., Kneis, D., de la Cruz Barron, M., Kunze, S., et al. (2022). Multiwalled carbon nanotubes promote bacterial conjugative plasmid transfer. *Microbiol. Spectr.* 10, e00410–e00422. doi: 10.1128/spectrum.00410-22

Wei, L., Xiao, M., An, Z., Ma, B., Mason, A. S., Qian, W., et al. (2013). New insights into nested long terminal repeat retrotransposons in brassica species. *Mol. Plant* 6, 470–482. doi: 10.1093/mp/sss081

Wessler, S. R. (1996). Plant retrotransposons: turned on by stress. *Curr. Biol.* 6, 959–961. doi: 10.1016/S0960-9822(02)00638-3

Wessler, S. R., Bureau, T. E., and White, S. E. (1995). LTR-Retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr. Opin. Genet. Dev.* 5, 814–821. doi: 10.1016/0959-437X(95)80016-X

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982. doi: 10.1038/nrg2165

Wolfgruber, T. K., Sharma, A., Schneider, K. L., Albert, P. S., Koo, D.-H., Shi, J., et al. (2009). Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic loci shaped primarily by retrotransposons. *PloS Genet.* 5, e1000743. doi: 10.1371/journal.pgen.1000743

Woodrow, P., Pontecorvo, G., and Ciarmiello, L. F. (2012). Isolation of Ty1-copia retrotransposon in myrtle genome and development of s-SAP molecular marker. *Mol. Biol. Rep.* 39, 3409–3418. doi: 10.1007/s11033-011-1112-8

Xie, Z., Johansen, L. K., Gustafson, A. M., Kasschau, K. D., Lellis, A. D., Zilberman, D., et al. (2004). Genetic and functional diversification of small RNA pathways in plants. *PloS Biol.* 2, e104. doi: 10.1371/journal.pbio.0020104

Yañez-Santos, A. M., Paz, R. C., Paz-Sepúlveda, P. B., and Urdampilleta, J. D. (2021). Full-length LTR retroelements in capsicum annuum revealed a few species-specific family bursts with insertional preferences. *Chromosome Res.* 29, 261–284. doi: 10.1007/s10577-021-09663-4

Yang, T.-J., Lee, S., Chang, S.-B., Yu, Y., de Jong, H., and Wing, R. A. (2005). In-depth sequence analysis of the tomato chromosome 12 centromeric region: Identification of a large CAA block and characterization of pericentromere retrotranposons. *Chromosoma* 114, 103–117. doi: 10.1007/s00412-005-0342-8

Yang, D.-L., Zhang, G., Tang, K., Li, J., Yang, L., Huang, H., et al. (2016). Dicer-independent RNA-directed DNA methylation in arabidopsis. *Cell Res.* 26, 66–82. doi: 10.1038/cr.2015.145

Ye, R., Chen, Z., Lian, B., Rowley, M. J., Xia, N., Chai, J., et al. (2016). A dicer-independent route for biogenesis of siRNAs that direct DNA methylation in arabidopsis. *Mol. Cell* 61, 222–235. doi: 10.1016/j.molcel.2015.11.015

Yi, C., Zhang, W., Dai, X., Li, X., Gong, Z., Zhou, Y., et al. (2013). Identification and diversity of functional centromere satellites in the wild rice species oryza brachyantha. *Chromosome Res.* 21, 725–737. doi: 10.1007/s10577-013-9374-8

Yokthongwattana, C., Bucher, E., Čaikovski, M., Vaillant, I., Nicolet, J., Scheid, O. M., et al. (2010). MOM1 and pol-IV/V interactions regulate the intensity and specificity of transcriptional gene silencing. *EMBO J.* 29, 340–351. doi: 10.1038/emboj.2009.328

Zhang, W., Cao, Y., Wang, K., Zhao, T., Chen, J., Pan, M., et al. (2014). Identification of centromeric regions on the linkage map of cotton using centromere-related repeats. *Genomics* 104, 587–593. doi: 10.1016/j.ygeno.2014.09.002

Zhang, H., Cao, Y., Xu, D., Goh, N. S., Demirer, G. S., Cestellos-Blanco, S., et al. (2021). Gold-nanocluster-mediated delivery of siRNA to intact plant cells for efficient gene knockdown. *Nano Lett.* 21, 5859–5866. doi: 10.1021/acs.nanolett.1c01792

Zhang, T., Talbert, P. B., Zhang, W., Wu, Y., Yang, Z., Henikoff, J. G., et al. (2013). The CentO satellite confers translational and rotational phasing on cenH3 nucleosomes in rice centromeres. *Proc. Natl. Acad. Sci.* 110, E4875–E4883. doi: 10.1073/pnas.1319548110

Zhang, H., Tao, Z., Hong, H., Chen, Z., Wu, C., Li, X., et al. (2016). Transposon-derived small RNA is responsible for modified function of WRKY45 locus. *Nat. Plants* 2, 1–8. doi: 10.1038/nplants.2016.16

Zhou, M., Hu, B., and Zhu, Y. (2017a). Genome-wide characterization and evolution analysis of long terminal repeat retroelements in moso bamboo (Phyllostachys edulis). *Tree Genet. Genomes* 13, 43. doi: 10.1007/s11295-017-1114-3

Zhou, S.-S., Yan, X.-M., Zhang, K.-F., Liu, H., Xu, J., Nie, S., et al. (2021). A comprehensive annotation dataset of intact LTR retrotransposons of 300 plant genomes. *Sci. Data* 8, 1–9. doi: 10.1038/s41597-021-00968-x

Zhou, M., Zhu, Y., Bai, Y., Hänninen, H., and Meng, X. (2017b). Transcriptionally active LTR retroelement-related sequences and their relationship with small RNA in moso bamboo (Phyllostachys edulis). *Mol. Breed.* 37, 1–11. doi: 10.1007/s11032-017-0733-6

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 December 2022
DOI 10.3389/fpls.2022.1035383

OPEN ACCESS

EDITED BY
Yusuf Khan,
Oslo University Hospital, Norway

REVIEWED BY
Md. Mia Mukul,
Bangladesh Jute Research Institute, Bangladesh
Zeba I. Seraj,
University of Dhaka, Bangladesh
Mrinalini Manna,
National Institute of Plant Genome Research (NIPGR), India

*CORRESPONDENCE
Tahmina Islam
subarna.islam@gmail.com;
t.islam@du.ac.bd

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 02 September 2022
ACCEPTED 12 October 2022
PUBLISHED 14 December 2022

CITATION
Akhter S, Sami AA, Toma TI, Jahan B and Islam T (2022) *Caffeoyl-CoA 3-O-methyltransferase* gene family in jute: Genome-wide identification, evolutionary progression and transcript profiling under different quandaries.
Front. Plant Sci. 13:1035383.
doi: 10.3389/fpls.2022.1035383

Caffeoyl-CoA 3-O-methyltransferase gene family in jute: Genome-wide identification, evolutionary progression and transcript profiling under different quandaries

Saima Akhter, Asif Ahmed Sami, Tamanna Islam Toma, Bushrat Jahan and Tahmina Islam*

Plant Breeding and Biotechnology Laboratory, Department of Botany, University of Dhaka, Dhaka, Bangladesh

Jute (*Corchorus* sp.), is a versatile, naturally occurring, biodegradable material that holds the promising possibility of diminishing the extensive use of plastic bags. One of the major components of the cell wall, lignin plays both positive and negative roles in fiber fineness and quality. Although it gives mechanical strength to plants, an excess amount of it is responsible for the diminution of fiber quality. Among various gene families involved in the lignin biosynthesis, Caffeoyl-CoA 3-O-methyltransferase (*CCoAOMT*) is the most significant and has remained mostly unexplored. In this study, an extensive *in-silico* characterization of the *CCoAOMT* gene family was carried out in two jute species (*C. capsularis* L. and *C. olitoroius* L.) by analyzing their structural, functional, molecular and evolutionary characteristics. A total of 6 *CCoAOMT* gene members were identified in each of the two species using published reference genomes. These two jute species showed high syntenic conservation and the identified *CCoAOMT* genes formed four clusters in the phylogenetic tree. Histochemical assay of lignin in both jute species could shed light on the deposition pattern in stems and how it changes in response to abiotic stresses. Furthermore, expression profiling using qPCR showed considerable alteration of *CCoAOMT* transcripts under various abiotic stresses and hormonal treatment. This study will lay a base for further analysis and exploration of target candidates for overexpression of gene silencing using modern biotechnological techniques to enhance the quality of this economically important fiber crop.

KEYWORDS

CCoAOMT gene, jute, lignin, abiotic stress, functional validation, histochemical assay

Introduction

Jute, a dicotyledonous fiber yielding plant of the genus *Corchorus*, family Malvaceae, has been one of the major fiber yielding crops worldwide. Jute is the second most important bast fiber (fiber collected from the skin of the plant) after cotton in terms of usage, worldwide consumption, and availability, accounting for 84% of global output. Jute and jute products significantly benefit the environment by slowing down ecological degradation and preserving the natural environment and atmosphere. Jute fibers are eco-friendly materials because of their biodegradable properties and are easily recyclable. The bag and cloth industry are the largest consumer of jute fibers available on the market. Jute bags have gained popularity as an eco-friendly alternative to both non-biodegradable poly bags derived from petroleum and paper bags that require significant amounts of wood. To improve the quality of jute products, knowing the components of its fiber is needed. Jute products are mainly made from the outer portion of the stem, also known as bast fiber. This is a lignocellulosic fiber composed of 61-71.5% cellulose, 13.6-20.6% hemicellulose and 12-13% lignin (Sood and Dwivedi, 2018). However, up to 12-26% lignin was also recorded in jute fibers (Selver et al., 2018). The current status of lignin in Tossa pat variety O-9897 of Bangladesh is 29.50% and 13.46% in whole stem and in fiber respectively (Shafrin et al., 2017) and in O-72 it was measured as 29% in mature jute plant with a deposition ratio 0.21% per day (Tanmoy et al., 2015).

Lignin is a complex heteropolymer of cinnamyl alcohols with high molecular weight and is a key component of the jute cell wall. Also, it is the second most prevalent biopolymer on Earth after cellulose (Baucher et al., 1998; Liu et al., 2018). Dehydrogenative polymerization of three hydroxy-cinnamyl alcohols (monolignols), namely p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol, results in p-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units of lignin, respectively (Vanholme et al., 2010). As one of the key components of the cell wall, lignin metabolism has a variety of roles, most notably in battling numerous environmental stressors, as the plant cell wall serves as the first line of defense against external threats. Lignin has a crucial function in the structural integrity of cell walls, stem strength, water transport, and pathogen resistance (Mutuku et al., 2019). This high lignin content separates jute fiber from other non-wood bast fibers such as flax, hemp, and ramie, as well as kenaf (Del Río et al., 2009). The poor quality of jute fibers in comparison to other fiber crops is mostly owing to greater levels of lignin in its fiber (more than ramie and cotton), making it less appropriate for creating finer textiles and other value-added goods (Loumerem and Alercia, 2016). Thus, understanding the enzymes involved in the biosynthesis of lignin is a prerequisite for any type of analysis to alter the amount or components of lignin in plants.

Caffeoyl-CoA 3-O-methyltransferase (*CCoAOMT*) is one of the primary enzymes responsible for the biosynthesis of lignin in plants (Zhong et al., 1998; Meyermans et al., 2000). This enzyme is a type of S-adenosyl-L-methionine (SAM) methyltransferase that utilizes coffee acyl coenzyme A as a substrate. The enzyme is encoded by the *CCoAOMT* gene, hitherto, various techniques have been reported to lower the lignin contents by targeting the *CCoAOMT* gene in those respective plants. For instance, in *Pinus radiata*, *CCoAOMT* suppression alters lignin composition (Zhong et al., 1998; Kawaoka et al., 2006; Wagner et al., 2011). Lignin content was lowered in tobacco by suppressing two O-methyltransferase genes - caffeic acid O-methyltransferase (*COMT*) and *CCoAOMT* through antisense technology (Zhao et al., 2002; Vanholme et al., 2008; Tronchet et al., 2010). *CCoAOMT* suppression in *Nicotiana tabacum*, *Arabidopsis thaliana*, *Medicago sativa*, and *Populus* resulted in lignin reductions of 20-45% of total lignin (Meyermans et al., 2000; Zhong et al., 2000; Do et al., 2007). The *CCoAMT1* gene was downregulated in the low-lignin mutant of *Corchorus olitorius* and was found to be accountable for the mutant's low lignin content (Choudhary et al., 2017). Furthermore, overexpression of a jute *CCoAOMT1* gene in *Arabidopsis thaliana* resulted in higher lignin content compared to the non-transgenic plants (Zhang et al., 2014). This confirms the important role of *CCoAOMT* gene in the lignin biosynthesis pathway (Zhang et al., 2014). Lignin gives mechanical support to fight against biotic stresses, it also helps plants in combating various abiotic stresses (Cabane et al., 2012).

Members of the *CCoAOMT* gene family have already been found and studied in a variety of plant species, including Arabidopsis, rice, sorghum, citrus, and poplar (Hamberger et al., 2007; Xu et al., 2009; Liu et al., 2016; Rakoczy et al., 2018). But, it is yet to be characterized fully in jute. The present study tackles this gap through an *in-silico* analysis of all the members of the *CCoAOMT* gene family from two jute species (*C. capsularis* and *C. olitorius*). Here, we report 6 *CCoAOMT* members in each of the jute species. Lignin deposition in the jute stem under normal environmental conditions and in different abiotic stresses was also shown. Overall, the functional validation and expression profiling of this gene family provides a foundation for future genetic engineering and plant breeding programs to alter lignin composition in jute plants.

Materials and methods

Plant growth condition and stress treatments

Among the two jute species, BJRI Tossa Pat-8 (Robi-1) variety of *C. olitorius* was chosen due to its high germination rate for expression profiling, histochemical assay and for stress responsive amino acid and phenolic measurement. Uniformly developed jute seed of the BJRI Tossa Pat-8 variety was planted

in optimum greenhouse conditions for (16 h photoperiod, 28C temperature, 80% humidity). Plants were irrigated with Hogland solution on alternate days to create optimum growing conditions. Two months old plants were subjected to different abiotic stresses such as hormone (20 uM ABA), drought (20% (w/v) PEG 6000 solution) and saline (200 mM NaCl) treatment. For salinity treatment, two months old soil grown plants irrigated with 200mM NaCl in Hogland solution (Li et al., 2019). A solution of 20% Poly-Ethylene Glycol (PEG6000) in Hogland solution was used to create drought conditions. Similarly, 20 uM ABA was applied in the plant by spraying to mimic the hormonal treatment. The control plants were irrigated with hogland solution. Leaf and stem samples (three biological replicates) from control and treated plants were collected at 24-, 48- and 72-hours intervals (for expression profiling) and after 72 hours for other experiments, and instantly frozen in liquid nitrogen before being stored at -80°C. The untreated plants served as control compared to the treated plants.

Histochemical assay of lignin and fluorescence microscopy

Stem sample from 72h stresses two months old, lignified jute (BJRI Tossa Pat-8 variety of *C. olitorius*) were collected to elucidate any difference in the deposition of lignin in different regions (epidermis, bast region, pith). Hand sections of jute stems were dyed with phloroglucinol, a dye that combines with the cinnamaldehyde end groups of lignin to produce a crimson color in the lignified region and examined under a 10× magnification light microscope (Nikon ECLIPES 50i). The presence of lignin was observed when the tissues were dyed red. A CCD (charged-couple device) camera (Nikon DS-U3 DS Camera Control Unit; Ver.1.00) was used to capture the fluorescence images. No filter was used during the microscopy.

Identification, gene nomenclature of *CCoAOMT* gene family in two species of jute along with molecular attributes

According to Zhang et al. (2021) the sequences and all the required files of identified *CCoAOMT* genes in two species of jute (*C. capsularis* var. 'Huangma 179' (HM179) and *C. olitorius* var. 'Kuanyechangguo' (KYCG)) were retrieved from Genome Warehouse in the National Genomics Data Center. To confirm their function, the putative CCoAOMT sequences were analyzed using the Pfam database (Mistry et al., 2020). The gene names were assigned in descending order based on their chromosome position (Hasan et al., 2021). Information regarding locus ID, protein size and full-length CDS were obtained from the National Genomics Data Center. Expasy's Protparam tool was used to acquire information on molecular weight, theoretical pI, and number of amino acids (Gasteiger et al., 2005). CELLO v.2.5: subCELlular Localisation predictor (Yu et al., 2006) and WoLF pSORT4 software were used to predict protein subcellular localization. (Horton et al., 2007). The violin plot was constructed using GraphPad Prism 9.3.1 software (GraphPad Inc.,San Diego, California USA, www.graphpad.com)

Chromosomal localization of *CCoAOMT* genes

Information on the physical location of the *CCoAOMT* sequences from two jute species were collected from the corresponding GFF files, and TBtools (v.1.0971) (Chen et al., 2020) was used to visualize the distribution of *CCoAOMT* genes on each *C. capsularis* and *C. olitorius* chromosome. All the genes involved in the lignin production pathway are highlighted in black, while the *CCoAOMT* genes are highlighted in red. The gene density information was also depicted in the chromosomes.

Exon-intron distribution of *CCoAOMT* genes, domain architecture along with the amino acid contents in CCoAOMT proteins

The gene structure was created using the Gene Structure Display Server. Pfam was used to obtain domain information for each CCoAOMT protein (Mistry et al., 2020) for the confirmation of identifying domain of this family Methyltransferase-3. The domain architecture was constructed using DOG2.0 software (Ren et al., 2009). The amino acid content was obtained from the Protparam tool of Expasy (Gasteiger et al., 2005).

Analysis of Cis-regulatory elements

To identify the cis-regulatory element present in the promoter region of two species of Jute, a 2.0 kb upstream region of each gene was used. The potential upstream sequences were submitted to the PlantCARE database (Lescot, 2002) for identification and analysis of several cis-regulatory elements. Later, sum of all elements present in each jute species was visualized as Heatmap constructed using the "Heatmap Illustrator" option in TBtools (v.1.0971) (Chen et al., 2020) software. The total number of different cis-elements in each jute species were categorized into four groups according to their role under various condition (stress, hormone, growth and light) and were visualized as Venn diagram using TBtools (v.1.0971) software (Chen et al., 2020).

Phylogenetic analysis along with conserved motif

To get an overview of the evolutionary pattern of the *CCoAOMT* gene family in jute species, a phylogenetic tree was constructed using the sequences from *Arabidopsis thaliana*, *Oryza sativa*, *Sorghum bicolor*, *Medicago truncatula*, *Populus trichocarpa*, *Vitis vinifera* and *Camellia sinensis*. The peptide sequences of CCoAOMTs of *Camellia sinensis* were obtained from tea genome database (TIPA, http://tpia.teaplant.org/) (Xia et al., 2019) and the sequences from other species were from Phytozome (https://phytozome.jgi.doe.gov/) (Goodstein et al., 2011). The phylogenetic tree was constructed by IQ-Tree v.2.1.2 (Minh et al., 2020) with 1000 ultrafast bootstrap replicates using the alignment file generated through MAFFT tool v.7 (Katoh et al., 2019) with default parameters. The tree was visualized and edited using iTOL (v.6) (Letunic and Bork, 2019). Information of all the sequences used in this phylogenetic tree are listed in Supplementary Table 4. MEME SUITE (v.5.4.1) (Bailey et al., 2009) was used to identify the conserved motif with default parameters except the number of motifs was set to 15 in this case. The information on the conserved motif was uploaded in iTOL (v.6) as an additional dataset along with a phylogenetic tree.

In-silico expression analysis of *CCoAOMT* genes throughout different tissues and analysis of their stress responsive cis-regulatory elements

Multiple sequence alignment of all selected candidate protein sequences was performed using the online version of MAFFT program v.7 (Katoh et al., 2019) with default parameters to determine the sequence conservation and functional homology of CCoAOMTs within the two species. With the multiple sequence alignment files, a phylogenetic tree was generated using 1000 ultrafast bootstrap iterations using IQ-Tree v.2.1.2 (Minh et al., 2020). The tree was then visualized and edited using iTOL (v.6) (Letunic and Bork, 2019). The cis-regulatory elements and *in silico* expression data from different developmental stages were uploaded as an additional dataset in the required format. For the analysis of cis-regulatory elements 2.0 kb regions of each member were analyzed using PlantCARE database (Lescot, 2002).

Synteny analysis

The whole proteome of both jute species and Arabidopsis were compared using the blastp command of DIAMOND (v2.0.13) with default parameters (Buchfink et al., 2021). The resulting.blast file was used to detect Synteny blocks between each pair of species with the help of were detected using MCScanX (Wang et al., 2012). The blocks were and later visualized using SynVisio (Bandi and Gutwin, 2020).

RNA extraction and qRT-PCR analysis

The total RNA from the stressed leaf and stem tissues, as well as the control tissue, was extracted using the Monarch Total RNA Miniprep Kit (New England Biolabs, #T2010) according to the manufacturer's procedure. Following the manufacturer's protocol, the first strand cDNA was synthesized using a ProtoScript® First Strand cDNA Synthesis Kit (New England Biolabs, #T2010). The expression of each gene was measured with gene-specific primers through qRT-PCR analysis with CFX96 Touch Real-Time PCR Detection System (Bio-Rad, United States) and SYBR Green mixture. The relative expression of the genes was determined using the $2^{-\Delta\Delta\ Ct}$ method (Livak and Schmittgen, 2001), where ΔΔCt was the difference in the threshold cycles and the reference gene, which was *actin* for the expression analyses. The sequences of all gene-specific primers are presented in Supplementary Table 3.

Statistical analysis

Statistical data analysis was performed from three biological replicates under each treatment and time-point (n = 3). Statistical significance was determined using the Analysis of variance (ANOVA) test at P value ≤ 0.05 that were marked with different letters.

Results

Histochemical assay of lignin deposition and analysis of accumulation of stress responsive amino acid and phenolics in jute plants under stress

One of the key components of the jute cell wall, lignin, is expected to play an important role in combating numerous biotic and abiotic stressors. Histochemical assay of lignin was carried out using phloroglucinol. In control plants (Figures 1A, E), reddish deposition of lignin was found between the epidermis (E), vascular cambium (VC) and pith (P) region. After 48 hours of stress exposure, the deposition of lignin increased notably in stressed tissues (Figures 1B–D) which is evident based on the intensity of the staining. In contrast, 72 hours post-stress, the deposition of lignin was found to be decreased in stressed tissues (Figures 1F–H). Apart from that, plants depend on the amino acid proline for several functions. It shields plants from a variety of challenges and also promotes a quicker recovery time for

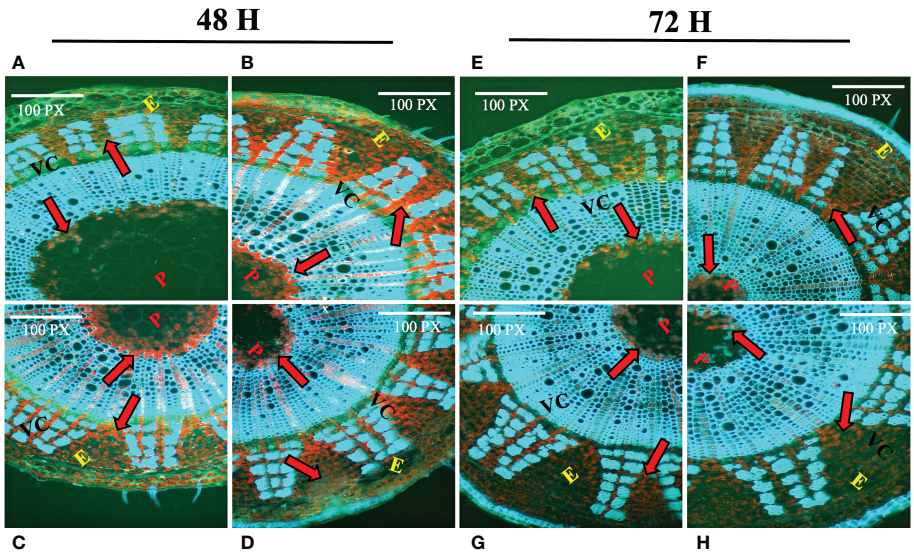


FIGURE 1
Histochemical assay of lignin deposition in two months old *C olitorius* stem under control and stressed conditions at different time intervals (48 hours and 72 hours). **(A–D)** Stem section of control and three stress treated plants **(A)** control **(B)** hormone **(C)** drought **(D)** salinity stress; after 48 hours; and **(E–H)** Stem section of control plant and three stress treated plants **(E)** control **(F)** hormone **(G)** drought **(H)** salinity stress; after 72 hours. Different letters indicate different regions of the stem (E, epidermis; VC, vascular cambium; P, pith). Red arrows indicate the deposition of lignin.

plants after experiencing stress. In response to stress, plants accumulate phenolic compounds as a defensive response. In this study, proline and phenolics accumulation of the farmer's popular variety, BJRI Tossa Pat-8 variety of *C. olitorius* was analyzed under different abiotic stresses (20 uM ABA, 200 mM NaCl and 20% PEG stress) and proline content was measured after 72 hours in leaf and stem tissues. The amount of proline accumulation was found to be higher in stem tissues compared to leaves and highest accumulation was found under drought stress (Supplementary Figure S1A). In contrast, a higher accumulation of phenolics was found in leaf tissues under all stresses (Supplementary Figure S1B). Because *CCoAoMT* gene family is engaged in the lignin production pathway, which is also a substantial component of jute stem bark, involvement of this gene family under diverse abiotic stressors will be an important study for selecting prospective gene members for various investigations.

Identification of *CCoAOMT* genes in jute

To identify the gene family members responsible for the synthesis of lignin, one of the important lignin bisysnthesis gene, *CCoAOMT* family were studied in jute. A total of six *CCoAOMT* genes were identified in each of the jute species. Genes were named *CcCCoAOMT* and *CoCCoAOMT* for *C. capsularis* and *C. olitorius*, respectively. All details of the identified genes are listed in Supplementary Table 1. Genes were numbered 1-6 based on their position on the chromosomes. The CDS length ranged from 1147 bp (*CcCCoAOMT2*) to 9315 bp (*CcCCoAOMT6*) in case of *C. capsularis* and from 1136 bp (*CoCCoAOMT3*) to 18963 bp (*CoCCoAOMT4*) in case of *C. olitorius* (Supplementary Table 1). Meanwhile, length of the proteins encoded by these genes ranged from 231 aa (CcCCoAOMT3) to 1624 aa (CcCCoAOMT6) in case of *C. capsularis* and 226 aa (CoCCoAOMT2) to 957 aa (CoCCoAOMT4) in case of *C. olitorius* (Supplementary Table 1). No noticeable variations were found in terms of CDS length or amino acid number between the two jute species (Figure 2). Additionally, the physical and chemical characteristics of CCoAOMT proteins were examined, and molecular mass and isoelectric points showed no discernible differences (Figure 2). Most of the CCoAOMTs were predicted to be localized in cytoplasm except for CcCCoAOMT6 (Nuclear and mitochondrial) and CoCCoAOMT4 (Cytoplasm and Chloroplast) (Supplementary Table 1).

Chromosomal localization of *CCoAOMT* genes

Gene families those are crucial for lignin biosynthesis include- *Phenylalanine ammonia-lyase (PAL), Cinnamic acid 4-hydroxylase (C4H), 4-Coumarate-coenzyme A ligase (4CL), caffeic acid O-methyltransferase (COMT), hydroxycinnamoyltransferase (HCT), Cinnamoyl-CoA reductase (CCR), cinnamyl alcohol dehydrase*

(CAD), caffeoyl coenzyme A 3-O-metyltransferase (CCoAOMT), p-coumarate 3-hydroxylase (C3H) and *Coniferaldehyde 5-hydroxylase (F5H)* (Yoon et al., 2015). The basic number of chromosome set for two species of jute is n=7. To identify their distribution and position on the jute chromosomes, a chromosome map was generated. All the gene members of the above mentioned families are represented in Figure 3. Members of the *CCoAOMT* gene family are indicated in red, whereas other gene members are marked in black. The chromosomal localization of the identified 6 *CCoAOMT* genes of *C. capsularis* reveals that these genes are located in 4 out of the 7 chromosomes (Figure 3A), and in *C. olitorius*, all 6 genes are located on 3 chromosomes and 1 contig (Figure 3B). Chromosome 1 of *C. olitorius* is devoid of any lignin biosynthesis pathway gene.

Exon-Intron distribution of *CCoAOMT* genes, domain architecture along with amino acid contents in CCoAOMT proteins

The exon-intron distribution was examined to gain a better understanding of the structure and evolution of the *CCoAOMT* genes. The number of exons and introns in both jute species show little variation. Nine out of a total of twelve members from both species contain 5 exons (Figure 4A). The highest number of exons was found in *CoCCoAOMT4*, which is 21 (Figure 4A). Variation in the exon-intron distribution indicates the structure of this gene family members is evolutionarily conserved which puts light on their specific role in the lignin biosynthesis pathway. Moreover, all members of this gene family had the characteristic domain Methyltransferase 3, confirming their identity as a member of the *CCoAOMT* gene family. In concordance with the gene structure, the highest number of domains were found in *CoCCoAOMT4* (Figure 4B). The percentage of the composition of aa is also found to be highly similar in all 12 CCoAOMT proteins of two jute species (Figure 4C). Overall, only minor variations were observed in structural properties between the *CCoAOMT* in these two species.

Analysis of Cis-regulatory elements

To further understand the putative regulatory networks and the elements responsible for the specific function of the *CCoAOMT* family genes, cis-elements in the 2 kb upstream sequences of the 6 *CCoAOMT* genes of each jute species were evaluated using PlantCARE. There were 74 different types of cis-elements discovered, which were then classified into four categories: stress, hormone, growth and light. The total

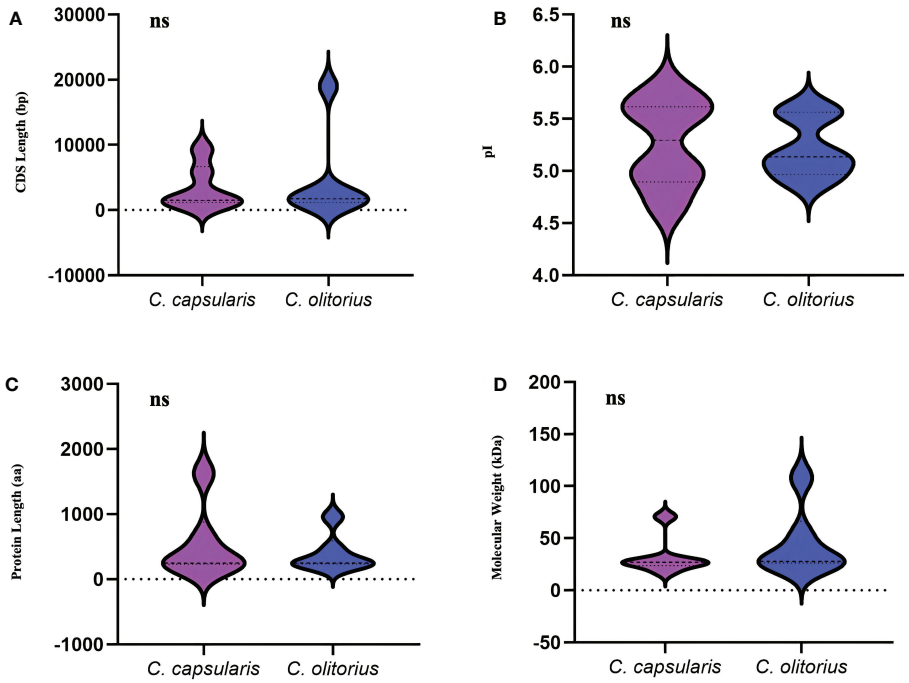


FIGURE 2
Overview of the physical and chemical properties of the *CCoAOMT* family members in two jute species. **(A)** CDS length of *CCoAOMT* genes and **(B)** Sum of amino acid residues in CCoAOMT proteins. **(C)** The molecular weight of CCoAOMT proteins. **(D)** The isoelectric point of CCoAOMT proteins. The 'ns' means no significant difference.

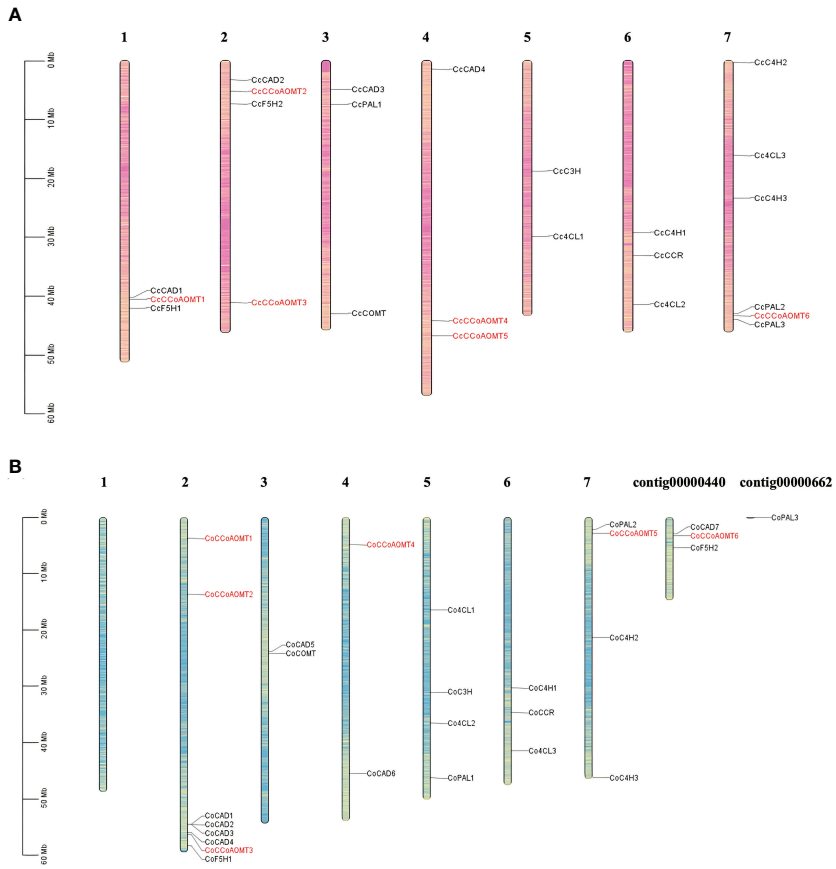


FIGURE 3
Distribution of the gene members involved in lignin biosynthesis of both *C capsularis* and *C olitorius* species. The *CCoAOMT* genes are indicated in red. Rest of the gene family members are marked in black. Each gene member pointed to the exact position of a specific chromosome, which could be estimated using the scale on the left. The scale in Megabase is represented by the left bar (Mb). Chromosome numbers are provided at the top of each chromosome bar. **(A)** Chromosomal localization of 6 *CcCCoAOMT* genes on 4 *C capsularis* chromosomes. **(B)** Chromosomal localization of 6 *CoCCoAOMT* genes on 3 *C olitorius* chromosomes and 1 contig.

number of various cis-elements under these four categories ranges from 4 (*CcCCoAOMT1*) to 20 (*CcCCoAOMT2*) in case of abiotic and biotic stress responsive elements, from 6 (*CcCCoAOMT6*) to 15 (*CcCCoAOMT1*) in case of hormone related elements; from 46 (*CcCCoAOMT1*) to 164 (*CcCCoAOMT4*) and from 6 (*CcCCoAOMT3*) to 16 (*CcCCoAOMT4*) in case of growth and light-responsive cis-elements, respectively in *C. capsularis* promoter region. In case of *C. olitorius* promoters, sum of different cis- elements ranges from 2 (*CoCCoAOMT6*) to 15 (*CoCCoAOMT1*) in case of stress-related elements, from 3 (*CoCCoAOMT6*) to 23 (*CoCCoAOMT1*); from 46 (*CoCCoAOMT6*) to 116 (*CoCCoAOMT3*) and from 4 (*CoCCoAOMT6*) to 20 (*CoCCoAOMT1*) in case of hormone, growth and light-responsive cis-elements, respectively. The lignin-associated element name Myb element (Patzlaff et al., 2003) was found in all the members except for *CcCCoAOMT1* and *CoCCoAOMT6*.

Phylogenetic analysis along with conserved motif

To analyze the evolutionary pattern, a phylogenetic tree was constructed based on a total of 69 CCoAOMT proteins, including 6 from *C. capsularis*, 6 from *C. olitorius*, 10 from *C. sinensis*, 7 from *A. thaliana*, 6 from *O. sativa*, 7 from *S. bicolor*, 10 from *M. truncatula*, 6 from *P. trichocarpa*, and 11 from *V. vinifera*. The CCoAOMTs are clustered into four groups in the phylogenetic tree (Group A, B, C and D). Group A contains two CCoAOMTs from *C. capsularis* and two from *C. olitorius* where most of the CCoAOMTs are associated with lignin biosynthesis. Except for Group D, the other three groups contain CCoAOMT proteins from two jute species. In concordance with Figure 5, the same members from *C. capsularis* and *C. olitorius* clustered into the same branches in this phylogenetic tree. To explore the difference among the protein structures, the MEME website was

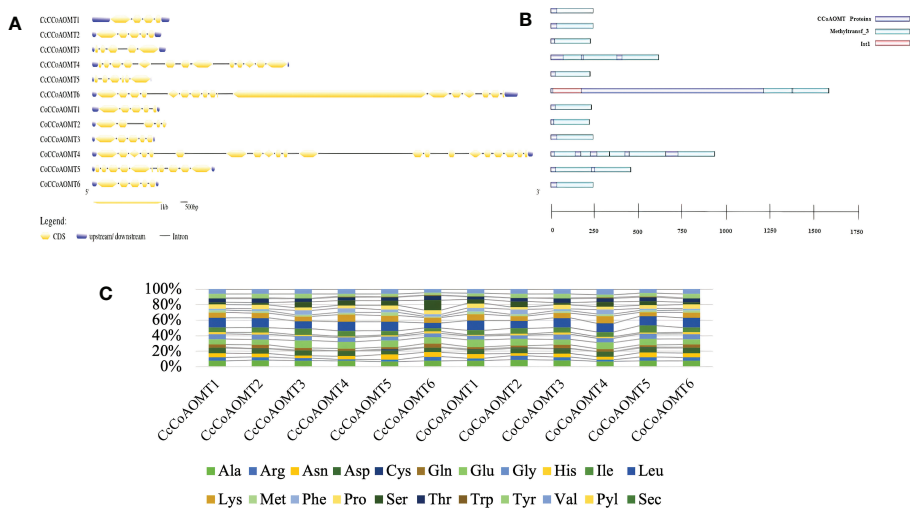


FIGURE 4
Overview of gene structure, domain architecture and amino acid composition of CCoAOMT members in jute. **(A)** Exon-intron distribution of all *CCoAOMT* genes discovered in *C capsularis* and *C olitorius*. Exon and UTR regions are represented by yellow and green boxes, respectively, while introns are represented by black lines. **(B)** Domain analysis of CCoAOMT proteins from the two species. Green boxes represent the identifying domain Methyltransf_3 and pink box Ist1 domain. **(C)** A stacked bar plot depicting the amino acid (aa) composition of CCoAOMT proteins in both species. Each amino acid's percentage content is displayed in a different color.

used to examine the variations in protein structures. Typically, most of the closely related individuals within a group share similar motif compositions (Figure 6). Motifs 1, 2, 3, 4 and 6 were found to be the most prevalent in all CCoAOMT proteins except for a few. Some motifs were exclusively found in a certain group in the phylogenetic tree (Motif 10, 11, 13 and 14). These unique motifs may contribute to the distinct roles of CCoAOMT proteins. In the case of the CCoAOMT proteins of *C. olitorius* and *C. capsularis*, motifs 1, 2 and 4 were found in all the members.

In-silico expression analysis of *CCoAOMT* genes in different tissues and analysis of their stress responsive cis-regulatory elements

To decipher the role of *CCoAOMT* genes their expression profile in two plant parts- stem bark and leaf, were analyzed along with the presence of different cis-elements in two jute species. The phylogenetic analysis of *CCoAOMT* genes across multiple plant species showed that genes with similar functions from different species cluster together in the same clade. Here, gene members with high expression values are clustered together in the phylogenetic tree. One of the stress-responsive cis-elements, Myb showed the most frequent abundance among the gene family members with higher expression that might function in the biosynthesis of lignin in plants. The number of Myb regulatory elements was found more prevalent in highly expressive genes, and genes with low expression showed minimum presence of Myb element. Presence of other elements could not be directly correlated with the expression data of 12 genes in stem bark and leaf (Figure 7).

Synteny between the two jute genomes

A synteny analysis was carried out to get an overview of the similarities between the two jute genomes. Expectedly, the results show that there is a high degree of syntenic conservation between the two genomes. In almost all cases, each chromosome in one species is seen to be in high synteny with a similar chromosome in the other species. However, all chromosomes except Chr06, show evidence of chromosomal inversion (Figure 8A and Supplementary Figure S2). Moreover, in case of Chr03, there appear to be two separate incidences of inversion. In contrast to this, when compared to the *A. thaliana* genome, both jute species exhibit less synteny; with the syntenic regions being much smaller than the regions between the two jute genomes (Figure 8A). Furthermore, for each jute chromosome, the synteny blocks are dispersed among all five *A. thaliana* chromosomes. This points to a higher number of genomic rearrangements throughout the evolutionary history between the two jute species and *A. thaliana*.

As previously mentioned, all the methyltransferase genes in this study were distributed over four chromosomes in both jute species (Figure 3). Out of the six genes, five were found to be present in synteny blocks, with one-to-one conservation between the two jute genomes (Figure 8B and Supplementary

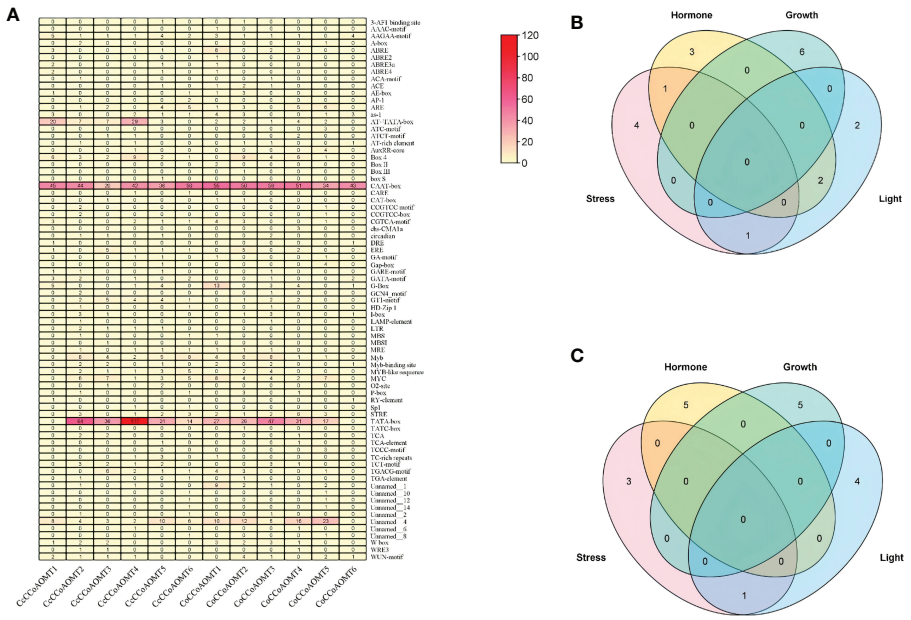


FIGURE 5
Study of the cis-regulatory elements present in the 2kb upstream putative promoter region of *CCoAOMT* genes in jute. **(A)** Total number of different cis-elements present in two species of Jute. The total number of cis-elements in each *CCoAOMT* gene of *C capsularis* **(B)** and *C olitorius* *CCoAOMT* in four categories is depicted by Venn diagram **(C)**.

Table 2). Interestingly, one of the *C. capsularis* genes (*CcCCoAOMT5*) appears to be a part of a syntenic block that extends from Chr04 to Chr02 of *C. olitorius*. This likely indicates a translocation event at this genomic region, leading to both homologous genes (*CcCCoAOMT5* and *CoCCoAOMT1*) being located on different chromosomes.

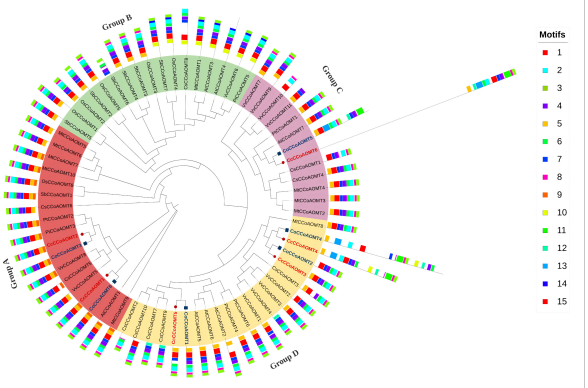


FIGURE 6
Phylogenetic relationship of CCoAOMTs of different plant species along with conserved motifs. A maximum likelihood method using 1000 ultrafast bootstrap replicates was used to construct the tree. The CCoAOMTs of *C. capsularis* and *C. olitorius* are labeled in red. The phylogenetic tree clustered into four major groups depicted in different colors. The 15 identified conserved motif is also added as an additional dataset represented in different colored boxes.

Expression profiling of *CoCCoAOMT* gene family under hormone, drought and salinity stresses

As the level of lignin is altered in response to different abiotic stresses (Figure 1), it is crucial to assess the expression of all 6 *CoCCoAOMT* genes experimentally for correlation purpose. Thus, we carried out quantitative RT-PCR analysis of 6 *CoCCoAOMT* genes in leaf and stem tissues after hormonal treatment of abscisic acid (ABA), polyethylene glycol (PEG) and 200mM NaCl for 24h, 48h and 72h. Most of the genes are found to be upregulated in response to all three conditions compared with respective control samples with few exceptions (Figure 9). Most of the analyzed genes showed significant upregulation in response to ABA treatment at both 24h and 72h in leaf samples with a fluctuating pattern in 48h. An opposite pattern was observed in the stem samples where all 6 genes showed clear upregulation in 48h of ABA treatment. Consistently, most of the genes were upregulated under PEG and NaCl treatment in all three-time points. Interestingly, the transcript abundance of all 6 genes was found to be decreased gradually in response to salinity in the stem samples.

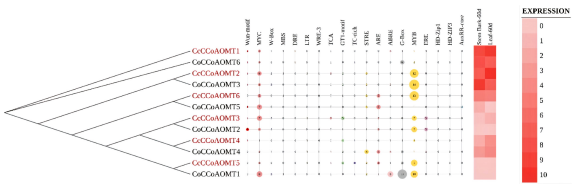


FIGURE 7
Phylogenetic relationship among two jute species along with their different stress-responsive cis-elements to demonstrate their regulatory response upon stress and *in-silico* expression profiles of different tissues (60 days old stem bark and leaf).

Discussion

Lignification is one of the key processes in higher plants that results in the deposition of lignin in the plant cell wall, which protects the plant against numerous biotic and abiotic agents (Cesarino, 2019) Several studies have found that after 12 days of water retention, decreased anionic peroxidase activity and increased cationic peroxidase activity, as well as an increase in free lignin precursors in the xylem sap, could be an indication that water stress reduces lignin biosynthesis in maize (Alvarez et al., 2008). Although various tissues in different places may behave differently under stress; for instance, specific regions of maize roots may react distinctly to water deficit (Fan et al., 2006; Yang et al., 2006; Yoshimura et al., 2008). To understand the lignin deposition status in jute plant, a histochemical assay of lignin was performed in this study for both control and stressed plants (post 48 and 72 hours) (Figure 1). The stain intensity suggests that the deposition of lignin increased after 48 hours of stress exposure but decreased after the stress has been induced for a longer time i.e., 72 hours. As multiple gene families are involved in the lignin synthesis of plants, the relative expression of specific genes involved in the pathway might influence the ultimate lignification of jute under stress conditions. Along with the secondary cell wall component lignin, quantity of certain amino acids and secondary metabolites also found to be increased under stress conditions. The accumulation of proline and phenolics in stressed jute plants were measured and found to be accumulated significantly in response to stress conditions compared with untreated control sample (Supplementary Figure 1). Proline causes plants under stress to develop more quickly and exhibit other positive physiological traits when administered exogenously to the plants. In addition to that, plants play a crucial function in the management of diverse environmental stresses and accumulate phenolic compounds in their tissues as an adaptive response to unfavorable environmental conditions (Naikoo et al., 2019). Here, higher deposition of lignin and high accumulation of proline and phenolic compounds was found in stressed plants which confirms the plants response mechanism under abiotic stresses.

In this study, we focused on the gene members involved in the formation of lignin in jute plants. The *CCoAOMT* gene family is one of the least investigated of the many gene families involved in the production of lignin in plants. Discovering the number and role of genes involved in the manufacture of lignin, which impacts the quality of fibers, is critical for fiber bearing crops such as jute. Here,

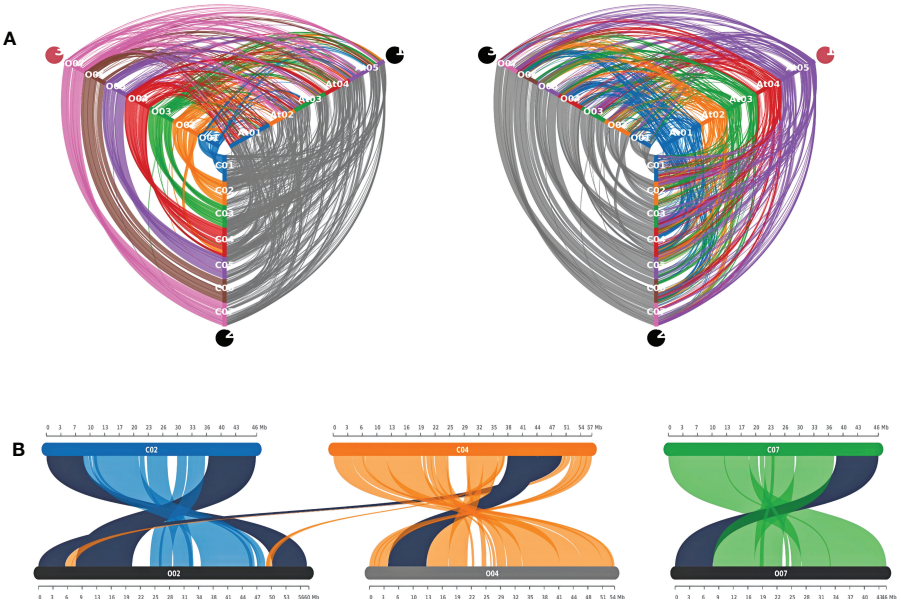


FIGURE 8
Synteny analysis between *C capsularis*, *C olitorius*, and *A thaliana* genomes. **(A)** Hive plots showing the syntenic conservation among the three genomes. **(B)** Syntenic *CCoAOMT* genes between the two jute genomes. Synteny pairs are highlighted in black.

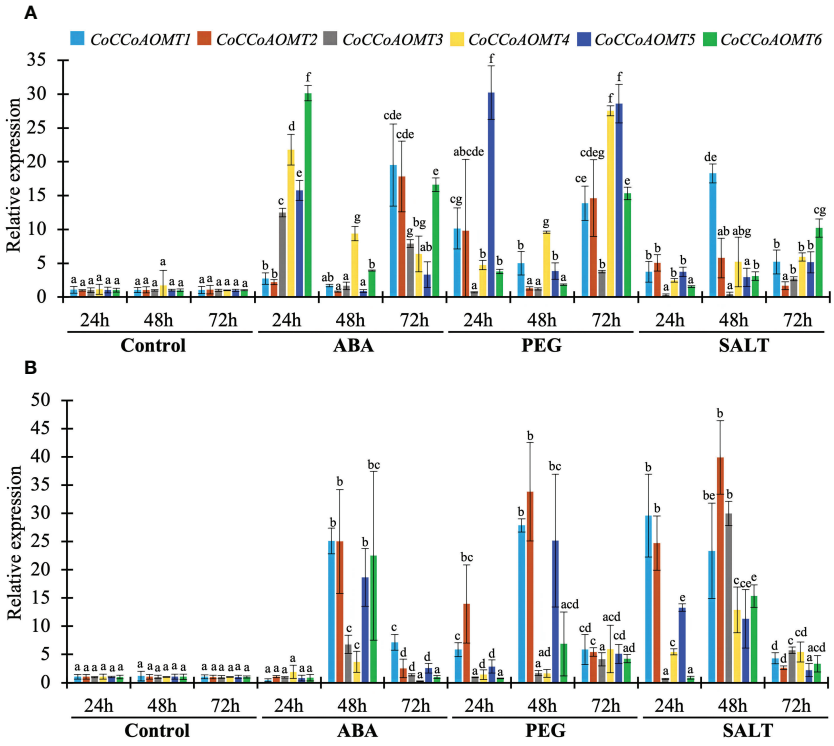


FIGURE 9
Expression profiles of *CoCCoAOMT* genes against control, hormonal, drought, and saline conditions. The left margin scale depicts the relative expression of *CoCCoAOMT* gene under control and various stress conditions. The results were reported as the mean value of the relative expression with standard deviation, using the jute Actin gene as a reference gene. Relative expression of *CoCCoAOMT* gene in *C olitorius* under control, ABA, PEG, and NaCl treatments in leaf **(A)** and stem **(B)** tissue. The expression level of each *CoCCoAOMT* gene at all three times under control condition was set to 1. Columns at each observation and time point followed by the different letters (such as a, b, c, d, e, f, or g) are statistically significantly different using the Analysis of variance test at P value ≤ 0.05, while columns with the same letter are statistically not significant (P value > 0.05).

two species of jute (*C. capsularis* and *C. olitorius*) were investigated to uncover the abundance, structure, relative expression and role of the *CCoAOMT* gene family that is involved in lignin biosynthesis in plants. The *CCoAOMT* gene family has been previously studied in several plants, including *A. thaliana, S. bicolor, P. trichocarpa, V. vinifera*, and *C. sinensis*, with members of the gene family reported to be 7, 7, 6, 11, and 10; respectively (Raes et al., 2003; Hamberger et al., 2007; Giordano et al., 2016; Rakoczy et al., 2018; Lin et al., 2020). According to previous reports, the *CCoAOMT* gene family consists primarily offewer family members in various plant species. We found both the species of jute were found to contain 6 *CCoAOMT* genes each which is in harmony with previous report.

In terms of the physical features and molecular attributes of the *CCoAOMT* genes these two jute species showed very little variation which indicates the functional conservation of this gene family in both jute species. Synteny analysis also revealed great conservation between the two jute genomes with an incidence of chromosomal inversion. Most of the *CCoAOMT* genes were found to have a one-to-one ortholog in the two species. Although the products of lignin biosynthesis and other phenylpropanoid routes are generally found in vacuoles and the cell wall, the pathways themselves are found in the cytoplasm, suggesting that the enzymes involved in the processes may be found there as well (Hrazdina and Jensen, 1992; Schmitt and An, 2017). Previous research has demonstrated that CCoAOMTs are cytoplasmic localization proteins (Chen et al., 2000; Zhang and Erickson, 2012) which aligns with our present result. Cis-Regulatory elements in promoter regions regulate the function of specific genes by binding to the right transcription factors (Hernandez-Garcia and Finer, 2014). For this reason, cis-elements in promoter regions are crucial for controlling gene expression. As a result, the role of *CCoAOMT* genes in lignin biosynthesis was explored by investigating the cis-acting regulatory components related to lignin biosynthesis in the promoter regions of *CCoAOMT*s. Studies suggest that by interacting with the P-box element, the Myb Plant motif found in the promoters of genes involved in phenylpropanoid biosynthesis controls the production of lignin (Tamagnone et al., 1998). Except for *CcCCoAOMT1* and *CoCCoAOMT6*, all of the genes in our investigation had the lignin-associated element Myb (Patzlaff et al., 2003) in their promoter region, indicating that these members are involved in the production

of lignin. In addition, all cis-elements were divided into four categories: those that relate to growth, hormones, stress, and light. The role played by this gene family in each of these four sectors is evidenced by the prevalence of cis elements in all four of these categories. The presence of various stress-responsive elements, such as LTR (low temperature), MBS (drought), TC-rich repeats (stress and defense), WUN-motif (wound), etc., were also found, which is consistent with the earlier study on *CCoAOMTs* in tea (Lin et al., 2020). Moreover, some of the highly expressed members (*CcCCoAOMT2, CcCCoAOMT6* and *CoCCoAOMT3*) contain a higher number of Myb elements in their respective promoter region (Figure 7). It could be assumed that these members are actively involved in the deposition of lignin due to their higher expression in the lignin depositing tissues. The role of these genes in the production of secondary cell wall components and in protecting plants from various biotic and abiotic stressors need to be studied further.

Based on phylogenetic analysis of 69 CCoAOMT protein sequences from jute plants and other species, CCoAOMTs can be clustered into four groups. No CCoAOMT proteins of jute plants were found in group B. CcCCoAOMT1, CcCCoAOMT2, CoCCoAOMT3 and CoCCoAOMT6 proteins clustered into group A which contains the proteins that are involved in lignin biosynthesis pathway (Guo et al., 2001; Day et al., 2009; Lin et al., 2020). Group A also contains AtCCoAOMT1 protein, which is considered the true CCoAOMT gene, while the other classes were considered CCoAOMT-like genes (Rakoczy et al., 2018). Proteins clustered in the same clade generally exhibit similar functions. Thus, CcCCoAOMT1, CcCCoAOMT2, CoCCoAOMT3 and CoCCoAOMT6 may play major roles in lignin biosynthesis. Similar motif patterns were also found in the same groups of the phylogenetic tree (Figure 6) which further confirms their function in the biosynthesis of lignin. In concordance with Figure 6, the phylogenetic tree within the jute species also showed clustering of CcCCoAOMT1, CoCCoAOMT6 and CcCCoAOMT2, CoCCoAOMT3 together in the same branch. The *in-silico* expression value of these four members was also found high in stem bark and leaf tissues. These results further put light on the importance of these gene members in transcriptional modulation of the *CCoAOMT* gene family in jute. To validate the *in-silico* expression result and to elucidate the expression profile of the *CCoAOMT* gene family under different abiotic stresses (hormone, drought and salinity), two months old *C. olitorius* seedling was examined. Surprisingly, leaf and stem tissues showed a differential expression patterns. In leaf tissues, most of the gene members showed up-regulation in 24- and 72-hours time points, whereas in stem tissues most genes were up-regulated after 48 hours of stress exposure. The reason behind this, in fiber crops most of the lignin is deposited in the stem tissues and when the plants are exposed to sudden stress shock, most of the lignin and other secondary metabolites work to combat the stress. During this process, the expression level of *CoCCoAOMTs* in stem tissues increased gradually from 24 hours up to 48 hours of stress. After a certain period, the expression level again reduces, as there is a certain trade-off between the growth and defense mechanism in plants (Ha et al., 2021). Contrarily, under stress conditions, leaf tissues readily act to fight the stresses and gradually decreases their expression to focus on growth. But during long exposure to stress, the expression level was again found to be increased in leaf tissues. From the expression profile of *CoCCoAOMTs* under stress, some members were more upregulated than others. According to Rakoczy et al., 2018, clade 1 characteristics are demonstrated by the high expression of *SbCCoAOMT1* in leaves and stems with high lignification. In plants, the down-regulation of *CCoAOMT* results in a striking reduction in lignin content, indicating *CCoAOMT* is involved in the up-regulation of lignin production (Jianhua et al., 2001; Zhao et al., 2002). In the *in-silico* expression analysis, two members- *CoCCoAOMT3* and *CoCCoAOMT6* were found to be highly expressive in two important tissues- stem and leaf (Figure 7). Similarly, both *CoCCoAOMT3* and *CoCCoAOMT6* transcripts showed significant upregulation under all three stresses in both leaf and stem tissues (Figure 9). Apart from these two, *CoCCoAOMT4* and *CoCCoAOMT5* showed significant upregulation in leaf tissues; and *CoCCoAOMT1* and *CoCCoAOMT2* in stem samples in response to all three stresses. Taken together, this study confirms the function of the *CCoAOMT* gene family members in the lignin biosynthesis as well as stress modulation pathways of jute. Therefore, further efforts should be made to overexpress or silence jute *CCoAOMT* genes using a transgenic technique employing *Arabidopsis* or other model plants to understand their molecular functions.

Conclusion

Overall, this study provides a thorough evaluation of the *CCoAOMT* gene family in the fiber crop jute on a genome wide scale. We were able to locate 6 *CCoAOMT* genes from two jute species in total, which is in harmony with previous studies. This study sheds light on *CCoAOMT* gene regulation under various stresses in jute. The multi-stress responsiveness of the prospective jute *CCoAOMT* genes was also confirmed by qRT-PCR. However, further study is needed to determine the functional role of these genes by creating overexpressing or knockout lines of *CCoAOMT* genes individually or collectively. This study will facilitate further investigation of the *CCoAOMT* gene family in jute and other plants.

Data availability statement

The original contributions presented in the study are included in the article/supplementary materials. Further inquiries can be directed to the corresponding author.

Author contributions

TI conceived the idea and designed the experiments. SA, AS and TT performed all the experiments. TI, SA and AS analysed of data. SA and AS wrote the manuscript. TT and BJ helped in preparing manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by Centennial Research Grant, University of Dhaka, Grant no. Reg/Admin-3/70919.

Acknowledgments

TI acknowledges University of Dhaka for providing Centennial Research Grant and publication fee. Authors acknowledge Plant Breeding and Biotechnology Laboratory, Department of Botany, University of Dhaka for technical support.

Conflict of interest

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

The reviewer ZS declared a shared affiliation with the authors to the handling editor at the time of the review.

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.2022.1035383/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Proline and phenolic content in *C. olitorius* in response to abiotic stress were measured. The accumulation was assessed in two-month-old jute seedlings under various stress conditions, such as salt (200 mM NaCl), drought (20% PEG), and hormone (20 uM ABA), as well as related control samples. The mean value and standard deviation (n = 3) were used to represent the results. **(A)** proline content in leaf and stem tissue under stress. **(B)** phenolic content in leaf and stem tissue in response to stress. Columns followed by the different letters are statistically significantly different according to P value ≤ 0.05.

SUPPLEMENTARY FIGURE 2
Syntenic relationship between two jute genomes (*C. capsularis* and *C. olitorius).*

References

Alvarez, S., Marsh, E. L., Schroeder, S. G., and Schachtman, D. P. (2008). Metabolomic and proteomic changes in the xylem sap of maize under drought. *Plant Cell Environ.* 31, 325–340. doi: 10.1111/j.1365-3040.2007.01770.x

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, W202–W208. doi: 10.1093/nar/gkp335

Bandi, V., and Gutwin, C. (2020). *Interactive exploration of genomic conservation. in proceedings of the 46th graphics interface conference on proceedings of graphics interface 2020 (GI'20)* (Waterloo, CAN: Canadian Human-Computer Communications Society).

Baucher, M., Monties, B., Van Montagu, M., and Boerjanm, W. (1998). Biosynthesis and genetic engineering of lignin. *Crit. Rev. Plant Sci.* 17, 125–197. doi: 10.1080/07352689891304203

Buchfink, B., Reuter, K., and Drost, H. G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* 18, 366–368. doi: 10.1038/s41592-021-01101-x

Cabane, M., Afif, D., and Hawkins, S. (2012). Lignins and abiotic stresses. advances in botanical research. In L Jouanin and C Lapierre editors. *Advances in Botanical Research* (Academic Press), 61, 219–262. doi: 10.1016/b978-0-12-416023-1.00007-0

Cesarino, I. (2019). Structural features and regulation of lignin deposited upon biotic and abiotic stresses. *Curr. Opin. Biotechnol.* 56, 209–214. doi: 10.1016/j.copbio.2018.12.012

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

Chen, C., Meyermans, H., Burggraeve, B., De Rycke, R., Inoue, K., De Vleesschauwer, V., et al. (2000). Cell-specific and conditional expression of caffeoyl-coenzyme a-3-O-methyltransferase in poplar. *Plant Physiol.* 123 (3), 853–867. doi: 10.1104/pp.123.3.853

Choudhary, S. B., Chowdhury, I., Singh, R. K., Pandey, S. P., Sharma, H. K., Kumar, A., et al. (2017). Morphological, histobiochemical and molecular characterisation of low lignin phloem fibre (llpf) mutant of dark jute (*Corchorus olitorius* l.). *Appl. Biochem. Biotechnol.* 183 (3), 980–992. doi: 10.1007/s12010-017-2477-5

Day, A., Neutelings, G., Nolin, F., Grec, S., Habrant, A., Crônier, D., et al. (2009). Caffeoyl coenzyme A O-methyltransferase down-regulation is associated with modifications in lignin and cell-wall architecture in flax secondary xylem. *Plant Physiol. Biochem.* 47 (1), 9–19. doi: 10.1016/j.plaphy.2008.09.011

Del Río, J. C., Rencoret, J., Marques, G., Li, J., Gellerstedt, G., Jiménez-Barbero, J., et al. (2009). Structural Characterization of the Lignin from Jute (Corchorus capsularis) Fibers. *J. Agric. Food Chem.* 57 (21), 10271–10281. doi: 10.1021/jf900815x

Do, C. T., Pollet, B., Thévenin, J., Sibout, R., Denoue, D., Barrière, Y., et al. (2007). Both caffeoyl coenzyme a 3-o-methyltransferase 1 and caffeic acid O-methyltransferase 1 are involved in redundant functions for lignin, flavonoids and sinapoyl malate biosynthesis in arabidopsis. *Planta* 226, 1117–1129. doi: 10.1007/s00425-007-0558-3

Fan, L., Linker, R., Gepstein, S., Tanimoto, E., Yamamoto, R., and Neumann, P. M. (2006). Progressive inhibition by water deficit of cell wall extensibility and growth along the elongation zone of maize roots is related to increased lignin metabolism and progressive stelar accumulation of wall phenolics. *Plant Physiol.* 140, 603–612. doi: 10.1104/pp.105.073130

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., et al. (2005). Protein identification and analysis tools on the ExPASy server. In J. M. Walker Editor. *The Proteomics Protocols Handbook. Springer Protocols Handbooks* (Humana Press) 52, 571–607. doi: 10.1385/1-59259-890-0:571

Giordano, D., Provenzano, S., Ferrandino, A., Vitali, M., Pagliarani, C., Roman, F., et al. (2016). Characterization of a multifunctional caffeoyl-CoA O-methyltransferase activated in grape berries upon drought stress. *Plant Physiol. Biochem.* 101, 23–32. doi: 10.1016/j.plaphy.2016.01.015

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2011). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40 (Database issue), D1178–D1186. doi: 10.1093/nar/gkr944

Guo, D., Chen, F., Inoue, K., Blount, J. W., and Dixon, R. A. (2001). Downregulation of caffeic acid 3-O-methyltransferase and caffeoyl CoA 3-O-methyltransferase in transgenic alfalfa. impacts on lignin structure and implications for the biosynthesis of G and S lignin. *Plant Cell.* 13 (1), 73–88. doi: 10.1105/tpc.13.1.73

Hamberger, B., Ellis, M., Friedmann, M., Souza, C. A., Barbazuk, B., and Douglas, C. J. (2007). Genome-wide analyses of phenylpropanoid-related genes in populus trichocarpa, *Arabidopsis thaliana*, and *Oryza sativa*: the *Populus* lignin toolbox and conservation and diversification of angiosperm gene families. *Can. J. Bot.* 85 (12), 1182–1201. doi: 10.1139/B07-098

Ha, C. M., Rao, X., Saxena, G., and Dixon, R. A. (2021). Growth–defense trade-offs and yield loss in plants with engineered cell walls. *New Phytol.* 231, 60–74. doi: 10.1111/nph.17383

Hasan, M. S., Singh, V., Islam, S., Islam, M. S., Ahsan, R., Kaundal, A., et al. (2021). Genome-wide identification and expression profiling of glutathione s-transferase family under multiple abiotic and biotic stresses in *Medicago truncatula* l. *PloS One* 16, e0247170. doi: 10.1371/journal.pone.0247170

Hernandez-Garcia, C. M., and Finer, J. J. (2014). Identification and validation of promoters and cis-acting regulatory elements. *Plant Sci.* 217, 109–119. doi: 10.1016/j.plantsci.2013.12.007

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucl. Acids Res.* 35, W585–W587. doi: 10.1093/nar/gkm259

Hrazdina, A., and Jensen, R. A. (1992). Spatial organization of enzymes in plant metabolic pathways. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 43 (1), 241–267. doi: 10.1146/annurev.pp.43.060192.001325

Jianhua, ,. W., Huayan, Z., Jingyu, Z., Huirong, L., and Yanru, S. (2001). Cloning of cDNA encoding CCoAOMT from Populus tomentosa and down-regulation of lignin content in transgenic plant expressing antisense gene. *Acta Botanica Sinica.* 43 (11), 1179–1183. http://europepmc.org/abstract/CBA/354763.

Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20, 1160–1166. doi: 10.1093/bib/bbx108

Kawaoka, A., Nanto, K., Ishii, K., and Ebinuma, H. (2006). Reduction of lignin content by suppression of expression of the LIM domain transcription factor in eucalyptus camaldulensis. *Silvae Genetica* 55, 269–277. doi: 10.1515/sg-2006-0035

Lescot, M. (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. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Li, Z., Lei, L., Kehai, Z., Yihao, Z., Yanpeng, D., Xiaoyang, G., et al. (2019). *GhWRKY6* acts as a negative regulator in both transgenic arabidopsis and cotton during drought and salt stress. *Front. Genet.* 10. doi: 10.3389/fgene.2019.00392

Lin, S.-J., Yang, Y.-Z., Teng, R.-M., Liu, H., Li, H., and Zhuang, J. (2020). Identification and expression analysis of caffeoyl-coenzyme A O-methyltransferase family genes related to lignin biosynthesis in tea plant (*Camellia sinensis*). *Protoplasma* 258, 115–127. doi: 10.1007/s00709-020-01555-4

Liu, Q., Luo, L., and Zheng, L. (2018). Lignins: Biosynthesis and biological functions in plants. *Int. J. Mol. Sci.* 19 (2), 335. doi: 10.3390/ijms19020335

Liu, X., Luo, Y., Wu, H., Xi, W., Yu, J., Zhang, Q., et al. (2016). Systematic analysis of O -methyltransferase gene family and identification of potential members involved in the formation of O -methylated flavonoids in Citrus. *Gene* 575 (2), 458–472. doi: 10.1016/j.gene.2015.09.048

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

Loumerem, M., and Alercia, A. (2016). Descriptors for jute (Corchorus olitorius l.). *Genet. Resour Crop Evol.* 63, 1103–1111. doi: 10.1007/s10722-016-0415-y

Meyermans, H., Morreel, K., Lapierre, C., Pollet, B., De Bruyn, A., Busson, R., et al. (2000). Modifications in lignin and accumulation of phenolic glucosides in poplar xylem upon down-regulation of caffeoylcoenzyme a O-methyltransferase, an enzyme involved in lignin biosynthesis. *J. Biol. Chem.* 275, 36899–36909. doi: 10.1074/jbc.M006915200

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). Corrigendum to: IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic Era. *Mol. Bio. Evol.* 37, 2461–2461. doi: 10.1093/molbev/msaa131

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2020). Pfam: the protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913

Mutuku, J. M., Cui, S., Hori, C., Takeda, Y., Tobimatsu, Y., Nakabayashi, R., et al. (2019). The structural integrity of lignin is crucial for resistance against striga hermonthica parasitism in rice. *Plant Physiol.* 179 (4), 1796–1809. doi: 10.1104/pp.18.01133

Naikoo, M. I., Dar, I. M., Raghib, F., Jaleel, H., Ahmed, B., Raina, A., et al. (2019). Plant signaling molecules || role and regulation of plants phenolics in abiotic stress tolerance: An overview. In MIR. Khan, PS. Reddy, A Ferrante and NA Khan Editor. *Plant Signaling Molecules* (Woodhead Publishing), 157–168. doi: 10.1016/B978-0-12-816451-8.00009-5

Patzlaff, A., Mcinnis, S., Courtenay, A., Surman, C., Newman, L. J., Smith, C., et al. (2003). Characterisation of a pine MYB that regulates lignification. *Plant J.* 36 (6), 743–754. doi: 10.1046/j.1365-313x.2003.01916.x

Raes, J., Rohde, A., Christensen, J. H., De Peer, Y. V., and Boerjan, W. (2003). Genome-wide characterization of the lignification toolbox in arabidopsis. *Plant Physiol.* 133 (3), 1051–1071. doi: 10.1104/pp.103.026484

Rakoczy, M., Femiak, I., Alejska, M., Figlerowicz, M., and Podkowinski, J. (2018). Sorghum CCoAOMT and CCoAOMT-like gene evolution, structure, expression and the role of conserved amino acids in protein activity. *Mol. Gen. Genomics* 293 (5), 1077–1089. doi: 10.1007/s00438-018-1441-6

Ren, J., Wen, L., Gao, X., Jin, C., Xue, Y., and Yao, X. (2009). DOG 1.0: illustrator of protein domain structures. *Cell Res.* 19 (2), 271–273. doi: 10.1038/cr.2009.6

Schmitt, D. L., and An, S. (2017). Spatial organization of metabolic enzyme complexes in cells. *Biochemistry* 56 (25), 3184–3196. doi: 10.1021/acs.biochem.7b00249

Selver, E., Ucar, N., and Gulmez, T. (2018). Effect of stacking sequence on tensile, flexural and thermomechanical properties of hybrid flax/glass and jute/glass thermoset composites. *J. Ind. Textil* 48, 494–520. doi: 10.1177/1528083717736102

Shafrin, F., Ferdous, A. S., Sarkar, S. K., Ahmed, R., Amin, A.-, Hossain, K., et al. (2017). Modification of monolignol biosynthetic pathway in jute: Different gene, different consequence. *Sci. Rep.* 7 (1), 39984. doi: 10.1038/srep39984

Sood, M., and Dwivedi, G. (2018). Effect of fiber treatment on flexural properties of natural fiber reinforced composites: A review. *Egyptian J. Petroleum* 27 (4), 775–783. doi: 10.1016/j.ejpe.2017.11.005

Tamagnone, L., Merida, A., Parr, A. J., Mackay, S., Culianezmacia, F. A., Roberts, K., et al. (1998). The AmMYB308 and AmMYB330 transcription factors from antirrhinum regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco. *Plant Cell* 10 (2), 135–154. doi: 10.1105/tpc.10.2.135

Tanmoy, A., Alum, M., Islam, M., Farzana, T., and Khan, H. (2015). Jute (*corchorus olitorius* var. O-72) stem lignin: variation in content with age. *Bangladesh J. Bot.* 43 (3), 309–314. doi: 10.3329/bjb.v43i3.21603

Tronchet, M., Balagué, C., Kroj, T., Jouanin, L., and Roby, D. (2010). Cinnamyl alcohol dehydrogenases c and d, key enzymes in lignin biosynthesis, play an essential role in disease resistance in arabidopsis. *Mol. Plant Pathol.* 11, 83–92. doi: 10.1111/j.1364-3703.2009.00578.x

Vanholme, R., Demedts, B., Morreel, K., Ralph, J., and Boerjan, W. (2010). Lignin biosynthesis and structure. *Plant Physiol.* 153, 895–905. doi: 10.1104/pp.110.155119

Vanholme, R., Morreel, K., Ralph, J., and Boerjan, W. (2008). Lignin engineering. *Curr. Opin. Plant Biol.* 11, 278–285. doi: 10.1016/j.pbi.2008.03.005

Wagner, A., Tobimatsu, Y., Phillips, L., Flint, H., Torr, K., Donaldson, L., et al. (2011). *CCoAOMT* suppression modifies lignin composition in pinus radiata. *Plant J.* 67 (1), 119–129. doi: 10.1111/j.1365-313x.2011.04580.x

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

Xia, E., Li, F., Tong, W., Li, P., Wu, Q., Zhao, H., et al. (2019). Tea Plant Information Archive (TPIA): A comprehensive genomics and bioinformatics platform for tea plant. *Plant Biotechnol. J.* 17 (10), 1938–1953. doi: 10.1111/pbi.13111

Xu, Z., Zhang, D., Hu, J., Zhou, X., Ye, X., Reichel, K. L., et al. (2009). Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. *BMC Bioinform.* 10 (Suppl 11), S3. doi: 10.1186/1471-2105-10-s11-s3

Yang, L., Wang, C. C., Guo, W. D., Li, X. B., Lu, M., and Yu, C. L. (2006). Differential expression of cell wall related genes in the elongation zone of rice roots

under water deficit. *Russian J. Plant Physiol.* 53, 390–395. doi: 10.1134/s1021443706030150

Yoon, J., Choi, H., and An, G. (2015). Roles of lignin biosynthesis and regulatory genes in plant development. *J. Integr. Plant Biol.* 57 (11), 902–912. doi: 10.1111/jipb.12422

Yoshimura, K., Masuda, A., Kuwano, M., Yokota, A., and Akashi, K. (2008). Programmed proteome response for drought avoidance/tolerance in the root of a c-3 xerophyte (wild watermelon) under water deficits. *Plant Cell Physiol.* 49, 226–241. doi: 10.1093/pcp/pcm180

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

Zhang, J., and Erickson, L. (2012). Harvest-inducibility of the promoter of alfalfa s-adenosyl-l-methionine: trans-caffeoyl-CoA3-O methyltransferase gene. *Mol. Biol. Rep.* 39 (3), 2489–2495. doi: 10.1007/s11033-011-1000-2

Zhang, G., Zhang, Y., Xu, J., Niu, X., Qi, J., Tao, A., et al. (2014). The CCoAOMT1 gene from jute (Corchorus capsularis l.) is involved in lignin biosynthesis in arabidopsis thaliana. *Gene* 546 (2), 398–402. doi: 10.1016/j.gene.2014.05.011

Zhang, L., Ma, X., Zhang, X., Xu, Y., Ibrahim, A. K., Yao, J., et al. (2021). Reference genomes of the two cultivated jute species. *Plant Biotechnol. J.* 19 (11), 2235–2248. doi: 10.1111/pbi.13652

Zhao, H. Y., Wei, J. H., Zhang, J. Y., Liu, H. R., Wang, T., and Song, Y. R. (2002). Lignin biosynthesis by suppression of two O-methyltransferases. *Chin. Sci. Bull.* 47, 1092–1095. doi: 10.1360/02tb9245

Zhong, R., III, W. H. M., Negrel, J., and Ye, Z. H. (1998). Dual methylation pathway in lignin biosynthesis. *Plant Cell* 10, 2033–2045. doi: 10.1105/tpc.10.12.2033

Zhong, R., Morrison, W. H.III, Himmelsbach, D. S., Poole, F. L.II, and Ye, Z. H. (2000). Essential role of caffeoyl coenzyme a O-methyltransferase in lignin biosynthesis in woody poplar plants. *Plant Physiol.* 124, 563–578. doi: 10.1104/pp.124.2.563

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 December 2022
DOI 10.3389/fpls.2022.1043261

OPEN ACCESS

EDITED BY
Yusuf Khan,
Oslo University Hospital, Norway

REVIEWED BY
Kelvin Kamfwa,
University of Zambia, Zambia
Tahmina Islam,
University of Dhaka, Bangladesh

*CORRESPONDENCE
Qinghua Liu
liuqinghua642@163.com
Zhichun Zhou
zczhou_risf@163.com

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 13 September 2022
ACCEPTED 17 November 2022
PUBLISHED 14 December 2022

CITATION
Nie Z, Li W, Deng L, Gao K, Liu Q and Zhou Z (2022) Comprehensive analysis of LRR-RLKs and key gene identification in *Pinus massoniana* resistant to pine wood nematode. *Front. Plant Sci.* 13:1043261. doi: 10.3389/fpls.2022.1043261

Comprehensive analysis of LRR-RLKs and key gene identification in *Pinus massoniana* resistant to pine wood nematode

Ziyan Nie[1,2], Wenhua Li[1,2], Lili Deng[1,2], Kai Gao[1], Qinghua Liu[1]* and Zhichun Zhou[1]*

[1]Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Engineering Research Center of Masson Pine of State Forestry and Grassland Administration, Key Laboratory of Tree Breeding of Zhejiang Province, Hangzhou, Zhejiang, China, [2]Nanjing Forestry University, Nanjing, China

Pinus massoniana is a pioneer tree widely planted for afforestation on barren hills in southern China where the total planted area is 8.04 million ha. The invasive pine wood nematode (*Bursaphelenchus xylophilus*) poses a serious threat to the survival of *P. massoniana*. Plant resistance genes encoded by leucine-rich repeat-containing transmembrane-receptor proteins play important roles in plant defense. Leucine-rich repeat receptor-like kinases (LRR-RLKs), the largest subfamily of the RLK protein family, play an important role in sensing stress signals in plants. However, the LRR-RLKs of *P. massoniana* have not been characterized previously, and their role in resistance to *B. xylophilus* is unknown. In this study, 185 members of the LRR-RLK subfamily were identified in *P. massoniana* and were categorized into 14 subgroups. Transcriptomic and quantitative real-time RT-PCR analyses showed that *PmRLKs32* was highly expressed in the stem tissue after inoculation with *B. xylophilus*. The gene exhibited high homology with *AtFLS2* of *Arabidopsis thaliana*. *PmRLKs32* was localized to the plasma membrane and was significantly upregulated in nematode-resistant and nematode-susceptible individuals. The transient expression of *PmRLKs32* resulted in a burst of reactive oxygen species production in *P. massoniana* and *Nicotiana benthamiana* seedlings. These results lay a foundation for further exploration of the regulatory mechanism of LRR-RLKs in response to biotic stress in *P. massoniana*.

KEYWORDS

***pinus massoniana*, receptor-like kinase, LRR-RLKs, expression pattern, PTI**

Introduction

Plants are exposed to diverse biotic stresses in nature. A plant must respond to adverse biotic factors for survival and to maintain growth. Pests are a biotic stress that seriously threaten the growth and development of trees. Plants rely on their defense system, which predominantly includes pattern-triggered immunity (PTI) and effector-triggered immunity (ETI), to resist pathogen attack (Naveed et al., 2020). Recent studies have shown that PTI and ETI synergistically enhance disease resistance in plants, and the disease response pathways of each system show a certain degree of overlap. The ETI response activates the MAPK signaling pathway and the production of reactive oxygen species (ROS) (Yuan et al., 2021b). The activation of ETI requires the participation of pattern-recognition receptors (PRRs) and co-receptors. The ETI signal can mobilize the PTI pathway and, in turn, amplify the PTI response (Ngou et al., 2021; Pruitt et al., 2021; Tian et al., 2021; Yuan et al., 2021a). In addition, the concept of the "resistosome" has been proposed (Bi et al., 2021). These new views and concepts indicate that research in the field of plant immunity is entering a new phase and lay a crucial theoretical foundation for targeted improvement of the plant immune system.

The PTI pathway involves recognition of conserved pathogen-associated molecular patterns (PAMPs), microbe-associated molecular patterns (MAMPs), and damage-associated molecular patterns (DAMPs) by PRRs (Abdul Malik et al., 2020). The PRRs, which mainly comprise receptor-like kinases (RLKs) and receptor-like proteins (RLPs), play an important role in the first-tier basal immunity of plants that is triggered by the recognition of external pathogens (Lin et al., 2020). The PRRs elicit downstream cellular responses, including defense gene expression, ROS production, and callose deposition, by recognizing highly conserved molecular structures or characteristics of pathogens and pests (Albert et al., 2019). However, the recognition mechanism and the cell surface and intracellular synergistic mechanisms remain unclear.

The leucine-rich repeat receptor-like kinase (LRR-RLK) subfamily, the largest group within the plant RLK family, is mediated by many cellular signal transduction pathways. The proteins of this family contain three functional domains. The extracellular domain (ECD) perceives signals that contain varying numbers of LRR repeats. The transmembrane domain connects the internal and external cellular compartments. The intracellular kinase domain (KD) can transmit a signal through phosphorylation (Chakraborty et al., 2019)—for example, *Arabidopsis thaliana FLAGELLIN-SENSITIVE 2* (*AtFLS2*) senses the flagellin elicitor flg22 to activate a relatively low-level defense response, and *EF-TU RECEPTOR* (*AtEFR*) mediates plant resistance to a bacterial pathogen. *RECEPTOR-LIKE KINASE 902* (*AtRLK902*) may directly bind to and phosphorylate the downstream gene *BRASSINOSTEROID-SIGNALING KINASE 1* (*BSK1*), thereby transmitting an immune signal downward (Zhao et al., 2019). *OsLRR-RLK1* is involved in the regulation of pest defense responses in rice (Ye et al., 2019). Thus, the LRR-RLK family plays important roles in responses to abiotic and biotic stresses.

Pinus massoniana is the main timber species grown in southern China and is adaptable to diverse topography and environments. The pine wood nematode (*Bursaphelenchus xylophilus*) feeds on cells in the vascular bundles of the tree. The resulting pine wilt disease is devastating and threatens the survival of *P. massoniana*. In previous studies, we identified several *P. massoniana* accessions that showed high resistance or susceptibility to *B. xylophilus*, from which we selected representative resistant and susceptible individuals for full-length transcriptome sequencing. Through transcriptomic analysis of gene families of *P. massoniana*, in this study we identified candidate LRR-RLK genes and selected the most representative gene for preliminary functional verification. The results provide a foundation for the future studies of pine wood nematode resistance.

Materials and methods

Identification of LRR-RLK family members

Transcriptome data for *P. massoniana* were derived from the previously reported transcriptome of resistant and susceptible individuals post-inoculation with *B. xylophilus* (Liu et al., 2017). TBtools software was used to extract all LRR-RLK genes from the transcriptome that contained the LRR domain (PF13855). The default parameters were used for screening, and the E-value was set to E < 10−20. HMMER (https://www.ebi.ac.uk/Tools/hmmer/results/032E22B6-BB2E-11EC-94AB-5DE9DBC3747A/score) and CD-Search (https://www.ncbi.nlm.nih.gov/cdd/) were used to screen protein sequences of *P. massoniana* containing the complete LRR domain and beginning with a Met residue. A Perl script in the Linux system was used to delete the duplicated sequences among the selected *P. massoniana* resistance protein sequences. Based on previous reports and functional annotations of *A. thaliana* LRR-RLK proteins, the LRR-RLK proteins that included the LRR and protein kinase motifs were downloaded from Phytozome v13 (https://phytozome-next.jgi.doe.gov/) (Xi et al., 2019). The selected LRR-RLK proteins were further screened with CD-Search.

Sequence analysis and phylogenetic analysis

TBtools software was used to predict the molecular weight and isoelectric point of the LRR-RLK proteins in batches. CELLO (http://cello.life.nctu.edu.tw/) and PSORT (https://

psort.hgc.jp/) were used for subcellular localization prediction (Chen et al., 2020). An evolutionary tree was reconstructed using the maximum-likelihood method based on a multiple sequence alignment of the LRR-RLK protein sequences of *P. massoniana* and *A. thaliana* with MEGA software. Support for the topology of the phylogeny was assessed by performing a bootstrap analysis with 1,000 replicates. The Multiple Expectation Maximization for Motif Elicitation (MEME) tool (http://meme-suite.org/tools/meme) was used to analyze the conserved motifs of the LRR-RLK proteins with the following parameters: minimum and maximum motif widths of six and 50, respectively, and maximum pattern number 10. The fragments per kilobase of exon per million mapped fragment values of the *P. massoniana* transcripts were used in conjunction with the read map per thousand bases exon model per million values for the nematode to estimate the abundance of *P. massoniana* LRR-RLK gene transcripts. The relative expression level of the control was set to "1".

Plant material and treatments

Two-year-old *P. massoniana* seedlings were obtained from the Linhai forest farm (28°40′–29°04′ N, 120°49′–121°41′ E), Linhai Province, China. To study the expression level of LRR-RLK genes, we selected three seedlings of uniform growth for nematode treatment. Each plant was inoculated with 1,000 heads. Samples of stem tissue were collected at 0, 1, 15, and 30 days after inoculation.

RNA extraction and quantitative real-time reverse transcription PCR

Total RNA was extracted from samples using the EASYspin Plus Complex Plant RNA Kit (RN53, Aidlab, Beijing, China). The RNA concentration and purity were measured with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), and the RNA integrity was estimated by agarose gel electrophoresis. The first-strand cDNA was synthesized using the HiScript III All-in-one RT SuperMix Perfect for qPCR (R333, Vazyme, Hangzhou, China). Primers for quantitative real-time reverse transcription PCR (qRT-PCR) were designed using Primer 5.0 (Supplementary Table S1). The Taq Pro Universal SYBR qPCR Master Mix (Q712, Vazyme) was used to amplify the target sequence. Each PCR mixture (20 μl) contained 2 μl diluted cDNA (10× dilution), 10 μl SYBR Green Real-time PCR Master Mix, 0.4 μl of each primer (10 μM), and 7.2 μl ddH_2O. The RT-PCR program accorded with the kit manufacturer's instructions. The β-tubulin gene was used as an internal reference, and three biological replicates and three technical replicates were analyzed for each sample. Information on the primers used is presented in Supplementary Table S1. The relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method. The IBM SPSS Statistics for Windows, version 19.0 (IBM Corp, Armonk, NY, USA), was used to estimate the significance of differences between means with Student's *t*-test ($^*P < 0.05$, $^{**}P < 0.01$).

Subcellular localization assay

The construct 35S:GFP-*PmRLKs32* was generated by homologous recombination. The 35S:GFP-*PmRLKs32* construct containing the green fluorescent protein (GFP) was transiently transformed into the leaves of *Nicotiana benthamiana* following a previously described method (Liu et al., 2017). The fluorescence signals were observed with a laser scanning microscope (LSM880, ZEISS, Jena, Germany).

Transient transformation

Four-week-old *N. benthamiana* seedlings and young *P. massoniana* tissue-cultured seedlings were used as materials for transient transformation. We used the agroinfiltration method for transient expression of GFP-fused proteins in *N. benthamiana* leaves, and 35S:GFP was used as the control. First, 3-ml LB cultures supplemented with rifampicin and kanamycin in a capped 10-ml conical flask were inoculated with the plasmid. The cultures were incubated at 28°C under agitation (200 rpm) for 4 h. In the second step, the cells were pelleted by centrifugation for 15 min at 5,000 rpm. After aspiration of the supernatant, the pellet was resuspended in 3 ml of infiltration medium (10 mM MES, 10 mM $MgCl_2$, and 200 μM acetosyringone). The samples were diluted to OD600 = 1 with the infiltration buffer. After letting it sit for 1 h, a 1-ml needleless syringe was filled with the suspension and injected into *N. benthamiana* leaves, which were incubated for 48 h in the dark. For *P. massoniana*, the suspension samples were introduced into the plant by vacuum infiltration. After incubation for 2 days, nitro blue tetrazolium (NBT) staining was performed, and RNA was also extracted from the plants for quantitative analysis.

Results

Verification of LRR-RLK genes in *P. massoniana*

A total of 185 candidate genes corresponding to LRR-RLK genes were screened from the transcriptome data of *B. xylophilus*. The statistics for these LRR-RLK genes are shown in Supplementary Table S2. The number of amino acid residues predicted to be encoded by these genes ranged from 470 to 1,460.

The LRR-RLK proteins comprising more than 900 amino acids were the most abundant, whereas 54 LRR-RLK proteins comprised fewer than 900 amino acids. The relative molecular weights of the LRR-RLK proteins were 52–160 kDa. Prediction of the subcellular localization suggested that the LRR-RLK proteins were widely distributed within the cell: 112 LRR-RLK proteins were predicted to be mainly localized in the plasma membrane, 14 mainly in the nucleus, 18 in the vacuole, 10 in the extracellular matrix, three in the endoplasmic reticulum, four in the cytoplasm, and 24 predominantly in the chloroplasts. Based on these observations, we identified the LRR-RLK genes that encoded proteins mainly localized to the plasma membrane and selected for further analysis. Other predicted physical and chemical properties of the LRR-RLK proteins are shown in Supplementary Table S2.

Motifs and phylogenetic analysis of LRR-RLK proteins

Phylogenetic analysis revealed that the 185 LRR-RLK proteins of *P. massoniana* could be divided into 14 groups consistent with the LRR-RLK protein classification for *A. thaliana*. The conserved shared motifs were determined using the full-length amino acid sequences of the LRR-RLK proteins with the MEME tool. Ten motifs were identified in the 185 LRR-RLK protein sequences of *P. massoniana* ranging in length from 16 to 50 amino acids. The genes of *A. thaliana* and *P. massoniana* placed in the same lineage were summarized, and a motif structure map was generated. Motif 1 was the main feature (Figure 1). The functions of most of the motifs remain to be elucidated. In combination with the results of the phylogenetic analysis, the conservative composition of LRR domains strongly supported the reliability of the classification of the proteins into groups. Certain genes of the RLK subgroup had a higher homology with genes of *A. thaliana*, but the number of genes differed significantly among the subgroups, which may be the result of functional evolution. The number of *P. massoniana* genes was more than six times that of *A. thaliana* genes in subgroups of the LRR XII group (Table 1). This may be due to environmental selection that has resulted in gene duplication in *P. massoniana*, and these genes may play an important role in disease resistance pathways. Importantly, *PmRLKs32* was classified in the LRR XII group; this protein is induced by *B. xylophilus* and affects plant secondary metabolism (Liu et al., 2021).

Transcriptome analysis of *P. massoniana*

The sequence information and heat map data for this analysis were based on the transcriptome. Combining the heat

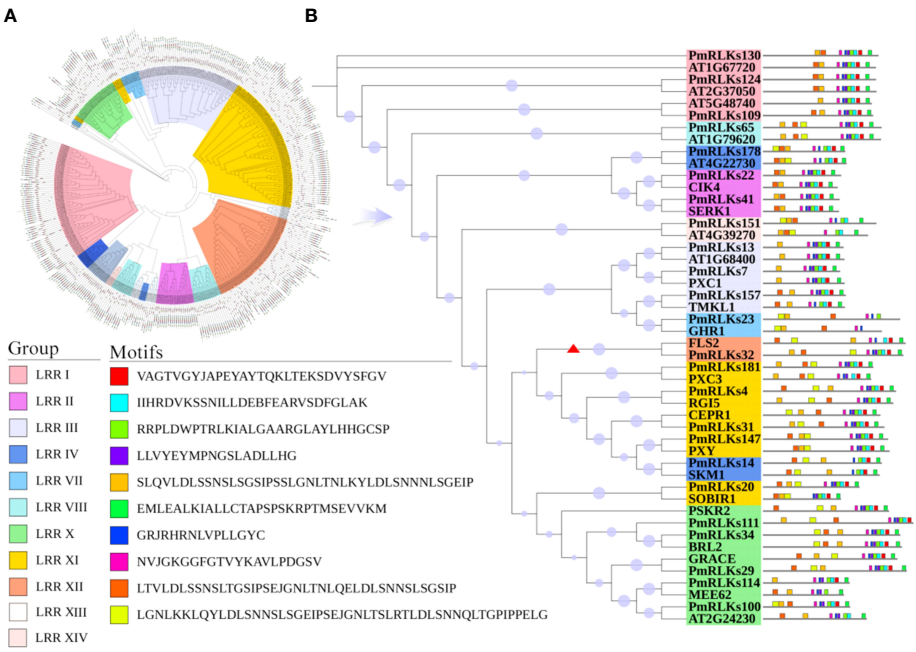


FIGURE 1
Motifs and phylogenetic tree for PmRLK proteins. **(A)** Phylogenetic tree based on the full-length amino acid sequences using the maximum likelihood method. The different subgroups are indicated by different colors. **(B)** Motifs and phylogenetic tree for *P. massoniana and A. thaliana* homologs. The red arrowhead shows the genes that were selected for further analysis.

10.3389/fpls.2022.1043261

TABLE 1 Number of genes in the subgroups of LRR-RLK genes from *Pinus massoniana* and *Arabidopsis thaliana*.

Subgroup	Number of genes	
	P. massoniana	*A. thaliana*
LRR I	20	46
LRR II	8	14
LRR III	15	44
LRR IV	1	3
LRR V	4	9
LRR VI	7	11
LRR VII	5	9
LRR VII	7	21
LRR IX	5	4
LRR X	12	16
LRR XI	46	41
LRR XII	51	8
LRR XIII	2	0
LRR XIV	2	2

map (Figure 2) results and weighted gene co-expression network analysis of the transcriptome, 22 candidate *PmRLK* genes were identified (Supplementary Figure S1).

A qRT-PCR analysis was conducted to determine the expression levels of the *PmRLK* genes. Consistent with the transcriptome data, the expression levels of *PmRLKs32* were significantly upregulated after infection. In the highly resistant sample, *PmRLKs32* was significantly upregulated at 1 day after inoculation, and the relative expression level exceeded the highest level observed for susceptible samples. On day 15 after inoculation, the *PmRLKs32* expression level decreased to less than the original level, and after 30 days it recovered to the original level. *PmRLKs39* was significantly upregulated at 1 day after inoculation. The trends for *PmRLKs153* and *PmRLKs159* were consistent with that for *PmRLKs32*, but the relative expression levels of the former two genes were slightly lower. The trend for *PmRLKs23* was opposite to that of *PmRLKs32*. Other *PmRLK* genes did not show a significant variation in relative expression levels in response to inoculation (Figure 3). The qRT-PCR results indicated that the samples with high resistance responded quickly to inoculation and maintained internal homeostasis. The *PmRLK* expression levels at 15 days post-inoculation were lower than those prior to inoculation, which may be due to negative feedback of plant immunity to prevent an excessive immune response from causing damage to the plant itself.

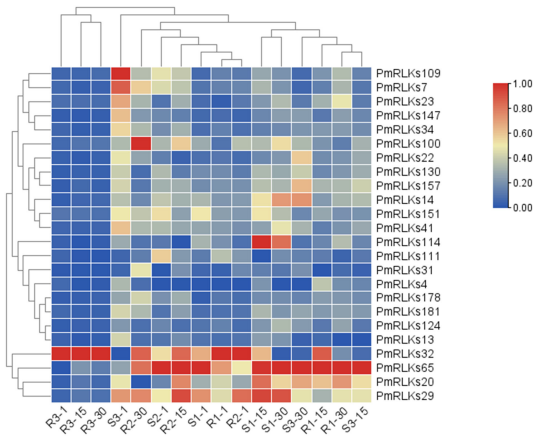


FIGURE 2
Heat map of PmRLK proteins. The PmRLK family members in *Pinus massoniana* that had high homology with the proteins of *Arabidopsis thaliana* are shown, and the heat map includes data for samples collected at 1, 15, and 30 days after inoculation. The expression level of the sample is relative to that of the control. The color scale represents the relative fold change in expression level.

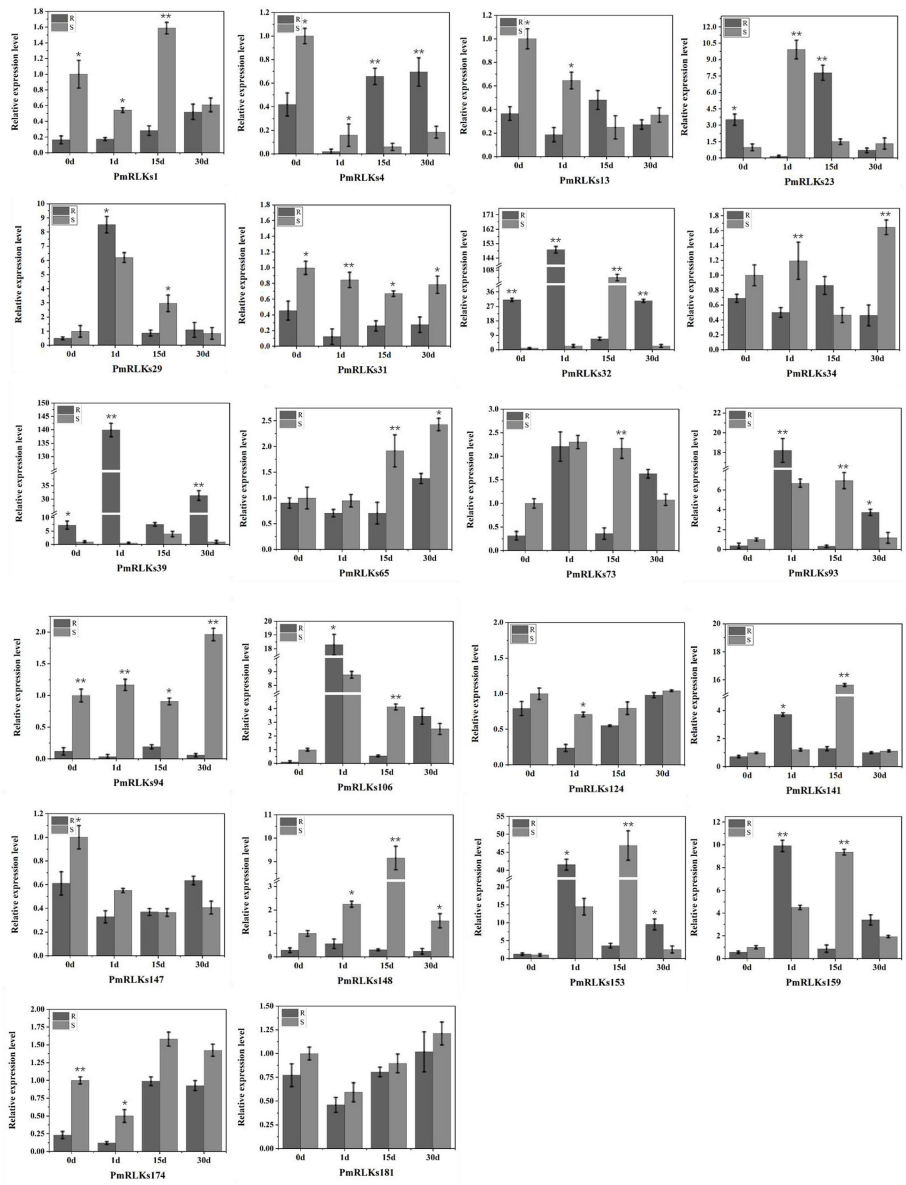


FIGURE 3
Relative expression of *PmRLK* genes of *P. massoniana* in response to nematode treatment. The relative expression levels are the $\log_2$ ($2^{-\Delta\Delta Ct}$) values. The asterisks indicate significant differences between each time point of highly resistant (R) and susceptible (S) accessions of *P. massoniana*. *$p < 0.05$, **$p < 0.01$.

Analysis of subcellular localization of *PmRLKs32*

The *PmRLK* genes mainly participate in signal transduction. The predictions for their subcellular localization showed that they mainly play a role in the plasma membrane. To verify the prediction, *PmRLKs32* was selected for a subcellular localization experiment because it showed the highest relative expression level (Figure 4). Consistent with the prediction, *PmRLKs32* was localized to the plasma membrane. The PmRLKs32 protein may bind to extracellular effectors to activate immunity and stimulate the synthesis of relevant hormones to resist various stresses.

Transient transformation with *PmRLKs32*

To further clarify the function of *PmRLKs32*, a transient transformation assay with *N. benthamiana* and *P. massoniana* was performed. The function of *PmRLKs32* was confirmed by qRT-PCR and histochemical visualization of ROS accumulation.

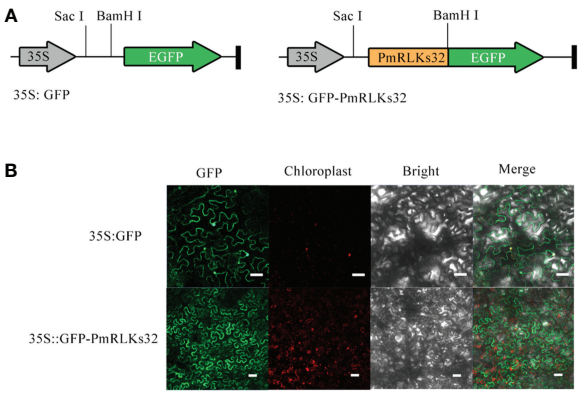


FIGURE 4
Subcellular localization of *PmRLKs32* in *Nicotiana benthamiana* leaf cells. **(A)** Schematic diagrams of the 35S:GFP vector and 35S:GFP-*PmRLKs32* construct. **(B)** Fluorescence from 5S:GFP and 35S:GFP-*PmRLKs32* transiently expressed in 4-week-old *N. benthamiana* leaves. The GFP signals were observed with a ZEISS LSM880 confocal microscope. Bar, 50 μm.

The qRT-PCR results showed that *PmRLKs32* was successfully expressed in *N. benthamiana* and *P. massoniana*. The expression levels of downstream related genes and resistance-related transcription factors were detected in the transient overexpression plants. The results revealed that the expression of downstream genes of the PTI resistance pathway was upregulated, whereas the expression of ETI resistance pathway-related transcription factors was downregulated in the transient overexpression plants (Figure 5). In addition, NBT staining indicated that ROS and callose were significantly accumulated in the overexpression plants compared with those of the control. These results confirmed that *PmRLKs32* could activate the PTI pathway and enhance plant resistance through ROS production (Figure 6).

Discussion

With plant evolution from unicellular eukaryotic algae to multicellular angiosperms, the number of RLK family members increased rapidly. In the present phylogenetic analysis, few genes showed a high homology between *P. massoniana* and *A. thaliana*, and the majority were phylogenetically well separated. The present results indicated that the *PmRLK* genes of *P. massoniana* may have undergone segmental duplication, tandem duplication, or polyploidization to adapt to various stresses in the environment. The tomato RLK proteins SlSERK3A and SlSERK3B enhance root-knot nematode resistance and interact with FLS2 *in vivo* (Peng and Kaloshian, 2014). FLS2 homologous genes have been reported in tobacco, *A.*

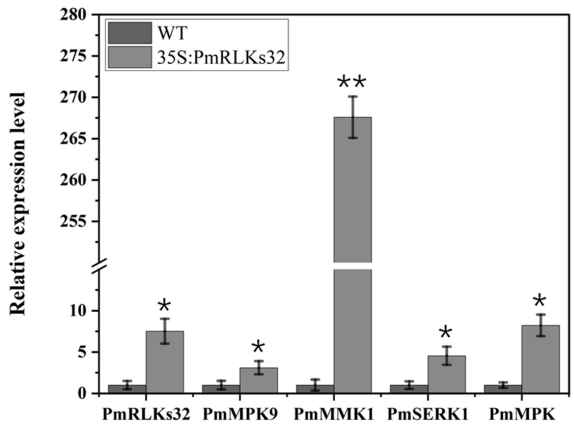


FIGURE 5
Gene expression downstream of *PmRLKs32*. The scale represents relative expression levels based on the $\log_2$ ($2^{-\Delta\Delta Ct}$) values. The asterisks indicate significant differences between each time point of the wild-type (WT) and 35S:*PmRLKs32*-overexpression plants. *$p < 0.05$, **$p < 0.01$.

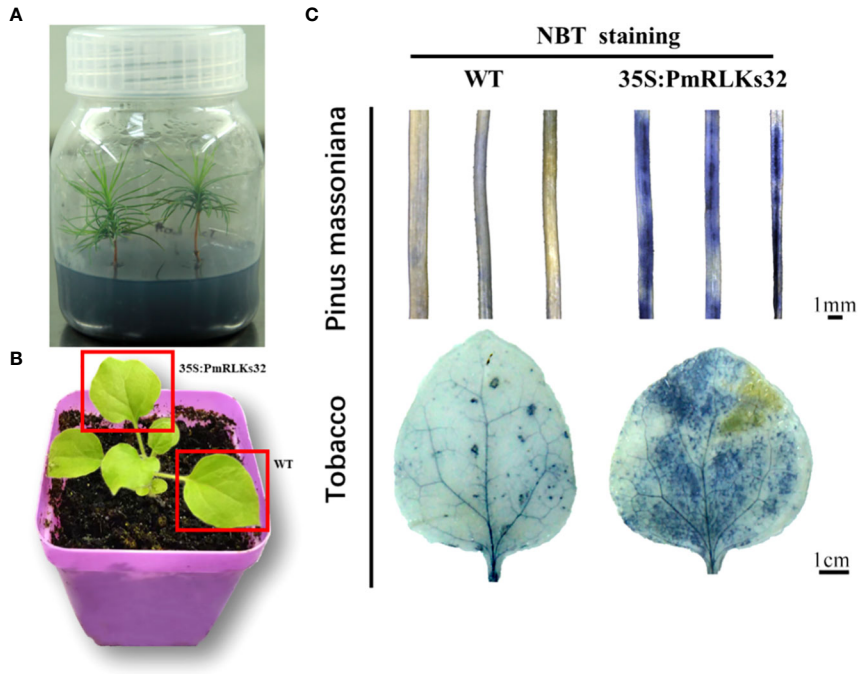


FIGURE 6
Staining of reactive oxygen species with nitro blue tetrazolium (NBT). **(A)** Tissue-cultured seedlings of *Pinus massoniana*. **(B)** Material of *Nicotiana benthamiana*: leaves used for transient transformation with the 35S:*PmRLKs32* construct are indicated by red boxes. **(C)** NBT staining of needles of *P. massoniana* (upper) and leaves of *N. benthamiana* (lower).

thaliana, and tomato (Chinchilla et al., 2006; Hao et al., 2016; Zhou and Zeng, 2018). We identified the *PmRLKs32* protein in *P. massoniana*, which had a high homology to *AtFLS2* and was classified in the subgroup LRR XII, and demonstrated that overexpression of *PmRLKs32* triggered a burst of ROS production. As a consequence of stress metabolism, ROS not only play a role in plant defense mechanisms but also participate in biotic stress responses as signal molecules (Tarfeen et al., 2022). After the transient overexpression of *PmRLKs32* in *P. massoniana* tissue-cultured seedlings, the downstream genes *MPK9*, *MPK*, *MMK1*, and *SERK1* were also up-regulated. Therefore, overexpression of *PmRLKs32* stimulated the immune response in *P. massoniana*, but the specific mechanism is still unclear and requires further investigation.

Some redundant and truncated RLK genes currently without functional annotations are often ignored. In the phylogenetic tree, the number of *P. massoniana* genes in subgroup LRR XII exceeded the number of *A. thaliana* genes. The functions of these extra *P. massoniana* genes in subgroup LRR XII are currently unknown, and the proteins may act independently or as accessory proteins. In *A. thaliana*, *AtVRLK1* redundantly regulates secondary cell wall thickening (Huang et al., 2018). An additional family of resistance proteins, LRR-RLPs, may also play an indispensable role. The difference between RLK and RLP proteins is only in the C-terminus, and RLP proteins may even be truncated variants of RLK proteins (He et al., 2018). These RLP genes may have acquired novel functions as a result of the structural changes. The *RLP2* and *RLP3* genes only lack a C-terminal kinase domain compared with LRR-RLK genes containing the sulfate tyrosine 1 receptor (*PSY1R*) (Oehlenschlæger et al., 2017). The LRR-RLP gene *CLAVATA 2* has the same function as the LRR-RLK gene *CLAVATA 1* (Hanemian et al., 2016; Pan et al., 2016).

Comprehensive identification of *P. massoniana* LRR-RLKs and resolution of their phylogenetic relationships will facilitate the further exploration of genes with potentially redundant functions. The PTI and ETI pathways are the major defense pathways in plants. The diversity of genes is closely associated with the ability to adapt to stress. The present analysis revealed the potential function of *PmRLKs32* in resistance to pine wilt disease.

Conclusions

This study comprehensively analyzed the RLK genes of *P. massoniana* and identified 185 *PmRLK* genes classified into 14 groups. Previous studies have shown that *RLK* genes play an important role in the plant defense system. A phylogenetic analysis revealed that *PmRLKs32* showed a high homology to *AtFLS2* of *A. thaliana*. This gene may play a crucial role in the defense regulatory network in *P. massoniana*. By performing qRT-PCR assays and

transcriptome analysis, we confirmed that *PmRLKs32* was highly expressed in pine wood nematode-resistant individuals in response to nematode inoculation. Transient transformation with *PmRLKs32* revealed the accumulation of callose and ROS in the needles of *P. massoniana*. Hydrogen peroxide may activate the jasmonate pathway as a signal molecule, thereby enhancing plant resistance. The results further implicate *PmRLKs32* involvement in the defense regulatory network of plants. In addition, the present findings provide new insights into the function of *RLK* genes and provide a basis for resistance breeding and further study of the defense regulatory network in *P. massoniana*.

Data availability statement

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

Author contributions

ZN: Conceptualization, methodology, software, experiment, investigation, data analysis, and writing—original draft; WL: data curation and writing—original draft; LD: experiment and data analysis; KG: visualization, writing—review and editing, resources, and supervision; QL: software and validation; ZZ: funding acquisition and supervision. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Zhejiang Science and Technology Program (2020C02007), the National Natural Science Foundation of China (31470670), the Zhejiang Science and Technology Major Program on Agricultural New Variety Breeding (2021C02070-5-2), and Forestry Science and Technology Innovation Special Project of Jiangxi Forestry Bureau (2021, no. 13).

Acknowledgments

We are deeply grateful to our supervisor for his constant encouragement and also indebted to all the other tutors and teachers for their direct and indirect help for us.

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.2022.1043261/full#supplementary-material

References

Abdul Malik, N. A., Kumar, I. S., and Nadarajah, K. (2020). Elicitor and receptor molecules: Orchestrators of plant defense and immunity. *Int. J. Mol. Sci.* 21, 963. doi: 10.3390/ijms21030963

Albert, I., Hua, C., Nürnberger, T., Pruitt, R. N., and Zhang, L. (2019). Surface sensor systems in plant Immunity[OPEN]. *Plant Physiol.* 182, 1582–1596. doi: 10.1104/pp.19.01299

Bi, G., Su, M., Li, N., Liang, Y., Dang, S., Xu, J., et al. (2021). The ZAR1 resistosome is a calcium-permeable channel triggering plant immune signaling. *Cell* 184, 3528–41. e12. doi: 10.1016/j.cell.2021.05.003

Chakraborty, S., Nguyen, B., Wasti, S. D., and Xu, G. (2019). Plant leucine-rich repeat receptor kinase (LRR-RK): Structure, ligand perception, and activation mechanism. *Molecules* 24, 3081. doi: 10.3390/molecules24173081

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

Chinchilla, D., Bauer, Z., Regenass, M., Boller, T., and Felix, G. (2006). The arabidopsis receptor kinase FLS2 binds flg22 and determines the specificity of flagellin perception. *Plant Cell* 18, 465–476. doi: 10.1105/tpc.105.036574

Hanemian, M., Barlet, X., Sorin, C., Yadeta, K. A., Keller, H., Favery, B., et al. (2016). Arabidopsis CLAVATA 1 and CLAVATA 2 receptors contribute to ralstonia solanacearum pathogenicity through a miR169-dependent pathway. *New Phytol.* 211, 502–515. doi: 10.1111/nph.13913

Hao, G., Pitino, M., Duan, Y., and Stover, E. (2016). Reduced susceptibility to xanthomonas citri in transgenic citrus expressing the FLS2 receptor from *Nicotiana benthamiana*. *Mol. Plant-Microbe Interact.* 29, 132–142. doi: 10.1094/MPMI-09-15-0211-R

He, Y., Zhou, J., Shan, L., and Meng, X. (2018). Plant cell surface receptor-mediated signaling – a common theme amid diversity. *J. Cell Sci.* 131(2):jcs209353. doi: 10.1242/jcs.209353

Huang, C., Zhang, R., Gui, J., Zhong, Y., and Li, L. (2018). The receptor-like kinase AtVRLK1 regulates secondary cell wall thickening. *Plant Physiol.* 177, 671–683. doi: 10.1104/pp.17.01279

Lin, X., Armstrong, M., Baker, K., Wouters, D., Visser, R. G., Wolters, P. J., et al. (2020). RLP/K enrichment sequencing; A novel method to identify receptor-like protein (RLP) and receptor-like kinase (RLK) genes. *New Phytol.* 227, 1264–1276. doi: 10.1111/nph.16608

Liu, B., Liu, Q., Zhou, Z., Yin, H., Xie, Y., and Wei, Y. (2021). Two terpene synthases in resistant *Pinus massoniana* contribute to defence against *Bursaphelenchus xylophilus*. *Plant Cell Environ.* 44, 257–274. doi: 10.1111/pce.13873

Liu, Q., Wei, Y., Xu, L., Hao, Y., Chen, X., and Zhou, Z. (2017). Transcriptomic profiling reveals differentially expressed genes associated with pine wood nematode resistance in masson pine (*Pinus massoniana* lamb.). *Sci. Rep.* 7, 1–14. doi: 10.1038/s41598-017-04944-7

Naveed, Z. A., Wei, X., Chen, J., Mubeen, H., and Ali, G. S. (2020). The PTI to ETI continuum in phytophthora-plant interactions. *Front. Plant Sci.* 11, 593905. doi: 10.3389/fpls.2020.593905

Ngou, B. P. M., Ahn, H.-K., Ding, P., and Jones, J. D. (2021). Mutual potentiation of plant immunity by cell-surface and intracellular receptors. *Nature* 592, 110–115. doi: 10.1038/s41586-021-03315-7

Oehlenschlæger, C. B., Gersby, L. B., Ahsan, N., Pedersen, J. T., Kristensen, A., Solakova, T. V., et al. (2017). Activation of the LRR receptor-like kinase PSY1R requires transphosphorylation of residues in the activation loop. *Front. Plant Sci.* 8, 2005. doi: 10.3389/fpls.2017.02005

Pan, L., Lv, S., Yang, N., Lv, Y., Liu, Z., Wu, J., et al. (2016). The multifunction of CLAVATA2 in plant development and immunity. *Front. Plant Sci.* 7, 1573. doi: 10.3389/fpls.2016.01573

Peng, H.-C., and Kaloshian, I. (2014). The tomato leucine-rich repeat receptor-like kinases SlSERK3A and SlSERK3B have overlapping functions in bacterial and nematode innate immunity. *PloS One* 9, e93302. doi: 10.1371/journal.pone.0093302

Pruitt, R. N., Locci, F., Wanke, F., Zhang, L., Saile, S. C., Joe, A., et al. (2021). The EDS1–PAD4–ADR1 node mediates arabidopsis pattern-triggered immunity. *Nature* 598, 495–499. doi: 10.1038/s41586-021-03829-0

Tarfeen, N., Nisa, Q., and Kahlief, K. (2022). "Antioxidant defense system in plants against biotic stress," in *Antioxidant defense in plants* (Singapore: Springer), 383–395.

Tian, H., Wu, Z., Chen, S., Ao, K., Huang, W., Yaghmaiean, H., et al. (2021). Activation of TIR signalling boosts pattern-triggered immunity. *Nature* 598, 500–503. doi: 10.1038/s41586-021-03987-1

Xi, L., Wu, X. N., Gilbert, M., and Schulze, W. X. (2019). Classification and interactions of LRR receptors and co-receptors within the arabidopsis plasma membrane–an overview. *Front. Plant Sci.* 10, 472. doi: 10.3389/fpls.2019.00472

Ye, M., Glauser, G., Lou, Y., Erb, M., and Hu, L. (2019). Molecular dissection of early defense signaling underlying volatile-mediated defense regulation and herbivore resistance in rice. *Plant Cell* 31, 687–698. doi: 10.1105/tpc.18.00569

Yuan, M., Jiang, Z., Bi, G., Nomura, K., Liu, M., Wang, Y., et al. (2021a). Pattern-recognition receptors are required for NLR-mediated plant immunity. *Nature* 592, 105–109. doi: 10.1038/s41586-021-03316-6

Yuan, M., Ngou, B. P. M., Ding, P., and Xin, X.-F. (2021b). PTI-ETI crosstalk: an integrative view of plant immunity. *Curr. Opin. Plant Biol.* 62, 102030. doi: 10.1016/j.pbi.2021.102030

Zhao, Y., Wu, G., Shi, H., and Tang, D. (2019). RECEPTOR-LIKE KINASE 902 associates with and phosphorylates BRASSINOSTEROID-SIGNALING KINASE1 to regulate plant immunity. *Mol. Plant* 12, 59–70. doi: 10.1016/j.molp.2018.10.008

Zhou, B., and Zeng, L. (2018). The tomato U-box type E3 ligase PUB13 acts with group III ubiquitin E2 enzymes to modulate FLS2-mediated immune signaling. *Front. Plant Sci.* 9, 615. doi: 10.3389/fpls.2018.00615

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 16 December 2022
DOI 10.3389/fpls.2022.1059804

OPEN ACCESS

EDITED BY
Hussain Touseef,
Matimate Agromart Pvt. Ltd. (Sevama AgriClinic Laboratory), India

REVIEWED BY
Prakit Somta,
Kasetsart University, Thailand
Chandra Mohan Singh,
Banda University of Agriculture and Technology, India

*CORRESPONDENCE
Huanxiu Li
huanxiuli62@163.com

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 02 October 2022
ACCEPTED 28 November 2022
PUBLISHED 16 December 2022

CITATION
Liang L, Zhang J, Xiao J, Li X, Xie Y, Tan H, Song X, Zhu L, Xue X, Xu L, Zhou P, Ran J, Sun B, Huang Z, Tang Y, Lin L, Sun G, Lai Y and Li H (2022) Genome and pan-genome assembly of asparagus bean (*Vigna unguiculata* ssp. *sesquipedialis*) reveal the genetic basis of cold adaptation. *Front. Plant Sci.* 13:1059804. doi: 10.3389/fpls.2022.1059804

Genome and pan-genome assembly of asparagus bean (*Vigna unguiculata* ssp. *sesquipedialis*) reveal the genetic basis of cold adaptation

Le Liang[1], Jianwei Zhang[1], Jiachang Xiao[1], Xiaomei Li[1,2], Yongdong Xie[3], Huaqiang Tan[4], Xueping Song[1], Li Zhu[1], Xinru Xue[1], Linyu Xu[1], Peihan Zhou[1], Jianzhao Ran[1], Bo Sun[1], Zhi Huang[1], Yi Tang[5], Lijin Lin[5], Guochao Sun[1], Yunsun Lai[5] and Huanxiu Li[5]*

[1]College of Horticulture, Sichuan Agricultural University, Chengdu, Sichuan, China, [2]Vegetable Germplasm Innovation and Variety Improvement Key Laboratory of Sichuan, Chengdu, China, [3]Institute for Processing and Storage of Agricultural Products, Chengdu Academy of Agriculture and Forestry Sciences, Chengdu, China, [4]Horticulture Research Institute, Chengdu Academy of Agriculture and Forestry Sciences, Chengdu, China, [5]Institute of Pomology and Olericulture, Sichuan Agricultural University, Chengdu, Sichuan, China

Asparagus bean (*Vigna unguiculata* ssp. *sesquipedialis*) is an important cowpea subspecies. We assembled the genomes of Ningjiang 3 (NJ, 550.31 Mb) and Dubai bean (DB, 564.12 Mb) for comparative genomics analysis. The whole-genome duplication events of DB and NJ occurred at 64.55 and 64.81 Mya, respectively, while the divergence between soybean and *Vigna* occurred in the Paleogene period. NJ genes underwent positive selection and amplification in response to temperature and abiotic stress. In species-specific gene families, NJ is mainly enriched in response to abiotic stress, while DB is primarily enriched in respiration and photosynthesis. We established the pan-genomes of four accessions (NJ, DB, IT97K-499-35 and Xiabao II) and identified 20,336 (70.5%) core genes present in all the accessions, 6,507 (55.56%) variable genes in two individuals, and 2,004 (6.95%) unique genes. The final pan genome is 616.35 Mb, and the core genome is 399.78 Mb. The variable genes are manifested mainly in stress response functions, ABC transporters, seed storage, and dormancy control. In the pan-genome sequence variation analysis, genes affected by presence/absence variants were enriched in biological processes associated with defense responses, immune system processes, signal transduction, and agronomic traits. The results of the present study provide genetic data that could facilitate efficient asparagus bean genetic improvement, especially in producing cold-adapted asparagus bean.

KEYWORDS

De novo* assembly, evolution, pan-genome, presence/absence variation, *Vigna unguiculata* ssp. *sesquipedialis

1 Introduction

Cowpea (*Vigna unguiculata*), a diploid (2n = 22) in the family Fabaceae, is a staple food crop globally. The plant was originally domesticated in sub-Saharan Africa and is closely related to several protein-rich warm-season beans (Lonardi et al., 2019). Asparagus bean (*Vigna unguiculata* ssp. *sesquipedialis*) is a cowpea subspecies mainly grown for its 50 −100-cm tender green pods. The pods are high in protein, organic acids, polyols, monosaccharides, and fatty acids.

Asparagus bean is subject to various abiotic stresses, including heat and drought (Xu et al., 2015), salt (Pan et al., 2019), and cold stresses (Heidarvand et al., 2017), affecting primary metabolism, respiration rate, and energy supply for growth. Cold stress is a major limiting factor affecting plant distribution and growth and can significantly alter plant cell characteristics (Heidarvand et al., 2017). Asparagus bean planted in early spring is often challenged by cold stress, which greatly affects subsequent yield and quality. Huang et al. (2018) performed high-throughput specific-locus amplified fragment sequencing of Ningjiang 3 (hereafter NJ) and Dubai bean (hereafter DB), and 5,225 single nucleotide polymorphism (SNP) markers were developed to construct a genetic map (consisting of 11 linkage groups) with a total length of 1,850.81 cM with an average distance of 0.35 cM between markers. In that study, NJ seedlings exhibited better cold tolerance than DB seedlings. Interestingly, the cold tolerance of these two cultivars at the pod setting stage was opposite to that at the seedling stage (Tan et al., 2016). Comparative transcriptome analysis of NJ and DB seedlings under 5 °C cold stress and 25 °C recovery conditions revealed that the differentially expressed genes were enriched mainly in the Gene Ontology (GO) terms for protein phosphorylation, photosynthesis, and redox processes (Miao et al., 2022).

A 519-Mb haploid genome of IT97K-499-35, with 722 scaffolds and 11 chromosomes, has previously been assembled, a phylogenetic tree constructed, and the genomes of some superfamilies analyzed (Lonardi et al., 2019). Additionally, Xia et al. (2019) assembled a 632.8-Mb genome (549.81 Mb non-N size) of the asparagus bean (Xiabao II) based on whole-genome shotgun sequencing and generating a linkage map. The genome assembly has facilitated greater insights into the diversity of cowpea species at the SNP level (Della Coletta et al., 2021). With the development of various sequencing technologies, a wider range of structural variants (SVs), presence/absence variants (PAVs), and copy number variations (CNVs) have been discovered (Sun et al., 2022). The SVs affect the gene density and gene copy number of the genome, highlighting the limitations of a single genome (Li et al., 2014; Willson, 2020; Della Coletta et al., 2021). Pan-genome analysis is mining re-sequenced individual variant genes, including core and variable genes, where variable genes may produce different traits (biotic or abiotic resistance, flowering, and cold resistance) (Razzaq et al., 2021). PAVs are one of the focus points in pan-genome research (Zanini et al., 2022). Recently, pan-genomics research has been carried out in rice (Qin et al., 2021), maize (Springer et al., 2018), cotton (Li et al., 2021), eggplant (Barchi et al., 2021), cucumber (Li et al., 2022), strawberry (Qiao et al., 2021), and legumes, i.e., soybean (Torkamaneh et al., 2021; Ruperao et al., 2021), barrel medic (Zhou et al., 2017), and pigeon pea (Zhao et al., 2020). Sequencing and *de novo* assembly of seven geographically representative accessions in the soybean pan-genome identified lineage-specific genes and genes with large effect mutations. Some showed positive selection and may lead to variation in various agronomic traits (Li et al., 2014). Based on 15 *de novo* assemblies of barrel medi, the dispensable genes had greater variability than the core genes. Additionally, 22% of the genome was involved in large structural changes, expanding the reference genome by 16% (630 Mb). Rapidly evolving gene families are mainly enriched in regions related to biological interactions and stress responses (Zhou et al., 2017). Zhao et al. (2020) constructed the first pigeon pea pan-genome in which the annotation of variable genes suggested that they were associated with self-fertilization and response to disease and identified 225 SNPs associated with nine important agronomic traits. Although the genome of the asparagus bean has been assembled, there is no report on its origin and pan-genome.

In this study, the genomes of NJ and DB were assembled and compared with the genomes of 11 other plants to analyze the evolution and origin of the asparagus bean. A pan-genome of NJ, DB, IT97K-499-35 and Xiabao II was constructed using NJ as a reference genome. Small (SNPs and InDels) and large (SVs, CNVs, and PAVs) variants in the genome were mined, a high-density genetic map was reconstructed. This information will provide a theoretical basis for the molecular mechanism of intraspecific variation and cold tolerance of *Vigna* and further research on gene regulation.

2 Materials and methods

2.1 Genome sequencing, assembly, and evaluation

Two varieties of asparagus bean, Ningjiang 3 (NJ) and Dubai bean (DB), were used for genome sequencing. High-quality genomic DNA was extracted from the leaves of three-week-old plants using a modified CTAB method. The extracted DNA was sequenced with Illumina Novaseq 6000 (Illumina Inc., San Diego, CA, USA) and PacBio Sequel II (Pacific Biosciences of California, Menlo Park, CA, USA) platforms. The long reads from the PacBio platform were used for genome assembly. The short reads from the Illumina platform were quality filtered using HTQC v1.92.310 and used to estimate the genome size,

level of heterozygosity, and repeat content. The assembled genome was assessed and Cegma v2.5 was used to evaluate the integrity of the final genome assembly by aligning short reads from the Illumina platform to IT97K-499-35 (Lonardi et al., 2019). The uniquely mapped data were retained to perform assembly by using LACHESIS software.

2.2 Transposable elements and tandem repeat annotations

TEs were identified by a combination of homology-based and *de novo* approaches. We customized a *de novo* repeat library of the genome using repeatModeler, automatically executing RECON v1.08 and repeat Scout to find two *de novo* repeats. Subsequently, the high-quality intact full-length long terminal repeat retrotransposons (fl-LTR-RTs) and non-redundant LTR library were identified using both LTRharvest, LTR_finder, and LTR_retriever. A non-redundant species-specific TE library was constructed by combining the *de novo* TE sequences library above with the known Repbase v19.06, REXdb v3.0, and Dfam v3.2 databases. Final TE sequences were identified and classified by homology search against the library using RepeatMasker v4.10. Tandem repeats were annotated by Tandem Repeats Finder and MIcroSAtellite identification tool MISA v2.1.

2.3 Gene prediction and annotation

We integrated three approaches, namely, *de novo* prediction (Augustus v2.4 and SNAP); homology search (GeMoMa v1.7); and transcript-based assembly (RNA-sequencing data were mapped to the reference genome (IT97K-499-35) using Hisat v2.0.4 and assembled by Stringtie v1.2.3) to annotate protein-coding genes in the genome. GeneMarkS-T v5.1 was used to predict genes based on the assembled transcripts. Gene models from the different approaches were combined using EVM v1.1.1 (Haas et al., 2008) and updated using PASA. The final gene models were annotated by searching the GenBank Non-Redundant (NR, 20200921), TrEMBL (202005), Pfam (v33.1), SwissProt (202005), GO (20200615), and KEGG (20191220) databases.

2.4 Annotation of non-coding RNA and pseudogenes

We used tRNAscan-SE v2.0.9 algorithms and Barrnap v0.9, both with default parameters, to identify the genes associated with tRNA and rRNA, respectively. MiRNAs and snRNAs were identified using Infernal v1.1.1 software against the Rfam v14.5 database with default parameters. After masking predicted functional genes, the GenBlastA v1.0.4 program was used to scan the whole genomes. Putative candidates were then analyzed by searching for non-mature and frame-shift mutations using GeneWise v2.4.1. References and software data are supplied as supplementary information.

2.5 Gene family identification

We used proteins from the longest transcripts of each gene from NJ, DB, and *Oryza sativa*, *Spinacia oleracea*, *Solanum tuberosum*, *Raphanus sativus*, *Lupinus albus*, *Cajanus cajan*, *Medicago truncatula*, *Glycine max*, *Vigna angularis*, *V. radiata*, and *V. unguiculata* to cluster the protein-coding gene families. The protein-coding sequences were compared using OrthoFinder v2.4.0. The PANTHER v15 database was used to annotate the gene families and perform GO and KEGG enrichment analyses for the species-specific gene family.

2.6 Phylogenetic analysis

The protein sequences of the 469 single-copy orthologous genes were aligned using the MAFFT v7.205 program. The corresponding CDS alignments were generated and concatenated using protein alignment. IQ-TREE v1.6.11 was used to construct the phylogenetic tree. Using *O. sativa* as the outgroup, the divergence time was calculated using PAML v4.9i (supplied with MCMCTREE). The final evolutionary tree with differentiation time was obtained using MCMCTreeR v1.1. Geological periods used were Permian (Pe), Triassic (Tr), Jurassic (Ju), Cretaceous (Cr), Paleogene (Pa), Neogene (Ne), and Quarternary (Qu). References and software data are supplied as supplementary information. From TimeTree (http://www.timetree.org/), we obtained a species evolutionary tree with divergence time.

2.7 Gene family expansion and contraction analysis

We used CAFÉ to analyze the expansion and contraction of gene families. The criteria for significant expansion or contraction of a gene family are that both family-wide P-values and viterbi *P*-values are < 0.05. GO and KEGG enrichment analyses were performed on the expansion and contraction gene families of NJ and DB. References and software data are supplied as supplementary information. Positively selected genes and whole genome duplication analysis.

We mainly used the CodeML module in PAML for positive selection analysis. Firstly, to obtain single-copy gene families

among *V. angularis*, *V. radiata*, *V. unguiculata*, DB, and NJ, we used MAFFT (parameter: –localpair –maxiterate 1000) to compare the protein sequences of each gene family. Subsequently, we reverted to codon-aligned sequences using PAL2NAL. Finally, we used CodeML based on the Branch-site model to pair the two models, and the significant differences were evaluated ($P < 0.05$). The Bayesian method was used to obtain the posterior probability of a site considered positively selected (usually > 0.95 indicates a significantly positively selected site). GO and KEGG enrichment analyses for positively selected genes were then performed. We used WGD v1.1.0 and a custom script to identify WGD events.

2.8 Pan-genome construction and gene family analysis

Four accessions (NJ-the reference genome, DB, IT97K-499-35 and Xiabao II) were used for pan-genome construction based on *de novo* alignment. The genome sequences of the NJ and DB parents were aligned with the reference genome sequences using MUMmer v4.0, and the pan-genome was constructed using ppsPCP. The gene families were defined as follows: (i) Core gene, gene family shared by the four species, (ii) Dispensable gene, gene family shared by two or three species, and (iii) Unique gene, the specific gene family for each species. The protein sequences of the three species were classified using OrthoFinder v2.3.7 software. The distribution map of the presence and absence of gene families was constructed according to the gene family clustering results. The GO and KEGG enrichment analysis was performed for the core, dispensable, and unique gene families. References and software data are supplied as Supplementary material.

2.9 Sequence variation analysis

Whole-genome alignment was performed using MUMmer 4.0. Synteny and Rearrangement Identifier (SyRI) detects variants using default parameters and identifies collinearity regions, structural rearrangements (inversions, translocations, and duplications), local variations (SNPs, InDels, SVs, PAVs, CNVs), and regions not aligned. A variation with a sequence length > 50 bp that does not exist in the NJ is defined as "presence". Variation that exists only in the NJ is defined as "absence". After detection, SNPs and InDels were annotated using the ANNOVAR software toolkit. We performed GO and KEGG enrichment analysis for the variable genes. References, software data, and other content are supplied as Supplementary material.

2.10 ABC transporter analysis

The sequence of the resulting ABC transporter was analyzed using the HMMERSEARCH software of the Pfam domain database, identifying all proteins from NJ and DB that contained an ABC transporter with $P < 0.05$. The gene IDs encoding the ABC transporter were obtained, and the CDS regions of the genes responding in NJ and DB were extracted with TBtools and translated into protein sequences (Chen et al., 2020). Phylogenetic trees were constructed using MEGA 7.0 and NJ and DB's ABC transporter sequences. The gene IDs were aligned to determine subgene families in NCBI's Conserved Domain Database, using *Arabidopsis* and IT97K-499-35 as references. Mutated genes in the pan-genome were determined based on gene IDs.

2.11 Linkage map construction

The modified logarithm of odds (MLOD) scores between markers were calculated to confirm the robustness of the markers for each linkage groups (LGs). Markers with MLOD scores < 5 were filtered before ordering. To ensure efficient construction of the high-density and high-quality map, a newly developed HighMap strategy was used to order the SLAF markers and correct genotyping errors within LGs (Liu et al., 2014). Map distances were estimated using the Kosambi mapping function. References and software data are supplied as Supplementary material.

3 Results

3.1 Genome assembly

The predicted genome sizes of NJ and DB were 479.14 and 494.50 Mb, respectively (Supplementary Figure 1). We obtained 77 contigs with a total size of 550.31 Mb for NJ and 129 contigs with a length of 564.12 Mb for DB, and sequencing depths of 31.33 X and 32.95 X, respectively (Table 1). The N50 contig is 24.19 Mb for NJ and 27.56 Mb for DB (Table 1). When assessing the integrity of the genome assembly, 98.69% and 98.69% of the CEGMA genes and 96.28% and 96.10% of the BUSCO conserved single-copy genes were present in NJ and DB, respectively (Supplementary Table 1). The completeness of the assembled genome of NJ and DB is 99.16% and 98.96% when assessed by comparison with the IT97K-499-35 genome (Table 1, Supplementary Table 1). We then linked the contigs into scaffolds based on Hi-C data, NJ and DB yielded approximately 105.32 G and 212.23 G clean data, respectively

TABLE 1 Assembly features of the NJ and DB genomes.

Genome feature	NJ	DB
Total Assembly length (Mb)	550.31	564.12
Total Contig number (> 1 Kb)	77	129
Contig N50 size (bp)	24,188,537	27,857,029
Total Scaffold number (> 1 Kb)	46	105
Scaffold N50 size (bp)	49,109,670	48,657,559
GC content (%)	33.41	33.86
Gap length (Gap number)	3,100	2,400
Percent assembly anchored to 11	99.78%	98.93%
Pseudogene	71	77
Repeat sequence (%)	41.19	41.41
Number of protein-coding genes	28,370	28,425
miRNA	129	131
tRNA	1,657	1730
rRNA	6,756	9,165

(Supplementary Table 2). In the assembled NJ genome, 549.09 Mb sequences can be mapped on 11 chromosomes and the sequence length determining the sequence and direction is ~546.82 Mb, accounting for 99.59% of the total mapped length (Table 1). In DB, ~558.07 Mb of sequences can be mapped to 11 chromosomes. The sequence length that determines the sequence and direction is ~545.77 Mb, accounting for 97.80% of the total mapped length (Table 1). Finally, we obtained 46 scaffolds in NJ, with a scaffold N50 size of 49.11 Mb and GC content of 33.41%. In DB, 105 scaffolds with a scaffold N50 size of 48.66 Mb and GC content of 33.86% were obtained (Table 1). The heatmaps of the Hi-C assembled chromosomes for NJ and DB showed that the interaction strength between adjacent sequences (diagonal positions) was high. The interaction signal strength between the non-adjacent sequences (off-diagonal positions) was weak (Supplementary Figure 2). From the circle, we observed that the GC content on each chromosome was inversely proportional to transposable elements (TEs) and gene density, and TE density was proportional to gene density (Supplementary Figure 3). Collectively, the results support a high-quality assembly for the two asparagus bean genomes.

3.2 Genome annotation

Approximately 226.70 Mb of TEs were found in the NJ genome, accounting for 41.20% of the genome (Supplementary Table 3), and ~233.57 Mb in DB, accounting for 41.41% of the genome (Supplementary Table 4). Additionally, the LTR retrotransposons of NJ and DB contained ~2.5-fold more elements in the Gypsy superfamily than Copia elements and about 1.5-fold more in IT97K-499-35. In the IT97K-499-35, CATA (5.7%) and hAT (2.4%) are the major groups of classical "cut-and-paste" transcripts in the DNA transposon (Lonardi et al., 2019). These two elements had the highest proportion in the NJ and DB genomes but a lower proportion than that in IT97K-499-35 (Supplementary Table 3 and Supplementary Table 4). Approximately 60.52 and 73.59 Mb of tandem duplications were identified in the NJ and DB genomes, accounting for 11.00% and 13.50% of the genomes, respectively (Supplementary Table 5). Differences in motifs between NJ (1,279 motifs and 30,228 domains) and DB (1,283 motifs and 30,281 domains) suggest that the two genomes differ in molecular function, structural properties, and gene families (Supplementary Table 6).

We predicted 28,370 and 28,425 protein-coding genes in NJ and DB, respectively, and many were derived from homogeneous transcriptome and homology predictions (Table 1, Supplementary Figure 4, and Supplementary Table 7). We compared the gene number, exon number, CDS number, and intron number of related species (*P. vulgaris*, *V. radiata*, *V. unguiculata* (IT97K-499-35), and *G. max*) and found that NJ and DB are more closely related to *V. radiata* (Supplementary Figure 5). Approximately 98.27% and 97.52% of the complete BUSCO conserved genes are present in NJ and DB genomes, respectively (Supplementary Table 8). Approximately 92.57% and 91.77% of the transcriptome data were mapped to exons of NJ and DB, respectively, demonstrating high accuracy for our prediction model (Supplementary Figure 6). For the NJ and DB genomes, 96.9% of the 28,370 genes and 96.88% of the 28,425 genes could be annotated, respectively, using various databases (Supplementary Table 9). GO analysis revealed that the genes were annotated with functions related to cellular anatomical entity, catalytic activity, metabolic process, and cellular process (Supplementary Figure 7). Additionally, 1,657 tRNA, 6,756 rRNA, 129 miRNA, and 71 pseudogenes were annotated for NJ, and 1,730 tRNA, 9,165 rRNA, 131 miRNA, and 77 pseudogenes for DB (Table 1).

3.3 Evolutionary history of asparagus bean and syntenic comparison with IT97K-499-35

Based on 469 conserved single-copy genes, we estimated the evolutionary relationships among NJ, DB, and 11 other plants, with *O. sativa* as an outgroup (Figure 1B). The differentiation age between *G. max* and *V. unguiculata* was ~23.82 Mya. From the mean Ks value of the corresponding peaks (0.27), we infer a silent mutation rate (r) of 5.67×10^{-9} substitutions/sites/year (Figure 1A). Based on similar comparisons, the separation of NJ and DB from *V. unguiculata*, *V. radiata*, and *G. max* was estimated to have occurred ~4, ~16, and ~48 Mya, respectively (Figures 1A, B). We also estimated the divergent ages of the cowpea phylogeny, calibrating the divergent ages of the 13 species using 6 fossils with the synaptic morphology of their respective clades as crown groups. The divergence between soybean and *Vigna* occurred in the Paleogene (65–23.3 Mya) (Figure 1B). The whole-genome duplication (WGD) events experienced by *G. max*, *V. unguiculata*, and *V. radiata* during genome evolution were 18.25, 64.81, and 65.25 Mya, respectively (Figure 1A). The WGD events of DB and NJ occurred 64.55 Mya (Ks = 0.732) and 64.81 Mya (Ks = 0.735), respectively (Figure 1A).

We analyzed synteny between NJ and DB and IT97K-499-35 (Supplementary Figure 8). There were 43,036 collinear genes in IT97K-499-35 and NJ, with collinearity of 75.92%; 43,073 genes were collinear between DB and IT97K-499-35, reaching 75.91%. From the 56,795 genes analyzed, 86.07% were collinear for the NJ and DB genomes (Supplementary Figure 8). The collinearity plot shows that NJ and DB have many chromosomal rearrangements compared with IT97K-499-35, and there are obvious breaks on chromosomes 01, 02, 03, 04, 05, 09, and 10 (Figure 1D). It was found that there was a 90.43 Mb inversion in chromosome 01 in NJ, but not in DB (Supplementary Figure 9). Most LTRs in the 13 species were inserted into the genome in the last 5 million years (Figure 2A). The pigeon bean had the earliest LTR insertion event, about 27,000 years ago, and the density of the LTR was the highest. The time of the LTR insertion event was 227,700 years ago in cowpea, 322,100 years ago in NJ, and 278,800 years ago in DB. The insertion times of the LTRs into the potato, spinach, and radish genomes were 4.25, 1.28, and 0.13 Mya, respectively (Figure 2A).

3.4 Gene family analyses

NJ and DB share many gene families with IT97K-499-35 (Figures 2B, C). There were 532 expansions and 553 contractions for NJ, and 550 expansions and 559 contractions for DB (Figure 2B). According to the GO enrichment, expanding gene families in NJ and DB may contribute to translation, metabolism and cellular process, response to stimulus, and reproduction (Supplementary Figure 10). Shared metabolic pathways from the KEGG analysis include photosynthesis,

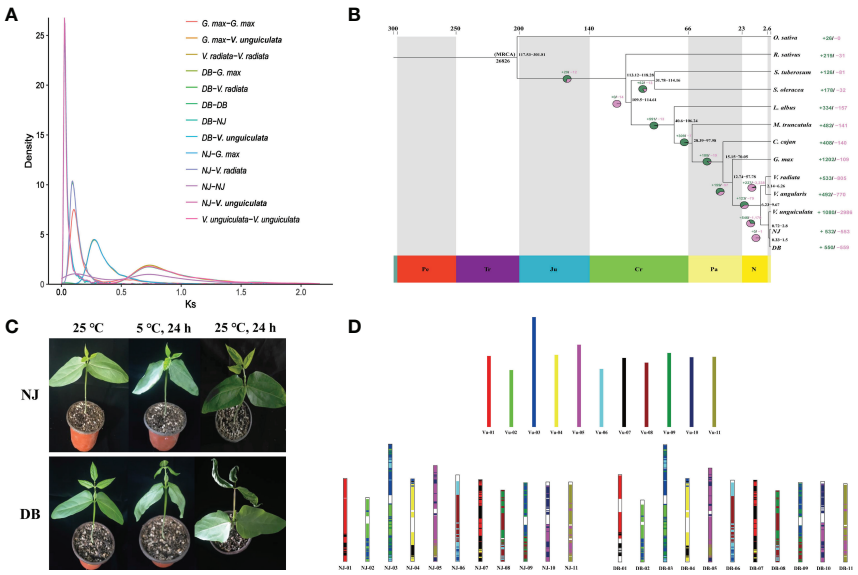


FIGURE 1
Phylogenetic relationships of asparagus bean and its genomic comparison with IT97K-499-35. **(A)**. Frequency distribution of Ks values between syntenic genes of the compared genomes. **(B)** Estimated differentiation time between asparagus bean and other plants, with O. sativa as the outgroup. Expanding and contracting gene families were identified using CAFE, and 26,826 families were deduced to exist in the most recent common ancestor (MCRA). Gene families that expand (green in the pie chart) or shrink (red in the pie chart) are plotted in the pie chart. **(C)**. Comparison of NJ and DB phenotypes under low-temperature stress (5 °C). **(D)**. Collinearity of chromosomes of NJ and DB with IT97K-499-35.

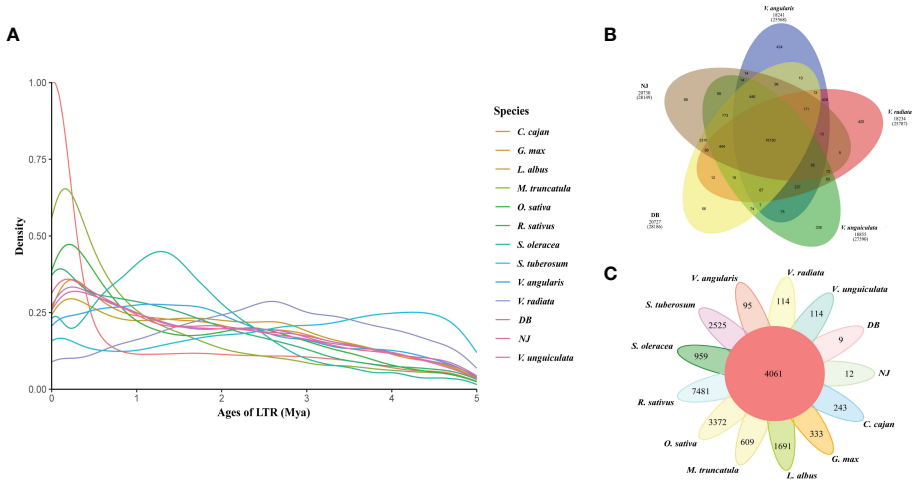


FIGURE 2
Number of orthogroups containing species and the LTR insertion time. **(A)**. LTR insertion time. The formula for calculating time is T = K/(2 × r), where the molecular clock r is chosen as 7×10^{-9}. **(B)**. Venn diagram represents shared and unique gene families of five legumes (NJ, DB, *V. unguiculata*, *V. angularis*, and *V. radiata*). **(C)**. Petal map of shared and unique gene families across 13 species.

oxidative phosphorylation, and flavonoid biosynthesis (Supplementary Figure 11). In the 13 plant genomes, 44,394 orthologous gene clusters were identified, out of which 4,061 orthogroups were shared by the species. Gene family copy number differences were identified between NJ and DB (Supplementary Figure 12), confirming that the DB genome was larger than that of NJ. Compared with the other 13 species, NJ (12) and DB (9) had fewer species-specific orthologous gene clusters (Figure 2C). The Venn diagram of the gene families of NJ, DB, *V. unguiculata*, *V. angularis*, and *V. radiata* shows that the legumes share 16,130 gene families. In comparison, 69 and 66 gene families are specific to NJ and DB, respectively (Figure 2B). NJ species-specific gene family GO annotated terms were in response to stimulus, biological regulation, cellular component organization, or biogenesis, with only oxidative phosphorylation enriched in the KEGG database (Supplementary Figure 13). GO terms in DB species-specific gene families were mainly involved in transporter activity, membrane part and binding, whereas KEGG enrichment was mainly reflected in photosynthesis (Supplementary Figure 14).

A total of 142 genes showed positive selection in NJ. The genes belonged to several GO categories ($P < 0.05$) and KEGG pathways ($P < 0.05$), with a higher representation of RNA-related genes, such as the mRNA cleavage factor complex, RNA methyltransferase activity, and mRNA surveillance pathway (Supplementary Table 10 and Supplementary Table 11). A total of 156 genes showed positive selection in DB. They were mainly enriched in DNA-dependent DNA replication, tRNA processing, and DNA repair in the biological process terms of the GO database (Supplementary Table 10), and in mismatch repair and DNA replication pathways in the KEGG metabolic pathway (Supplementary Table 11).

3.5 Construction of cowpea pan-genome

The pan-genome was constructed using the whole-genome alignments of the four varieties (NJ, DB, IT97K-499-35 and Xiabao II). The reference genome is 550.31 Mb, the final pan genome is 616.35 Mb, and the core genome is 399.78 Mb. Statistical analysis of core and pan gene families of the four varieties showed that the number of core gene families decreased gradually with an increase in the number of species, and the number of gene families in the pan-genome increased with an increase of the number of varieties (Figure 3A, B). In the pan-genome, with NJ as the reference, core gene families accounted for 70.5%, variable gene families for 22.56%, and unique gene families for 6.95% (Figure 3B). Through functional analysis, they contained 20,336, 6,507, and 2,004 gene families, respectively. NJ and DB have similar numbers of core and variable gene families, and Xiabao II and IT97K-499-35 have similar results. However, there are large differences among the four varieties in the number of unique gene families, including 9,259 in Xiabao II, 1,751 in IT97K-499-35, 28 in DB and 53 in NJ (Supplementary Table 12). The UpSet figure is used to show the clustering results of gene families within four species, including 1,736 in Xiabao II, 244 in IT97K-499-35, 9 in DB and 12 in NJ (Figure 3C). The KEGG analysis was carried out for the unique gene families of four varieties. Xiabao II was mainly enriched in zeatin biosynthesis and propoate metropolis, IT97K-499-35 was mainly enriched in dieterpenoid biosynthesis and linoleic acid metropolis, DB was mainly enriched in inorganic animion exchange activity and calcium-dependent phospholipid binding, and in the

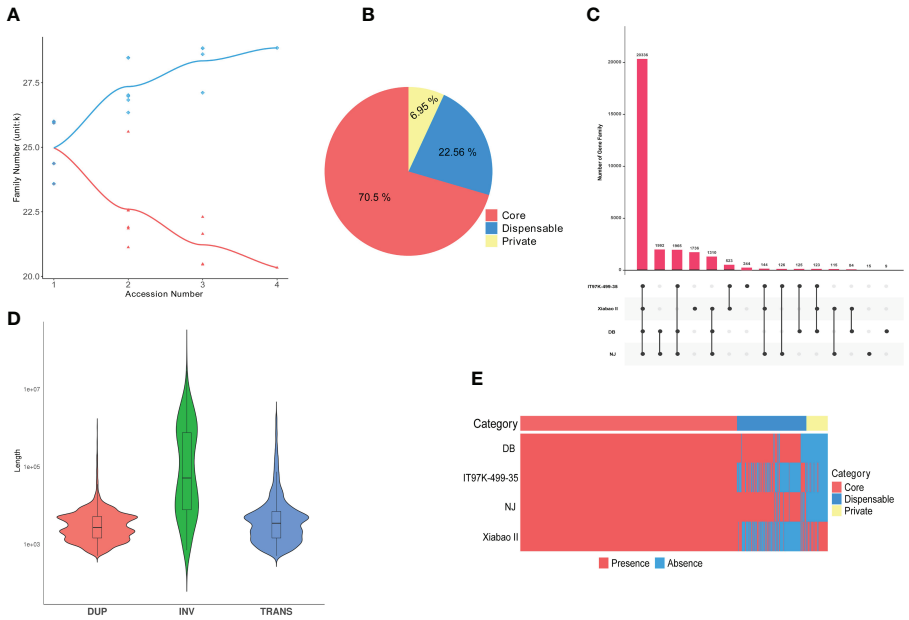


FIGURE 3
Cowpea pan-genome. **(A)**. Gene family accumulation curves for the pan-genome (blue) and core genome (orange). **(B)**. Pie charts show the proportion of each core, variable, and unique gene family. **(C)**. Quantitative dendrograms and dumbbell diagrams of core, variable, and unique gene families in the pan-genome. **(D)**. Violin plot of length statistics for each type of SV (TRANS: chromosomal translocation; INV: inversion; DUP: duplication). **(E)**. Presence/absence information of pan gene families in the NJ, DB, =IT97K-499-35 and Xiabao II.

unique gene family of NJ, it is mainly enriched in photogenic electron transport in photosystem and response to herbicide (Supplementary Figure 15).

3.6 Sequence variation analyses in the pan-genome

The three genomes (Xiabao II, IT97K-499-35 and DB) assemblies were blasted against the NJ genome to detect variations. Between each pair of genomes, 8.74×10^5 (155,245 genes involved) single-nucleotide polymorphisms (SNPs) for Xiabao II, 1.62×10^6 (432,899 genes involved) SNPs for IT97K-499-35, and 1.54×10^6 (294,414 genes involved) SNPs for DB (Supplementary Table 13). The location information of the mutation sites on the Xiabao II, IT97K-499-35 and DB genome showed that the upstream 1 kb region had more SNPs than the UTR-5' and the downstream 1 kb region (Supplementary Table 14). More SNPs were observed in the introns than the exons (Supplementary Table 14). Exons are the main carriers of gene function, but SNPs can cause various gene mutations, such as amino acid changes, loss of start sites or stop codons, or the premature generation of stop codons. The ratio of SNPs that lead to non-synonymous single nucleotide variants (SNVs) and synonymous SNV is roughly 1:1, and the effect of SNPs that lead to amino acid changes on exon function is mainly focused on the mutation generating stop codons (Supplementary Table 15). We observed 2–50 bp variant insertions and deletions (InDels) including 307,078 insertions and 107,962 deletions in Xiabao II, 245,319 insertions and 272,218 deletions in IT97K-499-35, and 175,558 insertions and 157,924 deletions in DB. Structural annotation of InDels followed a trend similar to that of the SNP structural annotation (Supplementary Table 14). The ratio of frameshift mutation and non-frameshift mutation caused by InDels in DB is close to 1:1, and the number of stop codons generated at the mutation site is greater than the sum of the promoter and terminator elimination caused by mutations (Supplementary Table 15). In Xiabao II and IT97K-499-35, InDels caused more frameshift mutations than non-frameshift mutations (Supplementary Table 15). In the circle plot, the frequencies of the SNPs and InDels on 11 chromosomes are similar, but their relationships with gene density are partially negatively correlated (Supplementary Figure 16).

There were 9,516 "presence" variants and 8,199 "absence" variants in the pan-genome (Supplementary Table 16). The numbers of presence/absence variants (PAVs) of Xiabao II and IT97K-499-35 were all higher than those of DB (Supplementary Table 16, Figure 3E). The PAV-affected genes are enriched in biological processes related to biotic and abiotic stress, such as response to stimulus, immune system process, defense response, signal transduction, and biological regulation (Supplementary Table 17).

Large-scale structural variations (SVs, 50 bp), including chromosomal inversions (INVs), chromosomal translocations (TRANs), and tandem duplications (DUPs, Figure 3D), are abundant in the genomes of Xiabao II and IT97K-499-35. A total of 209 INVs, 3,374 TRANs, and 12,209 DUPs were discovered (Supplementary Table 16). Genes affected by SVs are represented in the KEGG database by ribosome, translation, photosynthesis, oxidative phosphorylation, chloroplast thylakoid membrane, and NADH dehydrogenase (ubiquinone) activity terms (Supplementary Figure 17). CNVs variation is attributed to genome rearrangements, which generally refers to an increase or decrease in copy numbers of large genome fragments with a length > 1 kb (Supplementary Table 16). Such genes were enriched in response to stimulus, cellular component organization or biogenesis, detoxification, pentose and glucuronate interconversions, ABC transporters, and other terms (Supplementary Figure 18).

3.7 Variation in ABC transporters

There were 74 and 109 SNP-induced and InDel-induced ABC transporters, respectively. The number of ABC transporters affected by large fragments, such as SVs and CNVs, was relatively low, at 2 (including *ABCC* and *ABCG*) and 3 (including *ABCB* and *ABCC*), respectively. There were 7 ABC transporter subgene families (*ABC*) with "presence" variants and two (*ABCC*) with "absence" variants in the NJ *vs.* DB genomes. The genes encoding ABC transporters in the NJ and DB genomes with $P < 0.05$ were used to construct an evolutionary tree (Figure 4). These genes were clustered in the B and C subgene families, and SNP, InDel and SV variants in the genes were analyzed. In the DB genome, Vun04G000790.1 contains SNP and InDel mutations, Vun04G013310.1 and Vun09G001410.1 both have SNP and InDel mutations, and also have SV mutations. Phylogenetic tree shows that Vun04G000790.1 belongs to B subgene family, and Vun04G013310.1 and Vun09G001410.1 belong to C subgene family.

3.8 High-density genetic map construction

NJ, DB, and 100 F_2 generations, were analyzed using the locus-specific amplified fragment (SLAF) sequencing strategy. A total of 345.25 Mb of clean reads were generated, and an average of 92.58% of the reads had a Q30 quality score, indicating good sequencing quality. We obtained 284,020 SLAF tags, including 29,893 (10.52%) polymorphic SLAF tags. The average sequencing depth of the SLAF-tagged parents was 26.31 ×, and the average sequencing depth of progeny was 9.32 ×. After filtering, 9,986 SLAF tags were obtained to construct 11 linkage groups. After filtering out the MLOD values of each pair of labels below 3, the number of markers was 4,257. The result was a total map distance of 1,253.11 cM and an average genetic distance of 0.3 cM (Figure 5A). The number of SLAF-tags on each chromosome is shown in Figure 5B, Supplementary Table 18 and Supplementary Table 19. We obtained 7,539 SNP markers in the graph; 2,331 SNPs were transitions, and 5,208 SNPs were transversions (Supplementary Table 20). The heat map shows that the linkage relationship between adjacent markers in each group is very strong, and the linkage relationship between markers weakens gradually with an increase in distance (Supplementary Figure 19).

4 Discussion

Asparagus beans are heat-, drought-, and cold tolerant. NJ has better cold resistance than DB in the seedling stage, whereas the opposite is true in the mature stage (Miao et al., 2022); therefore, NJ and DB can be used as relatively extreme materials for studying the cold resistance of asparagus beans. Through the assembly, evaluation, and analysis of the genomes of the two asparagus beans, we inferred the origin of the asparagus bean, its time of differentiation, the timing of WGD, and the rearrangement of the 11 chromosomes. The genome of DB is larger than NJ can be explained from the number and annotation of tandem repeats, this is similar to the conclusion of Dai et al. (2022). We observed that genes that improve cold tolerance were greatly expanded through gene family expansion and positive selection, improving NJ's capacity to adapt to cold environments. Although much genetic information can be obtained from a single genome, analysis of the impact of intra-genome variation on species diversity and agronomic traits is limited (Pinosio et al., 2016; Dolatabadian et al., 2020). The construction of the pan-genome enabled us to unravel the sequence variation within the germplasm and identify genomic regions associated with temperature-adaptive traits. The results enhance our understanding of genome evolution and provide clues for genome variation and cold adaptation in asparagus beans from a pan-genome perspective.

4.1 Evolutionary research on asparagus beans

The assembly, annotation, and comparative genomics analyses showed that soybean and cowpea diverged ~23.82 Mya, and the fossil divergence time demarcated its origin in the Neogene period, ~23 Mya. The timings of WGD events for the NJ and DB genomes were similar to those for the *V. unguiculata* and *V. radiata* genomes (~60 Mya). A peak with a Ks value close to 0 was observed in NJ, which should be the peak generated by the increase of collinear gene pairs caused by the

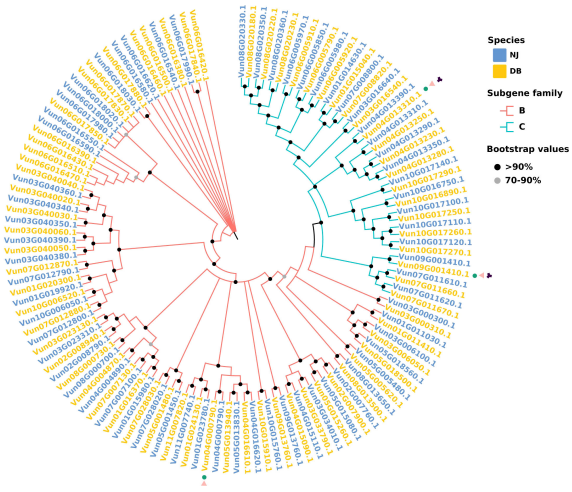


FIGURE 4
Phylogenetics of the NJ and DB ABC transporter gene families. Neighbor-joining NB-ARC tree containing gene models from both species ($P < 0.05$). The red branch represents the B subgene family of the ABC transporter, and the blue branch represents the C subgene family. The red squares represent the NJ gene model, and the blue squares represent the DB gene model. In the outer circle of the phylogenetic tree, green circles represent SNP mutations in the pan-genome, =pink triangles represent InDel mutations and black diamonds represent SV mutations.

insertion of long fragments. Using *Arabidopsis thaliana* as an outgroup, Xu et al. (2020) found a relatively new polyploid event in legumes, a WGD event (~8.62 Mya) shared by *G. max* and its wild relative *Glycine soja*. The WGD event for *G. max* in our research occurred at ~18.25 Mya. The large gaps are probably due to the species used in the phylogenetic tree and outgroup. Using *A. thaliana* as outgroup, a phylogenetic tree showed that cowpea and soybean diverged ~21.20 Mya (Moghaddam et al., 2021); similar to our results (~23.82 Mya). Our data included spinach, radish, and potato to the evolutionary tree. However, this increased the breadth of species comparison, which may allow for relatively large time gaps in species divergence. By elucidating the insertion timing of plant-enriched LTRs, genome expansion can be timed, providing an important basis for species differentiation. It may be that the divergence time with IT97K-499-35 genome is limited, and the expansion and contraction frequency of NJ and DB gene families are also limited.

4.2 NJ adaptation to cold

Changes in the functions of cell membrane enzyme systems are the main manifestations of plant chilling damage (Zhu, 2016). Cold stress reduces the photosynthesis capacity of plants, reduces aerobic respiration, increases anaerobic respiration, accelerates the decomposition of proteins, causes

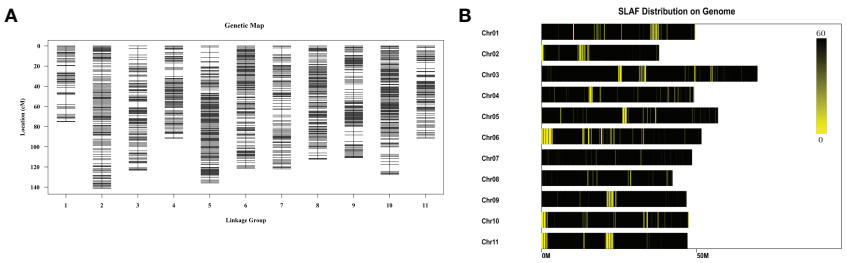


FIGURE 5
Genetic linkage map. **(A)**. Genetic map. The 9,986 SLAF tags screened out were divided into 11 linkage groups by mapping with the reference genome, and the tags whose MLOD values between the two tags were both > 3 were filtered out, and a total of 4,257 tags were markers. Each chromosome is a linkage group, and the linkage group is used as the unit to map according to the size of the linkage group map distance. **(B)**. Distribution of SLAF tags across chromosomes. The abscissa is the length of the chromosome, each yellow band represents a chromosome, and the genome is divided according to the size of 1M. The darker the area in the figure is the area where the SLAF label is concentrated.

protein deficiency and the accumulation of toxic hydrolysates, and ultimately seriously affects plant growth and development (Guo et al., 2018). In our previous study, NJ outperformed DB at 5°C, and NJ showed less low-temperature damage (Figure 1C). By functionally enriching the unique NJ and DB gene families that underwent expansion and positive selection, we found more gene families related to environmental fitness clustered in NJ. They included oxidative phosphorylation, flavonoid synthesis, and photosynthesis genes in the KEGG database. GO terms included responses to hydrogen peroxide, reactive oxygen species, accumulate of toxic substances, defensive responses, and ATP synthesis. Such species-specific gene functions enhance NJ's adaptation to low-temperature environments.

4.3 Pan-genome construction

The asparagus bean genome and comparative genomics with other species have shown that the genetic variation information for the species has not been captured comprehensively, suggesting the need to construct a pan-genome. In addition to the conservation of the core genes (70.50%), there was diversity in the dispensable genes (22.56%) and species-specific genes (6.95%) across the pan-genome, which can facilitate the discovery of new genes to significantly improve the reference genome. Compared with, for example, white lupin (Hufnagel et al., 2021), pigeon pea (Zhao et al., 2020), soybean (Torkamaneh et al., 2021), eggplant (Barchi et al., 2021), cotton (Li et al., 2021), and rice (Qin et al., 2021), the pan-genome of cowpea has a high content of core genes, which may reflect its short domestication history. Functional analysis of the dispensable genes showed their association with enhanced stress adaptation, signal transduction, and seed storage. *Vigna* is an adaptable and resilient crop variety, and current breeding programs aim to enhance food security in the wake of climate change risks. The functional enrichment of such variable genes could facilitate the achievement of this aim (Vincent et al., 2013; Zuluaga et al., 2021).

4.4 Functional analysis of variant genes in the pan-genome

Small variants (SNPs and InDels) and large structural variants (SVs, PAVs, and CNVs) are central to the pan-genome. Among the variants caused by SNPs and InDels, many genes have similar functions. The *ICE1-CBF-COR* is a classic cold response pathway. *ICE1* (SNP/InDel-affected) is an MYC-type bHLH transcription factor that can activate *DREB/CBFs* (SNP/InDel-affected). This reaction up-regulates the downstream cold stress-related *COR* (InDel-affected) genes and ultimately improves the tolerance response of asparagus bean to cold signals and stress (Supplementary Table 21). Small variations have also been found in genes associated with flowering, such as the RNA-binding protein FCA, encoding a strong promoter for the transition to flowering in Arabidopsis (Wang et al., 2020), that exhibits SNP variation and is associated with early flowering-related gene protein ELF4-Like, which exhibits InDel variations (Lin et al., 2019) (Supplementary Table 16). The central roles of heat stress transcription factor and drought-inducible protein 19 (Liu et al., 2013) under various abiotic stresses are influenced by SNPs and InDels, respectively, influencing the drought tolerance of asparagus bean. Numerous helix-loop-helix DNA-binding domain genes were found in the SNP and InDel variants, many encoding phytochrome-interacting factors (*PIFs*). Studies in *A. thaliana* found that *PIFs* play a central role in plant photomorphogenesis, inhibiting seed germination and regulating photogenesis in seedlings. It was also shown that abiotic stress and hormonal signal pathways converge to regulate *PIF* activity (Wang et al., 2022). The researchers also pointed out that the role of the light-induced signal network and the related molecular mechanisms will promote future research on light signal transcription factors and stress tolerance in other crops (Wang et al., 2020). This information should enable breeders to increase asparagus bean yields and develop climate-tolerant asparagus beans that can adapt to changing environmental conditions.

Seed color diversity is one of the main characteristics of *Vigna*, and several gene models identified proteins involved in the late regulation of the flavonoid biosynthesis pathway as candidate genes for seed color by QTL mapping, including the basic helix-loop-helix gene at the C locus, the WD-repeat gene, and the E3 ubiquitin ligase gene (Herniter et al., 2019). When comparing NJ as a reference, we did not find any genes related to flavonoid synthesis in the PAVs. These genes may be highly conserved in *Vigna*, but genes encoding the E3 ubiquitin-protein ligases that may indirectly affect the accumulation of seed coat color were found in the PAVs. Some PAV-affected genes have key roles in cold and heat stress and important agronomical traits. These include *CRPK1* and *14-3-3*, two modules that transmit cold signals from the plasma membrane to the nucleus to regulate CBF stability and, in turn, asparagus bean response to cold stress (Liu et al., 2017). Genes affected by PAV are included in the protease inhibitor/seed storage/LTP family (Meitei et al., 2018), which improves seed utilization, resulting in cowpea asparagus beans high in calories and protein.

4.5 Variations of the ABC transporters

ABC transporters or ATP-binding cassette proteins are an ancient and large family of ATP-driven pumps that play important roles in resistance to biotic pathogens, abiotic stress, and growth and development (Do et al., 2018). The ABC transporters are divided into eight subfamilies in plants: *ABCA-ABCG* and *ABCI*. The members of this family are involved in plant hormone transport, heavy metal ion efflux, plant response to

environmental stress, and transportation of secondary metabolites (Dahuja et al., 2021). Due to the broad roles of ABC transporters, we searched on a pan-genome basis. There were *ABCA* and seed dormancy genes in DB among the "presence" variants. *ABCA* induces the accumulation of triglycerides in seeds, thereby prolonging the seed dormancy period. This partly explains that DB appeared three days later than NJ. Additionally, there were "absence" variants in the DB genome within the *ABCC* subgene family. Whose primary function is multidrug resistance-associated transporters, and which are localized in the tonoplast in pathogenic microbial responses and play important roles in secondary metabolism and can improve NJ stress resistance. These results partly explain the better stress resistance of NJ.

5 Conclusions

The dissection of the asparagus bean genome sequence provides insights into its evolution and cold tolerance and is an important resource for genomic sequence variation analysis. Comparative evolutionary studies suggest that genes involved in the expansion and active selection of cold stress responses and plant-pathogen interactions may facilitate the production of plants with improved environmental adaptation and enhanced agricultural traits. The analysis of the asparagus bean pan-genome provides new insights into the size of the core and dispensable genes and differences between the NJ and DB genomes. These results provide a more comprehensive and important resource for future functional studies of asparagus beans and the development of cold-tolerant legumes.

Data availability statement

The data presented in the study are deposited in the National Center for Biotechnology Information (NCBI) BioProject database, accession number PRJNA869326.

Author contributions

LeL made significant contributions to the conception, data analysis and interpretation. JZ, JX, XL, HT, XS, LZ, XX, LX, PZ and JR participated in critical revisions of important intellectual content and contributed to the previous material; BS, ZH, YT, LiL, GS and YL reviewed and edited the manuscript. HL provided supervision and project management. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Sichuan Vegetable Innovation Team Post Expert Project (Grant Number: sccxtd-2021-05).

Acknowledgments

We would like to thank Biomarker Technologies Co, LTD (http://www.biomarker.com.cn/) for technical services and Editage (www.editage.cn) for English language editing.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1059804/full#supplementary-material

References

Barchi, L., Rabanus-Wallace, M. T., Prohens, J., Toppino, L., Padmarasu, S., Portis, E., et al. (2021). Improved genome assembly and pan-genome provide key insights into eggplant domestication and breeding. *Plant J.* 107, 579–596. doi: 10.1111/tpj.15313

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

Dahuja, A., Kumar, R. R., Sakhare, A., Watts, A., Singh, B., Goswami, S., et al. (2021). Role of ATP-binding cassette transporters in maintaining plant homeostasis under abiotic and biotic stresses. *Physiol. Plant* 171, 785–801. doi: 10.1111/ppl.13302

Dai, S. F., Zhu, X. G., Hutang, G. R., Li, J. Y., Tian, J. Q., Jiang, X. H., et al. (2022). Genome size variation and evolution driven by transposable elements in the *genus oryza*. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.921937

Della Coletta, R., Qiu, Y., Ou, S., Hufford, M. B., and Hirsch, C. N. (2021). How the pan-genome is changing crop genomics and improvement. *Genome Biol.* 22, 1–19. doi: 10.1186/s13059-020-02224-8

Dolatabadian, A., Bayer, P. E., Tirnaz, S., Hurgobin, B., Edwards, D., and Batley, J. (2020). Characterization of disease resistance genes in the *Brassica napus* pangenome reveals significant structural variation. *Plant Biotechnol. J.* 18, 969–982. doi: 10.1111/pbi.13262

Do, T. H. T., Martinoia, E., and Lee, Y. (2018). Functions of ABC transporters in plant growth and development. *Curr. Opin. Plant Biol.* 41, 32–38. doi: 10.1016/j.pbi.2017.08.003

Guo, X., Liu, D., and Chong, K. (2018). Cold signaling in plants: Insights into mechanisms and regulation. *J. Integr. Plant Biol.* 60, 745–756. doi: 10.1111/jipb.12706

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Bio.* 9, R7. doi: 10.1186/gb-2008-9-1-r7

Heidarvand, L., Millar, A. H., and Taylor, N. L. (2017). Responses of the mitochondrial respiratory system to low temperature in plants. *Crit. Rev. Plant Sci.* 36, 217–240. doi: 10.1080/07352689.2017.1375836

Herniter, I. A., Lo, R., Muñoz-Amatriaín, M., Lo, S., Guo, Y. N., Huynh, B. L., et al. (2019). Seed coat pattern QTL and development in cowpea (*Vigna unguiculata* [L.] walp.). *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01346

Huang, H., Tan, H., Xu, D., Tang, Y., Niu., Y., Lai, Y., et al. (2018). High-density genetic map construction and comparative genome analysis in asparagus bean. *Sci. Rep.* 8, 4836. doi: 10.1038/s41598-018-23173-0

Hufnagel, B., Soriano, A., Taylor, J., Divol, F., Kroc, M., Sanders, H., et al. (2021). Pangenome of white lupin provides insights into the diversity of the species. *Plant Biotechnol. J.* 19, 2532–2543. doi: 10.1111/pbi.13678

Lin, K., Zhao, H., Gan, S., and Li, G. (2019). Arabidopsis ELF4-like proteins EFL1 and EFL3 influence flowering time. *Gene* 700, 131–138. doi: 10.1016/j.gene.2019.03.047

Liu, Z., Jia, Y., Ding, Y., Shi, Y., Li, Z., Guo, Y., et al. (2017). Plasma membrane *CRPK1*-mediated phosphorylation of 14–3-3 proteins induces their nuclear import to fine-tune *CBF* signaling during cold response. *Mol. Cell* 66, 117–128.e5. doi: 10.1016/j.molcel.2017.02.016

Liu, D., Ma, C., Hong, W., Huang, L., Liu, M., Liu, H., et al. (2014). Construction and analysis of high-density linkage map using high-throughput sequencing data. *PloS One* 9, e98855. doi: 10.1371/journal.pone.0098855

Liu, W. X., Zhang, F. C., Zhang, W. Z., Song, L. F., Wu, W. H., and Chen, Y. F. (2013). Arabidopsis Di19 functions as a transcription factor and modulates PR1, PR2, and PR5 expression in response to drought stress. *Mol. Plant* 6, 1487–1502. doi: 10.1093/mp/sst031

Li, H., Wang, S., Chai, S., Yang, Z., Zhang, Q., Xin, H., et al. (2022). Graph-based pan-genome reveals structural and sequence variations related to agronomic traits and domestication in cucumber. *Nat. Commun.* 13, 682. doi: 10.1038/s41467-022-28362-0

Li, J., Yuan, D., Wang, P., Wang, Q., Sun, M., Liu, Z., et al. (2021). Cotton pan-genome retrieves the lost sequences and genes during domestication and selection. *Genome Biol.* 22, 119. doi: 10.1186/s13059-021-02351-w

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

Lonardi, S., Muñoz-Amatriaín, M., Liang, Q., Shu, S., Wanamaker, S. I., Lo, S., et al. (2019). The genome of cowpea (*Vigna unguiculata* [L.] walp.). *Plant J.* 98, 767–782. doi: 10.1111/tpj.14349

Meitei, A. L., Bhattacharjee, M., Dhar, S., Chowdhury, N., Sharma, R., Acharjee, S., et al. (2018). Activity of defense related enzymes and gene expression in pigeon pea (*Cajanus cajan*) due to feeding of helicoverpa armigera larvae. *J. Plant Interact.* 13, 231–238. doi: 10.1080/17429145.2018.1466373

Miao, M., Tan, H., Liang, L., Huang, H., Chang, W., Zhang, J., et al. (2022). Comparative transcriptome analysis of cold-tolerant and -sensitive asparagus bean under chilling stress and recovery. *PeerJ* 10, e13167. doi: 10.7717/peerj.13167

Moghaddam, S. M., Oladzad, A., Koh, C., Ramsay, L., Hart, J. P., Mamidi, S., et al. (2021). The tepary bean genome provides insight into evolution and domestication under heat stress. *Nat. Commun.* 12, 2638. doi: 10.1038/s41467-021-22858-x

Pan, L., Yu, X., Shao, J., Liu, Z., Gao, T., Zheng, Y., et al. (2019). Transcriptomic profiling and analysis of differentially expressed genes in asparagus bean (*Vigna unguiculata* ssp. *sesquipedalis*) under salt stress. *PloS One* 14, e0219799. doi: 10.1371/journal.pone.0219799

Pinosio, S., Giacomello, S., Faivre-Rampant, P., Taylor, G., Jorge, V., Paslier, M. C. L., et al. (2016). Characterization of the poplar pan-genome by genome-wide identification of structural variation. *Mol. Biol. Evol.* 33, 2706–2719. doi: 10.1093/molbev/msw161

Qiao, Q., Edger, P. P., Xue, L., Qiong, L., Lu, J., Zhang, Y., et al. (2021). Evolutionary history and pan-genome dynamics of strawberry (*Fragaria* spp.). *Proc. Natl. Acad. Sci. U. S. A.* 118, e2105431118. doi: 10.1073/pnas.2105431118

Qin, P., Lu, H., Du, H., Wang, H., Chen, W., Chen, Z., et al. (2021). Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* 184, 3542–3558. e16. doi: 10.1016/j.cell.2021.04.046

Razzaq, A., Kaur, P., Akhter, N., Wani, S. H., and Saleem, F. (2021). Next-generation breeding strategies for climate-ready crops. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.620420

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. doi: 10.3389/fpls.2021.666342

Springer, N. M., Anderson, S. N., Andorf, C. M., Ahern, K. R., Bai, F., Barad, O., et al. (2018). The maize w22 genome provides a foundation for functional genomics and transposon biology. *Nat. Genet.* 50, 1282–1288. doi: 10.1038/s41588-018-0158-0

Sun, Y., Wang, J., Li, Y., Jiang, B., Wang, X., Xu, W. H., et al. (2022). Pan-genome analysis reveals the abundant gene presence/absence variations among different varieties of melon and their influence on traits. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.835496

Tan, H., Huang, H., Tie, M., Tang, Y., Lai, Y., and Li, H. (2016). Transcriptome profiling of two asparagus bean (*Vigna unguiculata* subsp. *sesquipedalis*) cultivars differing in chilling tolerance under cold stress. *PloS One* 11, e0151105. doi: 10.1371/journal.pone.0151105

Torkamaneh, D., Lemay, M. A., and Belzile, F. (2021). The pan-genome of the cultivated soybean (PanSoy) reveals an extraordinarily conserved gene content. *Plant Biotechnol. J.* 19, 1852–1862. doi: 10.1111/pbi.13600

Vincent, H., Wiersema, J., Kell, S., Fielder, H., Dobbie, S., Castañeda-Álvarez, N. P., et al. (2013). A prioritized crop wild relative inventory to help underpin global food security. *Biol. Conserv.* 167, 265–275. doi: 10.1016/j.biocon.2013.08.011

Wang, P., Abid, M. A., Qanmber, G., Askari, M., Zhou, L., Song, Y., et al. (2022). Photomorphogenesis in plants: The central role of phytochrome interacting factors (PIFs). *Environ. Exp. Bot.* 194, 104704. doi: 10.1016/j.envexpbot.2021.104704

Wang, Y., Tao, Z., Wang, W., Filiault, D., Qiu, C., Wang, C., et al. (2020). Molecular variation in a functionally divergent homolog of FCA regulates flowering time in arabidopsis thaliana. *Nat. Commun.* 11, 5830. doi: 10.1038/s41467-020-19666-0

Wang, W., Wang, X., Zhang, J., Huang, M., Cai, J., Zhou, Q., et al. (2020). Salicylic acid and cold priming induce late-spring freezing tolerance by maintaining cellular redox homeostasis and protecting photosynthetic apparatus in wheat. *Plant Growth Regul.* 90, 109–121. doi: 10.1007/s10725-019-00553-8

Willson, J. (2020). Resolving the roles of structural variants. *Nat. Rev. Genet.* 21, 507. doi: 10.1038/s41576-020-0264-6

Xia, Q., Pan, L., Zhang, R., Ni, X., Wang, Y., Dong, X., et al. (2019). The genome assembly of asparagus bean, vigna unguiculata ssp. *sesquipedialis*. *Sci. Data.* 6, 124. doi: 10.1038/s41597-019-0130-6

Xu, P., Moshelion, M., Wu, X., Halperin, O., Wang, B., Luo, J., et al. (2015). Natural variation and gene regulatory basis for the responses of asparagus beans to soil drought. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00891

Xu, W., Zhang, Q., Yuan, W., Xu, F., Aslam, M. M., Miao, R., et al. (2020). The genome evolution and low-phosphorus adaptation in white lupin. *Nat. Commun.* 11, 1–13. doi: 10.1038/s41467-020-14891-z

Zanini, S. F., Bayer, P. E., Wells, R., Snowdon, R. J., Batley, J., Varshney, R. K., et al. (2022). Pangenomics in crop improvement-from coding structural variations to finding regulatory variants with pangenome graphs. *Plant Genome* 15, e20177. doi: 10.1002/tpg2.20177

Zhao, J., Bayer, P. E., Ruperao, P., Saxena, R. K., Khan, A. W., Golicz, A. A., et al. (2020). Trait associations in the pangenome of pigeon pea (*Cajanus cajan*). *Plant Biotechnol. J.* 18, 1946–1954. doi: 10.1111/pbi.13354

Zhou, P., Silverstein, K. A. T., Ramaraj, T., Guhlin, J., Denny, R., Liu, J., et al. (2017). Exploring structural variation and gene family architecture with *de novo* assemblies of 15 medicago genomes. *BMC Genomics* 18, 261. doi: 10.1186/s12864-017-3654-1

Zhu, J. K. (2016). Abiotic stress signaling and responses in plants. *Cell.* 167, 313–324. doi: 10.1016/j.cell.2016.08.029

Zuluaga, D. L., Lioi, L., Delvento, C., Pavan, S., and Sonnante, G. (2021). Genotyping-by-sequencing in *Vigna* unguiculata landraces and its utility for assessing taxonomic relationships. *Plants (Basel)* 10, 1–13. doi: 10.3390/plants10030509

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 22 December 2022
DOI 10.3389/fpls.2022.1076229

OPEN ACCESS

EDITED BY
Yusuf Khan,
Oslo University Hospital, Norway

REVIEWED BY
Namo Dubey,
Institute of Himalayan Bioresource Technology (CSIR), India
Preeti Arya,
Institute of Microbial Technology (CSIR), India

*CORRESPONDENCE
Ramya Parakkunnel
ramyakurian@gmail.com

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 21 October 2022
ACCEPTED 02 December 2022
PUBLISHED 22 December 2022

CITATION
Parakkunnel R, Naik K B, Vanishree G, C S, Purru S, Bhaskar K U, Bhat KV and Kumar S (2022) Gene fusions, micro-exons and splice variants define stress signaling by AP2/ERF and WRKY transcription factors in the sesame pan-genome.
Front. Plant Sci. 13:1076229.
doi: 10.3389/fpls.2022.1076229

Gene fusions, micro-exons and splice variants define stress signaling by AP2/ERF and WRKY transcription factors in the sesame pan-genome

Ramya Parakkunnel[1]*, Bhojaraja Naik K[1], Girimalla Vanishree[1], Susmita C[2], Supriya Purru[3], Udaya Bhaskar K[1], KV Bhat[4] and Sanjay Kumar[2]

[1]ICAR- Indian Institute of Seed Science, Regional Station, Gandhi Krishi Vigyana Kendra (GKVK) Campus, Bengaluru, India, [2]ICAR- Indian Institute of Seed Science, Mau, Uttar Pradesh, India, [3]ICAR- National Academy of Agricultural Research Management, Hyderabad, Telengana, India, [4]Division of Genomic Resources, ICAR- National Bureau of Plant Genetic Resources, New Delhi, India

Evolutionary dynamics of AP2/ERF and WRKY genes, the major components of defense response were studied extensively in the sesame pan-genome. Massive variation was observed for gene copy numbers, genome location, domain structure, exon-intron structure and protein parameters. In the pan-genome, 63% of AP2/ERF members were devoid of introns whereas >99% of WRKY genes contained multiple introns. AP2 subfamily was found to be micro-exon rich with the adjoining intronic sequences sharing sequence similarity to many stress-responsive and fatty acid metabolism genes. WRKY family included extensive multi-domain gene fusions where the additional domains significantly enhanced gene and exonic sizes as well as gene copy numbers. The fusion genes were found to have roles in acquired immunity, stress response, cell and membrane integrity as well as ROS signaling. The individual genomes shared extensive synteny and collinearity although ecological adaptation was evident among the Chinese and Indian accessions. Significant positive selection effects were noticed for both micro-exon and multi-domain genes. Splice variants with changes in acceptor, donor and branch sites were common and 6-7 splice variants were detected per gene. The study ascertained vital roles of lipid metabolism and chlorophyll biosynthesis in the defense response and stress signaling pathways. 60% of the studied genes localized in the nucleus while 20% preferred chloroplast. Unique cis-element distribution was noticed in the upstream promoter region with MYB and STRE in WRKY genes while MYC was present in the AP2/ERF genes. Intron-less genes exhibited great diversity in the promoter sequences wherein the predominance of dosage effect indicated variable gene expression levels. Mimicking the NBS-LRR genes, a chloroplast localized WRKY gene, Swetha_24868, with additional domains of chorismate mutase, cAMP and voltage-dependent potassium channel was found to act as a master regulator of defense signaling, triggering immunity and reducing ROS levels.

KEYWORDS

AP2/ERF, defense, duplication, evolution, gene fusion, micro-exon, sesame, WRKY

1 Introduction

Transcription Factors (TF) are an important class of genes involved in the regulation of plant response under many biotic and abiotic stress conditions. APETALA2/ETHYLENE RESPONSIVE FACTOR (AP2/ERF) and WRKY genes are major components of complex regulatory networks in plants during developmental processes and defense responses (Abdullah-Zawawi et al., 2021; Li et al., 2021). The AP2/ERF transcription factors contain a conserved AP2/ERF domain of about 60 to 70 amino acids, and consist of five subfamilies, AP2, ERF, DREB (Dehydration Responsive Element-Binding), RAV (Related to ABI3/VP1) and Soloist based on the number of AP2/ERF domains and the presence of other DNA binding domains (Dossa et al., 2016). The differential expression of AP2/ERF genes under multiple stresses of heat, drought, cold and salinity has been characterized in wheat (Riaz et al., 2021), *Brassica napus* (Ghorbani et al., 2020), pear (Li et al., 2018) and sesame (Dossa et al., 2016) while secondary metabolite biosynthesis was studied in eggplant (Li et al., 2021). WRKY TF family is the seventh largest and contains the signature domain of 60-70 amino acids representing WRKYGQK/WRKYGKK at the N-terminus and a Zn-finger domain at the C-terminus (Yang et al., 2017). Genome-wide characterization of WRKY genes has been reported in soybean (Yang et al., 2017) against cyst nematode, sorghum (Baillo et al., 2020) against multiple stress responses, the biotic and abiotic stress response in sunflower (Liu et al., 2020), the abiotic stress response in apple (Qin et al., 2022) and sesame (Li et al., 2017).

Sesame, (*Sesamum indicum* L.) belonging to the family Pedaliaceae is an ancient oilseed crop cultivated in the tropical and sub-tropical regions of the world by poor and marginal farmers. Majority of the wild species of the genus *Sesamum* are native to sub-Saharan Africa however, domestication happened in India (Bedigian, 2003). Recently, a sesame pan-genome assembly of 554.05Mb comprising modern cultivars and landraces was developed including 26472 orthologous gene clusters (Yu et al., 2019). In order to exploit the full potential of genetic diversity present in the germplasm of the crop plants, trait-based investigations in the different cultivars of the same crop, adapted to widely different agro ecological conditions are imperative. In this context, the pan-genome offers a viable alternative presenting researchers with useful genetic variation in a number of component genomes as against a single reference genome. Particularly in crops like sesame where domestication syndrome is evident in the genome for many useful traits, the constructed pan-genomes become a valuable tool facilitating researchers in mining natural variation for molecular breeding (Yu et al., 2019). In addition, the fine dissection of homologs and paralogs at exonic, intronic and promoter sequence levels attune evolutionary studies with limitless possibilities. In the present study, accelerated evolution under multiple stress conditions is discussed in cultivars adapted to wider climatic niches and parts of the sesame pan-genome. The homologs for AP2/ERF and WRKY genes are studied in relation to evolutionary adaptations, gene duplications, gene fusions, variations in cis-element architecture and variations in splicing machinery involved in defense response and development.

2 Materials and methods

2.1 Identification of AP2/ERF and WRKY genes from the sesame pan-genome

The pan-genome assembly include *S. indicum* var Zhongzhi-13, *S. indicum* var Yuzhi-11, *S. indicum* var Baizhima, *S. indicum* var Mishouzhima (all from China) and the Indian variety *S. indicum* var Swetha. Here after the component genomes will be referred to as Zhongzhi-13, Yuzhi-11, Baizhima, Mishouzhima and Swetha respectively. From the sesame pan-genome (Yu et al., 2019); the CDS, protein and gff files were used to identify sequences corresponding to Pfam ids PF00847 and PF03106 representing AP2/ERF and WRKY genes. The 'gff' files were processed with excel and exon-intron size was identified. The genes were mapped onto the chromosome using 'gene location visualize' tool from TB tools (Chen et al., 2020). The exon/intron structures were determined by the gene structure display server (Hu et al., 2015), (http://gsds.cbi.pku.edu.cn/). The different domains were categorized by SMART (Letunic and Bork, 2018) tool (http://smart.embl-heidelberg.de/). The conserved domains in the sesame CDS were identified using NCBI-CDD database search tool (Marchler-Bauer et al., 2017). Protein parameters were worked out using 'ProtParam' tool (https://web.expasy.org/protparam/). The exon, intron, and micro-exon distribution for AP2/ERF and WRKY genes from sesame pan-genome was visualized with an online version of CIRCOS available at (http://mkweb.bcgsc.ca/tableviewer/). The package 'ggplot2' was used for the visualization of all other data in R.

2.2 Phylogenetic analysis

The initial phylogenetic analysis of Swetha protein sequences, Arabidopsis and rice was carried out through NJ algorithm in MEGA X (Kumar et al., 2018b) using the Jones-Taylor-Thornton distance matrix with 500 bootstrap replications. Multiple sequence alignment was done using CLUSTAL X ver. 2.1. Arabidopsis and rice homologs were identified from The Arabidopsis Information Resource (TAIR) available at https://www.arabidopsis.org/ and Plant Transcription Factor Data Base (PlantTFDB ver.5.0) available at http://planttfdb.gao-lab.org/ respectively. The reported classification of Arabidopsis and rice was used for classifying

Swetha AP2/ERF and WRKY genes. This classification was further extended to the sesame pan-genome.

The pan-genome protein sequences were aligned using CLUSTAL X ver. 2.1 and were subjected to Bayesian phylogenetic inference using MCMC by BEAST ver. 2.6.6 (Bouckaert et al., 2019). The input 'XML' files were generated using BEAUti interface (Drummond et al., 2012) with the model 'GTR+I+G' and the 'Yule speciation process' under a strict clock model. Two independent runs of 10000000 generations of MCMC chains were produced and sampled after every 5000 generations. TRACER ver1.7.1 (Rambaut et al., 2018) was used for combining the files and the plotted posterior estimates were inspected. The first 10,000 trees were discarded as burn-in, and the rest of the samples were summarized in a maximum clade credibility tree using TreeAnnotator ver. 2.6.6 with a posterior probability limit of 0.5. Means and 95% higher posterior densities (HPDs) obtained from the combined output of TRACER were used for the construction of trees using FigTree ver.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) with median heights.

2.3 Synteny and collinearity

Based on phylogeny genes were ordered as exon-intron sequences and the micro-exon sequences were identified. The 200bp upstream and downstream region including the micro-exon was analyzed for the presence of protein-coding domains using BLASTX search. The exon-intron size was estimated in excel. Based on BEAST phylogeny, gene pair files were created and used to calculate the non-synonymous/synonymous (Ka/Ks) mutation ratio with TB tools from the respective CDS, protein and gff data (Chen et al., 2020). *Arabidopsis thaliana* and *Oryza sativa* ssp *indica* genomes were downloaded from the Phytozome (https://phytozome-next.jgi.doe.gov/). The syntenic relationships between sesame, Arabidopsis and rice genomes were probed with MCScanX using TB tools. Based on the results of MCScanX sesame genes were classified as WGD or segmental duplicates. The evolutionary time in million years ago (MYA) for each orthologous pair was calculated using the formula, T= Ks/2r (Moghaddam et al., 2021); where 'r' the rate of mutation was kept as $1.5x10^{-9}$ based on the age of divergence of Zhongzhi-13 and Swetha genomes (Yu et al., 2019).

2.4 Alternative splicing and protein-protein interaction

The multi-exon homologs were probed for intron-exon size variation and such pairs were selected for identification of splice site and SRP protein site mutations (Kharabian, 2010; Karlik, 2021) through ESEfinder2.0 (http://krainer01.cshl.edu/tools/ESE2/). The cut-off for splice donor and splice acceptor sites was kept at 6.9 and for branch site was 2.0. The splice SRP protein sequences (SF1 & SF2) of ESE finder was used as a rough guideline for prediction and the identified sites were compared among the component genomes for probable mutations. The deviation in position and score of splice sites and SRP proteins were noted for protein homologs and the splicing mechanism was devised based on a comparison with the exon-intron data. The conserved motifs in the sesame proteins were identified using the MEME program (https://meme-suite.org/meme/tools/meme) using parameters: maximum number of motifs = 10; optimum width of motifs = 15–50. The identified motifs were subjected to 'GoMo' scan to identify 'GO' terms associated with the biological function (http://meme-suite.org/tools/gomo). Protein-protein interaction network was visualized with the help of STRING ver. 11.5 (https://string-db.org/) and plotted with the help of Cytoscape ver. 3.9.1. Prediction of protein sub-cellular localization was done with the help of WoLF PSORT tool (https://wolfpsort.hgc.jp/). The cis-element identification was done by subjecting upstream 2000bp from the start codon of selected AP2/ERF and WRKY sequences from Swetha and Zhongzhi13 genomes to PLANT CARE (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) and comparing with the reported Arabidopsis cis-elements.

2.5 Expression profiles of AP2/ERF and WRKY genes

Microarray data of AP2/ERF and WRKY genes were obtained from NCBI-Gene Expression Omnibus (GEO) database under the accession numbers GSE81039, GSE102714, GSE81325, GSE49418, GSE55835 and GSE167174. The data were properly grouped as per study objectives and was analyzed through GEO2R. After processing the transcriptome data, heat maps were constructed in R using the adjusted P-values for AP2/ERF and WRKY genes having significant logFC or F-statistics (more than two groups defined) for each accession.

3 Results

3.1 Identification of AP2/ERF and WRKY genes from sesame pan-genome

A total of 704 AP2/ERF genes and 387 WRKY genes were identified in the sesame pan-genome (Table 1). The lowest number of AP2/ERF genes was observed in the Yuzhi-11 genome (131) whereas the genomes of the Chinese landrace 'Mishouzhima' and the Indian variety 'Swetha' contained 145 each. As for WRKY genes, the Chinese cultivar (Yuzhi-11) and the landrace (Baizhima) contained 73 genes each whereas Swetha contained 89 genes.

TABLE 1 Summary statistics of identified AP2/ERF and WRKY genes in sesame pan-genome.

Genome	Genes	Intron	Exon	Micro_Ex	Exon/ Gene	Exon/ Intron	Micro_ex/ Gene	Micro_ex/ Exon	Intron/ Gene
AP2/ERF									
Baizhima	144	210	354	41	2.458333	1.685714	0.284722	0.115819	1.458333
Mishouzhima	145	208	353	34	2.434483	1.697115	0.234483	0.096317	1.434483
Swetha	145	356	501	51	3.455172	1.407303	0.351724	0.101796	2.455172
Yuzhi11	131	196	327	32	2.496183	1.668367	0.244275	0.097859	1.496183
Zhongzhi13	139	213	352	37	2.532374	1.652582	0.266187	0.105114	1.532374
Total	704	1183	1887	195	2.680398	1.595097	0.276989	0.103339	1.680398
WRKY									
Baizhima	73	221	294	4	4.027397	1.330317	0.054795	0.013605	3.027397
Mishouzhima	76	231	307	6	4.039474	1.329004	0.078947	0.019544	3.039474
Swetha	89	355	444	12	4.988764	1.250704	0.134831	0.027027	3.988764
Yuzhi11	73	216	289	4	3.958904	1.337963	0.054795	0.013841	2.958904
Zhongzhi13	76	221	297	6	3.907895	1.343891	0.078947	0.020202	2.907895
Total	387	1244	1631	32	4.21447	1.311093	0.082687	0.01962	3.21447

The numbers of genes, exons, introns and micro_exons are given for both the families for all the five component genomes.

3.1.1 Chromosomal location of AP2 and WRKY genes

The AP2/ERF and WRKY genes were distributed all over the 13 chromosomes with variations in individual genomes. Maximum AP2/ERF genes were located on chr-1 and chr-6 while the least numbers were observed on chr-5 and chr-11. For WRKY genes, chr-6 had the highest number including 10-13 genes from individual genomes whereas chr-5 contained a single gene in all the genomes. Moreover, 82 AP2/ERF and 39 WRKY genes were not mapped to any chromosome. Details are in Figure 1A, B), SI-1A , SI-2A while SI-18 gives the chromosomal location of mapped genes.

3.2 Phylogenetic analysis of AP2/ERF and WRKY genes

The phylogenetic analysis of AP2/ERF and WRKY genes of sesame was conducted using the multiple sequence alignment results of 'Swetha' protein sequences along with Arabidopsis homologs. Bayesian phylogeny trees were constructed for each gene family and the individual members were classified based on already published Arabidopsis gene classification. Afterward, the newly defined classifications of Swetha proteins were extended to the whole of the sesame pan-genome. The 145 AP2/ERF genes identified in Swetha genome were further classified as belonging to different subfamilies of DREB, ERF, and AP2. The ERF subfamily had the maximum share (70), followed by DREB (32), AP2 (31), RAV (9) and Soloist (3). The ERF family was further classified as different groups B1-B6 and contained 16, 7, 22, 7, 5 and13 genes respectively. The DREB subfamily contained 5 groups A1, A2, A4, A5 and A6 including 3, 4, 13, 9 and 4 members respectively. The WRKY genes were also classified as belonging to subclasses T1, T2 and T3 based on the number of WRKY domains and the type of zinc-finger motif present. Among the 89 genes present in Swetha T2 had the highest share (66), followed by T1 (19) while T3 contained only 4 genes. The details of gene classification in 'Swetha' are given in Figures 2A, B), SI-1B, SI-2B.

The ERF, AP2, RAV and Soloist domains retained similarity all through the pan-genome with frequent domain changes noticed in closely related clusters of proteins. The ERF-B3 domain exhibited sequence similarity to ERF-B1, AP2, Soloist, DREB-A5 and DREB-A4 domain proteins. The A4 domain genes showed sequence similarity to ERF-B4 and ERF-B1 genes along with AP2 genes. DREB-A2 domain genes in turn were found to be related to ERF-B6 and ERF-B1 domain genes. Among the DREB subfamily, A1 was more conserved where group-specific clustering was observed. The frequent domain changes, segmental duplication and exonic changes made the phylogeny reconstruction quite tedious in AP2/ERF family. However, for WRKY genes sequence conservation was noticed all through the pan-genome. Among the 387 WRKY genes identified the T1, T2 and T3 groups had 80, 285 and 22 genes respectively. T3 genes in the pan-genome shared sequence similarity with T2D and T2A whereas Swetha T3 genes were more related to T2D and T2E groups. One set of T3 genes was found solely in the Chinese accessions. T1 genes of the pan-

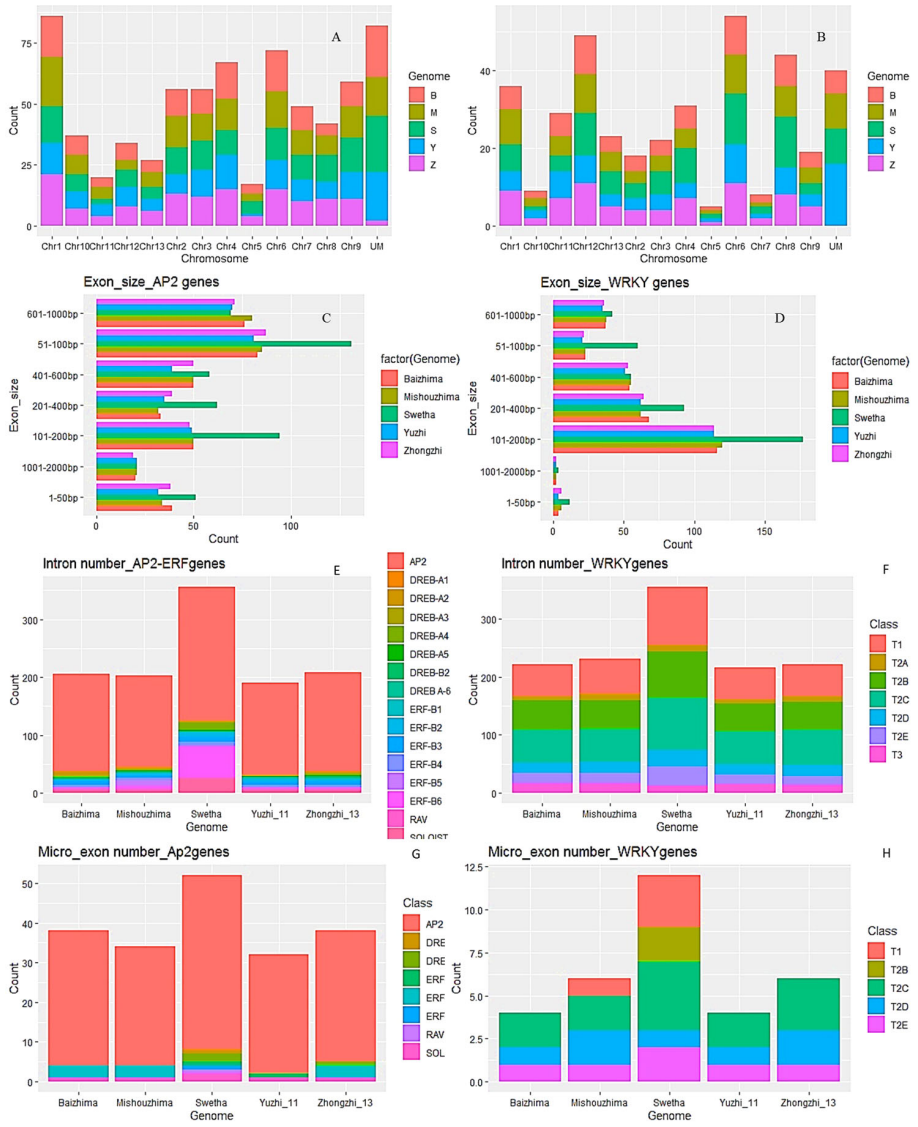


FIGURE 1
Salient features of AP2/ERF and WRKY genes in the sesame pan-genome. Chromosomal distribution **(A, B)** while exon size of AP2/ERF and WRKY genes in individual genomes **(C–F)** represent intron numbers; **(G, H)** represent micro-exon numbers of AP2/ERF and WRKY genes in individual genomes as per individual genome, subfamily and class wise.

genome were found more related to T2E genes whereas the Swetha genes were closer to the T2C genes. Details are given as SI-19 and SI- 20.

3.3 Gene structure of homologous genes in pan-genome

Based on multiple sequence alignment and phylogeny, the homologs were identified for AP2/ERF and WRKY genes from individual genomes. The exon-intron structure and sizes of exons and introns were identified for each homologous set (SI-4 and SI-6).

3.3.1 Intron number and size variants

In the AP2/ERF gene family, 443 genes were found to be devoid of introns (Figure 1E). Mishouzhima (99) had the largest while Swetha (77) had the least number of intron-less genes. 68 genes of Swetha contained introns while for Mishouzhima and Yuzhi genomes, only 47 genes had introns. 36 introns were present in the gene Swetha_28474, whereas Swetha_02835 had 20 introns. The mRNAs of these genes spanned 30 kb and 17kb

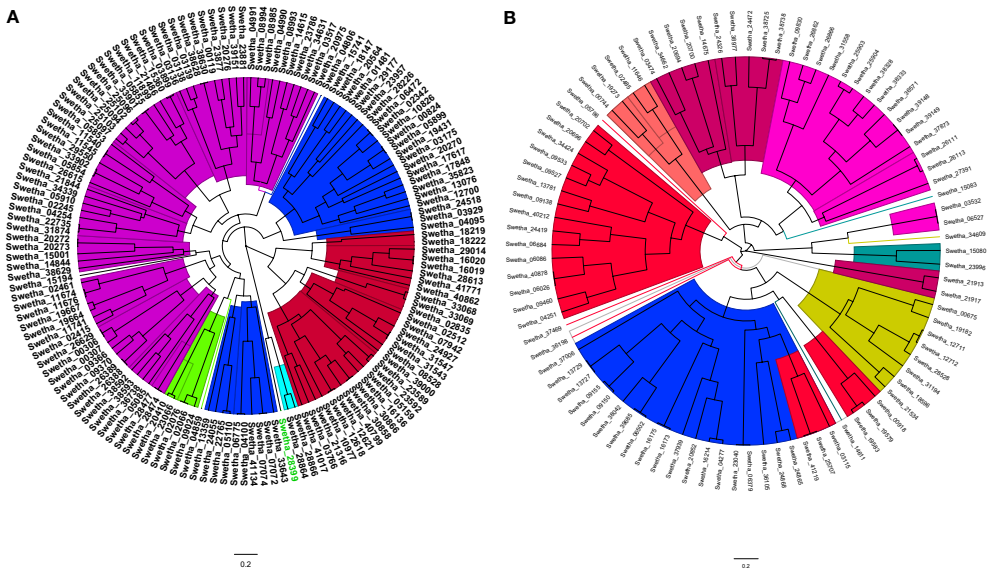


FIGURE 2
The MCMC phylogeny tree of **(A)** AP2/ERF and **(B)** WRKY families in the Swetha genome based on Arabidopsis and rice classification. In **(A)** the colour codes are as follows: Blue=DREB; Magenta=ERF; Red=AP2; Green=RAV and Cyan=SOLOIST. In **(B)** the different subfamilies are as follows; T1=Blue; T2A= Orange; T2B=Magenta, T2C=Red; T2D=Lime yellow; T2E= Purple and T3= Olive.

respectively. In the AP2/ERF family, 92 genes had single intron, 36 genes had 7 introns, 32 genes had 5 introns while 3 and 4 intron genes were less frequent. The intron size varied from 34bp in Mishuozhima_01646 (SI-4) located in chromosome-1 to 43.6kb in Swetha_11741 in chromosome-3 (SI-4). Two other genes (Swetha_11742 and Swetha_11743) were found nested in this huge intron coding for Alpha-amylase inhibitor and phospholipase D respectively. Another intron of size 28kb was observed in Yuzhi11 _12343 in chromosome-8. However, this large intronic region did not harbor any additional genes.

In the WRKY family, except for T2C genes, Swetha_15083 and Baizhima_02279; all the others had introns (Figure 1F). The largest number of introns noticed in a single gene was 17 in Swetha_24868 belonging to the T1 group while Swetha_09533 had 14 introns. All the genomes shared a common gene with 11 introns belonging to T2C. The smallest intron noticed was of size 31bp and was present in Zhongzhi13_04758 and its homologs in Baizhima, Mishouzhima and Yuzhi11. This gene present in Chinese accessions is worth noticing for its sequence conservation and exonic as well as intronic number and size conservation. The largest intron noticed was 35kb in the gene Swetha_09138 in chromosome-2 and the mRNA spanned 38kb in length. This huge intronic region was found to harbor two additional genes namely Swetha_09139 and Swetha_09140, coding for AB hydrolase1 and pentatricopeptide repeat-containing protein respectively. Another Swetha gene, Swetha_02485 in chromosome-1 also harbored a huge intron of size ~19kb. A Methyladenine glycosylase gene (Swetha_02486) was found nested in this intronic region. 153 WRKY genes from the sesame pan-genome contained 2 introns, while 79 had 4 and 72 had 3 introns. The variation in intron numbers in the individual genomes is represented in (SI-1B, SI-1C, SI-2B and SI-2C).

3.3.2 Exon number and size variants

The individual genomes differed greatly in exon number and size. In the AP2/ERF family, the total number of exons detected varied widely although gene number was comparable. Swetha genome contained a large number of multi-exon genes with total exons of 486 against 351 and 352 in Baizhima and Mishouzhima respectively. The number of exons in Zhongzhi-13 is 373 whereas 327 exons were found in Yuzhi-11. 51-100bp exons were most common while exons of size >1kb were least common. In the Swetha genome, 101-200bp exons were the second most common against 601-1000bp exons in all the Chinese accessions. The smallest exon noticed was of 3bp present in all the genomes while the largest exon was 1262bp in Yuzhi11 genome.

In the WRKY family, 443 exons were detected in Swetha against 297 in Zhongzhi-13. 101-200bp size exons were most common in all the genomes followed by 201-400bp and 401-600bp exons. The smallest exon detected was of 3bp (Swetha_06086) in Swetha genome, whereas among WRKY genes from the Chinese accessions the 3 bp exon was observed only in Mishuozhima_16023. The largest exon detected was of 1287bp, present as a single exon gene conserved in the pan-

genome. Details of exon number and size distribution are given in Figure 1C, D, Table 1, SI-1C, 1D. In addition, a number of gene duplication events were found unique among WRKY genes in the Swetha genome resulting in increased gene copy numbers. The duplicated gene was found positioned in the same chromosome with a different location or in a different chromosome. These genes differed in intron number (Swetha_20694 and Swetha_20700), exon size (Swetha_21913 and Swetha_21917; Swetha_03532 and Swetha_06527), conversion of exon into micro-exon (Swetha_38738 and Swetha_38725), transposon induced insertion or deletion (Swetha_21534, Swetha_18596 and Swetha_00675) to name a few. Although such duplicates occurred in the AP2-ERF family also, the genome-wise distribution was more or less equal. The details are given in SI-1A, SI-2A, SI-4, 5, 6, 7 and 8.

3.3.3 Micro-exons in sesame pan-genome

The exonic fragments of length <51bp were classified as micro-exons (Song et al., 2020) and we found 227 micro-exons in the sesame pan-genome varying in size from 3bp to 50 bp (SI-3A, SI-3B, SI-4 and SI-6). AP2/ERF gene family had 195 micro-exons distributed along 133 genes with Swetha contributing a major share of 36. Baizhima had 26 micro-exon genes whereas the numbers in Mishouzhima, Yuzhi11 and Zhongzhi-13 were 25, 21 and 25 respectively. Micro-exon genes were present in all subfamilies in Swetha genome Swetha had a total of 51 micro-exons with multiple micro-exons noticed in many genes. The gene Swetha_18222 had 5 micro exons out of the total 11 exons and had the largest micro exon count for a single gene. The 6th exon was the most preferred position for micro-exons whereas after the 10th exon the presence of micro-exons becomes very rare. Among the micro-exon containing genes, 38 genes (6 sets) including duplicates were found to have exonic and intronic sequence conservation across the pan-genome while 7 sets (28 genes) were found to have sequence conservation across 4 genomes. 8 sets (40 genes) were found to have exonic divergence while retaining the micro-exon conservancy with the change noted particularly in the 1st or the last exon. The bulk of micro-exons (171) were contributed by the AP2 subfamily while presence was noticed in DREB-A2, DREB-A4, ERF-B1, ERF-B3, ERF-B6, RAV and SOLOIST families. 32 micro-exons were noticed in the WRKY family with a major share (12) contributed by Swetha. These were distributed into T1 (2 genes), T2B (1 gene with 2 micro-exons), T2C (5 genes), T2D (1 gene) and T2E (2 genes). The T2 WRKY genes of Chinese accessions contained 19 micro-exons while class T1 had a single micro-exon. Details are in Figure 1G, H), SI-3A, SI-3B. The micro-exonic region and the adjoining intronic sequences were found to share sequence similarity to many functional domains and genes such as glycosyltransferase, phospholipase (LCAT3), pectate lyase, ribonuclease3, ASGR-BBM like2, asparagine synthase(common in all the genomes), chromatin modification-related protein EAF-1, G-protein coupled receptor1, PAS domain S-box containing protein, TonB dependent receptor, transmembrane helix (common), aquaporin, integrase as well as transposon 'Tpn104'. Details are given in Table 2 while SI-8 represents transposon distribution in coding sequences.

3.4 Protein diversity of AP2/ERF and WRKY homologs

The shortest protein observed in the AP2/ERF family was of length105AA and belonged to the ERF sub-family gene, Swetha_03899 which also had the lowest molecular weight (11.8kDa). The largest protein among AP2/ERF family also belonged to another ERF gene, Swetha_28474 with1980 AA and a molecular weight of 220.5kDa. Among the AP2 sub-family, Swetha_02835 measured a length and weight of 1937AA and 217.2kDa respectively. A lot of protein variants were found among homologs, where although the protein length was conserved; AA mutations in the functional domains gave way to altered protein parameters. Based on the occurrence of similar length homologs in all or at least 4 genomes the proteins were classified as all same (5 genomes), all different, 4 same and 4 different. We found 36 genes to have all the parameters conserved across all the 5 gnomes whereas conservation in 4 genomes was noticed for 29 genes. Even with the same protein length, significant variation was observed in the molecular weight and iso-electric point and was categorized as differing for all 5 cases (17 genes) or all 4 cases (13 genes). Among WRKY genes, the shortest protein was of length 129AA and was present in all the Chinese accessions homologous to Zhongzhi13_22905 and belonged to the T1 group. Another T1 gene, Swetha_24868 recorded the largest and heaviest WRKY protein with a length and weight of 1261AA and 142kb respectively. The second largest protein was common in all five genomes and belonged to T2C with 1141 AA and 11 introns, homologous to Zhongzhi13_28134. In the WRKY subfamily, the Chinese accessions showed a greater conservation pattern with regard to length and other protein parameters. We found 10 genes to have all the protein parameters conserved in all 5 genomes whereas 4 similar homologs were found in 22 genes majorly including Chinese accessions. Details in Figure 3D, E, SI-1A, SI-2A. The diversity of pan-genome is represented as circos plot (Figure 3F).

3.5 Multi-domain genes in sesame pan-genome

In addition to the main AP2/ERF and WRKY domains, we found additional domains in 55 genes in the sesame pan-genome possibly as a result of gene fusion. WRKY family had 39 multi-domain genes whereas AP2/ERF had 16 such genes. Maximum cases of multi-domain genes were noticed in the Swetha genome

TABLE 2 The functional domains identified in the 200bp upstream and downstream region including the micro-exon in the sesame pan-genome. The reported functions and the references are also given.

Sl. No	Domain name	Function	Reference
1	glycosyl transferase	1) glucosylation of lignans in sesame seed 2) maintenance of cell membrane integrity during abiotic stress	Ono et al., 2020; Shi et al., 2020
2	lecithin: cholesterol acyltransferase (LCAT3)	1) Lipid metabolism and production of specialized fatty acids 2) Defense response against *Podosphaera xanthii* in cucumber	Xu et al., 2020; Ming et al., 2022
3	Pectate lyase	1) pectin degradation 2) plant defense response and apoptopsis 3) ROS accumulation	Uluisik and Seymour, 2020; Chen et al., 2021; He et al., 2021
4	Ribonuclease3	1)RNA maturation, modification and splicing 2) antiviral defense	Olmedo and Guzman, 2008; Aguado and tenOever, 2018
5	ASGR-BBM like	1)Apomixis	Worthington et al., 2019
6	asparagine synthase	1)Multiple nutrient stress response 2) drought and nutritional stress in wheat	Curtis et al., 2018
7	EAF-1	1)Important component of chromatin remodeling complex NuA4 in Arabidopsis 2) Regulation of plant stress response	Wang et al., 2019;
8	G-protein coupled receptor1	1)Multiple abiotic stresses like salinity, drought, extreme temperature and high light intensity	Wu and Urano, 2018
9	PAS domain S-box containing protein	1) Circardian clock 2) Ecological adaptation to diverse stress stimulii	Vogt and Schippers, 2015; Tischkau, 2020.
10	TonB receptor	1)Metal resistance	Theriault and Nkongolo, 2017
11	Aquaporin	1)Stress tolerance and seed germination 2)Plant-pathogen interaction	Liu et al., 2013
12	Integrase	1) Retrotransposon introgression	Suh, 2021

including 20 WRKY and 12 AP2/ERF genes. The multi-domain genes found in genomes of Baizhima, Mishouzhima, Yuzhi11 and Zhongzhi-13 were 6, 6, 6 and 4 respectively. Details are given in Table 3 and Figure 3F. Among the WRKY genes, 4 sets of multi-domain genes were common and were present in all the genomes. In addition to the WRKY domain, these contained additional domains like Arginine/lysine/ornithine decarboxylase, ATP-dependent metalloprotease FtsH, eukaryotic translation initiation factor 5A, DUF3084 and Lung-7-transmembrane receptor. Moreover end to end fusion of genes resulting in multiple functional AP2/ERF and WRKY domains was also common and up to 4 copies of the active domain were noticed for both the families. Other common domains associated were mostly enzymes like kinases, reverse transcriptases, hydrolases, peroxidases, carboxylases, methyl transferases, etc. The inclusion of additional domains resulted in larger-sized genomes with added exons which completely altered gene structure and splicing mechanism. Fusion genes were noticed as novel genes in a single genome or were present in multiple genomes. Based on the location of parental and fusion genes in the genome a detailed classification was made and given in Figure 4G.

3.6 Alternative splicing of AP2/ERF and WRKY genes

The occurrence of multi-domain genes, protein variants and variable exon-intron structures of homologs in the sesame pan-genome prompted a thorough study of splice junctions to identify the splice variants in the pan-genome. After comparing exon-intron structure and the splice junction scores between the most common homolog and the identified variables we categorized the splice variants into different alternative splicing events. (SI-5, SI-7). We found the occurrence of the following splice events in the pan-genome namely intron gain or loss, alternative exon ends, alternative 5' and 3' ends, mutually exclusive exons, exon skipping and intron retention. We found splice variants in 70 and 52 genes of AP2/ERF and WRKY families respectively (Figure 5A). Multiple AS events were noticed in many cases and in AP2/ERF gene family intron gain or loss was most common followed by alternative 5' and alternative 3' events. The least common was intron retention and mutually exclusive introns. In the WRKY family, alternative 3' followed by alternative 5' ends were the most preferred splice

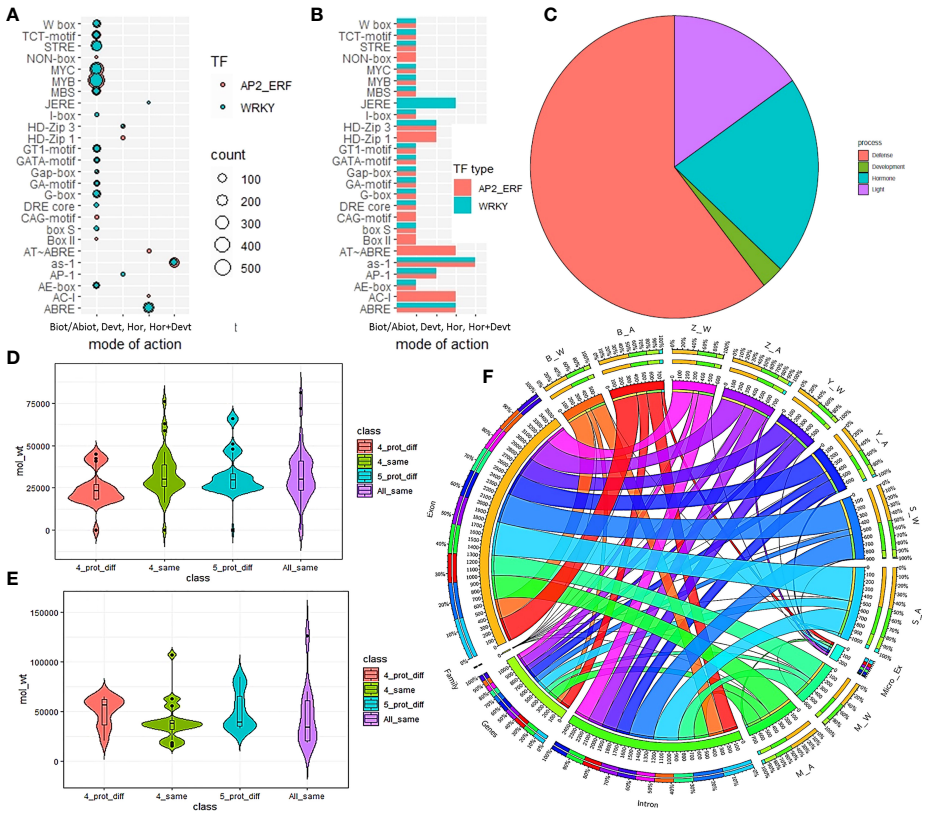


FIGURE 3
Distribution of cis-elements and protein parameters in the sesame pan-genome for AP2/ERF and WRKY genes. **(A, B)** represent the cis-element distribution and mode of action in the major categories of biotic/abiotic stress response, development, hormone as well as hormone and development. The pie chart **(C)** indicates the number of promoters involved in different functions. The difference in protein parameters in AP2/ERF and WRKY genes are given in **(D, E)** while **(F)** gives the CIRCOS plot for pan-genome.

variant. Here also intron retention and mutually exclusive introns were less common. (Figure 5B).

With respect to each of the variants, a corresponding change in both position and score was noticed in the splice acceptor site, splice donor site, branch site as well as different splicing enhancers. Although characterized majorly in human alternative splicing scenario, SF2 and SF1 are reported to have major roles in plants especially during salinity and irradiation stress (Stankovic et al., 2016; Jin, 2022). Among the AP2/ERF genes of sesame, splice donor site mutations were majorly reflected in alternate 5' ends, intron gain and exon skipping events in the genome. Spice acceptor site mutations drastically altered 3' ends added by intron gain and exon skipping. Branch site mutations complemented most of the AS events although majorly reflected in intron gain. Mutation in all of the splice enhancer element positions and sequences supported different AS events. Changes in SF1 majorly affected intron gain or loss whereas SF2 mutations helped in modifying 5' and 3' ends of exons in addition to intron gain. The mutations in SF2 favored alternate 5' ends. Alternate exons and mutually exclusive exons involved mutations in splice donor, acceptor and the branch sites (Figure 5C). In the WRKY family, the change in splice donor site was reflected in alternate 5'end, intron gain and exon skipping. The splice acceptor site change was noticeable in alternate 3' ends, intron gain and exon skipping. The branch site changes were observed in all the splice events recorded. Intron gain was associated with SF1and SF2 whereas alternative exon ends, exon skipping and alternative 5' ends were majorly associated with SF2. Mutually exclusive exons were comparatively less in the WRKY family and were associated with changes in splice acceptor, branch site and SF1 (Figure 5D).

3.7 Synteny and collinearity in sesame pan-genome

The comparison of the 'Swetha' and arabidopsis genomes resulted in the identification of 981 syntenic blocks involving 18614 genes with a collinearity percentage of 26.78 (Figure 4B). With rice, the collinearity percentage was 7.08 and 5947 collinear

TABLE 3 Multi-domain genes identified in the sesame pan-genome evolved through gene fusion.

Gene	Introns	Domains	Reported functions	Reference
Swetha_00824	4	AP2+exostosin	Defense, Salt stress, endomembrane organisation	Li et al., 2013;
Yuzhi11_06305	1	AP2+Alpha-Amylase Inhibitors (AAI), Lipid Transfer (LT) and Seed Storage (SS) Protein	Insect resistance; Abiotic stress response	Kumar et al., 2018a; Karray et al., 2022; Ming et al., 2022
Swetha_11540	4	40S ribosomal protein S15a, DNA translocase FtsK, AP2	Growth regulator; Chromosome segregation, Cell division; Oxidative stress	Mishra et al., 2022
Swetha_02245	5	Translation initiation factor 2B subunit, eIF-2B alpha/beta/delta family+ AP2	Plant virus resistance	Shopan et al., 2017
Swetha_02461	11	AP2+plasma-membrane proton-efflux P-type ATPase	Stomatal opening; Stress response	Ren et al., 2021; Mishra et al., 2022
Mishuozhima_25500	9	AP2+mito chondrial carrier prot (3 no.)	Stress recovery; Osmotic stress response	Monne et al; 2019
Swetha_28474	36	AP2+ Importin repeats (4 sets)+ HEAT like repeat+ Karyopherin (importin) beta	Autoimmunity; Pathogen response; Abiotic stress	Xu et al., 2020; Ludke et al., 2021
Baizhima_11916	5	AP2+ zinc-binding in reverse transcriptase	Adaptive evolution; Stress response	Galván-Gordillo et al., 2016; Lanciano and Mirouze,2018
Swetha_41017	8	AP2+AP2+Solute carrier families 5 and 6-like	Amino acid transport; Stress response	Hrmova and Gilliham, 2018
Yuzhi11_07044	16	alpha/beta hydrolases;+AP2+Ap2	Pathogenecity; Plant immune responses; Defense	Mindrebo et al., 2016; Jiao and Peng, 2018
Swetha_16136	9	alpha/beta hydrolases;+AP2+Ap2	Pathogenecity; Plant immune responses; Defense	Mindrebo et al., 2016; Jiao and Peng, 2018
Swetha_07942	11	AP2(4 dom)		
Swetha_02835	20	Protein FAR-RED ELONGATED HYPOCOTYL 3 (2 dom)+ FAR1 DNA-binding domain (Zn binding-2 dom)+ AP2	Negative regulation of carbon starvation and leaf senescence	Ma and Li, 2021; Tian et al., 2021
Swetha_18222	10	AP2+PWWP domain	Chromatin methylation reader; Stress response	Kenzior and Folk, 2015; Rona et al., 2016
Swetha_28399	14	Helicase+AP2	Plant stress response	Raikwar et al., 2015; Pandey et al., 2020
Swetha_28866	6	AP2+Tim17	Germination and Stress response	Chaudhuri et al., 2020
Swetha_03474	4	2 WRKY		
Swetha_14675	10	suppressor of G2 allele of SKP1+WRKY	Defense signalling and plant immunity	Yu et al., 2020
Swetha_21913	2	WRKY+MALA s1 propellar blade	Pathogenecity, Symbiosis	
Swetha_23996[#]	9	WRKY+Lung-7-transmembrane receptor	Plant immunity, Salt stress, Pathogenesis	Wu and Urano, 2018; Lu et al., 2019.
Swetha_18596	10	(WRKY+plant Zn cluster) 2 doms+ Signal recognition particle 9 kDa protein (SRP9)	Protection of mRNA degradation, Pathogen response	Bedassa et al., 2019; Kellogg et al., 2021
Swetha_00744	3	Protein kinases+C-terminal regulatory domain of Calcineurin B-Like (CBL)-interacting protein kinases+WRKY	Abiotic stresses like salt, drought, alkali	Pandey et al., 2015; Luo et al., 2017
Swetha_26866	6	Ubiquitin-protein ligase+WRKY	Abiotic stress response; Autophagy during stress response and development	Shu and Yang, 2017; Xu and Xue, 2019; Su et al., 2020
Swetha_26113[(Y)]	3	Arginine/lysine/ornithine decarboxylase+ WRKY	Abiotic stress response	Upadhyaya et al., 2021;
Swetha_39148	7	activating enzymes (E1) of the ubiquitin-like proteins+WRKY	Plant immunity, Autophagy during stress response	Su et al., 2020

(Continued)

TABLE 3 Continued

Gene	Introns	Domains	Reported functions	Reference
Swetha_25903[#]	5	WRKY+DUF 3084		
Swetha_38328	10	Tellurite-resistance/Dicarboxylate Transporter (TDT) family_ SLAC1+ pepsin retropepesin -like aspartate proteases+WRKY	Regulation of stomatal movement; Stress adaptation; Plant immunity	Zhang et al., 2018; Deng et al., 2019; Sun et al., 2019
Swetha_36198[#]	11	ATP-dependent metalloprotease FtsH+ FtsH Extracellular+WRKY	Photosystem II repair; Photo-oxidative stress	Nixon et al., 2010; Kato and Sakamoto, 2018; Pu et al., 2022
Swetha_37469	9	RWP-RK domain+ Prot. Kinase+WRKY	Nitrate signalling pathways; Nitrogen stress; Nodulation	Ge et al., 2018; Mu and Luo, 2019
Swetha_09150	6	eukaryotic translation initiation factor 5A +WRKY	Pathogen response	
Swetha_09155	9	eukaryotic translation initiation factor 5A +WRKY+WRKY		
Swetha_13727[(Z)]	3	eukaryotic translation initiation factor 5A +WRKY+WRKY		
Swetha_24868	17	WRKY+WRKY+Chorismate mutase typeII+ effector domain of the CAP family of transcription factors+Voltage-dependent potassium channel	Salinity and drought, Negative regulation of salicylic acid pathway; Transcription activation; Electrical signalling; Stress response	Busby and Ebright, 1999; Wang et al., 2018; Poveda, 2020; Musavizadeh et al., 2021; Dreyer et al., 2021
Swetha_37939	9	WRKY+WRKY+WRKY+WRKY		
Swetha_19579	7	WRKY+Cyclopropane fatty-acyl-phospholipid synthase and related methyltransferases	Lipid transport and metabolism; Protection from herbivory	Yu et al., 2011; Okada et al., 2020
Swetha_04251	1	WRKY+peptide synthase	Oxidative stress response; Pathogen response; Plant-pathogen interaction	Golomb et al., 2018
Swetha_09533	14	galactokinase+plant heme-dependent peroxidase (Class-III)+WRKY	Carbohydrate metabolism, Abiotic stress signalling	Xiao et al., 2015; Stein and Granot, 2018

The fused domains, the number of introns in the gene, reported functions and references are given. The symbols [#,Y] and [Z] indicate present in all, present in all except Yuzhi11 and present in all except Zhongzhi13 respectively.

genes were detected in 451 blocks (Figure 4C). The rice chromosomes 11 and 12 did not contain any syntenic homologs of sesame WRKY and AP2/ERF genes. Moreover, 103 whole genome or segmentally duplicated AP2/ERF and WRKY genes are retained as syntenic blocks in Arabidopsis compared to 48 of rice. The comparison of genomes of Zhongzhi-13 and Swetha revealed extensive synteny and collinearity among sesame genes. Of the total 78048 genes present in the two genomes, 48729 were found to be collinear. The percentage of collinearity was 62.43 and 956 syntenic blocks were present (Figure 4A). WGD/segmental duplication was found in 175 genes including 73 WRKY and 103 AP2/ERF genes in Swetha. The segmental duplication genes in Zhongzhi-13 included 108 AP2/ERF and 69 WRKY genes. Synteny analysis revealed the presence of three single-copy WRKY genes in the Chinese accessions including the homologs of Zhongzhi13_22905, Zhongzhi13_26827 and Zhongzhi13_29190. Details are in SI-9.

To study the selection pressure during evolution, Ka/Ks statistics were worked out. Significant selection effects were noticed on 70 pairs of AP2/ERF genes and 54 pairs of WRKY genes in the sesame pan-genome. In the AP2/ERF family 26 gene pairs were under purifying selection (Ka/Ks <1), 45 were under positive selection (Ka/Ks>1) and 2 pairs were under neutral selection (Figure 4D). Among the 54 duplicate genes under selection in the WRKY gene family, 29 were under purifying selection while positive selection effects were noticed in 25 pairs (Figure 4E). We also compared the selection effects of duplicated genes in Swetha and Zhongzhi-13, as well as the gene copy number varaiants of Swetha with each other (Figure 4F). Between the two genomes 52 AP2/ERF and 71 WRKY genes were under selection pressure. In the AP2/ERF family, 26 gene pairs were under positive selection while 26 were under purifying selection. The number of segmental duplicates under positive selection was much higher than dispersed genes whereas an equal distribution was found for

10.3389/fpls.2022.1076229

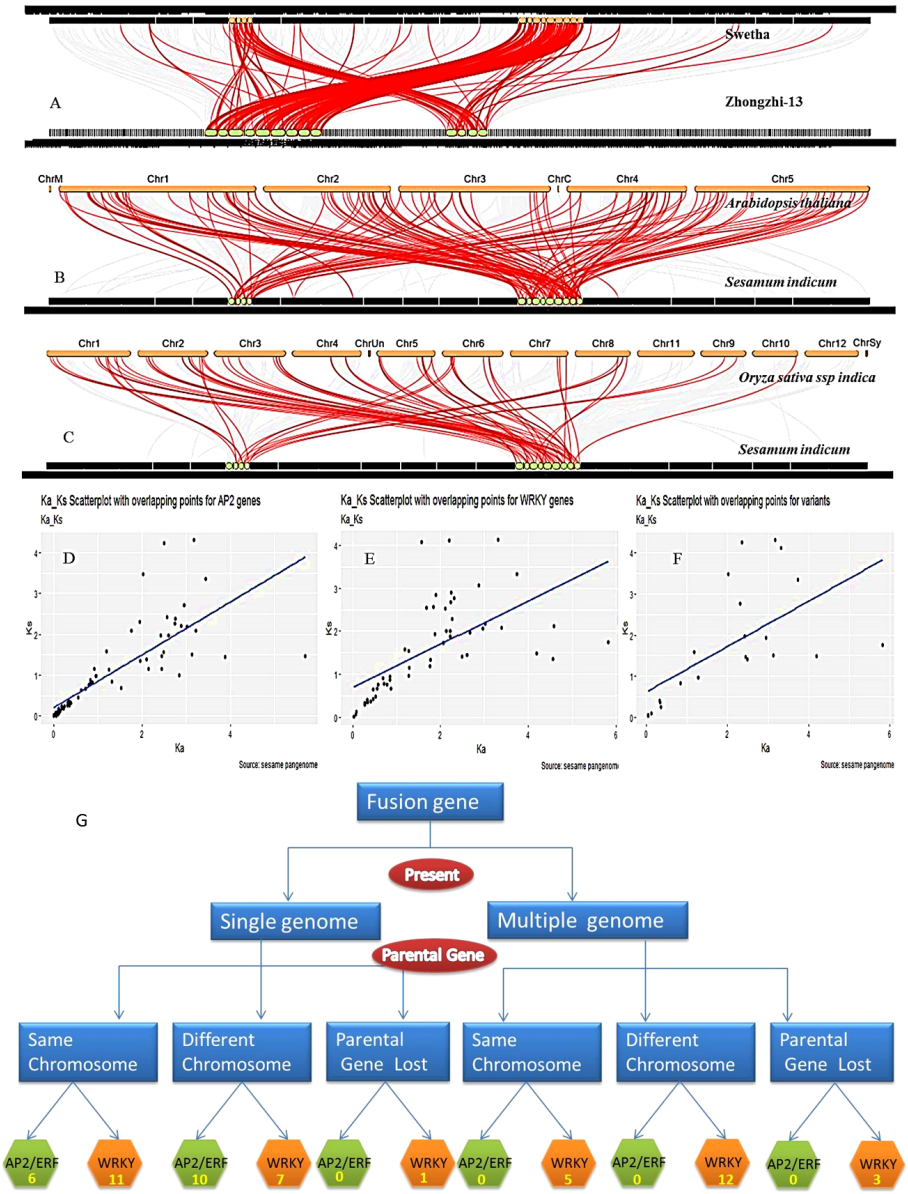


FIGURE 4
Synteny, collinearity and origin of gene fusions in pan-genome. **(A)** Synteny between Swetha and Zhongzhi genomes. **(B)** Sesame and Arabidopsis genomes **(C)** Sesame and rice genomes **(D–F)** specify selection pressure in AP2/ERF, WRKY and Swetha duplicates respectively. **(G)** Represent the origin of fusion genes in sesame pan-genome.

purifying selection. Among the WRKY genes, 28 gene pairs were under purifying selection while 43 were under positive selection. The segmental duplicated genes were under severe selection pressure in both categories. Among the 41 duplicated gene pairs identified in Swetha, selection effects were significant for 22 pairs. Here also the number of gene pairs under positive selection was higher than that under purifying selection SI-10 and SI-11. The evolutionary time period in million years for AP2/ERF genes was 0.4-143 and for WRKY genes was 0.8-137. The genes under neutral selection in AP2/ERF family were most recently evolved.

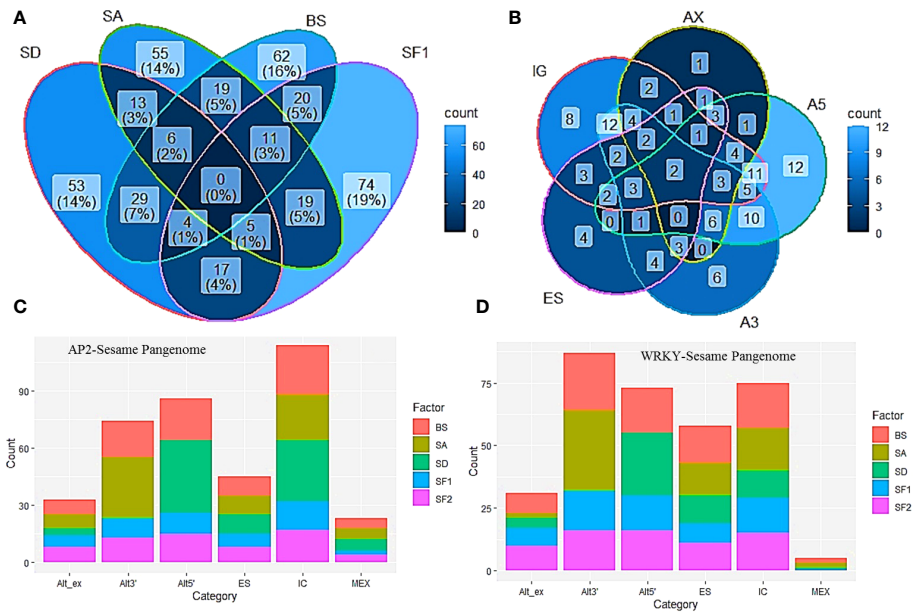


FIGURE 5
Alternative splicing in AP2/ERF and WRKY genes. **(A)** Venn diagram indicating major changes in Splice Acceptor (SA), Splice Donor (SD), Branch Site (BS) and the exonic splicing enhancer Serine/Arginine Splicing Factor 1 (SF1). **(B)** Venn diagram showing major splice mechanisms including Intron Gain (IG), Exon Skipping (ES), Alternative 5′ ends (A5), Alternative 3′ ends (A3) and Alternative exon ends (AX). **(C, D)** indicate changes observed in splice enhancers and branch sites in AP2/ERF and WRKY respectively.

3.8 Protein-protein interactions, functional domains and subcellular localization

The protein-protein interaction network visualized the major roles of AP2/ERF and WRKY transcription factors in defense mechanism, stress response, lipid metabolism and chlorophyll biosynthesis (Figures 6A, B; SI-16). The major interaction partners for defense and stress response included NIMIN family genes, TGA transcription factors, WRKY genes, MEDIATOR family genes including AT2G22370, PAD4 (Phytoalexin deficient 4), EDS1 (Enhanced Disease Susceptibility1), HKT1 (High-Affinity K^+ Transporter 1), bZIP family, MYC, ZAT family, Putative E3 ubiquitin-protein ligase RING1a, Cullin homolog 3 (CUL-3), ethylene activated signaling pathway genes like DREB, TINY, RAP, etc. In chlorophyll biosynthesis, the major interaction

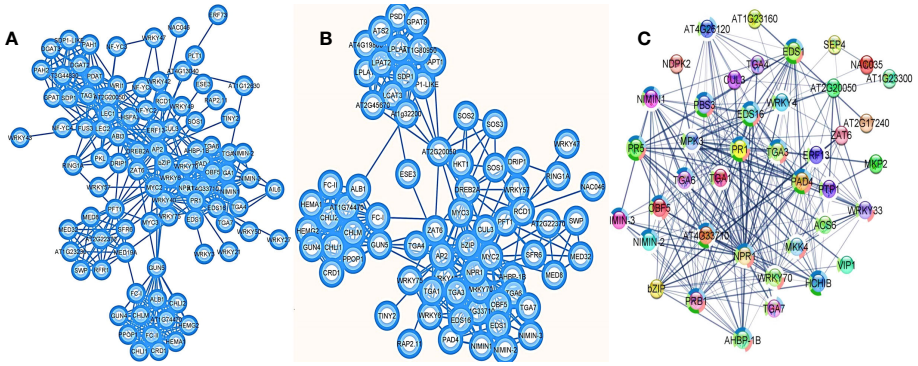


FIGURE 6
Protein-protein interactions of AP2/ERF and WRKY genes in **(A)** Arabidopsis homologs and **(B)** Sesame genome. In sesame defense response is closely related with lipid metabolism and chlorophyll biosynthesis. **(C)** Represent the interactions of WRKY fusion gene, Swetha_24868 as a master regulator of defense signaling in the absence of NBS-LRR genes in sesame.

partners were CRD1 (Copper Response Defect1), GUN (Genomes Uncoupled) 4 and 5, FC (Ferrochelatase) I and II, Albina1, Mg-protoporphyrin chelatase different sub-units (CHLI1, CHLI2, CHLM), Geranylgeranyl reductase (GGR, given as AT1G74470), Glutamyl-tRNA reductase 1 (HEMA1) and Protoporphyrinogen oxidase 1 (PPOP1). Among the lipid metabolism pathway genes, the major interactions identified were with lecithin–cholesterol acyltransferase (LCAT), lysophosphatidic acid acyltransferase (LPAT2), triacylglycerol lipase, Sugar-Dependent1 (SDP1), lysophosphatidylcholine acyltransferase (LPCAT), lyophosphatidylethanolamine acyltransferase (LPEAT1/AT1G80950 and LPEAT2/AT2G45670), Glycerol-3-phosphate acyltransferase 9 (GPAT9), phosphatidylserine decarboxylase (PSD1), Glycerol-3-phosphate acyltransferase (ATS1/AT1G32200), etc. All three pathways showed significant interaction with AT2G20050 representing cAMP-dependent protein kinase involved in the PKA signaling pathway. Swetha_24868, a gene functionally similar to WRKY 4 and Zinc dependent activator protein1 of Arabidopsis was found to act along with WRKY 70 and WRKY 33 initiating cascades of different defense responses (Figure 6C). Many interaction partners were observed to have significant roles in systemic acquired resistance (SAR), osmotic stress, hypoxia, cold stress and pathogenesis.

The significant GO terms associated with AP2/ERF and WRKY transcription factors include cellular response to stress, defense, immune response, intracellular signal transduction, MAPK cascade, regulation of the cellular process, response to abiotic stress, systemic acquired resistance, salicylic acid-mediated signaling pathway, plant-pathogen interaction, ribosome assembly, ribosome biogenesis, translation, etc. The conserved motifs identified through MEME (SI-13A) for WRKY family had molecular functions like transcription factor activity, ATPase activity coupled to transmembrane movement of substances and structural constituent of ribosome while associated with cellular components of the mitochondrion, ribosome and chloroplast (stroma, thylakoid and envelope). The biological process identified was translation. For the AP2/ERF family (SI-12A) the molecular functions attributed were transcription factor activities, structural constituent of ribosome, protein serine/threonine kinase activity and protein binding while being part of biological processes like translation, protein amino acid phosphorylation, transmembrane receptor protein tyrosine kinase signaling pathway and glycolysis. The cellular components identified were the nucleus, mitochondrion, chloroplast, ribosome and cullin-RING E3 ligases (CRLs) complex.

The subcellular localization was studied in detail to understand the regulatory functions. 56% of WRKY and 60% of AP2/ERF transcription factors had a high probability of being located in the nucleus. Among the WRKY genes 20% were predominantly localized in the chloroplast, 7.9% in the cytoplasm, and 6.4% in the mitochondria. For AP2/ERF family the statistics were 17%, 3.2% and 10.89% respectively (SI-12B). The WRKY gene Swetha_04277 and its three homologs were localized in the peroxisome whereas the Baizhima gene (Baizhima_17686) was localized in the chloroplast (SI-13B). Similarly, Zhongzhi13_00117 and three Chinese homologs were located in the extra-cellular space whereas Swetha_00675 was located in the nucleus. Among the splice variants, 64% of WRKY and 70% of AP2/ERF homologs showed differential sub-cellular localization.

3.9 Cis-element analysis in the promoter regions of AP2/ERF and WRKY genes

We examined the cis-element sequence distribution in selected single exon genes, all the splice and exonic variants of AP2/ERF genes and selected homologs from all the WRKY classes from Swetha and Zhongzhi-13. In addition to the common cis-acting elements CAAT box and TATA box, elements regulating phytohormonal response, development and stress response were found. More than 75% of identified cis-elements responded to abiotic and biotic stress responses including drought, salinity, light and pathogenesis. The most common cis-elements identified were ABRE, as-1, MBS, MYB, MYC and STRE. MYB and STRE were present in all the WRKY genes while MYC was present in all the AP2/ERF genes used for the study. Very few instances were found where the conservation existed all through the exon, intron and regulatory regions among the homologs as evident in Swetha_30858 & Zhongzhi13_23474 and Swetha_24927 & Zhongzhi13_18440 from AP2 family. Among gene duplicates in the same genome with difference in exon-intron size or numbers, one gene is found to retain similarity to parental regulatory sequences. Examples include Swetha_18219 & Swetha_18222, Swetha_28864 & Swetha_28866, Swetha_33069 & Swetha_33068, Zhongzhi13_33595 & Zhongzhi13_33551 and Swetha_6527 & Swetha_3532 (WRKY). Even in genes with a single intron and conserved protein structures, the regulatory landscape varied widely. In many cases, in spite of similar protein structures, a drastic change was noticed in the number of a core promoter like TATA to the extent of 2-3 folds, like in Swetha_04095 & Zhongzhi13_02825 and Swetha_41134 & Zhongzhi13_32245 indicating dosage effect. Details are in Figure 3A–C, SI-14, SI-15.

3.10 In-silico gene expression profiles of AP2/ERF and WRKY gene families

The GEO profiles targeting different biotic and abiotic stresses like drought, heat, salinity, osmotic stress, cold stress, wounding, etc. and primary cell wall thickening was selected with expression sites at stem, roots, leaves and seedlings. AP2/ERF genes ERF003 and ERF011 as well as WRKY48 were found to express under all types of abiotic stress and in all the tissues.WRKY48 was found up-regulated in the stem, seedlings and root while down-regulated in roots. 8 AP2/ERF and 12 WRKY genes were found down-regulated in the roots including ERF003, ERF005, WRKY76, WRKY62 and

WRKY24. Most of the AP2/ERF family genes were found to express in the leaves and stem (SI-17A). 6 AP2/ERF genes involved in primary cell wall thickening including ERF034 and ERF043 and were found to express in leaves and seedlings. 10 AP2/ERF and 5 WRKY genes including WRKY-4, 7 and 74 were found down-regulated in leaves during abiotic stress. As for biotic stress response, 22 differentially expressing AP2/ERF and WRKY genes were detected in response to wounding. Up-regulation was noticed in RAV2, ERF070, ERF060, RAP2.4, ERF011, ERF107, WRKY29, TEM1, ERF034, WRKY32, WRKY27, ERF118, RAP2.7, WRKY22, WRKY16, WRKY7 and ERF012 with more than 2 fold changes. 10 genes were down-regulated including WRKY39, WRKY69, ERF104 and ERF116. The expression levels of AP2/ERF and WRKY genes are given in Figure 7A, B, SI-17 B-D.

4 Discussion

Sesame grown in marginal environments across the world is subjected to various kinds of abiotic stresses like drought, water logging and salinity. Recently a pan-genome was developed by combining genomic sequences of five cultivars (Yu et al., 2019) including the reference genome Zhongzhi-13 (space mutant), Chinese landraces Baizhima and Mishouzhima, major Chinese domestic cultivar (Yuzhi11) and the Indian variety 'Swetha'. Pan-genome offers a better understanding of the evolutionary mechanisms that allow organisms to adapt faster to changing environments (Tranchant-Dubreuil et al., 2019).

Environmental adaptations change the genomic architecture and result in the introduction of new genetic diversity into elite cultivars which can be accessed through genome sequencing. Plant adaptations mainly rely on Structural Variations (SVs) including Presence/Absence Variations (PAVs) and Copy Number Variation (CNV), particularly for biotic and abiotic stress tolerance (Khan et al., 2020). Our focus was on the evolutionary adaptations pertaining to maximum fitness among the component genomes adapted to a wide ecological niche. In the major regulators of signal transduction and gene expression under biotic & abiotic stress conditions, AP2/ERF and WRKY, variation was detected in gene number, exon and intron numbers and size, protein characteristics, location in the genome, and promoter sequence architecture.

The number of AP2/ERF genes detected in wheat (322), sorghum (122), rice (139), Arabidopsis (122), *Brassica napus* (531) and sugarcane were comparable to sesame as per the ploidy level (Ghorbani et al., 2020; Riaz et al., 2021; Li et al., 2021). The numbers of RAV and soloist family members were much higher than those reported in soybean, rice and Arabidopsis while comparable with that of pear (Li et al., 2018). The number of AP2/ERF genes reported in the pan-genome (145) is higher than earlier reports (132) by Dossa et al., 2016 in sesame with wide difference in classification. In the WRKY family, we detected 89 genes in the pan-genome, much lesser than reported in sorghum (94), rice (104), Arabidopsis (74), apple (113), soybean (174) and sunflower (119) (Yang et al., 2017; Baillo et al., 2020; Liu et al., 2020; Abdullah-Zawawi et al., 2021; Qin et al., 2022). Like other crops, in

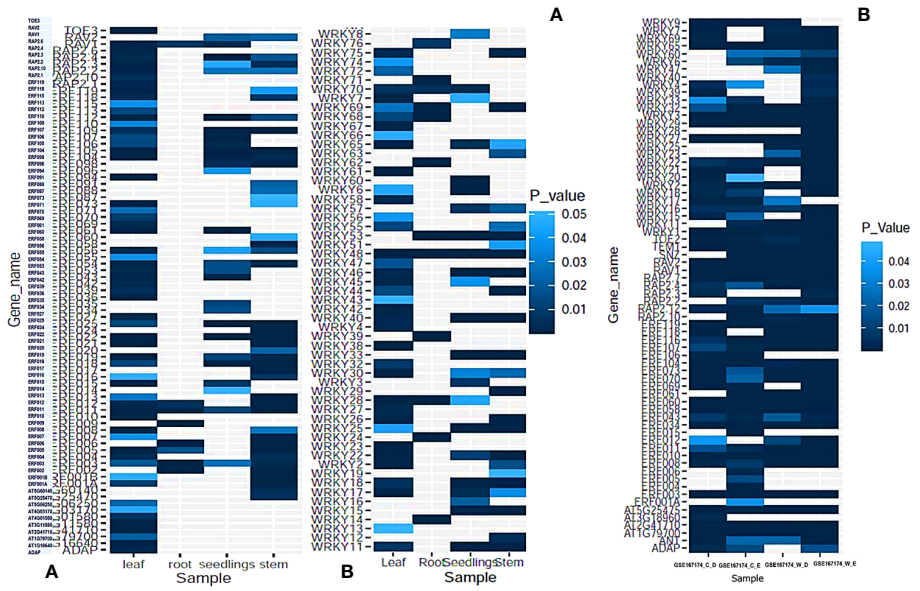


FIGURE 7
The expression profiles of AP2/ERF and WRKY genes during **(A)** Abiotic stress response in different tissues and seedlings. **(B)** During wounding response and biotic stress response in WIND1 mutants and wild type plants. P-values were calculated based on Welch's t-test and genes with P-value ≤0.05 is represented either up regulated or down regulated (> 2-fold or < 0.5-fold).

the sesame pan-genome class-II WRKY's dominated the family whereas the class-III members were very less, 5 in individual genome against 15 and 31 in rice and sorghum respectively. The earlier reports on sesame (Li et al., 2017) suggested 71 WRKY genes in sesame with 7 in class III.

Dossa et al., 2016 reported that 70% of AP2/ERF genes are intron-less and the exons detected were 1-10 in the sesame genome. In the pan-genome we found 63% of AP2/ERF genes to be intron-less with the numbers changing drastically with the individual genome. 53.1% of genes of Swetha genome were intron-less whereas the landraces, Baizhima and Mishouzhima had higher amounts, 67 and 68% respectively. More than 10 exons were detected in 16 genes covering all the genomes while the bulk was contributed by Swetha (7 genes). The number of exons also varied drastically across the genome with the Swetha genome having 3.5 exons/gene as against 2.6 in the pan-genome. In the sorghum pan-genome, 4.2 exons/genes were reported (Tao et al., 2021). Similarly in the WRKY family, the Swetha genome contributed 5 exons/gene as against 4.2 in the pan-genome. Li et al., 2017 reported sesame WRKY genes to consist of 1-11 introns whereas two genes with 14 and 17 introns were detected in the Swetha genome. Similarly, the introns/gene for WRKY and AP2 genes are 3.21 and 1.7 respectively whereas for Swetha it was 4.0 and 2.45 respectively, as against 4.15 introns/gene in plants (Frey and Pucker, 2020). The AP2 sub-family genes had 2-20 introns with the majority of genes having 5-8 introns. Dossa et al., 2016, reported the intron numbers to be 3-10 for AP2 and a single intronic gene was also identified. The ERF sub-family was found to be intron poor with 95% of members having 0-1 introns in accordance with earlier reports. The gene, Swetha_28474 had 36 introns as against the maximum of 9 introns reported earlier (Su et al., 2022). In the WRKY family, the intron numbers reported vary between 0-5 in rice (Abdullah-Zawawi et al., 2021), 0-11 in eggplant (Yang et al., 2020), 1-6 in barley (Zheng et al., 2021) and 1-11 in sesame (Li et al., 2017). One 11 intron T2C sub-family gene was found to be conserved in the genome without any change in exonic sequences although size variation was noticed for the 2nd intron. Enhanced gene copy number, as well as the predominance of large genes with multiple exons in the genome of Swetha, was reported earlier for TCP (Parakkunnel et al., 2020) and HSF (Parakkunnel et al., 2022) gene families in sesame.

A similar trend was found in the case of micro-exons wherein family-wise difference was quite wider with AP2/ERF family having more micro-exons than WRKY. The AP2/ERF family was reported to be micro-exon rich (Song et al., 2020) and in the sesame pan-genome, micro-exons totaled 10% of total exons as against 1.96% in the WRKY family. 88% of micro-exons were contributed by the AP2 subfamily in the sesame pan-genome with only 15 out of 128 identified genes lacking micro-exons. Song et al., 2020, also reported that AP2 micro-exon genes are highly conserved which we found only to be partially true in the pan-genome. 44% of micro-exon genes were conserved in the pan-genome in the AP2 sub-family whereas the extent was higher in the Chinese accessions particularly in landraces Baizhima and Mishouzhima, wherein 83% sequence conservation was observed. WRINKLED1 (*WRI1*) is an AP2 gene widely studied in Arabidopsis and higher plants (Ma et al., 2013) acting as a master regulator of fatty acid synthesis. The presence of 9bp long micro-exon coding for amino acids 'VYL' and its isoforms have been reported as essential for the *AtWRI1* gene. This micro-exon was missing in the sesame pan-genome. Instead, the WRI1 homologs of sesame may be alternative splice forms of the gene as reported in castor (Ji et al., 2018). The sesame homologs lacked 'VYL' sequence and formed five different clusters with intron numbers ranging from 4-20 although Yuzhi11 and Zhongzhi13 genomes shared the sequence (SI-13). The atypical splicing combined with the skipping of conserved micro-exons resulted in unique *WRI1* genes in rice (Mano et al., 2019). In the pan-genome (SI-4) exon skipping, gene fusion, mutually exclusive exon, as well as alternative 5' and 3' splice ends contribute to the generation of novel variants in *WRI1* genes. Micro-exons of size ≤15bp are considered as shortest and interestingly 38% of WRKY micro-exons belonged to this category whereas in the AP2/ERF family only 9% of micro-exons fulfilled this criterion (Yu et al., 2022). The most common micro-exons occurring in AP2/ERF genes are of sizes 9, 26, 31, and 45 bp according to Song et al., 2020. In sesame, the most common micro-exon is 50 bp occurring in 54% of AP2 genes at exonic positions 4, 5 or 6. An estimated 35% of AP2 genes without the 50bp micro-exon contained two micro exons of 25 and 30 bp size, occurring together at 2nd and 3rd exons in the same order. A 44bp micro-exon also occurred in the same group together with 25 and 30 bp micro-exons in 50% of cases. Only one case of exclusion of 25bp micro-exon was noticed in the sesame pan-genome wherein gene fusion involving PWWP domain with AP2 domain was observed in Swetha_18222 drastically altering genic architecture.

Among the 26 WRKY genes with micro-exons in the sesame pan-genome, 15 genes (all belonging to Chinese accessions) did not show any variation in exon-intron structure. However, alternative splice forms were noticed in 11 WRKY genes with micro-exons with more than 80% belonging to the Swetha genome. In the AP2/ERF family, alternative splice forms were noticed in 47 micro-exon genes (35%); while the majority (47%) again belonged to Swetha. Altogether, in the Swetha genome, 49% of genes with micro-exons were alternative splice forms. It can be understood (Table 2) that micro-exons are important functionaries of stress tolerance, circadian clocks, apomixis, defense response, the integrity of genetic material and transposon integration (Liu et al., 2013; Theriault and Nkongolo, 2017; Aguado and tenOever, 2018; Curtis et al., 2018; Wu and Urano, 2018; Worthington et al., 2019; Wang et al., 2019; Shi et al., 2020; Song et al., 2020; Tischkau, 2020; Chen et al., 2021; Ming et al., 2022.)

However, we do not agree with Song et al., 2020 that the micro-exons of AP2 genes are under negative selection. In the pan-genome, 4 genes were under neutral selection; 11 were under purifying selection, and 25 were under positive selection. Genes under positive selection included micro-exons with domain

similarities to lecithin: cholesterol acyltransferase (LCAT3), aquaporin and integrase while the ASGR-BBM-like micro-exon gene was under neutral selection. Most of the micro-exon genes (>65%) were also under WGD or segmental duplication.

Gene fusion was identified as a potent source of creation of evolutionary novelty wherein a new gene arises by joining parts from two or more genes and is controlled by the same regulatory region. It is widely reported in bacteria (Pasek et al., 2006) as the evolutionary force behind multi-domain proteins and in eukaryotes resulting in cancer-genesis mutations with deleterious consequences (Glenfield and Innan, 2021). In higher plant genomes gene fusion is a complex process with a poorly understood evolutionary mechanism. Gene fusion events have been reported in the biosynthesis of alkaloids in opium poppy (Catania et al., 2022), viral infection in maize (Zhou et al., 2022a) and in the evolution of new genes in the genus Oryza (Zhou et al., 2022b). In the sesame pan-genome, 2.3% of AP2/ERF genes and 10% of WRKY genes were products of gene fusion. Larger share of these novel variations was contributed by the Swetha genome where 8.3% of AP2/ERF and 24% of WRKY genes arose due to gene fusion. In the Oryza genus, the *O. japonica* genome had more fusion genes compared to *O. indica*, *O. barthii* and *O. glaberrima* (Zhou et al., 2022b) whereas differential gene fusions were observed among different Papaver species affecting alkaloid concentration (Catania et al., 2022). However, the gene fusions reported in sesame, particularly in Swetha are much higher than in maize (Zhou et al., 2022a) and Oryza (Zhou et al., 2022b). The duplication of parental genes prior to fusion was evident also in the sesame genome (Zhou et al., 2022b) while most of the added domains were products of horizontal transfer. In sesame, gene fusion and creation of novel multi-domain genes were reported in HSF genes (Parakkunnel et al., 2022) while the fusion with retrotransposons were reported for TCP genes (Parakkunnel et al., 2020). Selection effects were significant for 21 fusion genes in sesame wherein positive selection was noticed in >71% of genes in contrast to the rice genome (Zhou et al., 2022b).

Alternative Splicing (AS) increases the complexity of the transcriptome and proteome by generating multiple transcripts of the same gene through differential processing of introns and exons in pre-mRNA (Martin et al., 2021). Reports suggest that AS events are involved in the regulation of gene expression under a multitude of environmental and biotic stresses (Laloum et al., 2018: Martin et al., 2021). The identified spice variants in AP2/ERF and WRKY genes were 70 (48.27%) and 52 (58.43%) respectively with maximum splice variants per single gene of 6 and 7. Swetha genome contributed 40-50% of these variants.

Majority of the exonic additional domains and micro-exonic domains were involved in abiotic stress responses as well as hypersensitive responses against plant pathogens (Tables 2, 3). The maintenance of cell membrane and organelle integrity under stress is majorly dependent on the lipid and carbohydrate composition of the cell dynamically affecting the performance of cellular transporters (Rawat et al., 2021). Moreover, the epigenetic regulation of stress response through chromatin remodeling, regulatory RNA's and DNA methylation is manifested in the additionally acquired domains (Chinnusamy et al., 2013). Reactive Oxygen Species (ROS) are produced at higher levels in different organelles during abiotic stress and cause significant damage to the cell. Ironically ROS is also essential for stress response as they act as signaling molecules triggering signal transduction (Frederickson Matika and Loake, 2014). Plant FtsH4, an ATP-dependent mitochondrial protease is essential to preserve stem cell activity and temperature stress response throughout development and lack of which leads to the precocious cessation of growth (Huang et al., 2019). An important component of ROS signaling, FtsH4 is associated with photosystemII repair mechanism and mitigation of photo-oxidative stress (Khanna-Chopra, 2012) and manifested through apoptosis and leaf senescence. It is important that a fusion gene with WRKY and FtsH is conserved across the pan-genome indicating adaptive evolution under stress conditions. Similarly, E3 ubiquitin ligase essential for less accumulation of ROS and enhanced anti-oxidant capacity under drought stress, existed in fusion with the WRKY gene in sesame (Zhang et al., 2017). Cullin-RING E3 ligases (CRLs) were identified as a major cellular component in meme analysis. Oxidative burst; wherein ROS accumulation leads to the production of hydrogen peroxide (H_2O_2) occurs as a defense response against pathogens. Aquaporins help the diffusion of H_2O_2 through the plasma membrane to cytosol triggering MAMP-triggered immunity (Bigeard et al., 2015).

WRKY and AP2/ERF genes are expressed at all parts of the plant; root, stem, leaves and seedlings and in addition to stress response also function in the light-induced stomatal opening, redox homeostasis, callus formation, starch metabolism, cell wall biosynthesis, RNA regulation of transcription, hormone metabolism and lignin biosynthesis. The multi-domain fusion genes are involved in all the major defense pathways like SA-mediated signaling, and MAP kinase cascades (Eulgem and Somssich, 2007). WRKY genes are known to activate sugar-responsive genes through an epigenetic mechanism and a fusion gene Swetha_09533 containing the galactokinase domain is directly involved in sugar metabolism (Chen et al., 2019). Particularly the additional domains of the T1 WRKY gene Swetha_24868, chorismate mutase, cAMP and voltage-dependent potassium channel convert this gene into a master regulator. Recruiting other WRKY and AP2/ERF genes into the picture, Swetha_24868 mediates defense responses involving Salicylic acid-dependent signaling, thiol-based signaling and MAPK signaling. Chorismate mutase regulates defense mechanisms through enhancing the accumulation of SA, lignin and antioxidants (Jan et al., 2021). The cAMP (3′, 5′-cyclic adenosine monophosphate) is known as an important signaling molecule in defense responses in addition to roles in germination, stomatal opening, ion homeostasis and cell cycle progression (Blanco et al., 2020). Voltage-gated K+ channels are involved in high salinity stress and maintaining ion homeostasis in sweet potato (Zhu et al., 2022) and rice (Musavizadeh et al., 2021). Interestingly in the absence of NBS-LRR genes in sesame, the

pathway involves EDS-1, EDS-16, WRKY-4, NPR-1, PRB-1, PR-1, PR5, HCHIB and GH3 (AT1G23160) functioning in SA signaling pathway. In addition to triggering immunity, SA mediated pathway is also effective in regulating ROS levels as revealed by interaction with ERF-13 and CUL-3 (Figure 6). However, this gene and the parental gene, Swetha_24865 did not vary much in the promoter sequence except for the ABRE sequence in Swetha_24865. However, they did share DRE, MYB and MYC sequences. A single copy of DRE is needed for ABA independent induction of osmotic and cold stress genes and promoter sequence containing DRE without ABRE is found to work well under stress conditions (Yamaguchi-Shinozaki and Shinozaki, 1994). The localization of Swetha_24868 was found to be predominantly in chloroplast while Swetha_24865 was expressed equally in the nucleus and chloroplast. Class T3 of WRKY genes had W-box sequences in the promoter wherein protein localization, exonic and other cis-element sequences varied widely with individual genes. T2E genes, Swetha_21917 and Swetha_21913 as well as AP2 genes Swetha_33068 and Swetha_33069 in spite of sharing the same promoter sequences are located nearby each other and products of recent gene duplication localized differently in chloroplast and nucleus, respectively. The exon-intron, promoter sequence, localization and active domain diversity indicate the faster evolution scenario of defense response genes in sesame subjected to multiple stresses on account of its marginal growing conditions.

5 Conclusions

Breeding crop varieties for changing climate scenarios with the effective use of existing diversity is the primary challenge for food security. However, breeding techniques and adaptation to the environment significantly alter the genomic structure of crop plants. This was evident from the pan-genome study of sesame including varieties or landraces adapted for vastly varying climates of India and China. Although a certain level of gene conservation existed at the species level; evolution created different footprints on different genomes. The adaptive selection was evident in copy number variation and modification of function for most of the gene loci studied. Retention of ancient genes with the incorporation of extra functional domains to cope with extreme stress conditions was observed in *S. indicum*. Ecological adaptation was manifested in genome composition with geographical regions harboring variant forms of gene loci offering maximum fitness.

Data availability statement

The datasets presented in this study can be found in on line repositories. The pan-genome data (Yu et al. 2019) is available in public domain and all the other data can be found in the supplementary materials. The Swetha genome was sequenced by a team involving KVB and RP and is deposited as bio-project "PRJNA219369" and assembly ASM97556v1. The other genomes are available at accession numbers GCA_000512975.1, GCA_003268515.1 and GCA_026168435.1.

Author contributions

RP: Conceived the idea, extracted data, investigation, resources, softwares, analysed data, and drafted manuscript. BK: Investigation, review and drafted the manuscript. VG: Resources, review and writing. SC: Investigation, resources and drafted the manuscript. SP: Extraction of genic sequences. UK: Project administration, resources KB: Genomic resources, Swetha genome sequencing, manuscript finalization; SK: Supervision, project administration and manuscript finalization. All authors contributed to the article and approved the submitted version.

Acknowledgments

RP, BK, VG and SC are thankful to the director, ICAR-Indian Institute of Seed Science, Mau for providing the necessary facilities.

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.2022.1076229/full#supplementary-material

DATA SHEET 1
Contains SI-18 (Fig). The AP2/ERF and WRKY genes mapped to the 13 chromosomes of sesame genome. Here the WRKY and AP2/ERF genes from Swetha genome have been represented in 'blue' and 'red' coloured fonts respectively. Please adjust the zoom to 100% for clear visibility.

DATA SHEET 2
Contains SI-19 (Fig). The sub-families are coloured as per Swetha AP2/ERF classification given in Fig-2. The colour codes are as following: Blue=DREB; Magenta=ERF; Red=AP2; Green=RAV and Cyan=SOLOIST. Individual Swetha genes are coloured as per the initial classification of Fig-2 to reflect the extent of domain changes in the pan-genome during evolution.

DATA SHEET 3
Contains SI-20 (Fig). Different groups are coloured as per classification given in Fig-2. T1=Blue; T2A= Orange; T2B=Magenta, T2C=Red; T2D=Lime yellow; T2E= Purple and T3= Olive.

References

Abdullah-Zawawi, M. R., Ahmad-Nizammuddin, N. F., Govender, N., Harun, S., Mohd-Assaad, N., and Mohamed-Hussein, Z. A. (2021). Comparative genome-wide analysis of WRKY, MADS-box and MYB transcription factor families in arabidopsis and rice. *Sci. Rep.* 11, 1–18. doi: 10.1038/s41598-021-99206-y

Aguado, L. C., and tenOever, B. R. (2018). RNase III nucleases and the evolution of antiviral systems. *BioEssays* 40 (2), 1700173. doi: 10.1002/bies.201700173

Baillo, E. H., Hanif, M. S., Guo, Y., Zhang, Z., Xu, P., and Algam, S. A. (2020). Genome-wide identification of WRKY transcription factor family members in sorghum (*Sorghum bicolor* (L.) moench). *PloS One* 15 (8), e0236651. doi: 10.1371/journal.pone.0236651

Bedassa, S. B., Akkaya, M. S., and Ersoy, F. (2019). HvSRP72 silencing enhanced *Blumeria graminis f.* sp. *hordei* growth in compatible interaction with barley. *J. Plant Pathol.* 101, 91–96. doi: 10.1007/s42161-018-0145-4

Bedigian, D. (2003). Evolution of sesame revisited: domestication, diversity and prospects. *Genet. Resour. Crop Evol.* 50, 779–787. doi: 10.1023/A:1025029903549

Bigeard, J., Colcombet, J., and Hirt, H. (2015). Signaling mechanisms in pattern-triggered immunity (PTI). *Mol. Plant* 8, 521–539. doi: 10.1016/j.molp.2014.12.022

Blanco, E., Fortunato, S., Viggiano, L., and de Pinto, M. C. (2020). Cyclic AMP: A polyhedral signalling molecule in plants. *Int. J. Mol. Sci.* 21, 4862. doi: 10.3390/ijms21144862

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., et al. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PloS Comput. Biol.* 15, e1006650. doi: 10.1371/journal.pcbi.1006650

Busby, S., and Ebright, R. H. (1999). Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol.* 293, 199–213. doi: 10.1006/jmbi.1999.3161

Catania, T., Li, Y., Winzer, T., Harvey, D., Meade, F., Caridi, A., et al. (2022). A functionally conserved STORR gene fusion in papaver species that diverged 16.8 million years ago. *Nat. Commun.* 13, 1–11. doi: 10.1038/s41467-022-30856-w

Chaudhuri, M., Darden, C., Soto Gonzalez, F., Singha, U. K., Quinones, L., and Tripathi, A. (2020). Tim17 updates: A comprehensive review of an ancient mitochondrial protein translocator. *Biomolecules* 10, 1643. doi: 10.3390/biom10121643

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

Chen, Y., Li, W., Turner, J. A., and Anderson, C. T. (2021). PECTATE LYASE LIKE12 patterns the guard cell wall to coordinate turgor pressure and wall mechanics for proper stomatal function in arabidopsis. *Plant Cell* 33, 3134–3150. doi: 10.1093/plcell/koab163

Chen, X., Li, C., Wang, H., and Guo, Z. (2019). WRKY transcription factors: evolution, binding, and action. *Phytopathol. Res.* 1, 1–15. doi: 10.1186/s42483-019-0022-x

Chinnusamy, V., Dalal, M., and Zhu, J. K. (2013). "Epigenetic regulation of abiotic stress responses in plants," in *Plant abiotic stress*. Eds. M. A. Jenks and P. M. Hasegawa (John Wiley & Sons, Inc), 203–229. doi: 10.1002/9781118764374.ch8

Curtis, T. Y., Bo, V., Tucker, A., and Halford, N. G. (2018). Construction of a network describing asparagine metabolism in plants and its application to the identification of genes affecting asparagine metabolism in wheat under drought and nutritional stress. *Food Energy Secur.* 7, e00126. doi: 10.1002/fes3.126

Deng, Y., Wang, C., Wang, N., Wei, L., Li, W., Yao, Y., et al. (2019). Roles of small-molecule compounds in plant adventitious root development. *Biomolecules* 9, 420. doi: 10.3390/biom9090420

Dossa, K., Wei, X., Li, D., Fonceka, D., Zhang, Y., and Wang, L. (2016). Insight into the AP2/ERF transcription factor superfamily in sesame and expression profiling of DREB subfamily under drought stress. *BMC Plant Biol.* 16, 1–16. doi: 10.1186/s12870-016-0859-4

Dreyer, I., Sussmilch, F. C., Fukushima, K., Riadi, G., Becker, D., Schultz, J., et al. (2021). How to grow a tree: plant voltage-dependent cation channels in the spotlight of evolution. *Trends Plant Sci.* 26, 41–52. doi: 10.1016/j.tplants.2020.07.011

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

Eulgem, T., and Somssich, I. E. (2007). Networks of WRKY transcription factors in defense signaling. *Curr. Opin. Plant Biol.* 10, 366–371. doi: 10.1016/j.pbi.2007.04.020

Frederickson Matika, D. E., and Loake, G. J. (2014). Redox regulation in plant immune function. *Antioxid Redox Signal* 21, 1373–1388. doi: 10.1089/ars.2013.5679

Frey, K., and Pucker, B. (2020). Animal, fungi, and plant genome sequences harbor different non-canonical splice sites. *Cells* 9, 458. doi: 10.3390/cells9020458

Galván-Gordillo, S. V., Martínez-Navarro, A. C., Xoconostle-Cázares, B., and Ruiz-Medrano, R. (2016). Bioinformatic analysis of arabidopsis reverse transcriptases with a zinc-finger domain. *Biologia* 71, 1223–1229. doi: 10.1515/biolog-2016-0145

Ge, M., Liu, Y., Jiang, L., Wang, Y., Lv, Y., Zhou, L., et al. (2018). Genome-wide analysis of maize NLP transcription factor family revealed the roles in nitrogen response. *Plant Growth Regul.* 84, 95–105. doi: 10.1007/s10725-017-0324-x

Ghorbani, R., Zakipour, Z., Alemzadeh, A., and Razi, H. (2020). Genome-wide analysis of AP2/ERF transcription factors family in *Brassica napus*. *Physiol. Mol. Bio.l Plants* 26, 1463–1476. doi: 10.1007/s12298-020-00832-z

Glenfield, C., and Innan, H. (2021). Gene duplication and gene fusion are important drivers of tumourigenesis during cancer evolution. *Genes* 12, 1376. doi: 10.3390/genes12091376

Golomb, B. L., Yu, A. O., Coates, L. C., and Marco, M. L. (2018). The *Lactococcus lactis* KF 147 nonribosomal peptide synthetase/polyketide synthase system confers resistance to oxidative stress during growth on plant leaf tissue lysate. *Microbiologyopen* 7, e00531. doi: 10.1002/mbo3.531

He, D., Liang, R., Long, T., Yang, Y., and Wu, C. (2021). Rice RBH1 encoding a pectate lyase is critical for apical panicle development. *Plants* 10, 271. doi: 10.3390/plants10020271

Hrmova, M., and Gilliham, M. (2018). Plants fighting back: to transport or not to transport, this is a structural question. *Curr. Opin. Plant Biol.* 46, 68–76. doi: 10.1016/j.pbi.2018.07.006

Huang, H., Ullah, F., Zhou, D. X., Yi, M., and Zhao, Y. (2019). Mechanisms of ROS regulation of plant development and stress responses. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00800

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

Jan, R., Khan, M. A., Asaf, S., Lee, I. J., and Kim, K. M. (2021). Over-expression of chorismate mutase enhances the accumulation of salicylic acid, lignin, and antioxidants in response to the white-backed plant hopper in rice plants. *Antioxidants* 10, 1680. doi: 10.3390/antiox10111680

Jiao, J., and Peng, D. (2018). Wheat microRNA1023 suppresses invasion of *Fusarium graminearum via* targeting and silencing FGSG_03101. *J. Plant Interact.* 13, 514–521. doi: 10.1080/17429145.2018.1528512

Ji, X. J., Mao, X., Hao, Q. T., Liu, B. L., Xue, J. A., and Li, R. Z. (2018). Splice variants of the castor WRI1 gene up regulate fatty acid and oil biosynthesis when expressed in tobacco leaves. *Int. J. Mol. Sci.* 19, 146. doi: 10.3390/ijms19010146

Jin, X. (2022). Regulatory network of Serine/Arginine-rich (SR) proteins: the molecular mechanism and physiological function in plants. international journal of molecular sciences. *Int. J. Mol. Sci.* 23, 10147. doi: 10.3390/ijms231710147

Karlik, E. (2021). Why lncRNAs were not conserved? is it for adaptation? *Front. Life Sci. RT* 2, 103–110. doi: 10.51753/flsrt.1027595

Karray, A., Alonazi, M., Jallouli, R., Alanazi, H., and Ben Bacha, A. (2022). A proteinaceous alpha-amylase inhibitor from *Moringa oleifera* leaf extract: purification, characterization, and insecticide effects against *C. maculates* insect larvae. *Molecules* 27, 4222. doi: 10.3390/molecules27134222

Kato, Y., and Sakamoto, W. (2018). FtsH protease in the thylakoid membrane: physiological functions and the regulation of protease activity. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00855

Kellogg, M. K., Miller, S. C., Tikhonova, E. B., and Karamyshev, A. L. (2021). SRPassing co-translational targeting: the role of the signal recognition particle in protein targeting and mRNA protection. *Int. J. Mol. Sci.* 22, 6284. doi: 10.3390/ijms22126284

Kenzior, A., and Folk, W. R. (2015). *Arabidopsis thaliana* MSI4/FVE associates with members of a novel family of plant specific PWWP/RRM domain proteins. *Plant Mol. Biol.* 87, 329–339. doi: 10.1007/s11103-014-0280-z

Khan, A. W., Garg, V., Roorkiwal, M., Golicz, A. A., Edwards, D., and Varshney, R. K. (2020). Super-pangenome by integrating the wild side of a species for accelerated crop improvement. *Trends Plant Sci.* 25, 148–158. doi: 10.1016/j.tplants.2019.10.012

Khanna-Chopra, R. (2012). Leaf senescence and abiotic stresses share reactive oxygen species-mediated chloroplast degradation. *Protoplasma* 249, 469–481. doi: 10.1007/s00709-011-0308-z

Kharabian, A. (2010). An efficient computational method for screening functional SNPs in plants. *J. Theor. Biol.* 265, 55–62. doi: 10.1016/j.jtbi.2010.04.017

Kumar, M., Brar, A., Yadav, M., Chawade, A., Vivekanand, V., and Pareek, N. (2018a). Chitinases–potential candidates for enhanced plant resistance towards fungal pathogens. *Agriculture* 8, 88. doi: 10.3390/agriculture8070088

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

Laloum, T., Martín, G., and Duque, P. (2018). Alternative splicing control of abiotic stress responses. *Trends Plant Sci.* 23, 140–150. doi: 10.1016/j.tplants.2017.09.019

Lanciano, S., and Mirouze, M. (2018). Transposable elements: all mobile, all different, some stress responsive, some adaptive? *Curr. Opin. Genet. Dev.* 49, 106–114. doi: 10.1016/j.gde.2018.04.002

Letunic, I., and Bork, P. (2018). 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* 46, D493–D496. doi: 10.1093/nar/gkx922

Li, W., Guan, Q., Wang, Z. Y., Wang, Y., and Zhu, J. (2013). A bi-functional xyloglucan galactosyltransferase is an indispensable salt stress tolerance determinant in arabidopsis. *Mol. Plant* 6, 1344–1354. doi: 10.1093/mp/sst062

Li, D., He, Y., Li, S., Shi, S., Li, L., Liu, Y., et al. (2021). Genome-wide characterization and expression analysis of AP2/ERF genes in eggplant (*Solanum melongena* l.). *Plant Physiol. Biochem.* 167, 492–503. doi: 10.1016/j.plaphy.2021.08.006

Li, D., Liu, P., Yu, J., Wang, L., Dossa, K., Zhang, Y., et al. (2017). Genome-wide analysis of WRKY gene family in the sesame genome and identification of the WRKY genes involved in responses to abiotic stresses. *BMC Plant Biol.* 17, 1–19. doi: 10.1186/s12870-017-1099-y

Li, X., Tao, S., Wei, S., Ming, M., Huang, X., Zhang, S., et al. (2018). The mining and evolutionary investigation of AP2/ERF genes in pear (Pyrus). *BMC Plant Biol.* 18, 46. doi: 10.1186/s12870-018-1265-x

Liu, C., Fukumoto, T., Matsumoto, T., Gena, P., Frascaria, D., Kaneko, T., et al. (2013). Aquaporin OsPIP1; 1 promotes rice salt resistance and seed germination. *Plant Physiol. Biochem.* 63, 151–158. doi: 10.1016/j.plaphy.2012.11.018

Liu, A., Liu, C., Lei, H., Wang, Z., Zhang, M., Yan, X., et al. (2020). Phylogenetic analysis and transcriptional profiling of WRKY genes in sunflower (*Helianthus annuus* l.): Genetic diversity and their responses to different biotic and abiotic stresses. *Ind. Crops Prod.* 148, 112268. doi: 10.1016/j.indcrop.2020.112268

Lüdke, D., Roth, C., Kamrad, S. A., Messerschmidt, J., Hartken, D., Appel, J., et al. (2021). Functional requirement of the arabidopsis importin-α nuclear transport receptor family in autoimmunity mediated by the NLR protein SNC1. *Plant J.* 105, 994–1009. doi: 10.1111/tpj.15082

Lu, C., Liu, H., Jiang, D., Wang, L., Jiang, Y., Tang, S., et al. (2019). *Paecilomyces variotii* extracts (ZNC) enhance plant immunity and promote plant growth. *Plant Soil* 441, 383–397. doi: 10.1007/s11104-019-04130-w

Luo, M., Cheng, K., Xu, Y., Yang, S., and Wu, K. (20172147). Plant responses to abiotic stress regulated by histone deacetylases. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.02147

Ma, W., Kong, Q., Arondel, V., Kilaru, A., Bates, P. D., Thrower, N. A., et al. (2013). Wrinkled1, a ubiquitous regulator in oil accumulating tissues from arabidopsis embryos to oil palm mesocarp. *PloS One* 8, e68887. doi: 10.1371/journal.pone.0068887

Ma, L., and Li, G. (2021). Arabidopsis FAR-RED ELONGATED HYPOCOTYL3 negatively regulates carbon starvation responses. *Plant Cell Environ.* 44, 1816–1829. doi: 10.1111/pce.14044

Mano, F., Aoyanagi, T., and Kozaki, A. (2019). Atypical splicing accompanied by skipping conserved micro-exons produces unique WRINKLED1, an AP2 domain transcription factor in rice plants. *Plants* 8, 207. doi: 10.3390/plants8070207

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., et al. (2017). CDD/SPARCLE: functional classification of proteins *via* subfamily domain architectures. *Nucleic Acids Res.* 45, D200–D203. doi: 10.1093/nar/gkw1129

Martín, G., Márquez, Y., Mantica, F., Duque, P., and Irimia, M. (2021). Alternative splicing landscapes in *Arabidopsis thaliana* across tissues and stress conditions highlight major functional differences with animals. *Genome Biol.* 22, 35. doi: 10.1186/s13059-020-02258-y

Mindrebo, J. T., Nartey, C. M., Seto, Y., Burkart, M. D., and Noel, J. P. (2016). Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. *Curr. Opin. Struct. Biol.* 41, 233–246. doi: 10.1016/j.sbi.2016.08.005

Ming, Q., Wang, K., Wang, J., Liu, J., Li, X., Wei, P., et al. (2022). The combination of RNA-seq transcriptomics and data-independent acquisition proteomics reveals the mechanisms underlying enhanced salt tolerance by the ZmPDI gene in *Zoysia matrella* [L.] merr. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.970651

Mishra, P., Singh, A., and Roy, S. (2022). "Plasma membrane h+-ATPase in plants," in *Cation transporters in plants* (Academic Press), 357–373. doi: 10.1016/B978-0-323-85790-1.00012-9

Moghaddam, S. M., Oladzad, A., Koh, C., Ramsay, L., Hart, J. P., Mamidi, S., et al. (2021). The tepary bean genome provides insight into evolution and domestication under heat stress. *Nat. Commun.* 12, 1–14. doi: 10.1038/s41467-021-22858-x

Monné, M., Vozza, A., Lasorsa, F. M., Porcelli, V., and Palmieri, F. (2019). Mitochondrial carriers for aspartate, glutamate and other amino acids: A review. *Int. J. Mol. Sci.* 20, 4456. doi: 10.3390/ijms20184456

Mu, X., and Luo, J. (2019). Evolutionary analyses of NIN-like proteins in plants and their roles in nitrate signaling. *Cell. Mol. Life Sci.* 76, 3753–3764. doi: 10.1007/s00018-019-03164-8

Musavizadeh, Z., Najafi-Zarrini, H., Kazemitabar, S. K., Hashemi, S. H., Faraji, S., Barcaccia, G., et al. (2021). Genome-wide analysis of potassium channel genes in rice: expression of the OsAKT and OsKAT genes under salt stress. *Genes* 12, 784. doi: 10.3390/genes12050784

Nixon, P. J., Michoux, F., Yu, J., Boehm, M., and Komenda, J. (2010). Recent advances in understanding the assembly and repair of photosystem II. *Ann. Bot.* 106, 1–16. doi: 10.1093/aob/mcq059

Okada, K., Fujiwara, S., and Tsuzuki, M. (2020). Energy conservation in photosynthetic microorganisms. *J. Gen. Appl. Microbiol.* 66, 59–65. doi: 10.2323/jgam.2020.02.002

Olmedo, G., and Guzmán, P. (2008). Processing precursors with RNase III in plants. *Plant Sci.* 175, 741–746. doi: 10.1016/j.plantsci.2008.07.001

Ono, E., Waki, T., Oikawa, D., Murata, J., Shiraishi, A., Toyonaga, H., et al. (2020). Glycoside-specific glycosyltransferases catalyze regio-selective sequential glucosylations for a sesame lignan, sesaminol triglucoside. *Plant J.* 101, 1221–1233. doi: 10.1111/tpj.14586

Pandey, S., Prasad, A., Sharma, N., and Prasad, M. (2020). Linking the plant stress responses with RNA helicases. *Plant Sci.* 299, 110607. doi: 10.1016/j.plantsci.2020.110607

Pandey, P., Ramegowda, V., and Senthil-Kumar, M. (2015). Shared and unique responses of plants to multiple individual stresses and stress combinations: physiological and molecular mechanisms. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00723

Parakkunnel, R., Bhojaraja Naik, K., Susmita, C., Girimalla, V., Bhaskar, K. U., Sripathy, K. V., et al. (2022). Evolution and co-evolution: insights into the divergence of plant heat shock factor genes. *Physiol. Mol. Biol. Plants* 28, 1029–1047. doi: 10.1007/s12298-022-01183-7

Parakkunnel, R., Bindhani, N., Purru, S., Lakhanpaul, S., and Venkataramanna Bhat, K. (2020). Adaptive evolution and response to phytoplasma: A genome-wide study of TCP transcription factors in *Sesamum indicum* l. *Ann. Appl. Biol.* 176, 75–95. doi: 10.1111/aab.12561

Pasek, S., Risler, J. L., and Brézellec, P. (2006). Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. *Bioinformatics* 22, 1418–1423. doi: 10.1093/bioinformatics/btl135

Poveda, J. (2020). *Trichoderma parareesei* favors the tolerance of rapeseed (*Brassica napus* l.) to salinity and drought due to a chorismate mutase. *Agronomy* 10, 118. doi: 10.3390/agronomy10010118

Pu, L., Cheng, L., Li, A., Liang, S., Wei, Q., Wu, S., et al. (2022). Effects of clonal integration on allelopathy of invasive plant *Wedelia trilobata* under heterogeneous light conditions. *J. Plant Ecol.* 15, 663–671. doi: 10.1093/jpe/rtab028

Qin, Y., Yu, H., Cheng, S., Liu, Z., Yu, C., Zhang, X., et al. (2022). Genome-wide analysis of the WRKY gene family in *Malus domestica* and the role of MdWRKY70L in response to drought and salt stresses. *Genes* 13, 1068. doi: 10.3390/genes13061068

Raikwar, S., Srivastava, V. K., Gill, S. S., Tuteja, R., and Tuteja, N. (2015). Emerging importance of helicases in plant stress tolerance: characterization of *Oryza sativa* repair helicase XPB2 promoter and its functional validation in tobacco under multiple stresses. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.01094

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032

Rawat, N., Singla-Pareek, S. L., and Pareek, A. (2021). Membrane dynamics during individual and combined abiotic stresses in plants and tools to study the same. *Physiol. Plant* 171, 653–676. doi: 10.1111/ppl.13217

Ren, H., Su, Q., Hussain, J., Tang, S., Song, W., Sun, Y., et al. (2021). Slow anion channel GhSLAC1 is essential for stomatal closure in response to drought stress in cotton. *J. Plant Physiol.* 258, 153360. doi: 10.1016/j.jplph.2020.153360

Riaz, M. W., Lu, J., Shah, L., Yang, L., Chen, C., Mei, X. D., et al. (2021). Expansion and molecular characterization of AP2/ERF gene family in wheat (*Triticum aestivum* l.). *Front. Genet.* 12. doi: 10.3389/fgene.2021.63215

Rona, G. B., Eleutherio, E. C., and Pinheiro, A. S. (2016). PWWP domains and their modes of sensing DNA and histone methylated lysines. *Biophys. Rev.* 8, 63–74. doi: 10.1007/s12551-015-0190-6

Shi, Y., Phan, H., Liu, Y., Cao, S., Zhang, Z., Chu, C., et al. (2020). Glycosyltransferase OsUGT90A1 helps protect the plasma membrane during chilling stress in rice. *J. Exp. Bot.* 71, 2723–2739. doi: 10.1093/jxb/eraa025

Shopan, J., Mou, H., Zhang, L., Zhang, C., Ma, W., Walsh, J. A., et al. (2017). Eukaryotic translation initiation factor 2B-beta (eIF 2Bβ), a new class of plant virus resistance gene. *Plant J.* 90, 929–940. doi: 10.1111/tpj.13519

Shu, K., and Yang, W. (2017). E3 ubiquitin ligases: ubiquitous actors in plant development and abiotic stress responses. *J. Plant Physiol.* 58, 1461–1476. doi: 10.1093/pcp/pcx071

Song, Q., Bari, A., Li, H., and Chen, L. L. (2020). Identification and analysis of micro-exons in AP2/ERF and MADS gene families. *FEBS Open Bio* 10, 2564–2577. doi: 10.1002/2211-5463.12990

Stankovic, N., Schloesser, M., Joris, M., Sauvage, E., Hanikenne, M., and Motte, P. (2016). Dynamic distribution and interaction of the arabidopsis SRSF1 subfamily splicing factors. *Plant Physiol.* 170, 1000–1013. doi: 10.1104/pp.15.01338

Stein, O., and Granot, D. (2018). Plant fructokinases: evolutionary, developmental, and metabolic aspects in sink tissues. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00339

Suh, A. (2021). "Horizontal transfer of transposons as genomic fossils of host-parasite interactions," in *The evolution and fossil record of parasitism* (Springer, Cham), 451–463.

Su, T., Li, X., Yang, M., Shao, Q., Zhao, Y., Ma, C., et al. (2020). Autophagy: an intracellular degradation pathway regulating plant survival and stress response. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00164

Su, Y., Ma, Z., Mao, J., Li, W., Cao, X., and Chen, B. (2022). Genome-wide identification and characterization of the strawberry (*Fragaria vesca*) FvAP2/ERF gene family in abiotic stress. *Plant Mol. Biol. Rep.* 40, 646–660. doi: 10.1007/s11105-022-01343-9

Sun, L. R., Yue, C. M., and Hao, F. S. (2019). Update on roles of nitric oxide in regulating stomatal closure. *Plant Signal. Behav.* 14, e1649569. doi: 10.1080/15592324.2019.1649569

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

Theriault, G., and Nkongolo, K. K. (2017). Evidence of prokaryote like protein associated with nickel resistance in higher plants: horizontal transfer of TonB-dependent receptor/protein in betula genus or *de novo* mechanisms? *Heredity* 118, 358–365. doi: 10.1111/tpj.15486

Tian, P., Lin, Z., Lin, D., Dong, S., Huang, J., Huang, T., et al. (2021). The pattern of DNA methylation alteration, and its association with the changes of gene expression and alternative splicing during phosphate starvation in tomato. *Plant J.* 108, 841–858. doi: 10.1111/tpj.15486

Tischkau, S. A. (2020). Mechanisms of circadian clock interactions with aryl hydrocarbon receptor signalling. *Eur. J. Neurosci.* 51, 379–395. doi: 10.1111/ejn.14361

Tranchant-Dubreuil, C., Rouard, M., and Sabot, F. (2019). Plant pangenome: impacts on phenotypes and evolution. *Annu. Plant Rev.* 2, 453–78. doi: 10.1002/9781119312994.apr0664

Uluisik, S., and Seymour, G. B. (2020). Pectate lyases: Their role in plants and importance in fruit ripening. *Food Chem.* 309, 125559. doi: 10.1016/j.foodchem.2019.125559

Upadhyaya, D. C., Bagri, D. S., Upadhyaya, C. P., Kumar, A., Thiruvengadam, M., and Jain, S. K. (2021). Genetic engineering of potato (*Solanum tuberosum* l.) for enhanced α-tocopherols and abiotic stress tolerance. *Physiol. Plant* 173, 116–128. doi: 10.1111/ppl.13252

Vogt, J. H., and Schippers, J. H. (2015). Setting the PAS, the role of circadian PAS domain proteins during environmental adaptation in plants. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00513

Wang, J., Gao, S., Peng, X., Wu, K., and Yang, S. (2019). Roles of the INO80 and SWR1 chromatin remodeling complexes in plants. *Int. J. Mol. Sci.* 20, 4591. doi: 10.3390/ijms20184591

Wang, L., Zhu, J., Li, X., Wang, S., and Wu, J. (2018). Salt and drought stress and ABA responses related to bZIP genes from *V. radiata* and *V. angularis*. *Gene* 651, 152–160. doi: 10.1016/j.gene.2018.02.005

Worthington, M., Ebina, M., Yamanaka, N., Heffelfinger, C., Quintero, C., Zapata, Y. P., et al. (2019). Translocation of a parthenogenesis gene candidate to an alternate carrier chromosome in apomictic *Brachiaria humidicola*. *BMC Genom.* 20, 41. doi: 10.1186/s12864-018-5392-4

Wu, T. Y., and Urano, D. (2018). Genetic and systematic approaches toward G protein-coupled abiotic stress signaling in plants. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01378

Xiao, W., Chang, H., Zhou, P., Yuan, C., Zhang, C., Yao, R., et al. (2015). Genome-wide identification, classification and expression analysis of GHMP genes family in *Arabidopsis thaliana*. *Plant Syst. Evol.* 301, 2125–2140. doi: 10.1007/s00606-015-1219-9

Xu, W., Tang, W., Wang, C., Ge, L., Sun, J., Qi, X., et al. (2020). SiMYB56 confers drought stress tolerance in transgenic rice by regulating lignin biosynthesis and ABA signaling pathway. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00785

Xu, F. Q., and Xue, H. W. (2019). The ubiquitin-proteasome system in plant responses to environments. *Plant Cell Environ.* 42, 2931–2944. doi: 10.1111/pce.13633

Yamaguchi-Shinozaki, K., and Shinozaki, K. (1994). A novel cis-acting element in an arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. *Plant Cell* 6, 251–264. doi: 10.1105/tpc.6.2.251

Yang, Y., Liu, J., Zhou, X., Liu, S., and Zhuang, Y. (2020). Identification of WRKY gene family and characterization of cold stress-responsive WRKY genes in eggplant. *PeerJ* 8, e8777. doi: 10.7717/peerj.8777

Yang, Y., Zhou, Y., Chi, Y., Fan, B., and Chen, Z. (2017). Characterization of soybean WRKY gene family and identification of soybean WRKY genes that promote resistance to soybean cyst nematode. *Sci. Rep.* 7, 17804. doi: 10.1038/s41598-017-18235-8

Yu, J., Golicz, A. A., Lu, K., Dossa, K., Zhang, Y., Chen, J., et al. (2019). Insight into the evolution and functional characteristics of the pan-genome assembly from sesame landraces and modern cultivars. *Plant Biotechnol. J.* 17, 881–892. doi: 10.1111/pbi.13022

Yu, H., Li, M., Sandhu, J., Sun, G., Schnable, J. C., Walia, H., et al. (2022). Pervasive misannotation of microexons that are evolutionarily conserved and crucial for gene function in plants. *Nat. Commun.* 13, 820. doi: 10.1038/s41467-022-28449-8

Yu, M., Romer, K. A., Nieland, T. J., Xu, S., Saenz-Vash, V., Penman, M., et al. (2011). Exoplasmic cysteine Cys384 of the HDL receptor SR-BI is critical for its sensitivity to a small-molecule inhibitor and normal lipid transport activity. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12243–12248. doi: 10.1073/pnas.1109078108

Yu, G., Xian, L., Xue, H., Yu, W., Rufian, J. S., Sang, Y., et al. (2020). A bacterial effector protein prevents MAPK-mediated phosphorylation of SGT1 to suppress plant immunity. *PloS Pathog.* 16, e1008933. doi: 10.1371/journal.ppat.1008933

Zhang, J., De-oliveira-Ceciliato, P., Takahashi, Y., Schulze, S., Dubeaux, G., Hauser, F., et al. (2018). Insights into the molecular mechanisms of CO2-mediated regulation of stomatal movements. *Curr. Biol.* 28, R1356–R1363. doi: 10.1016/j.cub.2018.10.015

Zhang, N., Yin, Y., Liu, X., Tong, S., Xing, J., Zhang, Y., et al. (2017). The E3 ligase TaSAP5 alters drought stress responses by promoting the degradation of DRIP proteins. *Plant Physiol.* 175, 1878–1892. doi: 10.1104/pp.17.01319

Zheng, J., Zhang, Z., Tong, T., Fang, Y., Zhang, X., Niu, C., et al. (2021). Genome-wide identification of WRKY gene family and expression analysis under abiotic stress in barley. *Agronomy* 11, 521. doi: 10.3390/agronomy11030521

Zhou, Y., Lu, Q., Zhang, J., Zhang, S., Weng, J., Di, H., et al. (2022a). Genome-wide profiling of splicing and gene fusion during rice black-streaked dwarf virus stress in maize (*Zea mays* l.). *Genes* 13 (3), 456. doi: 10.3390/genes13030456

Zhou, Y., Zhang, C., Zhang, L., Ye, Q., Liu, N., Wang, M., et al. (2022b). Gene fusion as an important mechanism to generate new genes in the genus oryza. *Genome Biol.* 23, 1–23. doi: 10.1186/s13059-022-02696-w

Zhu, H., Yang, X., Li, Q., Guo, J., Ma, T., Liu, S., et al. (2022). The sweetpotato voltage-gated k+ channel β subunit, KIbB1, positively regulates low-k+ and high-salinity tolerance by maintaining ion homeostasis. *Genes* 13, 1100. doi: 10.3390/genes13061100

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 12 January 2023
DOI 10.3389/fpls.2022.1095781

OPEN ACCESS

EDITED BY
Pranav Pankaj Sahu,
Czech Academy of Sciences, Czechia

REVIEWED BY
Mehanathan Muthamilarasan,
University of Hyderabad, India
Santosh Kumar Gupta,
National Institute of Plant Genome Research (NIPGR), India

*CORRESPONDENCE
Ambika Baldev Gaikwad
ambika.gaikwad@icar.gov.in;
ambikabg@gmail.com

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 11 November 2022
ACCEPTED 15 December 2022
PUBLISHED 12 January 2023

CITATION
Gaikwad AB, Kaila T, Maurya A, Kumari R, Rangan P, Wankhede DP and Bhat KV (2023) The chloroplast genome of black pepper (*Piper nigrum* L.) and its comparative analysis with related *Piper* species. *Front. Plant Sci.* 13:1095781. doi: 10.3389/fpls.2022.1095781

The chloroplast genome of black pepper (*Piper nigrum* L.) and its comparative analysis with related *Piper* species

Ambika Baldev Gaikwad[1*], Tanvi Kaila[2], Avantika Maurya[1], Ratna Kumari[1], Parimalan Rangan[1], Dhammaprakash Pandhari Wankhede[1] and K. V. Bhat[1]

[1]Division of Genomic Resources, Indian Council of Agricultural Research (ICAR)-National Bureau of Plant Genetic Resources, New Delhi, India, [2]Indian Council of Agricultural Research (ICAR)-National Institute for Plant Biotechnology, New Delhi, India

Piper nigrum, also known as black pepper, is an economically and ecologically important crop of the genus *Piper*. It has been titled as the king of spices due to its wide consumption throughout the world. In the present investigation, the chloroplast genome of *P. nigrum* has been assembled from a whole genome sequence by integrating the short and long reads generated through Illumina and PacBio platforms, respectively. The chloroplast genome was observed to be 161,522 bp in size, having a quadripartite structure with a large single copy (LSC) region of 89,153 bp and a small single copy (SSC) region of 18,255 bp separated by a copy of inverted repeats (IRs), each 27,057 bp in length. Taking into consideration all the duplicated genes, a total of 131 genes were observed, which included 81 protein-coding genes, 37 tRNAs, 4 rRNAs, and 1 pseudogene. Individually, the LSC region consisted of 83 genes, the SSC region had 13 genes, and 18 genes were present in each IR region. Additionally, 216 SSRs were detected and 11 of these were validated through amplification in 12 species of *Piper*. The features of the chloroplast genome have been compared with those of the genus *Piper*. Our results provide useful insights into evolutionary and molecular studies of black pepper which will contribute to its further genetic improvement and breeding.

KEYWORDS

black pepper, chloroplast genome, gene content, gene order, phylogeny

Introduction

Chloroplasts are essential intracellular organelles having an independent genome with several genes responsible for the process of photosynthesis. Most higher plants harbor a single, circular, double-stranded chloroplast DNA with a quadripartite structure that includes a long single copy (LSC) region, a small single copy (SSC) region, and two

copies of large inverted repeat (IRa and IRb) regions (Bock, 2007). Chloroplast genomes are highly conserved in terms of their size, gene content, and structure with a few exceptions (Dong et al., 2013; Daniell et al., 2016). According to the Organelle Genome Resource-NCBI (www.ncbi.nlm.nih.gov/genome/browse#!/organelles/), 4,557 chloroplast genomes of land plants are reported until December 2020.

Magnoliidae is one of the largest and early-diverging Angiosperm clade representing four orders—Canellales, Piperales, Laurales, and Magnoliales—that comprise around 10,000 species in 20 families (Massoni et al., 2015; Chase et al., 2016; Sinn et al., 2018). So far, 384 chloroplast genomes belonging to Magnoliidae have been submitted to NCBI-Organelle Genome Resource of which 48 belong to Piperales and only 15 to the genus *Piper*. Piperaceae is considered to be the most diverse group of magnoliids with around more than 4,300 species. The genus *Piper* possesses more than 2,000 species (Divya et al., 2015) including the economically important crop black pepper (*Piper nigrum*).

Black pepper has been titled as the "king of spices" due to its wide consumption throughout the world for inducing a pungent flavor in culinary dishes. The Western Ghats region of India is known to be the center of origin and diversity for this species with around 30 countries growing it commercially, mainly India, Indonesia, Malaysia, China, Thailand, and Brazil (Kubitzki and Rohwer, 1993; Ravindran, 2000). The medicinal value of secondary metabolites produced by *P. nigrum* reveals its role in digestive disorders, hepatoprotection, antithyroidism, anti-inflammation, antidepressant, immunomodulator, antispasmodic, antioxidant, and many more (Ahmad et al., 2012).

In the present investigation, we report on the complete sequence of the *P. nigrum* chloroplast genome. Detailed characterization of the black pepper chloroplast including size, gene content, structure repeats, and GC content and comparative genomics analysis of the chloroplast genome provide insights into functional genomics (Li et al., 2019). Moreover, phylogenetic analysis of 58 genes and 37 taxa consisting of 34 angiosperms and 3 gymnosperms provides insight into the phylogenetic position of *P. nigrum* in accordance with other magnoliids, monocots, and eudicots.

Material and methods

DNA extraction and next-generation sequencing

Leaves of black pepper landrace 'Thottumuriyan' were used to extract high-quality genomic DNA that was assessed for its quality and quantity using Qubit. Two paired-end libraries were prepared from genomic DNA using TruSeq DNA PCR-Free LT sample Prep Kit (Illumina, San Diego, USA) as per standard protocol and quantified using Bioanalyzer (Agilent Technologies, USA) and qPCR (Kapa Biosystems, Wilmington, MA, USA) sequenced on HiSeq1000 and MiSeq (Illumina, San Diego, CA, USA) using 2 × 100 and 2 × 150 bases paired-end chemistry, respectively. Pacific Biosciences (PacBio) SMRT bell sequencing libraries were prepared for long read generation and were sequenced on PacBio Sequel. A total of 82.74- and 14.65-GB data were generated through Illumina sequencing and PacBio, respectively.

Chloroplast genome assembly and annotation

The FastQC software (Andrews, 2010) was used to check the quality of the sequenced reads, and then the Trimmomatic software (Bolger et al., 2014) was used to trim and filter the low-quality reads. Filtered data were mapped to the reference chloroplast genome using CLC Genomics workbench 9.0. Mapped reads of PacBio sequence data were extracted and assembled using the Organelle PBA software (Soorni et al., 2017) which was polished through the Pilon software (Walker et al., 2014) using Illumina sequence data.

For the preliminary annotation of the chloroplast genome, the D online program GeSeq (Tillich et al., 2017) offered by MPI-MP Chlorobox was used and the output was curated. The tRNA genes were identified by the ARAGORN v1.2.38 tRNA annotator (Laslett and Canback, 2004), which was available at the GeSeq server. The OGDraw (Greiner et al., 2019) online server of Chlorobox was used to construct a gene map of the chloroplast genome with default settings and checked manually. The black pepper chloroplast genome sequence was assembled to 161,522 bp and submitted to NCBI, GenBank accession No. MK883818 (Supplementary File 1).

Repeat structure and comparative genome analysis

Codon usage and GC content were analyzed using CodonW and Molecular Evolutionary Genetics Analysis (MEGA 7.0) (Kumar et al., 2016), respectively. To investigate the size and location of repeat sequences comprising forward, reverse, palindromic, and complement repeats, REPuter online server was used with a minimal size of 30 bp, >90% identity, and hamming distance of 3 between two repeat copies. Also, Tandem Repeats Finder (Benson, 1999) was used to extract the tandem repeats present in the chloroplast genome of *P. nigrum* and the other reported species of *Piper*. Simple sequence repeats (SSRs) were detected using the MISA software (Beier et al., 2017) with parameters of eight, four, four, three, three, and three repeat units for mono-, di-, tri-, tetra-, penta-, and hexanucleotide SSRs, respectively. mVISTA (Mayor et al., 2000) was used to

perform whole genome sequence alignment of *P. nigrum* along with other members of Piperaceae comprising *Piper kadsura*, *Piper laetispicum*, *Piper cenocladum*, and *Piper auritum*.

Validation of SSRs

A set of 11 primer pairs were synthesized using the software Primer3 (Untergasser et al., 2012) and validated in 21 accessions of *Piper* representing 12 different species (*Piper nigrum*, *Piper longum*, *Piper arboreum*, *Piper argyrophyllum*, *Piper attenuatum*, *Piper betel*, *Piper chaba*, *Piper hymenophyllum*, *Piper trichostachyon*, *Piper wallichi*, *Piper columbrinum*, and *Piper sarmentosum*). Genomic DNA was isolated from the leaf using the cTAB DNA extraction method. The PCR reaction consisted of 1× PCR buffer, 2.5 mM of $MgCl_2$, 1 µM of primer, 0.2 mM of each dNTP, 1 U of *Taq* DNA polymerase (NEB), and 15 ng of template DNA in a total volume of 20 µl and cycled at 95°C for 5 min followed by 35 cycles of denaturation at 95°C for 1 min, annealing at 55°C for 1 min and extension at 72°C for 1 min followed by a final extension at 72°C for 10 min. The amplification products were resolved on a QIAxcel multicapillary system using QIAxcel High Resolution Kit 1200 (QIAGEN, No. 929002, New Delhi,Qiagen India Pvt. Ltd.) 50-800 bp v2.0 QX DNA size marker (QIAGEN, No. 929561, New Delhi,Qiagen India Pvt. Ltd.) and 15 bp/1,000 bp Qx alignment marker (QIAGEN No. 929521, New Delhi,Qiagen India Pvt. Ltd.) with the high-resolution run method OM700. The allelic sizes of each sample were calculated in the form of gel profiles and peaks using the QIAxcel ScreenGel software (QIAGEN, v1.5).

Synonymous substitution

The numbers of synonymous substitutions per synonymous sites (*Ks*) and non-synonymous substitutions per non-synonymous sites (*Ka*) and functional polymorphism within magnoliids were calculated using the DnaSP software (Rozas et al., 2017).

Phylogenetic *a*nalysis

Phylogenetic analysis was performed with 58 genes common to all 37 complete chloroplast genomes downloaded from NCBI (Supplementary File 2). The 37 complete chloroplast genomes used in the study are those of species representing 31 orders of the phylum Tracheophyta. These include 3 orders from the gymnosperms; those of Amborella, Austrobaileyales, and Chloranthales; 4 orders of the magnoliids; the Alismatid, Lilioid, and Commelinid monocots; and 17 orders of the eudicots. Firstly, each gene from all 37 taxa was aligned individually using ClustalW 2.0 (Thompson et al., 1994). On an online server, gBlocks was used to create conserved blocks with default settings, eliminating the poorly aligned regions. Maximum likelihood analysis was performed with PhyML using the subtree pruning and regrafting (SPR) topological moves with Smart Model Selection (SMS) of the substitution model based on Akaike information criterion (AIC) and 100 bootstrapping.

Results and discussion

Features of the *Piper nigrum* chloroplast genome

With the increased use of next-generation sequencing, the number of reported angiosperm chloroplast genomes has increased. Nonetheless, only 15 chloroplast genomes belonging to the genus *Piper* have been sequenced. Recently, the complete chloroplast genome sequences of *Piper hancei* (Zhang et al., 2021) and *P. sarmentosum* (Geng et al., 2022) have been reported. A comparative analysis of eight *Piper* species with *P. nigrum* has been also reported (Li et al., 2022). In the present investigation, the chloroplast genome of *P. nigrum* was sequenced and analyzed. The complete assembled chloroplast genome of *P. nigrum* is a circular model with the length of 161,522 bp, having a typical quadripartite structure comprising an LSC region of 89,153 bp and an SSC region of 18,255 bp separated by a copy of inverted repeats (IRs) each 27,057 bp in length (Figure 1). *Piper nigrum* differs slightly from other reported *Piper* species in terms of the total length of the chloroplast genome. It is smaller than *P. laetispicum* (161,721 bp), but larger than *P. kadsura* (161,486 bp), *P. cenocladum* (160,624 bp), *P. auritum* (160,624 bp), and *P. hancei* (161,476 bp) in terms of length.

The *P. nigrum* chloroplast genome was observed to have 131 genes, which include 81 protein-coding genes, 37 tRNAs, 4 rRNAs, and 1 pseudogene (Table 1). On considering each region individually, it was observed that the LSC consists of 83 genes, the SSC consists of 13 genes, and 18 genes were present in the IR region. The duplication of tRNA genes in the LSC region has been observed in various angiosperms (Tsudzuki et al., 1994; Kaila et al., 2016; Kaila et al., 2017). However, as with other reported Piperaceae chloroplast genomes (Cai et al., 2006), the tRNA gene was not observed to be duplicated in *P. nigrum*. The gene *rps12* was observed to be trans-spliced with one end present in the LSC region and another duplicated end present in the IR region. Trans-splicing of the *rps12* gene has also been reported in other chloroplast genomes like *Pinus taeda*, *P. kadsura*, and *Fagus crenata* (Lee et al., 2016; Asaf et al., 2018; Worth et al., 2019). In consonance with the previous studies, the *trnK-UUU* gene was observed to be harboring the largest intron (2,548 bp) which includes the *matK* gene.

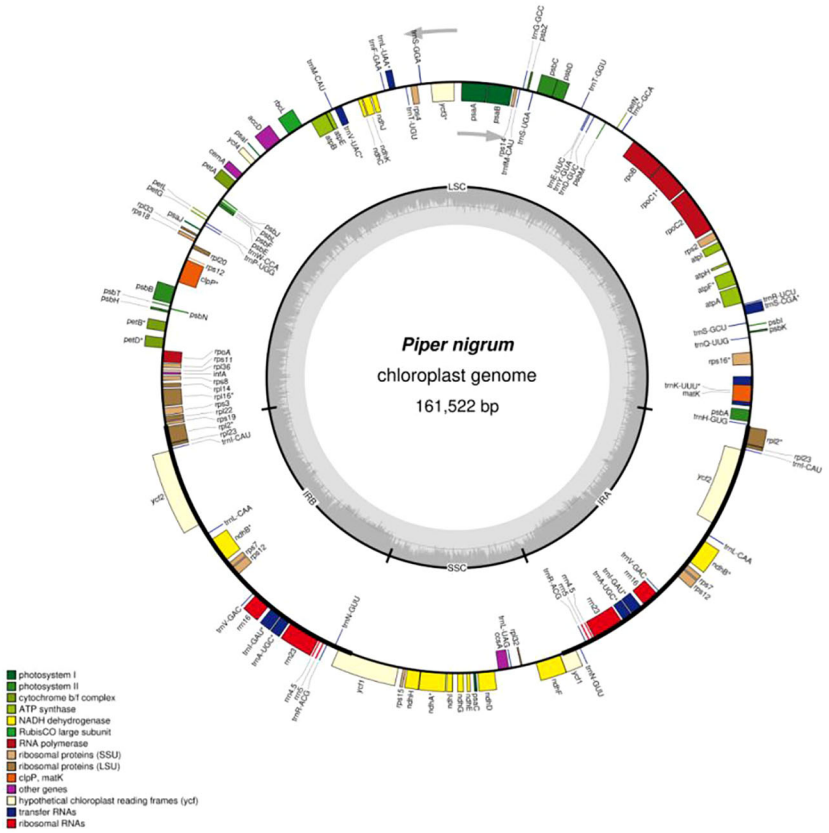


FIGURE 1
Circular map of the *Piper nigrum* chloroplast genome.

Pseudogenes are similar to functional genes except that they do not produce functional proteins. One reason for this may be a disrupted open reading frame due to the presence of an internal stop codon(s). In *P. nigrum*, the *ycf1* gene was identified to be the only pseudogene, similar to *P. kadsura* (Lee et al., 2016) and *P. laetispicum* (Wang et al., 2018), while *Cinnamomum camphora* and *Persea americana* are reported to have three (*rpl23*, *ycf1*, and *ycf2*) and two (*ycf1* and *ycf2*) pseudogenes, respectively (Li et al., 2019; Liu et al., 2019).

The average AT content of the *P. nigrum* chloroplast genome was observed to be 61.7% which is similar to other reported *Piper* species (Cai et al., 2006; Lee et al., 2016; Wang et al., 2018; Li et al., 2022). The single copy regions LSC (63.2%) and SSC (67.9%) possess higher AT content than the repeat regions (57.1%) (Table 2). The lower AT content of the IR region can be associated with the presence of rRNA genes, which possesses a decreased number of AT nucleotides, thereby contributing to genome stabilization and sequence complexity. Also, the AT content for the CDS region (protein-coding region) was observed to be 60.9% (Table 2). Coding regions have higher GC content and therefore lower AT content than non-coding regions. While if we consider the different regions of the chloroplast genome, the IR regions have the lowest AT content and the SSC regions have the highest AT content. This is due to the presence of NADH genes present in the SSC region that are reported to have the lowest GC content as compared with any class of genes (Cai et al., 2006).

The chloroplast genome of *P. nigrum* consists of 18 intron-containing genes, viz., *ycf3*, *rpoC1*, *atpF*, *rps16*, *rpl2*, *ndhB*, *ndhA*, *rpl16*, *petB*, *petD*, *clpP*, *rps12*, and 6 tRNA genes (Table 3). Three genes, viz., *ycf3*, *rps12*, and *clpP*, consist of two introns, while the rest have one intron. Similar observations have been made for other plants belonging to the Magnoliid clade like *P. cenocladum* and *Drimys granadensis* (Cai et al., 2006) which harbor 18 intron-containing genes with three genes spanning 2 introns. The recently reported chloroplast genome of *P. hancei* also harbors 18 intron-containing genes, 16 of which contain one intron and 2 genes (*clpP* and *ycf3*) possess two introns (Zhang et al., 2021). In contrast, a recent analysis of eight *Piper* chloroplast genomes reveal 14 intron-containing genes of which 10 protein-coding genes and 2 tRNA genes had a single intron and 2 genes had two introns (Li et al., 2022). Different from these results, the chloroplast genome of *P. sarmentosum* was reported to have 21 intron-containing genes, consisting of 8

TABLE 1 List of genes in the chloroplast genome of *Piper nigrum*.

Category of genes	Group of genes	Gene name	Total number
Photosynthesis	Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*	5
	Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*	15
	Cytochrome b/f complex	*petA*, *petB*[a], *petD*[a], *petG*, *petL*, *petN*	6
	ATP synthase	*atpA*, *atpB*, *atpE*, *atpF*[a], *atpH*, *atpI*	6
	NADH dehydrogenase	*ndhA*[a], *ndhB*[a] (×2) 1, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, ndhI, *ndhJ*, *ndhK*	12
	RubisCO large subunit	*rbcL* 1	1
Self-replication	RNA polymerase	*rpoA*, *rpoB*, *rpoC1*[a], *rpoC2*	4
	Ribosomal proteins	*rps2*, *rps3*, *rps4*, rps7 (×2), *rps8*, *rps11*, *rps12*[b] (×2), *rps14*, *rps15*, *rps16*[a], *rps18*, *rps19*, *rpl2*[a] (×2), *rpl14*, *rpl16*[a], *rpl20*, *rpl22*, *rpl23* (×2), *rpl32*, *rpl33*, *rpl36*	25
	Ribosomal RNAs	*rrn4.5* (×2), *rrn5* (×2), *rrn16* (×2), *rrn23* (×2) 8 (4)	8
	Transfer RNAs	37 tRNAs (6 contain an intron, 7 in the IRs)	37
Other genes	Hypothetical proteins	*ycf1*, *ycf2* (×2), *ycf3*[b], *ycf4*	5
	Translational initiation factor	*infA*	1
	Maturase	*matK*	1
	Protease	*clpP*[b]	1
	Envelope membrane protein	*cemA*	1
	Subunit of Acetyl-CoA-Carboxylase	*accD*	1
	C-type cytochrome synthesis gene	*ccsA*	1
	Putative pseudogene	*ycf1*	1
Total genes			131

[a]One intron.
[b]Two introns.

TABLE 2 Nucleotide composition in different regions of the chloroplast genome of *Piper nigrum*.

Regions	Position	T(U)	C	A	G	Length
IRa	–	28.8	22.1	28.3	20.8	27,057
IRb	–	28.3	20.8	28.8	22.1	27,057
LSC	–	32.2	18.8	31.0	18.0	89,153
SSC	–	33.9	15.3	34.0	16.9	18,255
Total	–	31.2	19.3	30.5	19.0	161,522
CDS region	–	30.9	18.2	30.3	20.7	79,099
	Position 1	32	16.5	30.2	21.0	26,366
	Position 2	26	19.4	30.6	23.6	26,366
	Position 3	34	18.6	29.9	17.4	26,366

TABLE 3 Length of exons and introns in chloroplast genes of *Piper nigrum*.

Gene	Location	Exon1	Intron1	Exon2	Intron2	Exon3
trnk-UUU	LSC	37	2,548	35		
trnS-CGA	LSC	31	736	62		
rps16	LSC	40	823	254		
atpF	LSC	145	752	410		
rpoC1	LSC	432	771	1,620		
ycf3	LSC	124	823	230	726	153
trnL-UAA	LSC	35	489	50		
trnV-UAC	LSC	39	599	34		
rps12	LSC	114	–	31	537	137
clpP	LSC	71	882	293	639	245
petB	LSC					
petD	LSC					
rpl16	LSC					
rpl2	IR	385	665	431		
ndhB	IR	777	700	756		
trnI-GAU	IR	37	947	35		
trnA-UGC	IR	38	822	35		
ndhA	SSC	552	1,129	540		

tRNA genes and 13 protein-coding genes of which *ycf3* and *clpP* possess two introns each and the others contain only one intron (Geng et al., 2022). Similarly, *C. camphora* (Li et al., 2019) which also belongs to the magnoliids consists of 17 intron-containing genes which include 3 genes having two introns. In the case of the *P. nigrum* chloroplast genome, the protein-coding region accounts for 49.6% of the genome, while the tRNA and rRNA regions comprise 1.7% and 5.5% of the whole genome, respectively (Table 4). Therefore, the remaining genome consists of introns, intergenic regions, and pseudogenes. Within the Piperaceae, no change is observed in the number of protein-coding genes, rRNA, and GC percentage. Except for *P. auritum* which is reported to have 36 tRNAs and no pseudogene (NC_034697, unpublished), *P. cenocladum* (Cai et al., 2006) that contains no pseudogene, and *P. hancei* (Zhang et al., 2021) that consists of 36 tRNAs, the rest of the *Piper* species contain 37 tRNAs and 1 pseudogene (*ycf1*), respectively.

Considering the other members of the Magnoliid clade, *Persea* was found to have 81 protein-coding genes, whereas *Cinnamomum* and *Calycanthus* of Laurales (Chen et al., 2018) and *Liriodendron* of Magnoliales (Park et al., 2019) consist of 84 protein-coding genes. *Cinnamomum* and *Persea* comprise 36 tRNA, whereas *Calycanthus* and *Liriodendron* have 37 tRNAs. *Drimys* of the order Canellales consists of 85 protein-coding genes and 44 tRNAs which is the maximum number of tRNAs present within magnoliids (Table 4). In comparison to other magnoliids, *ycf1* is present as a pseudogene at the SSC/IRa junction in *Piper* species, whereas in *Drimys*, *Calycanthus*, and *Liriodendron*, the 3' fragment of *ycf1* is present at the IRb/SSC junction and a complete copy of *ycf1* is present at the SSC/IRa junction.

Codon usage

Codon usage for the *P. nigrum* chloroplast genome was calculated with the CodonW software. It was observed that 61 codons encoded for 20 amino acids and 3 stop codons. A total of 26,366 codons represented 85 protein-coding genes, 79,099 nucleotides in length. Among the amino acids, serine was the most (2,687 codons) and tryptophan was the least (550 codons) encoded amino acids (Figure 2). Codon usage bias is a phenomenon in which certain codons are preferred repeatedly over the others. This usage varies for different genomes and genes. Biasness was also observed in the presentation of a nucleotide at the third codon position. The AT content varies according to codon position with the highest at the third codon position, followed by the first and second codon positions (Morton, 2003). The above observation is also supported by

TABLE 4 Gene content comparison of *Piper nigrum* chloroplast with reported *Piper* species and other magnoliids.

Species	Total length of the cp genome (bp)	GC %	Protein-coding genes	rRNA-coding genes	tRNA-coding genes	Total number of genes	Number of pseudogenes
Calycanthus chinensis	153,596	39.3	84	8	37	129	0
Cinnamomum camphora	152,570	39.1	84	8	36	130	2
Drimys granadensis	160,604	38.8	85	8	44	137	0
Liriodendron tulipifera	159,886	39.2	84	8	37	129	0
Persea americana	152,723	39.1	81	8	36	127	2
Piper auritum	159,909	38.3	85	8	36	129	0
Piper cenocladum	160,624	38.3	85	8	37	130	0
Piper kadsura	161,486	38.3	85	8	37	131	1
Piper laetispicum	161,721	38.3	85	8	37	131	1
Piper nigrum	161,522	38.3	85	8	37	131	1

relative synonymous codon usage (RSCU) values. It was found that 40.7% of codons ended with A and T, whereas 23.4% of codons ended with C and G. It is known that organellar proteins are primarily encoded by codons having A or U at the third codon position. Similar results have been reported for various plastid genomes in the past (Cai et al., 2006; Asaf et al., 2018).

Gene content

The chloroplast genomes have evolved over a period of time. On comparing the algal and embryophyte plastomes, it is shown that a number of genes (*tufA*, *ftsH*, *odpB*, and *rpl5*) have been lost or transferred to the nucleus and a few genes, namely, *matK*, *ycf1*, and *ycf2*, have been gained by the abovementioned plastomes (Turmel et al., 2006). There also exist differences in the gene content among the angiosperms as genes like *infA*, *accD*, and *rpl22* have been reported to be lost in legumes (Doyle et al., 1995), and on the other hand, *accD*, *ycf1*, and *ycf2* have been lost in *Poaceae* (Guisinger et al., 2010). Complete loss or pseudogenization of *ndh* genes has been shown in heterotrophic plants in the past. Nonetheless, events showing *ndh* gene loss have also been observed in Pinaceae, gnetophytes, and autotrophic orchids (Wakasugi et al., 1994; Braukmann et al., 2009; Kim et al., 2015).

Among the Mesangiospermae, magnoliids are the third largest group after the eudicots and monocots. Except for a few differences in gene content, the magnoliids are identical to tobacco and many unarranged angiosperms in terms of gene order. The *ACRS* gene, which has been reported in a number of plastid genomes as the *ycf68* gene, was identified in *Calycanthus* based on its similarity to a mitochondrial ACR sensitivity toxin gene of *Citrus jambhiri* (Ohtani et al., 2002). Due to the presence of internal stop codons, it has been reported as a pseudogene in a number of plastomes (Curci et al., 2015; Yan et al., 2015). However, it was not annotated in the *P. nigrum* chloroplast genome and hence observed to be absent from the plastid genome. Similar observations were made for *D. granatenis* and

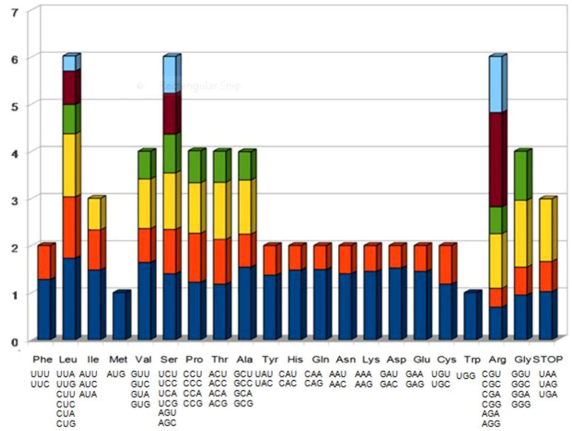


FIGURE 2
Codon usage of the CDS region of the chloroplast genome of *Piper nigrum*.

P. cenocladum (Cai et al., 2006), belonging to the magnoliid clade. Likewise, another gene, namely, *ycf15*, was not annotated in the *P. nigrum* plastid genome and hence was found to be absent. The *ycf15* has also been reported to be a non-functional protein-coding gene (Shi et al., 2013) and hence a pseudogene due to the presence of internal stop codons (Yan et al., 2015; Kaila et al., 2017). It is found to be present in the *Calycanthus* plastome but absent in the *D. granatenis* and *P. cenocladum* plastid genomes (Cai et al., 2006). If we consider the phenomenon of nuclear substitution, then one such gene which is well-studied is *rps16*. The *rps16* gene has been found to be present as a pseudogene in some plastid genomes (Roy et al., 2010; Kaila et al., 2016) or has been completely lost from the plastid genome due to non-functional protein sequence or incorrect splicing of intron due to mutation at the 5' or 3' splicing site. Like in certain legumes, the loss of *rps16* from the chloroplast genome and its replacement by nuclear copy have been reported (Ueda et al., 2008). However, in *P. nigrum*, its copy has been retained in the chloroplast genome similar also to *P. cenocladum*, *P. laetispicum*, and *P. auritum*. As mentioned above, the chloroplast genome of *Piper* is similar to tobacco, and it has been shown that *rps16* is indispensable for tobacco plastome function (Fleischmann et al., 2011). A better regulation of plastome *rps16* gene in certain conditions or the non-presence of chloroplast-targeting sequence in nuclear-encoded peptide has been proposed as the reasons behind the retention of plastid *rps16* gene in some plants (Keller et al., 2017). Some differences in gene content may also lead to the expansion of the IR region, which is a common phenomenon occurring in the plastid genome. One such example is duplication and inclusion of the *trnH* gene in the IR region of the *D. granatenis* chloroplast genome, hence leading to the expansion of the IR region (Cai et al., 2006). Such duplication of the *trnH* gene has not been observed in any other plant belonging to magnoliids or basal angiosperms. Similarly, such duplication was not observed in the *P. nigrum* plastid genome.

Repeat structures

Repeat structure analysis through the REPuter software (Kurtz et al., 2001) revealed that *P. nigrum* comprises 23 palindromic, 18 forward, 2 complement, and 1 reverse repeat along with 55 tandem repeats that were predicted using the Tandem Repeats Finder. The length of the largest forward and palindromic repeat was 52 bp each. Among the tandem repeats, 64% of the repeats were present in the intergenic region, 28% in the coding region, and 8% in the intronic region. Palindromic repeats were observed to be present in majority within all the reported species of *Piper* (Figure 3). Reverse and complement repeats were absent in *P. laetispicum*, whereas no complement repeat was present in *P. auritum*. *Piper auritum* had a minimum number of tandem repeats (Wang et al., 2018; NC_034697).

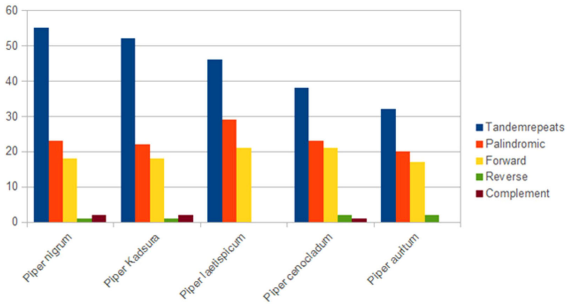


FIGURE 3
Repeat sequences in five *Piper* species plastid genomes calculated using REPuter and Tandem Repeats Finder.

Likewise, palindromic and forward repeats were the most abundant repeat types in other plant types like *P. americana* (Liu et al., 2019) and *Aristolochia contorta* (Zhou et al., 2017). These repeats have been reported to be responsible for chloroplast genome rearrangement, gene duplication, and gene expression (Do et al., 2014; Vieira et al., 2014). It is due to their role in genome rearrangement that they have proved to be helpful in phylogenetic studies (Cavalier-Smith, 2002). Improper recombination and slipped strand mispairing of these repeat units result in plastome rearrangement and sequence variation among different plastid genomes. These repeats are also responsible for substitutions and indels in the chloroplast genome (Yi et al., 2013). They have also proved to be an informational source for developing markers and therefore play an important role in population and phylogenetic studies (Nie et al., 2012).

Simple sequence repeats

A group of tandem repeats comprising DNA sequences of 1-6 nucleotides long are called microsatellites or SSRs. They are present in large numbers in both coding and non-coding regions. They are highly reproducible, abundant in number, polymorphic, and uniparentally inherited. Variation in SSRs at a specific locus contributes to its use as a molecular marker (Olmstead and Palmer, 1994). Therefore, they are used in evolutionary studies, phylogenetic relationships, and plant population genetics. A total of 216 SSRs were found using the MISA perl script [http://pgrc.ipk-gatersleben.de/misa/] (Beier et al., 2017). The number of SSRs extracted was similar to those reported in other *Piper* species like *P. kadsura* (215 SSRs), *P. laetispicum* (216 SSRs), *P. cenocladum* (198 SSRs), and *P. auritum* (189 SSRs). Among the abovementioned 216 SSRs, 65% were present in the LSC region, 13% in the SSC region, and 11% in the IR regions. It was also observed that 32% of the SSRs were present in the coding region and 13% in the non-coding region, and the maximum SSRs (53%) were present

in the intergenic region. Thus, the maximum number of SSRs is present in the non-coding region than in the coding region. Similar observations have also been made for *P. taeda* (Asaf et al., 2018) and *Cyamopsis tetragonoloba* (Kaila et al., 2017). If we look individually at the coding region, then it is observed that the *ycf1* gene had a maximum number of repeats. Previous studies have also reported the *ycf1* gene as the most variable locus, and similar observations have been made for *Glycine* species, *Cynara cardunculus*, and *Camellia* species (Huang et al., 2014; Curci et al., 2015; Ozyigit et al., 2015). Among the 216 SSRs found, the most abundant were mononucleotide repeats (147), followed by dinucleotide (49), trinucleotide (6), tetranucleotide (9), pentanucleotide (3), and hexanucleotide repeats (2). On comparison, it was found that all the repeat types were present in different *Piper* species except pentanucleotides and hexanucleotides, which were absent in *P. cenocladum* and *P. auritum* (Figure 4). In *P. nigrum*, the A/T repeats were found to be in majority with 138 in number out of 147 mononucleotide repeats present. The AT/AT was the most common among the dinucleotide repeats and five out of six trinucleotide repeats were AAT/ATT (Figure 5). Such A/T richness of mononucleotides has been observed earlier by Wheeler et al. (2014). This biasness in base composition and the presence of polyadenine (A) or polythymine (T) rich repeats have been reported earlier in species like *C. tetragonoloba* (Kaila et al., 2017), *P. taeda* (Asaf et al., 2018), *P. bournei* (Li et al., 2017), and *Aristolochia debilis* (Zhou et al., 2017).

Wet lab validation at 11 SSR loci across 12 *Piper* species (including *P. nigrum*) resulted in the amplification of 60 alleles (Figure 6). The 11 loci included di-, tri-, tetra-, and pentanucleotide repeat core motifs (Table 5). The number of alleles detected ranged from 4 to 11, with the maximum number of alleles detected at the locus at BPC8 with a pentanucleotide core repeat. Amplification at this locus indicates the presence of pentanucleotide repeats in *P. longum*, *P. arboreuum*, *P. argyrophyllum*, *P. attenuatum*, *P. betel*, *P. chaba*, *P. trichostachyon*, *P. sarmentosum*, *P. columbrinum*, *P. wallichi*, and *P. hymenophyllum*. The high degree of polymorphism detected across the 12 *Piper* species allows for the identification of hypervariable regions informative for DNA barcoding and is a source for further evolutionary and phylogenetic studies in *Piper* species.

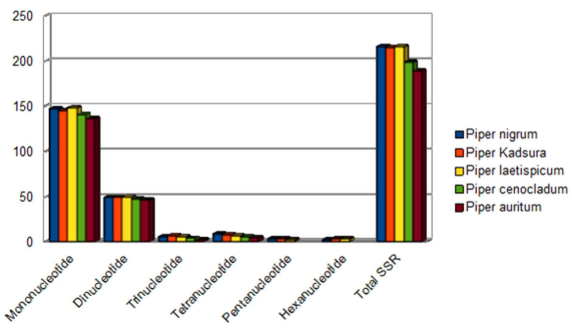


FIGURE 4
Simple sequence repeats (SSRs) in plastid genomes of five *Piper* species.

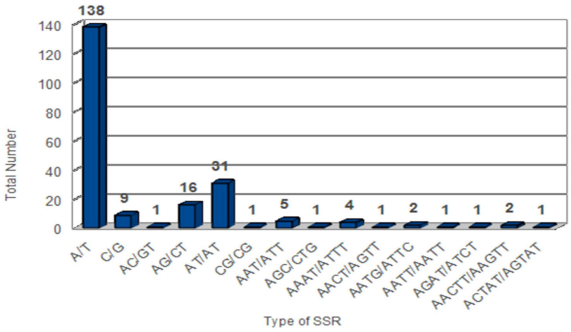


FIGURE 5
Distribution of SSRs based on their repeat type in *Piper nigrum*..

Comparison of IR boundaries

The IR regions provide stability to the chloroplast genome by exhibiting intramolecular recombination between the two IR copies (Palmer et al., 1987; Palmer, 1991). The expansion and contraction of IRs lead to the conversion of single-copy genes to duplicate copies and vice versa, respectively, thereby leading to variation in the plastome size among the angiosperms (Sabir et al., 2014). The chloroplast genomes of all *Piper* species are highly conserved, but expansion and contraction at the IR/SSC boundary regions leads to variation in the size of the genomes (Goulding et al., 1996). Here, we have compared the IR/SSC boundaries of *P. nigrum* with four other *Piper* species, namely, *P. kadsura*, *P. laetispicum*, *P. cenocladum*, and *P. auritum*. It was observed that 10 completely duplicated genes were present in the IR region of the *P. nigrum* chloroplast genome. In Figure 7, it can be seen that an intergenic region of 1,106 bp is present between *ndhF* and *ycf1* genes at the SSC/IRa junction in *P. cenocladum*, whereas *ndhF* and *ycf1* genes are overlapping at the SSC/Ira junction in other *Piper* species. The size of an overlapping fragment of the *ndhF* gene varies in different *Piper* species, hence contributing to fluctuating IR boundaries among different plastomes. The *ycf1* gene (5,474 bp) is present at the IRb/SSC junction in *P. nigrum*, *P. kadsura*, and *P. laetispicum*, whereas it is absent in *P. cenocladum* and *P. auritum*. A fragment of the *ycf1* gene of different sizes is present as a pseudogene at the SSC/IRa junction in all the species. The IR region of *P. nigrum* (27,057 bp) was observed to be larger than *P. kadusra* (24,657 bp), *P. laetispicum* (26,133 bp), and *P. cenocladum* (27,039 bp).

Gene order

The plastomes of five *Piper* species, viz., *P. laetispicum* (Wang et al., 2018), *P. auritum* (NC_034697), *P. cenocladum* (Cai et al.,

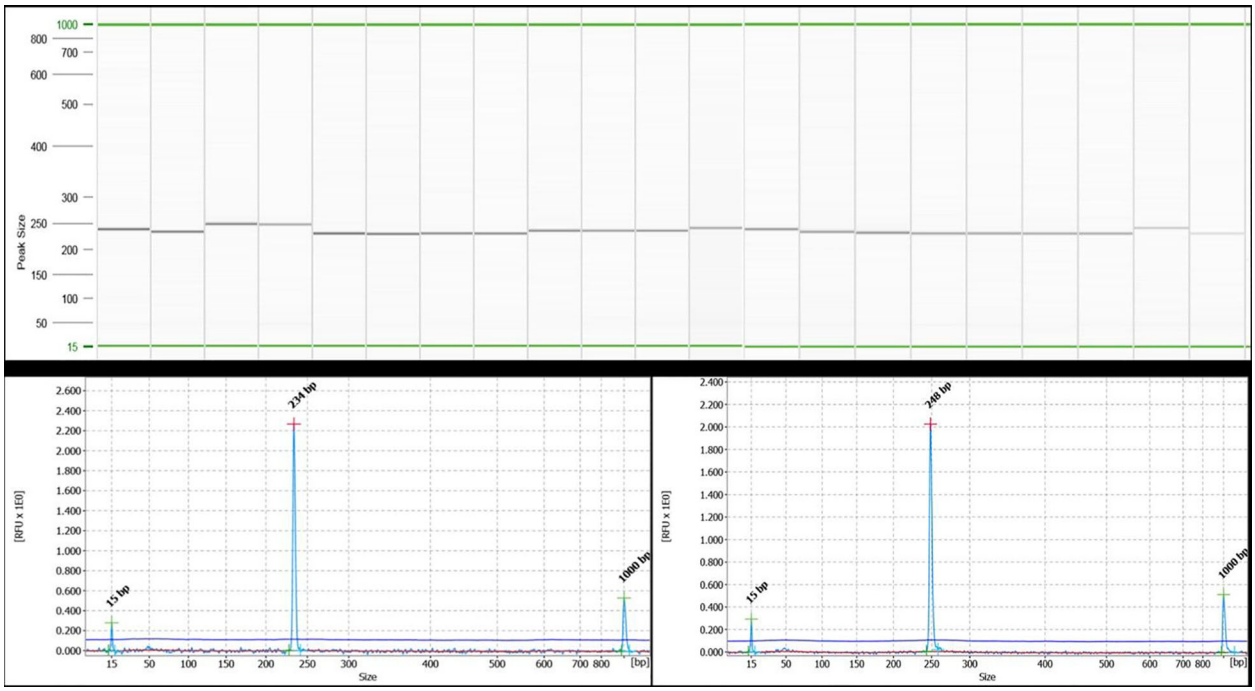


FIGURE 6
QIAxcel gel image of PCR amplification using the black pepper chloroplast SSR marker BPC5 in 21 accessions of 12 different species of black pepper as captured on the QIAxcel ScreenGel software. Lanes 1 and 2: *Piper longum*; lanes 3 and 4: *Piper arboreum*; lanes 5 and 6: *Piper argyrophylum*; lane 7: *Piper attenuatum*; lanes 8–10: *Piper betel*; lanes 11 and 12: *Piper chaba*; lane 13: *Piper hymenophyllum*; lanes 14 and 15: *Piper trichostachyon*; lane 16: *Piper wallichi*; lanes 17–19: *Piper nigrum*; lane 20: *Piper columbrinum*; lane 21: *Piper sarmentosum*. The lane marked as "M" is DNA molecular weight standard 50-800 bp v2.0 Qx DNA size marker. A representative electropherogram showing allele size of 234 bp for lane 2 and 248 bp for lane 3 has been shown below the gel image.

2006), *P. nigrum*, and *P. kadsura* (Lee et al., 2016) were compared with the help of the Mauve software (Darling et al., 2004). In the comparison of different *Piper* species, no change was observed in the gene order within Piperaceae except for the absence of the *ycf1* gene at the IRb/SSC junction in *P. cenocladum* and *P. auritum* (Figure 8). On analysis, it was found that the sequence of the *ycf1* gene was present at the IRb/SSC junction in both the *P. cenocladum* and *P. auritum* plastid genomes, but it is not being annotated. Therefore, it can be inferred that a significant level of synteny can be seen between the plastid genome of different *Piper* species. One difference that can be accounted for is the inclusion of the *ycf1* gene in the IR region, which varies among different plastid genomes and therefore contributes to variation in IR size and ultimately to the chloroplast genome size.

Phylogenetic analysis

Chloroplast genomes have proved to be useful in evolutionary, phylogenetic, and molecular systematic studies. They are an important resource for exploring evolutionary histories within and among different species. Various studies have compared intergenic spacers, protein-coding genes, and complete chloroplast genome sequences and have, therefore, contributed to our understanding of evolutionary relationships among different angiosperms and gymnosperms. In the present investigation, the phylogenetic data set included 58 protein-coding genes (38,262 bp after the removal of poorly aligned regions) from 37 taxa, comprising 34 angiosperms and 3 gymnosperms. Maximum likelihood analysis resulted in a single tree with higher bootstrap values (Figure 9). Twenty-eight out of 35 nodes showed 100% bootstrap value. The 100% bootstrap value strongly supports the monophyly of magnoliids. Magnoliids with 100% support value form two distinct clades: one with Canellales and Piperales and the other Magnoliales and Laurales. Piperales with four species of *Piper* got further divided into two clades with 100% bootstrap value. *Piper nigrum* is positioned with *P. kadsura* in one clade, and *P. auritum* and *P. cenocladum* were observed to be more similar. Our findings were in agreement with the results reported by Wang et al. (2018), where *P. nigrum* and *P. kadsura* were placed in one clade.

10.3389/fpls.2022.1095781

TABLE 5 Details of 11 chloroplast SSR primers validated in 12 *Piper* species (21 accessions).

Primer ID	Primer sequence	Repeat motif	No. of alleles	Annealing temp. (°C)	Allele size (bp)
BPC1	f-CGAATACACCAGCTACGCCT r-AGTTTCCGTCTGGGTATGCG	(CAG)4	4	55	220
BPC2	f-CTGGGCGCGAGGACTAAAAA r-TGAGTGGACTAAAGCGGCAG	(GA)4	7	55	280
BPC3	f-CCCCCAACTCAACTAGGTCG r-GGTATTTCGGCGAGTCCTCA	(AT)4	5	55	271
BPC4	f-TCCAAATCAATCCTGCGGGT r-GGCTTGGCCGAAGAACTTCA	(TC)4	5	55	178
BPC5	f-TTGGCCCACTCTTCATCGAC r-GTGATGGAGTTTTGTTTCCAGGA	(TG)5	7	55	231
BPC6	f-TTCCACTCACCCCATGGTTG r-TCAGATTCAAAACGGCTTGCT	(AATG)3	5	55	280
BPC7	f-GTCCTGCGAAGGGAAGGAAA r-ACTTTGGGATTGCTGAATTGCA	(AG)4	4	55	272
BPC8	f-GAACTTTTTGGTTTGGGGGCT r-GTGCTGATCGTTACAGGCCT	(ATAGT)3	11	55	278
BPC9	f-TGAGTTGGGCGCTTTAACCA r-ACTCCAAATTCGGGGGTCAA	(GA)4	4	55	273
BPC10	f-ACGACAGGCAAACCTCTAGA r-GCACCAATCCCGCTTTCTTG	(TTTA)3	4	55	26
BPC11	f-CGCCAGACAAAAGTCAGAACA r-ATGCGGTTAAACCAGCTGTG	(CT)4	4	55	210

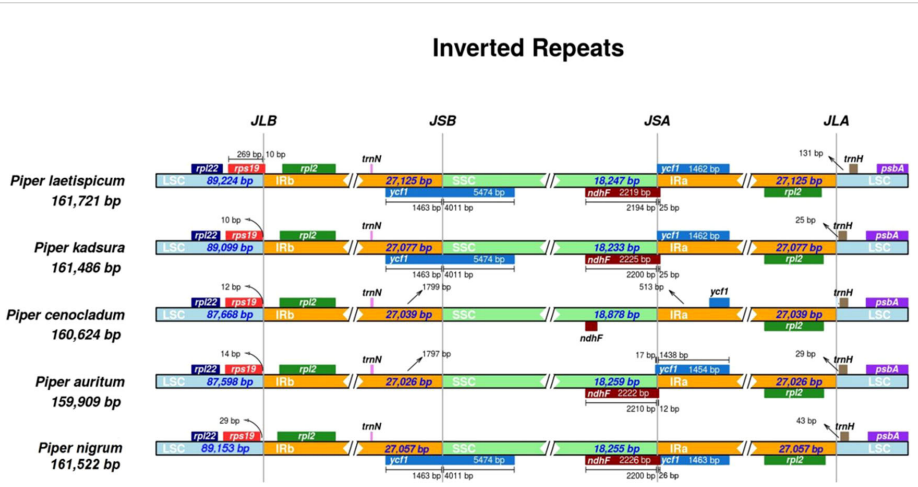


FIGURE 7
Contraction and expansion at junctions of four regions (LSC, SSC, IRa, IRb) of the plastid genome of five *Piper* species.

Conclusion

In the present investigation, the chloroplast genome of *P. nigrum* was sequenced by integrating two platforms: Illumina and PacBio. The sequenced chloroplast genome was assembled using the Organelle PBA software and the Pilon software. The chloroplast genome having 161,522 bp length showed a typical quadripartite structure with 81 protein-coding genes, 37 tRNAs,

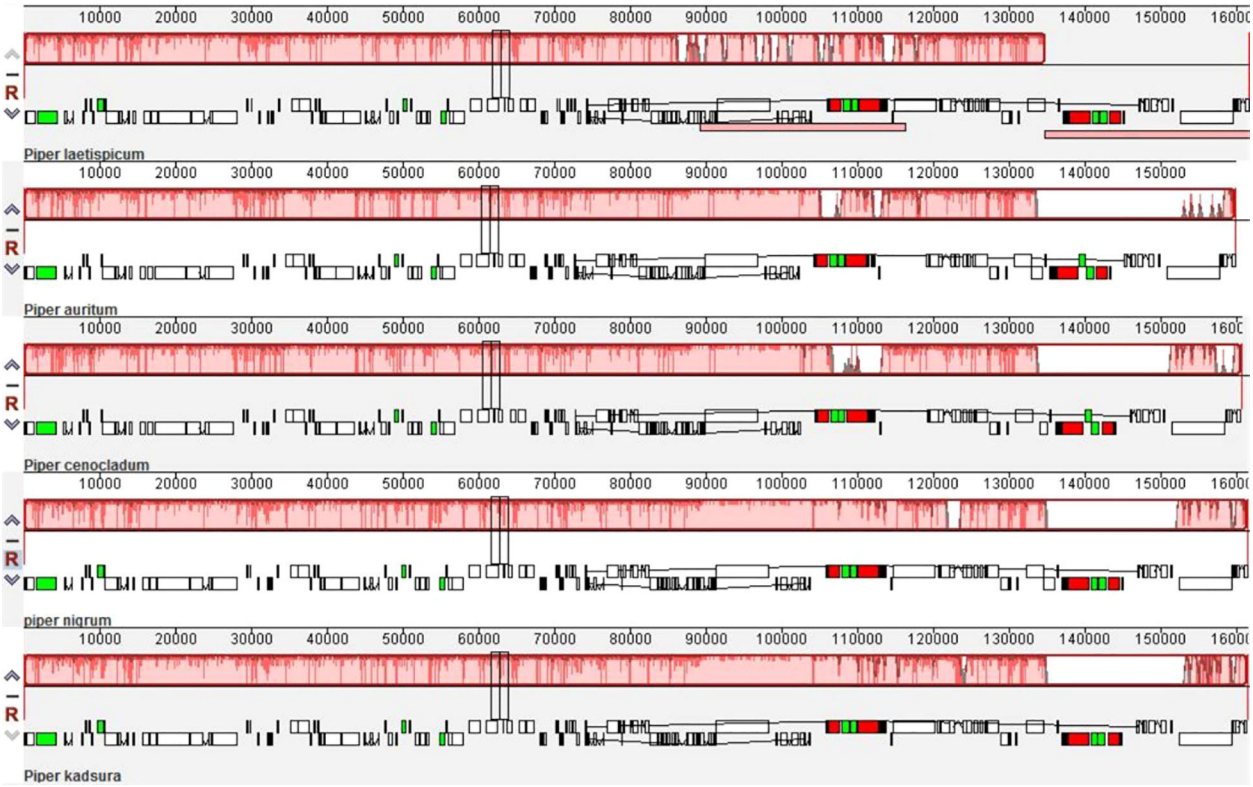


FIGURE 8
Gene order comparison of the chloroplast genome within Piperaceae.

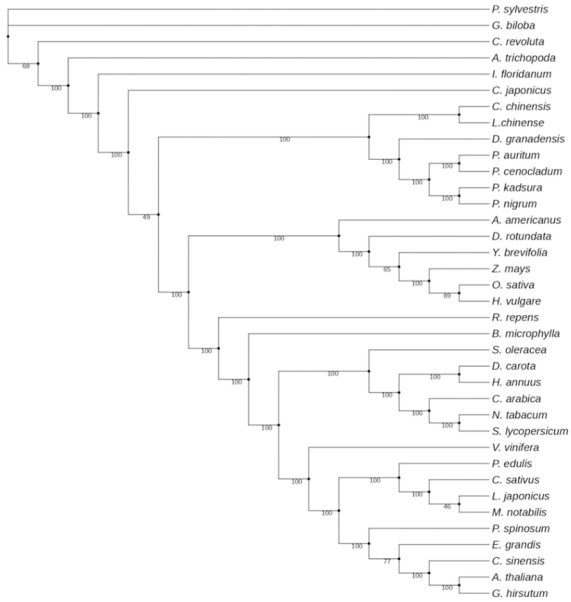


FIGURE 9
Phylogenetic tree of 37 different species based on 58 protein-coding genes.

4 rRNAs, and 1 pseudogene. Codon usage analysis revealed that serine was the most coded amino acid, and 61 codons were present for coding 20 amino acids. The gene *rps16*, which has been observed to be lost from some of the plastid genomes, is present in the *P. nigrum* chloroplast genome similar also to other *Piper* species. Repeat analysis was also done, and there were 23 palindromic, 18 forward, 2 complement, and 1 reverse repeat along with 55 tandem repeats that were predicted. Of the 216 SSRs identified, the majority were present in the intergenic region. Phylogenetic analysis done with 58 protein-coding genes for 37 taxa, comprising 34 angiosperms and 3 gymnosperms, revealed the presence of *P. nigrum* and *P. kadsura* in one clade. This study has laid the foundation for future evolutionary and molecular studies on Piperales species.

Data availability statement

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

Author contributions

Conceptualization and funding acquisition: AG and KB. Investigation and data analysis: AG, TK, AM, and RK. Manuscript preparation: AG, TK, PR and DW. All authors contributed to the article and approved the submitted version.

Funding

This work is funded by the grant received under the project Indian Council of Agricultural Research-Consortium Research Platform on Genomics (project number 1007341).

Acknowledgments

The authors acknowledge the financial support from the ICAR-Consortium Research Platform on Genomics and the facilities provided by the ICAR-National Bureau of Plant Genetics Resources, New Delhi.

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.2022.1095781/full#supplementary-material

References

Ahmad, N., Fazal, H., Abbasi, B. H., Farooq, S., Ali, M., and Khan, M. A. (2012). Biological role of *Piper nigrum* (Black pepper): A review. *Asian Pac. J. Trop. BioMed.*, 5 1945–1953. doi: 10.1016/S2221-1691(12)60524-3

Andrews, S. (2010) *FastQC: a quality control tool for high throughput sequence data*. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastq.

Asaf, S., Khan, A. L., Khan, M. A., Shahzad, R., Kang, S. M. L., Harrasi, A. A., et al. (2018). Complete chloroplast genome sequence and comparative analysis of loblolly pine (*Pinus taeda* l.) with related species. *PloS One* 13 (3), e0192966. doi: 10.1371/journal.pone.0192966

Beier, S., Thiel, T., Mu¨ nch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33 (16), 2583–2585. doi: 10.1093/bioinformatics/btx198

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27 (2), 573–580. doi: 10.1093/nar/27.2.573

Bock, R. (2007). "Structure, function, and inheritance of plastid genomes," in *Cell and molecular biology of plastids*. Ed. R. Bock (Berlin, Heidelberg: Springer), 29–63.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30 (15), 2114–2120. doi: 10.1093/bioinformatics/btu170

Braukmann, T. W. A., Kuzmina, M., and Stefanovic, S. (2009). Loss of all plastid ndh genes in gnetales and conifers: extent and evolutionary significance for the seed plant phylogeny. *Curr. Genet.* 55 (3), 323–337. doi: 10.1007/s00294-009-0249-7

Cai, Z., Penaflor, C., Kuehl, J. V., Leebens-Mack, J., Carlson, J. E., dePamphilis, C. W., et al. (2006).). complete plastid genome sequences of *Drimys, liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids. *BMC Evol. Biol.* 6, 77. doi: 10.1186/1471-2148-6-77

Cavalier-Smith, T. (2002). Chloroplast evolution: Secondary symbiogenesis and multiple losses. *Curr. Biol.* 12 (2), R62–R64. doi: 10.1016/S0960-9822(01)00675-3

Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., Soltis, D. E., et al. (2016). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc* 181 (1), 1–20. doi: 10.1111/boj.12385

Chen, X., Yang, J., Zhang, H., Bai, R., Zhang, X., Bai, G., et al. (2018). The complete chloroplast genome of *Calycanthus chinensis*, an endangered species endemic to china. conserv. *Genet. Resour.* 11, 55–58. doi: 10.1007/s12686-017-0966-z

Curci, P. L., De Paola, D., Danzi, D., Vendramin, G. G., and Sonnante, G. (2015). Complete chloroplast genome of the multifunctional crop globe artichoke and comparison with other asteraceae. *PloS One* 10, e120589. doi: 10.1371/journal.pone.0120589

Daniell, H., Lin, C. S., Yu, M., and Chang, W. J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17, 134. doi: 10.1186/s13059-016-1004-2

Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704

Divya, K. G., Nair, M. C., Shaji, P. K., and Nair, P. K. K. (2015). Pollen morphology of pepper cultivars and their wild allies from southern Western ghats, kerala, India. *Int. J. Adv. Res.* 3 (3), 344–353.

Do, H. D. K., Kim, J. S., and Kim, J. H. (2014). A trnI CAU triplication event in the complete chloroplast genome of *Paris verticillata* m. bieb. (Melanthiaceae, liliales). *Genome Biol. Evol.* 6 (7), 1699–1706. doi: 10.1093/gbe/evu138

Dong, W., Xu, C., Cheng, T., Lin, K., and Zhou, S. (2013). Sequencing angiosperm plastid genomes made easy: A complete set of universal primers and a case study on the phylogeny of saxifragales. *Genome Biol. Evol.* 5, 989–997. doi: 10.1093/gbe/evt063

Doyle, J. J., Doyle, J. L., and Palmer, J. D. (1995). Multiple independent losses of 2 genes and one intron from legume chloroplast genomes. *Syst. Bot.* 20 (3), 272–294. doi: 10.2307/2419496

Fleischmann, T. T., Scharff, L. B., Alkatib, S., Hasdorf, S., Schottler, M. A., and Bock, R. (2011). Nonessential plastid-encoded ribosomal proteins in tobacco: A developmental role for plastid translation and implications for reductive genome evolution. *Plant Cell.* 23, 3137–3155. doi: 10.1105/tpc.111.088906

Geng, X., Zhu, Y., Ren, Z., Chen, R., and Liu, Q. (2022). The complete chloroplast genome of *Piper sarmentosum* roxburg 1820 (Piperaceae). *Mitochondrial DNA Part B.* 7 (5), 854–855. doi: 10.1080/23802359.2022.2074805

Goulding, S. E., Wolfe, K. H., Olmstead, R. G., and Morden, C. W. (1996). Ebb and flow of the chloroplast inverted repeat. *Mol. Genet. Genom.* 252, 195–206. doi: 10.1007/bf02173220

Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic. Acids Res.* 47 (W1), W59–W64. doi: 10.1093/nar/gkz238

Guisinger, M., Chumley, T., Kuehl, J., Boore, J., and Jansen, R. (2010). Implications of the plastid genome sequence of typha (Typhaceae, poales) for understanding genome evolution in poaceae. *J. Mol. Evol.* 70 (2), 149–166. doi: 10.1007/s00239-009-9317-3

Huang, H., Shi, C., Liu, Y., Mao, S., and Gao, L. (2014). Thirteen camellia chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evol. Biol.* 14, 1–17. doi: 10.1186/1471-2148-14-151

Kaila, T., Chaduvla, P. K., Rawal, H. C., Saxena, S., Tyagi, A., Mithra, S., et al. (2017). Chloroplast genome sequence of clusterbean (*Cyamopsis tetragonoloba* l.): Genome structure and comparative analysis. *Genes* 8 (9), 212. doi: 10.3390/genes8090212

Kaila, T., Chaduvla, P. K., Saxena, S., Bahadur, K., Gahukar, S. J., Chaudhury, A., et al. (2016). Chloroplast genome sequence of pigeonpea (*Cajanus cajan* (L.) millspaugh) and *Cajanus scarabaeoides* (L.) thouars: Genome organization and comparison with other legumes. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01847

Keller, J., Rousseau-Gueutin, M., Martin, G. E., Morice, J., Boutte, J., Coissac, E., et al. (2017). The evolutionary fate of the chloroplast and nuclear rps16 genes as revealed through the sequencing and comparative analyses of four novel legume chloroplast genomes from lupinus. *DNA Res.* 24 (4), 343–358. doi: 10.1093/dnares/dsx006

Kim, H. T., Kim, J. S., Moore, M. J., Neubig, K. M., Williams, N. H., Whitten, W. M., et al. (2015). Seven new complete plastome sequences reveal rampant independent loss of the *ndh* gene family across orchids and associated instability of the inverted Repeat/Small single-copy region boundaries. *PloS One* 10 (11), e0142215. doi: 10.1371/journal.pone.0142215

Kubitzki, K., and Rohwer, J. G. (1993). "Betrich," in *Flowering plants, dicotyledons magnolid, hamamelid and caryophyliid families* (Berlin, Germany: Springer Verlag).

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29 (22), 4633–4642. doi: 10.1093/nar/29.22.4633

Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32 (1), 11–16. doi: 10.1093/nar/gkh152

Lee, J. H., Choi, I. S., Choi, B. H., Yang, S., and Choi, G. (2016). The complete plastid genome of *Piper kadsura* (Piperaceae), an East Asian woody vine. mitochondrial DNA a: DNA mapp. *Seq. Anal.* 27 (5), 3555–3556. doi: 10.3109/19401736.2015.1074216

Li, J., Fan, R., Xu, J., Hu, L., Su, F., and Hao, C. (2022). Comparative analysis of the chloroplast genomes of eight *Piper* species and insights into the utilization of structural variation in phylogenetic analysis. *Front. Genet.* 13, 925252. doi: 10.3389/fgene.2022.925252

Li, P., Jia, G., Xin, G., and Cai, X. (2019). The complete chloroplast genome of *Cinnamomum camphora* (L.) presl., a unique economic plant to China. *Mitochondrial DNA Part B.* 4 (2), 2511–2512. doi: 10.1080/23802359.2019.1640083

Li, Y., Xu, W., Zou, W., Jiang, D., and Liu, X. (2017). Complete chloroplast genome sequences of two endangered phoebe (Lauraceae) species. *Bot. Stud.* 58 (1), 37. doi: 10.1186/s40529-017-0192-8

Liu, J., Gong, L. D., Qi, L., Zheng, C., Niu, Y. F., and Shi, C. (2019). Complete plastid genome of *Persea americana* mill. (Ebenales) and phylogenetic analysis. *Mitochondrial DNA Part B.* 4 (1), 2033–2034. doi: 10.1080/23802359.2019.1617073

Massoni, J., Couvreur, T. L., and Sauquet, H. (2015). Five major shifts of diversification through the long evolutionary history of magnoliidae (angiosperms). *BMC Evol. Biol.* 15, 49. doi: 10.1186/s12862-015-0320-6

Mayor, C., Brudno, M., Schwartz, J. R., Poliakov, A., Rubin, E. M., Frazer, K. A., et al. (2000). VISTA: Visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* 16, 1046. doi: 10.1093/bioinformatics/16.11.1046

Morton, B. R. (2003). The role of context-dependent mutations in generating compositional and codon usage bias in grass chloroplast DNA. *Mol. Evol.* 56, 616–629. doi: 10.1007/s00239-002-2430-1

Nie, X., Lv, S., Zhang, Y., Du, X., Wang, L., Biradar, S. S., et al. (2012). Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PloS One* 7, e36869. doi: 10.1371/journal.pone.0036869

Ohtani, K., Yamamoto, H., and Akimitsu, K. (2002). Sensitivity to *Alternaria alternata* toxin in citrus because of altered mitochondrial RNA processing. *Proc. Natl. Acad. Sci. U.S.A.* 99, 2439–2444. doi: 10.1073/pnas.042448499

Olmstead, R. G., and Palmer, J. D. (1994). Chloroplast DNA systematics: a review of methods and data analysis. *Am. J. Bot.* 8, 1205–1224. doi: 10.2307/2445483

Ozyigit, I. I., Dogan, I., and Filiz, E. (2015). *In silico* analysis of simple sequence repeats (SSRs) in chloroplast genomes of glycine species. *Plant Omics.* 8, 24–29.

Palmer, J. D. (1991). "Plastid chromosomes: Structure and evolution," in *Molecular biology of plastids*. Ed. L. Bogorad (San Diego, CA, USA: Academic Press), 5–53.

Palmer, J. D., Osorio, B., Aldrich, J., and Thompson, W. F. (1987). Chloroplast DNA evolution among legumes: Loss of a large inverted repeat occurred prior to other sequence rearrangements. *Curr. Genet.* 11, 275–286. doi: 10.1007/BF00355401

Park, J., Kim, Y., Kwon, W., Xi, H., and Kwon, M. (2019). The complete chloroplast genome of tulip tree, *Liriodendron tulifipera* l. (Magnoliaceae): investigation of intra-species chloroplast variations. *Mitochondrial DNA Part B.* 4 (2), 2523–2524. doi: 10.1080/23802359.2019.1598822

Ravindran, P. (2000). *Black pepper* (London: CRC Press). doi: 10.1201/9780203303870

Roy, S., Ueda, M., Kadowaki, K., and Tsutsumi, N. (2010). Different status of the gene for ribosomal protein S16 in the chloroplast genome during evolution of the genus *Arabidopsis* and closely related species. *Genes Genet. Syst.* 85, 319–326. doi: 10.1266/ggs.85.319

Rozas, U., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of Large data sets. *Mol.Biol. Evol.* 34, 12, 3299–3302. doi: 10.1093/molbev/msx248

Sabir, J., Schwarz, E., Ellison, N., Zhang, J., Baeshen, N. A., Mutwakil, M., et al. (2014). Evolutionary and biotechnology implications of plastid genome variation in the inverted-repeat-lacking clade of legumes. *Plant Biotechnol. J.* 12, 743–754. doi: 10.1111/pbi.12179

Shi, C., Liu, Y., Huang, H., Xia, E. H., Zhang, H. B., and Gao, L. Z. (2013). Contradiction between plastid gene transcription and function due to complex post transcriptional splicing: an exemplary study of ycf15 function and evolution in angiosperms. *PloS One* 8, e59620. doi: 10.1371/journal.pone.0059620

Sinn, B. T., Sedmak, D. D., Kelly, L. M., and Freudenstein, J. V. (2018). Total duplication of the small single copy region in the angiosperm plastome: Rearrangement and inverted repeat instability in asarum. *Am. J. Bot.* 105 (1), 71–84. doi: 10.1002/ajb2.1001

Soorni, A., Haak, D., Zaitlin, D., and Bombarely, A. (2017). Organelle_PBA, a pipeline for assembling chloroplast and mitochondrial genomes from PacBio DNA sequencing data. *BMC Genomics* 18 (1), 49. doi: 10.1186/s12864-016-3412-9

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22 (22), 4673–4680. doi: 10.1093/nar/22.22.4673

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45 (W1), W6–W11. doi: 10.1093/nar/gkx391

Tsudzuki, J., Ito, S., Tsudzuki, T., Wakasugi, T., and Sugiura, M. (1994). A new gene encoding tRNAPro (GGG) is present in the chloroplast genome of black pine: A compilation of 32 tRNA genes from black pine chloroplasts. *Curr. Genet.* 26, 153–158. doi: 10.1007/BF00313804

Turmel, M., Otis, C., and Lemieux, C. (2006). The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. *Mol. Biol. Evol.* 23 (6), 1324–1338. doi: 10.1093/molbev/msk018

Ueda, M., Nishikawa, T., Fujimoto, M., Takanashi, H., Arimura, S. I., Tsutsumi, N., et al. (2008). Substitution of the gene for chloroplast *RPS16* was assisted by generation of a dual targeting signal. *Mol. Biol. Evol.* 25, 1566–1575. doi: 10.1093/molbev/msn102

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3–new capabilities and interfaces. *Nucleic Acids Res.* 40 (15), e115. doi: 10.1093/nar/gks596

Vieira, L., Faoro, H., Rogalski, M., Fraga, H., Cardoso, R. L. A., de Souza, E. M., et al. (2014). The complete chloroplast genome sequence of *Podocarpus lambertii*: genome structure, evolutionary aspects, gene content and SSR detection. *PloS One* 9 (3), e90618. doi: 10.1371/journal.pone.0090618

Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T., and Sugiura, M. (1994). Loss of all ndh genes as determined by sequencing the entire chloroplast genome of the black pine pinus thunbergii. *Proc. Natl. Acad. Sci. U.S.A.* 91 (21), 9794–9798. doi: 10.1073/pnas.91.21.9794

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One* 9 (11), e112963. doi: 10.1371/journal.pone.0112963

Wang, M. T., Wang, J. H., Zhao, K. K., Zhu, Z. X., and Wang, H. F. (2018). Complete plastome sequence of *Piper laetispicum* (Piperaceae): An endemic plant species in south China. *Mitochondrial DNA Part B.* 3 (2), 1035–1036. doi: 10.1080/23802359.2018.1511850

Wheeler, G. L., Dorman, H. E., Buchanan, A., Challagundla, L., and Wallace, L. E. (2014). A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. *Appl. Plant Sci.* 2, 1400059. doi: 10.3732/apps.1400059

Worth, J. R. P., Liu, L., Wei, F. J., and Tomaru, N. (2019). The complete chloroplast genome of *Fagus crenata* (subgenus *Fagus*) and comparison with *F. engleriana* (subgenus *Engleriana*). *PeerJ* 7, e7026. doi: 10.7717/peerj.7026

Yan, L., Lai, X., Li, X., Wei, C., Tan, X., and Zhang, Y. (2015). Analyses of the complete genome and gene expression of chloroplast of sweet potato [*Ipomoea batata*]. *PloS One* 10, e124083. doi: 10.1371/journal.pone.0124083

Yi, X., Gao, L., Wang, B., Su, Y. J., and Wang, T. (2013). The complete chloroplast genome sequence of *Cephalotaxus oliveri* (Cephalotaxaceae): Evolutionary comparison of cephalotaxus chloroplast DNAs and insights into the loss of inverted repeat copies in gymnosperms. *Genome Biol. Evol.* 5 (4), 688–698. doi: 10.1093/gbe/evt042

Zhang, L., Hu, X., and Liu, M. (2021). Complete chloroplast genome sequences of the medicinal plant *Piper hancei*. *Mitochondrial DNA Part B.* 6 (9), 2775–2776. doi: 10.1080/23802359.2021.1967217

Zhou, J., Chen, X., Cui, Y., Sun, W., Li, Y., Wang, Y., et al. (2017). Molecular structure and phylogenetic analyses of complete chloroplast genomes of two aristolochia medicinal species. *Int. J. Mol. Sci.* 18 (9), 1839. doi: 10.3390/ijms18091839

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 19 January 2023
DOI 10.3389/fpls.2022.1089109

OPEN ACCESS

EDITED BY
Yusuf Khan,
Oslo University Hospital, Norway

REVIEWED BY
Stanislav Valentinovich Isayenkov,
National Academy of Sciences of Ukraine (NAN Ukraine), Ukraine
Akihiro Ueda,
Hiroshima University, Japan

*CORRESPONDENCE
Zeba I. Seraj
zebai@du.ac.bd

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics,
a section of the journal Frontiers in Plant Science

RECEIVED 03 November 2022
ACCEPTED 19 December 2022
PUBLISHED 19 January 2023

CITATION
Haque US, Elias SM, Jahan I and Seraj ZI (2023) Functional genomic analysis of K^+ related salt-responsive transporters in tolerant and sensitive genotypes of rice.
Front. Plant Sci. 13:1089109.
doi: 10.3389/fpls.2022.1089109

Functional genomic analysis of K^+ related salt-responsive transporters in tolerant and sensitive genotypes of rice

Umme Sabrina Haque[1], Sabrina M. Elias[2], Israt Jahan[1] and Zeba I. Seraj[1]*

[1]Plant Biotechnology Laboratory, Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka, Bangladesh, [2]Department of Life Sciences, Independent University Bangladesh, Dhaka, Bangladesh

Introduction: Salinity is a complex environmental stress that affects the growth and production of rice worldwide. But there are some rice landraces in coastal regions that can survive in presence of highly saline conditions. An understanding of the molecular attributes contributing to the salinity tolerance of these genotypes is important for developing salt-tolerant high yielding modern genotypes to ensure food security. Therefore, we investigated the role and functional differences of two K^+ salt-responsive transporters. These are OsTPKa or Vacuolar two-pore potassium channel and OsHAK_like or a hypothetical protein of the HAK family. These transporters were selected from previously identified QTLs from the tolerant rice landrace genotype (Horkuch) and sensitive genotype (IR29).

Methods: In silico comparative sequence analysis of the promoter sequences of two these genes between Horkuch and IR29 was done. Real-Time expression of the selected genes in leaves and roots of IR29 (salt-sensitive), I-14 and I-71 (Recombinant Inbred Lines of IR29(♀)× Horkuch), Horkuch and Pokkali (salt-tolerant) under salt-stress at different time points was analyzed. For further insight, OsTPKa and OsHAK_like were chosen for loss-of-function genomic analysis in Horkuch using the CRISPR/Cas9 tool. Furthermore, OsTPKa was chosen for cloning into a sensitive variety by Gateway technology to observe the effect of gain-of-function.

Results: The promoter sequences of the OsTPKa and OsHAK_like genes showed some significant differences in promoter sequences which may give a survival advantage to Horkuch under salt-stress. These two genes were also found to be overexpressed in tolerant varieties (Horkuch and Pokkali). Moreover, a coordinated expression pattern between these two genes was observed in tolerant Horkuch under salt-stress. Independently transformed plants where the expression of these genes was significantly lowered, performed poorly in physiological tests for salinity tolerance. On the other hand, positively transformed T_0 plants with the OsTPKa gene from Horkuch consistently showed growth advantage under both control and salt stress.

Discussion: The poor performance of the transgenic plants with the down-regulated genes OsTPKa and OsHAK_like under salt stress supports the assumption that OsTPKa and OsHAK_like play important roles in defending the rice landrace Horkuch against salt stress, minimizing salt injury, and maintaining plant growth. Moreover, the growth advantage provided by overexpression of the vacuolar OsTPKa K^+ transporter, particularly under salt stress reconfirms its important role in providing salt tolerance. The QTL locus from Horkuch containing these two transporters maybe bred into commercial rice to produce high-yielding salt tolerant rice.

KEYWORDS

rice, CRISPR/Cas9, expression analysis, Horkuch, *OsTPKa*, *OsHAK_like*, salt-tolerant, salt-sensitive

1 Introduction

Rice is one of the most important food crops globally, with 154 million hectares under cultivation and humans consuming 85 percent of the total rice produced (UNDP, 1997). However, it is considered a salt-sensitive crop with its growth and yield greatly affected under stress (Munns and Tester, 2008). Salinity is considered one of the major environmental stresses which negatively affect plant growth and productivity across the world. It is known to influence about 20% of the Earth's land and 50% of irrigated land worldwide which includes about 30% of the rice areas (Ahmadizadeh et al., 2016). Moreover, nearly 1 million ha of coastal soil in Bangladesh, or nearly a ninth of the total cultivable area is affected by soil salinity, mostly due to sea-water intrusion (Jahiruddin and Satter, 2010). However, there are some salt-tolerant rice landraces, like Horkuch, endemic to the coastal region of Bangladesh, which are likely to harbor novel sources of salt tolerance, due to their allelic diversity, that can complement known salt tolerance genes (Razzaque et al., 2019). Understanding the factors that can provide salinity tolerance such as ensuring efficient ion homeostasis under stress to maintain their productivity, would provide the resources to develop stress-tolerant high yielding crops.

Rice is mostly sensitive at the early seedling and reproductive stages. Excess salt causes both osmotic stress and later on ionic toxicity in rice plants. Ionic stress causes chlorosis and necrosis which either impairs growth and development and/or accelerates senescence. Increased accumulation of Na^+ in the cytoplasm also adversely affects the physiology of rice plants, such as hindering the net photosynthesis (Pn), degradation of photosynthetic pigment, stomatal conductance (Gs) and transpiration rate (Tr) (Hussain et al., 2017). These deleterious effects are not just about an increase in Na^+; there is a consequential inhibition of K^+ uptake because of their antagonistic effect. This event is more pronounced in salt-sensitive than tolerant plant genotypes (Wu et al., 2018). The effects of a high Na^+/K^+ ratio unavoidably leads to growth retardation, metabolic disruption, and even cell death, as K^+ has a vital role in enzyme activation, osmotic adjustment, turgor generation, cell expansion, regulation of membrane electric potential, and pH homeostasis (Sharma et al., 2013). It is also a major factor in resistance to drought, salinity, and fungal diseases (Amtmann et al., 2008). However, there are some stress-tolerant cultivars in which several catalytic, DNA binding, transcription regulator activities, and metabolic pathways have been reported to be enriched (Yamaguchi-Shinozaki and Shinozaki, 2006) which can help the plant to have natural adaptation responses at physiological, molecular, and cellular levels to tolerate salinity stress.

The farmer popular rice landrace, Horkuch was previously characterized as salt-tolerant at the seedling and reproductive stages (Lisa et al., 2011). Functional enrichment for signal transduction, DNA-dependent regulation, and transport activities under stress conditions are assumed to complement salt tolerance in this landrace (Razzaque et al., 2019). In a previous study of identifying genetic loci for salinity tolerance of Horkuch, at two sensitive developmental stages, 14 QTLs for 9 traits were identified. Among them, 2 QTLs for total potassium (TK) were found to differ significantly among the Horkuch (tolerant donor) and IR29 (sensitive) at the seedling stage (Haque et al., 2022). Both the QTL intervals for TK were enriched with various transmembrane transporters including potassium ion transmembrane transporters. Vacuolar Two pore Potassium channel (*OsTPKa*) and HAK type transporter (*OsHAK_like*) reside in that QTL region (Haque et al., 2022).

Both the TPK and HAK potassium transporters are very important for K^+ homeostasis. The TPK potassium channels mainly reside in the vacuolar membrane and are the gateway for

K^+ exchange between the vacuole and cytoplasm as and when necessary. These highly K^+ selective TPK channels are likely involved in plant stress and developmental processes (Dabravolski and Isayenkov, 2022). The rice TPKa isoform is located in the tonoplast of the main lytic vacuole, while the TPKb is located in small vacuoles (Isayenkov et al., 2011a). The C-terminals of these isoforms determine their location. The TPK channels have the GYGD motif responsible for K^+ selectivity, as well as Ca^{2+} binding EF and 14-3-3 domains which may play regulatory roles (Isayenkov et al., 2011b). A third rice isoform TPKc has not been well characterized as yet. The constitutive overexpression of the TPKb isoform has been reported to confer osmotic and drought tolerance to rice and lower sensitivity to K^+ deficiency (Ahmad et al., 2016). A family of 27 HAK (often referred to as KT/HAK/KUP) transporters, most of which are located on the plasma membrane, have been reported in rice, and have been classified into 4 clusters based on their amino acid sequences (Gupta et al., 2008; Okada et al., 2018). Cluster I transporters function in high affinity K^+ uptake, while those in cluster II are low affinity transporters. The function of Cluster III and IV transporters are not well-characterized so far.

The salt tolerance in this endemic rice landrace is regulated by myriad correlated genes and complex mechanisms. Despite progress in understanding the molecular signaling and mechanisms of salt tolerance, much remains to be learned from the natural and adaptive salt tolerance observed in landraces. Being an essential nutrient for plants, K^+ improves the growth, development and yield of crops. K^+ can also promote growth under stress conditions through different defense responses. For example, the reactive oxygen species generated due to the osmotic stress in excess salinity, can be ameliorated *via* K^+ application. Also, K^+ helps in strengthening antioxidant enzyme activities, thus improving nitrogen use efficiency. Overall maintenance of K^+ homeostasis enhances a plant's salt tolerance by aiding in sustainable crop yield (Kumari et al., 2021). High Na^+ concentration in cytosol hampers cell metabolism, e.g., photosynthetic activities. Replacing K^+ as vacuolar osmoticum is a defense mechanism for the cell under excessive Na^+ load. K^+ is involved in electrical neutralization of ions and macromolecules, pH homeostasis, regeneration of cell osmotic pressure, increase in leaf area, activation of enzymes, stomatal movement and tropism and translocation of ions, e.g. Na^+, etc. Hence identifying potentially salt-responsive transporters and ion channels from the tolerant landrace Horkuch could provide a source of diverse genes for a comparative study with the existing pool of membrane transport proteins and for creating a novel artificial transporter. With a view to this, the study focuses on functional genomics analysis of some differentially expressed genes related to salt-responsive K^+ channel *OsTPKa* (LOC_Os03g54100) and putative K^+ transporter *OsHAK_like* (LOC_Os03g55370) from Horkuch.

In this study, we have carried out *in silico* comparative structural and functional analysis, differential gene expression, and loss and gain of function-based analysis of candidate transporter genes between tolerant Horkuch and sensitive IR29 to elucidate their function and likely mechanism in providing salt tolerance. If the function of these genes is proven, they can be used for selective breeding and gene editing to generate a high-yielding tolerant variety. Moreover, to have a cumulative effect, those genes can be recombined in a single variety to provide enhanced tolerance.

2 Methods and materials

2.1 Plant material

Five varieties: IR29, Horkuch, I-14, I-71 and Pokkali were selected based on their functional response towards salinity. Horkuch from the coastal region (Lisa et al., 2011) and Pokkali have shown tolerance to salt condition (Xie et al., 2000). IR29 is a modern rice variety from IRRI which is highly susceptible to salinity. I-14 and I-71 are Recombinant Inbred Lines of the F_7 generation from the reciprocal crosses between IR29 and Horkuch (Haque et al., 2022). I-14 behaves more like Horkuch in terms of maintaining K^+/Na^+ homeostasis as Horkuch is the donor plant of the Potassium-QTL on chromosome 3 for I-14. On the contrary, in most cases I-71 behaves more like IR29 as IR29 is the donor plant of Potassium-QTL for I-71. BRRI dhan28 (http://dhcrop.bsmrau.net/rice-variety-brri-dhan-28/), a salt sensitive high-yielding modern variety of rice was used to overexpress the *TPKa* gene. BRRI dhan28 was used for transformation because it is a commercially popular variety liked by farmers. Moreover, it is genotypically very similar to IR29 (Rahman et al., 2019).

2.2 *In silico* analysis of candidate genes

As the genome of Horkuch and IR29 is unannotated, the gene sequences are not available in rice databases. To extract the sequence from raw data of Horkuch and IR29 genome, (Unpublished work, Zeba I. Seraj) local BLAST was performed. In the BLAST program, the reference sequence of *Oryza sativa* Nipponbare was used as a reference to detect the chromosomal location of specific sequences of Horkuch and IR29. For the extraction of the sequences and the CDS (coding sequence), 5′ UTRs (5′-untranslated regions), 3′ UTRs (3′-untranslated regions) of candidate genes from salt-tolerant Horkuch and salt-sensitive IR29 transcriptome was extracted. Expasy's translate tool (https://web.expasy.org/translate/) was used to translate and predict amino acid sequences of the genes. To find out the functional differences of candidate

protein between Horkuch and IR29, Provean (http://provean.jcvi.org/seq_submit.php) and TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) and for detection of putative motif in promoter regions of candidate gene PLACE (http://www.dna.affrc.go.jp/PLACE/) was used.

2.3 RNA extraction, cDNA synthesis and quantitative RT-PCR

Total RNA was extracted from leaf and root of 18-day old seedlings after 0, 12, 24, 36, 48 and 72 h of 150 mM NaCl stress using TRIZOL reagent (Invitrogen, Carlsbad, CA, USA). Total RNA concentration and purity was determined using a Nanodrop ND1000 spectrophotometer (Thermo SCIENTIFIC). The cDNA was synthesized from 1.5µg of total extracted RNA using the Invitrogen Superscript III reverse transcription (RT)-PCR following the manufacturer's protocol (Invitrogen, USA) in a final volume of 20µl. The final cDNA products of 1500-1700 ng/µl concentration were diluted 4-fold prior to use in real-time PCR.

Primers for qRT-PCR were designed using Primer3web program (http://bioinfo.ut.ee/primer3). The primers are listed in Supplementary Table 1. In quantitative Real-Time PCR analysis, a total of 3 replicates for all the samples of both leaf and root were used. As a control for genomic DNA contamination, an equivalent amount of total RNA without reverse transcription was tested for each sample per gene. Quantitative PCR was done by a 6µl reaction using SYBR Green (Bio-Rad, USA) with gene-specific primers in CFX96 TM Real-Time PCR detection system (Bio-Rad, USA). Using the comparative cycle threshold method, the relative abundance of transcripts was calculated. Elongation Factor- α (EF- α) was used as the normalization control. The thermal profile of the reaction was 95°C for 3 min activation and denaturation, followed by 41 cycles of 95°C for 10 sec, and 62°C for 1 min. At, 62°C step the amplification was determined by SYBR Green. Finally, a dissociation curve was generated by increasing temperature starting from 65 to 95°C to create the melting temperature curve for amplicon validation. For statistical analysis mean variables of gene expression between different genotypes and at different time points were compared with the two-way analysis of variance (ANOVA) followed by Tukey's HSD *post hoc* test. The $p < 0.05$ and $p < 0.01$ were considered as significant and highly significant change with respect to control.

2.4 Construction of the sgRNA-Cas9 expression vector

Two single sgRNA were designed against *OsTPKa* (LOC_Os03g54100) and *OsHAK_like* (LOC_Os03g55370) using CRISPR-P web base resource (http://crispr.hzau.edu.cn/CRISPR2/) (Xie et al., 2014). The PAM sequence was set to be 5′-NGG-3′ (SpCas9 from Streptococcus pyogenes). The guide sequence length was defined to be 20 nucleotides and the locus tag of the above-mentioned genes were also provided. In the preliminary step, the sgRNA sequences (Supplementary Table 2) with an on-score of greater than 0.05, a GC content between (30-80%) and a lower off-target were chosen.

pRGEB32 plant binary expression vector was obtained from Addgene (Addgene plasmid # 63142) (Xie et al., 2015) which contains rice snoRNA U3 (Os U3p) promoter and UBI promoter (UBIp) for simultaneous sgRNA and Cas9 expression respectively. The 20 bp long multiple cloning site was removed by *BsaI* restriction enzyme and the designed sgRNA was inserted into the vector by restriction, digestion and ligation reaction according to provider's instructions (Xie et al., 2014). The *sgRNA_pRGEB32* constructs were transformed into *E.coli* using heat shock transformation.

2.5 Molecular confirmation of sgRNA constructs by PCR and sequencing

Specific sgRNA inserted plasmids were isolated from the *E.coli* colonies using Promega plasmid isolation kit protocol. Then the plasmids were confirmed by PCR reaction using Primer sets 4, 5, 6, 7 (Supplementary Table 1). PCR analyses were carried out in a 25 µl reaction mixture containing 100 ng of plant DNA, 100 µM of each dNTP, 2.4 ng each of primers, 1 unit of Taq DNA polymerase (Invitrogen, USA), 1.5 mM $MgCl_2$, 2.4% DMSO and 1 × PCR Buffer(-$MgCl_2$) (Invitrogen, USA). The optimized reaction was: Initial denaturation at 95°C for 5 min, 35 cycles of denaturation at 95°C for 1 min, annealing at 62.5°C for 1 min and extension at 72°C for 1 min following a final extension at 72°C for 10 min. The PCR-positive colonies were sequenced using primer sets-8 (Supplementary Table 1) by First Base DNA sequencing services from Malaysia.

2.6 *Agrobacterium*-mediated *in planta* transformation

The *pRGEB32_sgRNA* vector was transformed into *Agrobacterium tumefaciens* strain LBA4404 by applying standard protocols (Sambrook et al., 1989). The insertions were also confirmed by gene specific PCR. Transformed *Agrobacterium* strain (LBA4404) containing specific sgRNA construct was cultured and prepared for *in planta* transformation by following the standard protocol discussed in Parvin et al. (2015) and adapted according to Ahmed et al. (2018). Following the protocols mentioned by Amin et al., (2012), mature seeds of Horkuch were sterilized and two-days-old, germinated seeds were used for *Agrobacterium* inoculation. During the transformation events, each time embryonic apical

meristem of germinated seeds of Horkuch were infected with the transformed *Agrobacterium* strain. Infected seeds were then incubated in the dark at 28°C for 6–7 days. Later, the seedlings were treated with carbenicillin solution (250 mg/l) for 1 hour to kill and remove the remnants of *Agrobacterium*. After that, seedlings were kept in light for 16 hours and in dark for 8 hours at 28°C.

When the seedlings turned green, these were transferred to hydroponic solution. After 2-3 days, the hydroponic pots were transferred to a net-house. When the seedlings were 15-20 days old, they were transferred to soil and allowed to grow and set seeds.

2.7 *Hygromycin* resistance test

For the *hygromycin* resistant assay, the plant leaves were cut into 3-6 pieces with a length of about 2 cm at the heading stage and then immediately soaked in Petri dish containing selected solution (20 mg/L *hygromycin*). The plates were incubated under both dark and light condition (16h day/8 h night) at 25± 2°C.

2.8 Leaf disc senescence assay

To evaluate the stress tolerance potential of a plant, leaf disc senescence (LDS) assay was done after the maturation of plants. Flag leaf of around 1.0-cm dimension were cut from fully developed T_0 transgenic lines and wild type (WT) plants and then allowed to float in a 20 mL of 150 mM NaCl solution (Sanan-Mishra et al., 2005) for 3-7 days at 25°C temperature. There was no added K^+ or K^+ adjustment.

2.9 Seed germination test

To compare the germination level under salinity stress, both the non-transformed and *OsTPKa_sgRNA* and *OsHAK_like_sgRNA* transformed Horkuch seeds were subjected to 100 mM salt-stress at germination. After 8 days, shoot and root length were measured and compared.

2.10 Salt-stress screening at the seedling stage

The salt-stress screening of non-transformed and sgRNA transformed Horkuch was done according to the protocol described by Amin et al., (2012). 50mM salt-stress was applied to the 14 days old seedling with a gradual increment of 50 mM up to 150 mM stress. After 9 days of salt application the samples were collected and different parameters like electrolyte leakage, H_2O_2 concentration, chlorophyll content, root, shoot length, and weight were measured. These protocols were followed for salt stress screening of the overexpressed *TPKa* transgenic lines as well with gradual increment of 30 mM up to 120 mM salt stress.

2.10.1 Measurement of relative electrolyte leakage

0.1g of leaf sample were taken in a 25mL falcon tube in which there was deionized water and kept in a shaker for 2 hours. After that, the initial electrical conductivity (C1) was measured using a conductivity detector. Then the leaf segments in deionized water were autoclaved, cooled down to room temperature and final electrical conductivity (C2) was measured. The values of C1 and C2 were used to calculate relative electrolyte leakage.

2.10.2 Measurement of the chlorophyll concentration

0.1g leaf sample was weighed from both wild type and transgenic lines and was kept in a bottle submerged in 12mL 80% acetone solution. The leaf sample containing bottles were kept in dark for 72 hours, followed by absorbance measurements at 663nm and 645nm. The chlorophyll content was measured using the equation, A=ECd, (A is proportional to C, as E and d is constant) where A = observed absorbance; C = chlorophyll concentration (mg/mL); d = distance of the light path (= 1 cm); E = a proportionality constant (extinction Co-efficient) (=36 mL/cm).

2.10.3 Measurement of H_2O_2 level

0.3gm leaf sample from wild type and transgenic lines were collected and crushed with liquid nitrogen, followed by adding 5 mL of 0.1% (w/v) TCA. This mixture was then centrifuged at 12000rpm for 15 minutes at room temperature. Then 3mL supernatant was collected in screw cap tube and 0.5 mL of 1 M potassium phosphate buffer (pH 7.0) and 1 mL of potassium iodide (1 M) were added. Finally, the absorbance of the mixture was taken at 390nm using 1 mL of 0.1% (w/v) TCA and 1 mL of potassium iodide as blank control. The amount of H_2O_2 was calculated using the equation: H_2O_2 ($\mu mole.g^{-1}FW$) = 1 + ($227.8*OD_{390}$).

2.10.4 Measurement of Na^+ and K^+ content

Shoot and root from seedling of transgenic and wild type were washed and dried in oven. Dried leaves and roots were grinded and analyzed by a flame photometer after 48 hours extraction with 1N HCL according to the protocol of Yoshida (1976).

2.11 Construction of *pENTR-D-TOPO-OsTPKa*

Total RNA was isolated from 16 days-old salt (NaCl) stressed (100 mM) *O. sativa* cv. Horkuch using Trizol and first

strand cDNA was synthesized using oligodT following manufacturer's protocol (Invitrogen, USA). PCR was performed to amplify the target sequences of *OsTPKa*. Primer set 9 were designed for the whole fragments (1.0 kb) of *OsTPKa* (Supplementary Table 1). The forward primers were designed adding *CACC* overhang at the 5′ end to ensure correct orientation while cloning into the entry vector (*pENTR-D-TOPO*). PCR reactions for the target fragment was performed at 95°C for 5 min, 35 cycles of 1 min at 95°C, 30 s at 61°C, 2 min at 72°C followed by a final extension of 10 min at 72°C. The resulting fragments were gel purified using Qiagen Gel Purification system and later cloned to *pENTR-D-TOPO* vector following the manufacturer's protocol (Invitrogen, USA). Positive clones were initially selected based on the lysate PCR confirmation from *kanamycin* resistant clones of transformed *E. coli* cells. Plasmid isolation was done from confirmed colonies using Promega plasmid isolation kit (Promega, USA). Isolated plasmids were digested with *Pst*I to check the insertion of the desired fragments into the entry vector. Then, digestion-positive colonies were sequenced with M13 forward and reverse primers Primer set 12 by First Base DNA sequencing services from Malaysia. Clones with the confirmed sequences were then recombined with the destination vector (*pH7WG2.0*) by LR reaction (Invitrogen, USA) (Karimi et al., 2002). Positive clones were selected by gene specific PCR and restriction digestion. The constructs were then transformed into *Agrobacterium* and used to infect the embryonic apical meristem of germinated seeds of salt-sensitive variety BRRI dhan28 using *in planta* transformation method as described above.

3 Results

3.1 *OsTPKa* and *OsHAK_like* sequences showed structural and functional differences in tolerant and sensitive variety

No sequence variation in CDS (Coding Sequence) and amino acid sequence was found but a significant variation was observed for both genes in the promoter region between the tolerant and sensitive genotypes which may result in variation in the expression of these genes.

For functional characterization of the differences in the promoter region, promoter sequences isolated from salt tolerant Horkuch and sensitive IR29 were scanned using the PLACE (Higo et al., 1999) programs. For *OsTPKa*, differences in the pattern of motifs between these two varieties were observed, among which auxin, jasmonate, abscisic acid, water stress, and dehydration responsiveness motifs were found in higher amounts in tolerant Horkuch whereas a negative regulatory element for the inducible expression of WRKY18 was present in sensitive IR29. Similarly, for *OsHAK_like* promoter, auxin and dehydration responsive motifs were more abundant in Horkuch. Additionally, tissue-specific expression related motifs and gibberellin-responsive motif was found to be present in Horkuch but absent in IR29. As shown below, salt stress induces increased expression of *OsTPKa* and *OsHAK_like* in the root of the tolerant genotype.

3.2 Tissue-specific expression analysis

The expression pattern of *OsTPKa* and *OsHAK_like* was analyzed in leaf and root tissue of IR29, Horkuch, I-14, I-71 and Pokkali at 24 hours under 150 mM salt stress condition. Both genes showed higher expression in the root compared to the leaf region in all the genotypes. In addition, the expression in root was found to be much higher in salt-tolerant Horkuch, Pokkali and I-14 compared to IR29 under salt-stress (Figures 1A, B).

For a more detailed understanding of the differences in the expression pattern of *OsTPKa* and *OsHAK_like* gene between tolerant and sensitive genotypes, differential gene expression analysis was carried out at different times in both leaf and root of IR29 and Horkuch in 150 mM salt stress.

In leaf samples, for both *OsTPKa* and *OsHAK_like* genes no significant difference was observed between tolerant Horkuch and sensitive IR29. However, in the roots of IR29, *OsTPKa* showed a significant reduction in expression level at various time points under salt stress compared to non-stress, whereas a significant escalation in the expression was observed in Horkuch (Figure 1C). Similarly, *OsHAK_like* showed a reduction in expression level at all the observed time points under salt stress conditions compared to control in IR29. But for Horkuch, there was a gradual increase in expression compared to control at all time points under salt stress (Figure 1D).

3.3 *OsTPKa* and *OsHAK_like* showed coordinated expression pattern in tolerant genotype

The expression of *OsTPKa* significantly increased up to 24 hours under salt compared to *OsHAK*_like, followed by a sharp reduction in the next 48 hours. In contrast, there was a gradual increase in expression of *OsHAK_like* and it was significantly higher than that of *OsTPKa* at 48 and 72 hours of salt stress, indicating that *OsTPKa* and *OsHAK_like* may work in a coordinated pattern in Horkuch to provide salt tolerance (Figure 2A). In IR29, no such relationship was observed (Figure 2B). A proposed mechanism of how the coordinated function might help Horkuch to fight saline stress has been discussed (Figure 3).

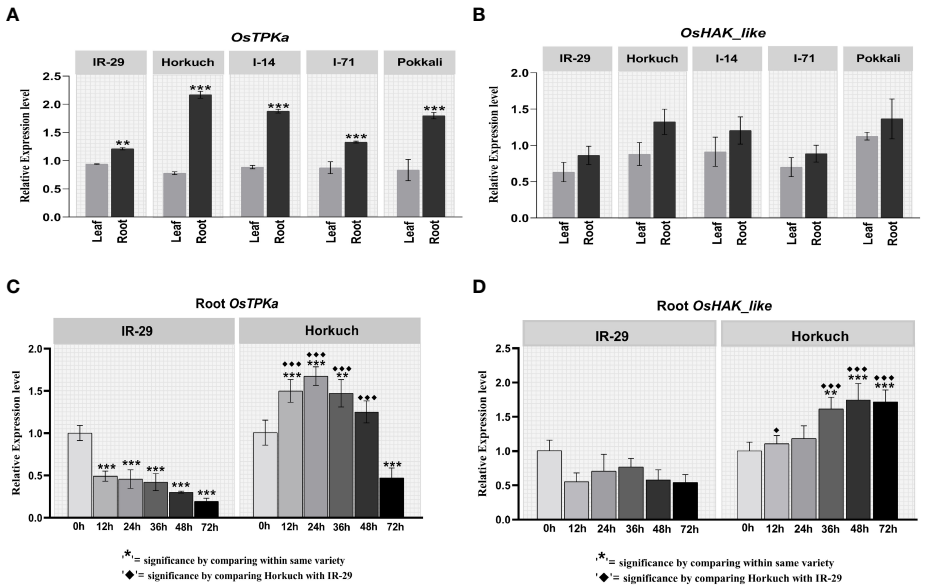


FIGURE 1
Differential expression analysis of *OsTPKa* and *OsHAK_like* by Real-Time quantitative PCR. **(A, B)** tissue specific expression analysis in IR29, Horkuch, I-14, I-71 and Pokkali at 24 hour under 150 mM salt-stress; **(C, D)** Salt-responsive genotype specific expression analysis in root of IR29 and Horkuch at 12 hour, 24 hour, 36 hour, 48 hour and 72 hour under 150 mM stress condition.

3.4 CRISPR/Cas9 mediated targeted mutagenesis of *OsTPKa* and *OsHAK_like* in the salt-tolerant variety Horkuch

The expression of *OsTPKa* and *OsHAK_like* is upregulated in tolerant Horkuch under salt-stress. Therefore, to ascertain the role of these two genes in conferring tolerance to Horkuch, a loss of function CRISPR/Cas9 mediated mutagenesis of *OsTPKa* and *OsHAK_like* was done in Horkuch. Successful transformation was achieved and is shown in Supplementary Figures 1, 2.

Transformation was confirmed by hygromycin resistant assay which was used as the selection marker. Due to absence of hygromycin resistant gene in wild type plant, leaf pieces from the positive transformed lines remained green, whereas the wild type ones turned yellow (Figure 4A). The transformation rate was 64.63% and 69.83% for *OsTPKa* and *OsHAK_like* respectively. For double confirmation, we performed PCR amplification of the hygromycin positive lines using the primer sets designed to specifically amplify *Cas9* and *hptII* sequences. 77.35% and 75% of the hygromycin positive lines contained the targeted sequences for *OsTPKa* and *OsHAK_like* respectively (Figure 4B).

Phenotypical assessment was performed at the T_0 stage using leaf disc senescence assay. Flag leaf pieces of non-inoculated Horkuch plants (wild type) remain mostly green after 5 days in salt solution. However, about 68% of *OsTPKa_sgRNA* and 63% of *OsHAK_like_sgRNA* transformed lines became almost yellow indicating sensitivity in salt (Figures 4C).

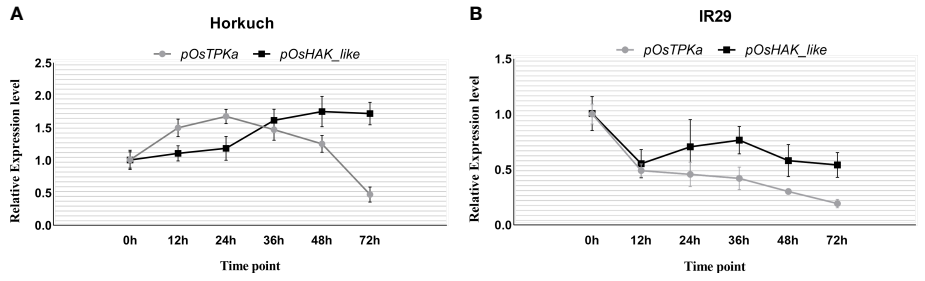


FIGURE 2
Coordinated expression analysis of OsTPKa and OsHAK_like in Horkuch **(A)** and IR29 **(B)**.

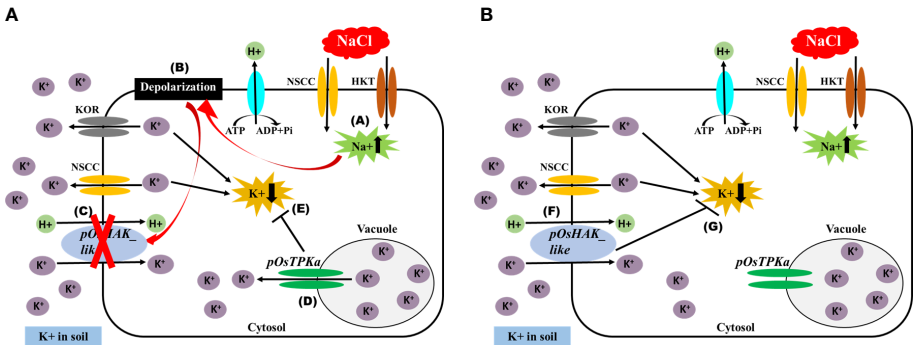


FIGURE 3
Schematic diagram of the proposed mechanism of maintaining K+/Na+ ratio in cytoplasm by coordinated function of OsTPKa and OsHAK_like. **(A)** Initially under salt stress condition, **(A)** there is influx of Na+ in cell, **(B)** which ultimately leads to membrane depolarization, **(C)** Depolarized membrane inhibits the function of HAK transporter, **(D)** Vacuolar K+ releases in cytoplasm by TPKa channel, **(E)** it helps to inhibit the decrease of K+ in cytoplasm. **(B)** at a later stage of salt stress when there is no depolarization of membrane, **(F)** HAK participates in uptaking K+ from soil and thus **(G)** helps to maintain K+ concentration in cytoplasm.

3.5 *OsTPKa_sgRNA* and *OsHAK_like* transformed plants showed poor phenotypical characteristics under salt-stress

The lines which performed poorly in LDS assay were selected for the salt-stress screening at the seedling stage. The 21-day-old seedlings of transformed Horkuch and wild type were exposed to 150 mM salt-stress for *OsTPKa_sgRNA* (Figure 5A), where more than 70% of the transformed plants performed poorly, showing yellowing and wilting. For the *OsHAK_like_sgRNA* (Figure 5B) about 55% resulted in visually wilted, rolled, and yellow leaves after 9 days. The score for seedling injury (SES score) was recorded after 9 days. The transformed plants had higher SES score compared to wild type plants (Figures 5C, D). The lines showed other salt-sensitive phenotypes with respect to several physiological parameters.

After 8 days of 100 mM salt-stress, the transgenic T_1 seeds showed higher sensitivity, poor germination rate, and significantly decreased shoot length for both *OsTPKa* (Figures 6A, C) and *OsHAK_like* (Figures 6B, D) compared to wild type indicating the loss in tolerance of transgenic Horkuch.

After 9 days of salt stress, it was observed that the percent increase in both electrolyte leakage and H_2O_2 content was significantly higher for 6 *OsTPKa* and 5 *OsHAK_like* mutated lines compared to wild type indicating stress injury (Figures 7A–D).

The percent increase in root length and weight was significantly lower whereas the percent reduction in shoot length and weight was higher in the mutated lines compared to wild type Horkuch indicating severe salt susceptibility (Figures 8A–D).

The chlorophyll content was greatly reduced in the transgenic lines compared to wild type under 150 mM salt-stress indicating mutation of these genes makes a tolerant plant susceptible to chlorophyll degradation under salt stress (Figures 9A, B).

The percentage increase in Na^+/K^+ ratio was higher in transgenic lines in comparison with wild type under salt stress which indicates that the lowering of *OsTPKa* and Os*HAK_like* gene expression disrupts the cytosolic balance of Na^+/K^+ (Figures 9C–F)

3.6 Gateway technology-based overexpression of *OsTPKa* in salt-sensitive variety

OsTPKa gene of Horkuch was selected for cloning into the salt sensitive modern high-yielding farmer popular genotype BRRI dhan28 to test for gain of function. Molecular confirmation of the transformed plants is provided in the supplementary section (Supplementary Figure 3). The transformed plants showed enhanced growth and improved phenology at the T_0 generation, compared to the wild type plants, both under 80 mM salt stress as well as without stress (Supplementary Figure 4A). Significant plant growth was observed in the putative transgenics in 9 days old plants but the difference between wild type and transgenics' height reduced over subsequent days (Supplementary Figure 4B). In terms of effective tiller, panicle length, flag leaf length, filled grain number and filled grain weight, the transgenic plants performed better compared to the wild type plant without stress. Transgenic status was confirmed with PCR using *hpt* gene primers from the vector construct. Leaf Disc Senescence (LDS) assay from T_0 plant flag leaves under 150 mM salt stress and PCR positive results were used to select T_1 seeds from specific panicles for generation advancement. Leaf sections that showed greener texture under

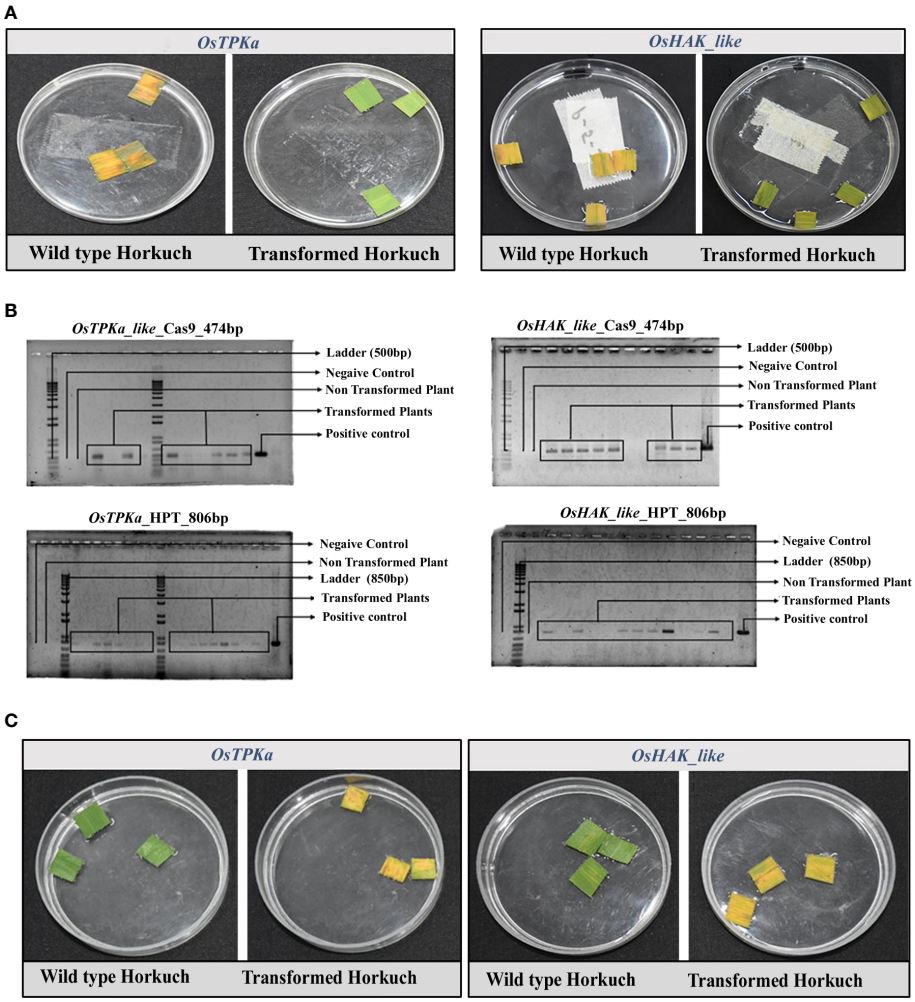


FIGURE 4
(A) Hygromycin resistance assay of T_0 transformants. The wild type became yellow, but the positive transgenic lines remain green in presence of hygromycin, **(B)** molecular confirmation of transformed plants by targeting *Cas9* and *hptII* gene. The lines with the PCR bands are the positive transgenic lines, wild type showed no such bands; **(C)** Leaf disc senescence assay of flag leaves from both *OsTPKa_sgRNA* and *OsHAK_like_sgRNA* T_0 transformants of Horkuch. sgRNA transformed transgenic lines performed poorly compared to non-transformed wild type.

salt stress for 3 days compared to the wild type, were considered positive.

3.7 Overexpressed lines of *OsTPKa* performed better under salt stress

Phenotypic screening of T_1 plants at seedling stage under 120 mM salt stress showed lower SES (salt injury), higher survival rate, lesser reduction in shoot and root length as well as shoot and root biomass (Figures 10A, B, 11A–D). Leaf tissue measurements of peroxide, reduction in chlorophyll content and electrolyte leakage, were found to be less but not significantly so in the transgenic plants compared to the wild type. However, in the putative transgenic plants' shoots, both wild type and transgenic plants showed increase in K^+/Na^+ ratio under salt stress but transgenic lines showed significant increase compared to no stress condition (Figure 11E). In root tissue, a significant reduction in K^+/Na^+ ratio was observed in the wild type plants under salt stress, which was opposite in case of the transgenic lines, specifically a higher increase was observed in line 39 (Figure 11F).

Relative expression analysis of the *TPKa* gene in the overexpresser T_1 plants RNA at seedling stage also showed higher expression of the potassium channel genes in transgenic lines 3, 21 and 39 under 150 mM salt stress for 24 hr (Supplementary Figure 5) compared to the wild type plants. LDS assay in 3 replicates of the T_1 plant flag leaves also confirmed better performance of the transgenic plants (Supplementary Figure 6). The selected transgenic plants are being advanced to T_2 generation and will be followed up in subsequent generations for their performance.

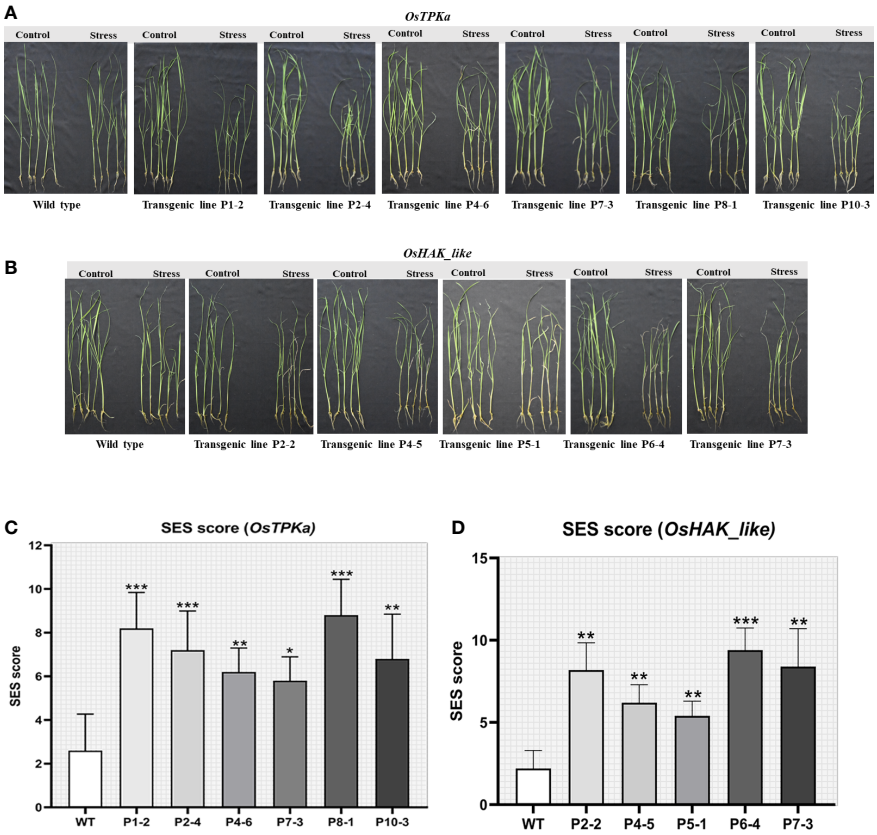


FIGURE 5
Phenotype of salt-tolerance in transformed lines. **(A, B)** The transformed and wild type seedlings growth at three leaf stage in hydroponics solution with 150 mM salt stress and no stress conditions for 9 days. **(C, D)** The comparison of seedling injury (SES score) under stress between wild type and both *OsTPKa_sgRNA* and *OsHAK_like_sgRNA* transformed plants. The stars in the graph indicates significant differences (***P < 0.001, **P < 0.01, *P < 0.1, Two way ANOVA test).

4 Discussion

Rice, as a salt-sensitive cereal crop plant, experiences environmental stresses including soil salinity which ultimately affects plant growth, yield and productivity. To make matters worse, the salinity levels in the Bangladeshi coast are gradually rising due to climate change and lack of freshwater permeation due to dams in the upper riparian regions. This challenges the food security of the ever-growing population all over the world and particularly in Bangladesh. To cope with this increased soil salinity and to facilitate the development of salinity-tolerant high-yielding rice genotypes, it is imperative to understand the physiological, biochemical, and molecular events which take place in rice, especially in the tolerant varieties like- Horkuch, which are indigenous to coastal Bangladesh (Lisa et al., 2011; Razzaque et al., 2017).

The availability of contrasting genotypes of rice, like the tolerant landrace, Horkuch, and sensitive high-yielding IR29 presents a good model to dissect the cellular response towards salinity. We focused on the elucidation of differences in the mechanism of some selected salt-responsive transporters between Horkuch and IR29. From a previous study of QTL analysis between a reciprocal population derived from Horkuch and IR29 (Haque et al., 2022), the transporters *OsTPKa* (Putative vacuolar two-pore K^+ channel) and *OsHAK_like* (Putative HAK like transporter) were targeted to elucidate their role in maintaining ionic balance under excessive Na^+ stress.

Since both genes reside in the salt tolerance QTL from chr 3 (Haque et al., 2022), our *OsHAK_like* is from chromosome 3. However, it is different from all the 27 OsHAK transporters reported earlier (Gupta et al., 2008), including those in Chr 3. It is truncated and matches with HAK24 and 25 based on their conserved potassium transport domain spanning 56 amino acids with 72% identity (OsHAK24) and 46 amino acids with 77% identity (OsHAK25). This gene is annotated in the MSU database (LOC_Os03g55370) as well as in Uniprot (Q10EU6)

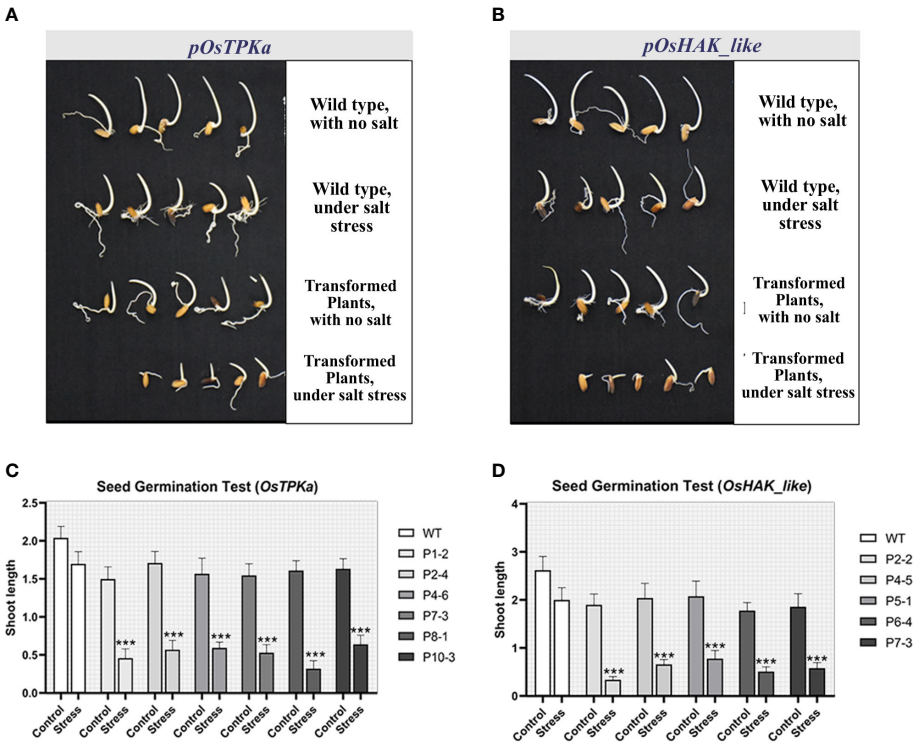


FIGURE 6
(A, B) Seed germination test under 100 mM salt condition. OsTPKa_sgRNA and OsHAK_like_sgRNA transformed Horkuch performed poorly under saline condition compared to their respective wild type; **(C, D)** Comparison of Shoot length of germinated seeds in seed germination test for both lines. The stars in the graph indicates significant differences (***P < 0.001, Two way ANOVA test).

and ENA (ABF99012.1) database which also shows 92% percent identity with a transporter in *Oryza barthii* (A0A0D3FPE3). Sequence based prediction with interproscan (Jones et al., 2014) suggests the role of this short protein as a putative potassium transporter. This is because it has the potassium transporter domain (interpro domain IPR003855) with two transmembrane helices and a C terminal end with hydrophilic chains residing inside (Supplementary Figures 7, 8). It does have the signature motif sequence **G**VVY**GK**VAMA**PLY** mentioned in Ragel et al. (2019) to be conserved as **G**VVY**GD**LGTS**PLY** (Ragel et al., 2019) where the bold amino acids are conserved for HAK members and OsHAK_like have the conserved region with lysine instead of aspartic acid, though both are hydrophilic amino acid. It is termed as hypothetical potassium transporter in the databases, but its functionality is quite evident from our results based on the seedling germination test of the transgenic plants where the gene was downregulated. On target specificity of the guideRNA for the CRISPR and primers for expression analysis was confirmed *via* genome wide blast, with no significant similarity other than our identified OsHAK_like gene on Chr 3. The separate downregulation of each of the genes, OsTPKa as well as OsHAK_like in the salt tolerant rice landrace resulted in its poorer performance. Yang et al. (2009) concluded in their study on HAK transporters that segmental duplication, tandem duplications, random translocation, and insertion events have contributed on the rice HAK family members diversity. The OsHAK_like protein under study can be an outcome of such events.

Salt tolerance has been found to be a function of a plant's ability to maintain a higher K^+/Na^+ ratio in barley (Shabala and Cuin, 2008). In a study of 70 barley genotypes, it was found that K^+ efflux was significantly correlated to salt sensitivity (Chen et al., 2007). Under salt stress, Na^+ enters the plasma membrane through non-selective cationic channels (NSCC) as well as HKT transporters, causing strong de-polarization. This in turn activates outward rectified K^+ channels such as KOR (Shabala et al., 2006). Simultaneously there is a rapid increase in cytosolic Ca^{2+} within seconds of exposure to NaCl (Knight et al., 1997). This in turn activates the SOS pathway, where Ca^{2+} binds SOS3 (calcium sensor protein or CBL4). SOS3 then binds and activates

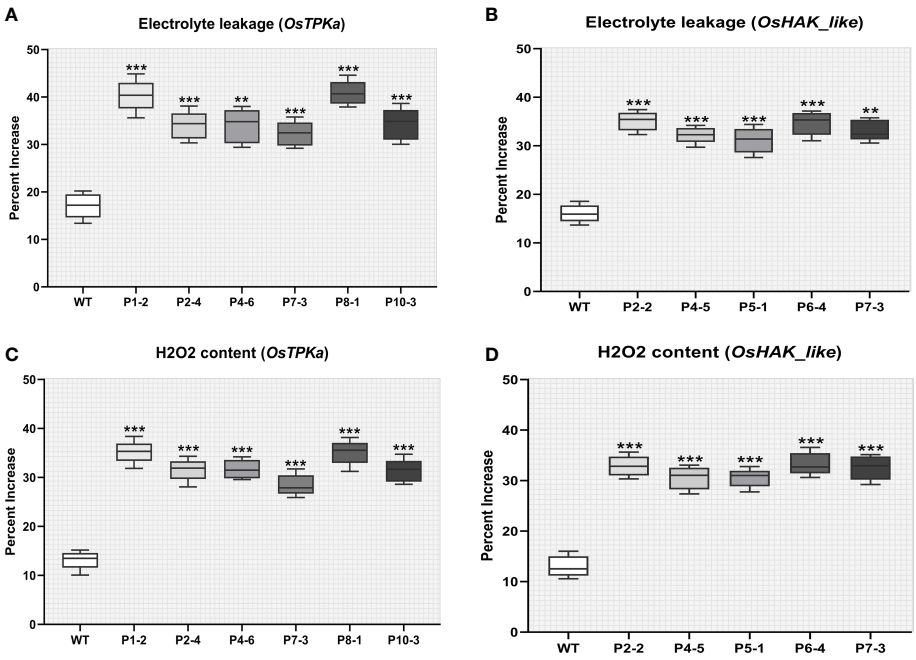


FIGURE 7
Effect of salt stress on phenotypical characteristics of transgenic plants. **(A–D)** Percentage increase in electrolyte leakage and H_2O_2 content; In all the parameters wild type Horkuch performed better compared to sgRNA inserted transgenic lines. The stars in the graph indicates significant differences (***P < 0.001, **P < 0.01, Ordinary one way ANOVA test).

SOS2, a CBL interacting kinase causing phosphorylation of the plasma membrane Na^+/H^+ antiporter SOS1, which extrudes Na^+ from the cell (Zhang et al., 2022). SOS2 also interacts with vacuolar H^+ATPase, enhancing its activity which in turn drives the vacuolar Na^+/H^+ antiporter (NHX1) to sequester Na^+ into the vacuole (Batelli et al., 2007). Salt stress also causes the activation of the plasma membrane H^+ATPase by an unknown osmosensor (Shabala and Cuin, 2008). Higher H^+ATPase activity and lowering of the cytoplasmic Na^+ load causes the plasma membrane to re-polarize.

So how do OsTPKa and OsHAK_like, the two K^+ transporters, which are the subject of this study fit into the salt perception and mitigation scenario? In Arabidopsis it was shown that different sets of CIPKs were activated by a Ca^{2+} signal targeting the vacuolar tonoplast transporters. As a result, the TPK channels cause efflux of K^+ out of the vacuole to cytoplasm (Dabravolski and Isayenkov, 2022). The Ca^{2+} dependent protein kinase 3 (CDPK3) has been shown to interact with TPK1 (two-pore potassium channel homologous protein from Arabidopsis) and activate it (Latz et al., 2013). HAKs are also activated to import K^+ across the cell into the cytoplasm. Plasma membrane CIPK-CBLs have been reported to be activated by Ca^{2+} as well as AKT1 and HAK transporters to import K^+ from outside of the cell (Isayenkov and Maathuis, 2019).

Ability to retain the K^+ in root and leaf tissue is an important trait for salt tolerance in plants. HAK/KUP K^+ transporter family plays the main role for High affinity K^+ uptake in a plants' response to salt stress (Wu et al., 2018). Shen et al. (2015) reported involvement of OsHAK21 in the maintenance of K^+ homeostasis. OsHAK5 overexpression improved shoot K^+ accumulation and biomass compared to wild type plants. (Yang et al., 2014). OsHAK1-D mutated lines resulted in significantly reduced K^+ content in shoot and root (Chen et al., 2015). These symporters need ATP for H^+ ATPase pumping of H^+ to maintain the H^+ gradient (Wu et al., 2018). Moreover, the increased plasma membrane H^+-pump activity can help K^+ uptake *via* the HAK transporters (Shabala and Cuin, 2008).

The cytosolic K^+ concentration needs to be restored after Na^+ induced signaling and membrane depolarization for continuing normal cellular metabolism. The balance between K^+ efflux and K^+ uptake determines tolerance to salt stress. These 2 processes are maintained by K^+ channels and transporters regulated in a cell and tissue specific manner. K^+ efflux is mainly mediated by GORK (guard cell outward rectifying K^+ channel) type and ROS-activated

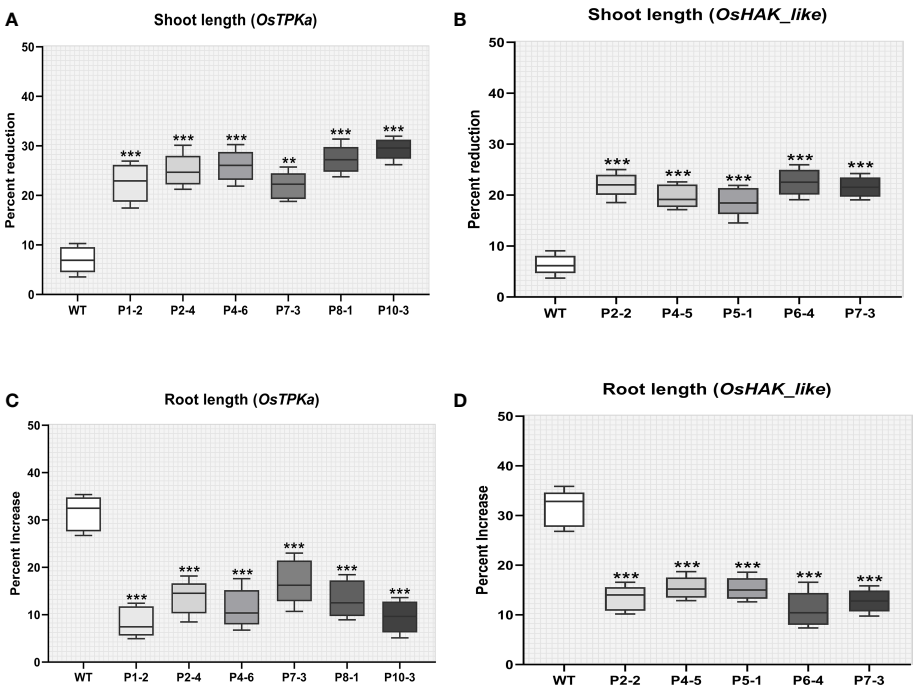


FIGURE 8
Effect of salt stress on phenotypical characteristics of transgenic plants. **(A, B)** Percentage increase in shoot length **(C, D)** Percentage reduction in root length. In all the parameters wild type Horkuch performed better compared to sgRNA inserted transgenic lines. The stars in the graph indicates significant differences (***P < 0.001, **P < 0.01, Ordinary one way ANOVA test).

NSCC-type channels (Jayakannan et al., 2013; Wu et al., 2015). Wu et al., (2018) suggested a probable coordination between TPK, TPC channels and 14-3-3 proteins, where increased expression of the latter activated the TPK but inactivated the TPC channels resulting in replenishing of the cytosolic K^+ from vacuolar pool. TPK channels can act as osmosensor in case of high external osmotic pressure, since they are osmo sensitive (MacRobbie, 2006). Hence activation of TPK channels ensures rapid release of K^+ from vacuole which is the main cellular depository of osmotica and water (Maathuis, 2011).

4.1 *In silico* analysis of *OsTPKa* and *OsHAK_like* between Horkuch and IR29 showed significant differences in stress-responsive motifs in promoter region

In the study of *In silico* promoter analysis, for both *OsTPKa* and *OsHAK_like*, there were significant differences in stress-related cis regulatory element in promoter region such as- presence of higher number of Auxin, Jasmonate, Gibberellin, Abscisic acid responsive and dehydration responsive motifs in Horkuch. Phytohormones like auxin, jasmonate, gibberellin, abscisic acid regulates plant growth adaptation under salt stress. Additionally, these hormones help plant build a defense system to fight against stress (Yu et al., 2020). Therefore, the presence of these motifs in higher number might account for the greater responsiveness of those genes to salt-stress stimuli in the tolerant rice landrace.

4.2 Tolerant genotypes might get survival advantage due to increased expression of *OsTPKa* and *OsHAK_like* transporter under salt-stress

In the tissue specific expression analysis study (Figure 1) and salt-responsive genotype specific expression analysis study (Figure 2), *OsTPKa* and *OsHAK_like*, a hypothetical protein of HAK (High affinity K^+ transporter) family were found to be upregulated in roots of the salt-tolerant Horkuch, Pokkali and I-14 compared to sensitive IR29, indicating that the increased expression of these 2 genes is perhaps important for the performance of tolerant varieties under saline conditions. During the onset of salt stress, the rapid increase in Ca^{2+} may lead to compensatory K^+ efflux into cytoplasm from the

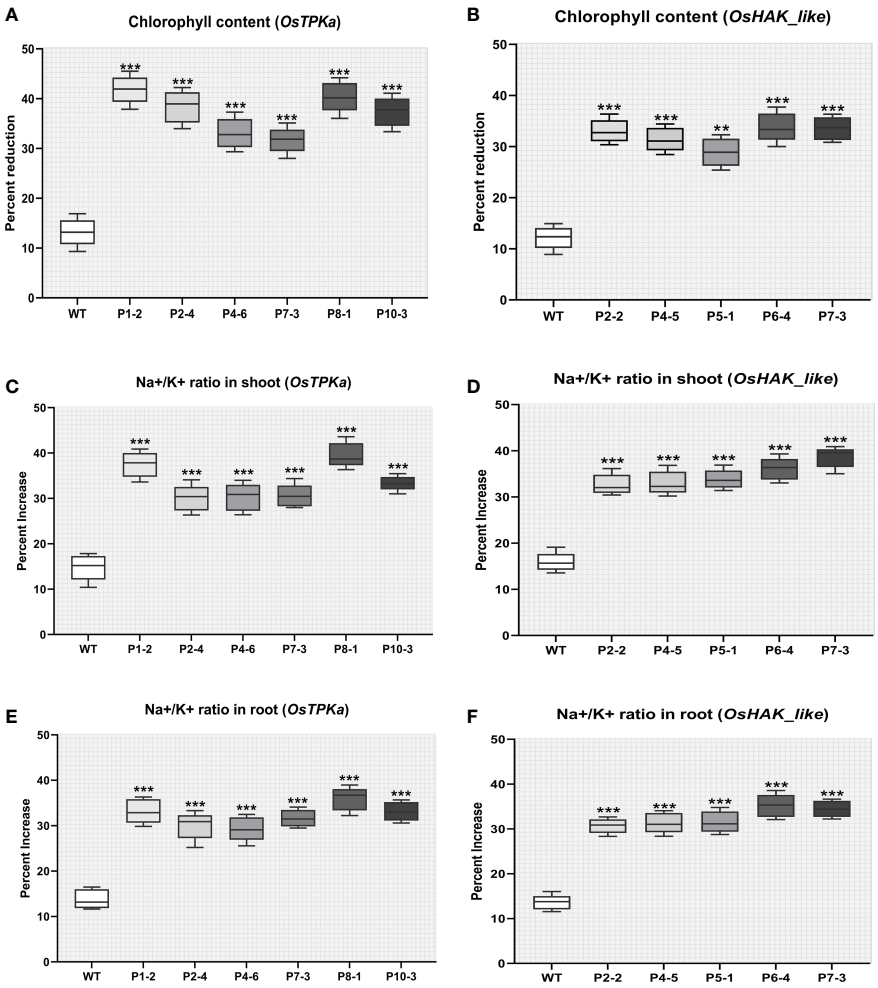


FIGURE 9
Effect of salt stress on phenotypical characteristics of transgenic plants. **(A, B)** Percent reduction in chlorophyll content; **(C, D)** Percentage increase in Na^+/K^+ ratio in shoot; **(E, F)** Percentage increase in Na^+/K^+ ratio in root under 150 mM salt condition. In all the parameters wild type Horkuch performed better compared to sgRNA inserted transgenic lines. The stars in the graph indicates significant differences (***P < 0.001, **P < 0.01, Ordinary one way ANOVA test).

FIGURE 10
(A) The transgenic plants (line 39 and line 21) showed healthy appearance under 120 mM salt stress compared to the wild type plants. **(B)** Transgenic Line 39 showed significant (P<0.01) low SES score (high tolerance) than the WT. **P < 0.01, *P < 0.1.

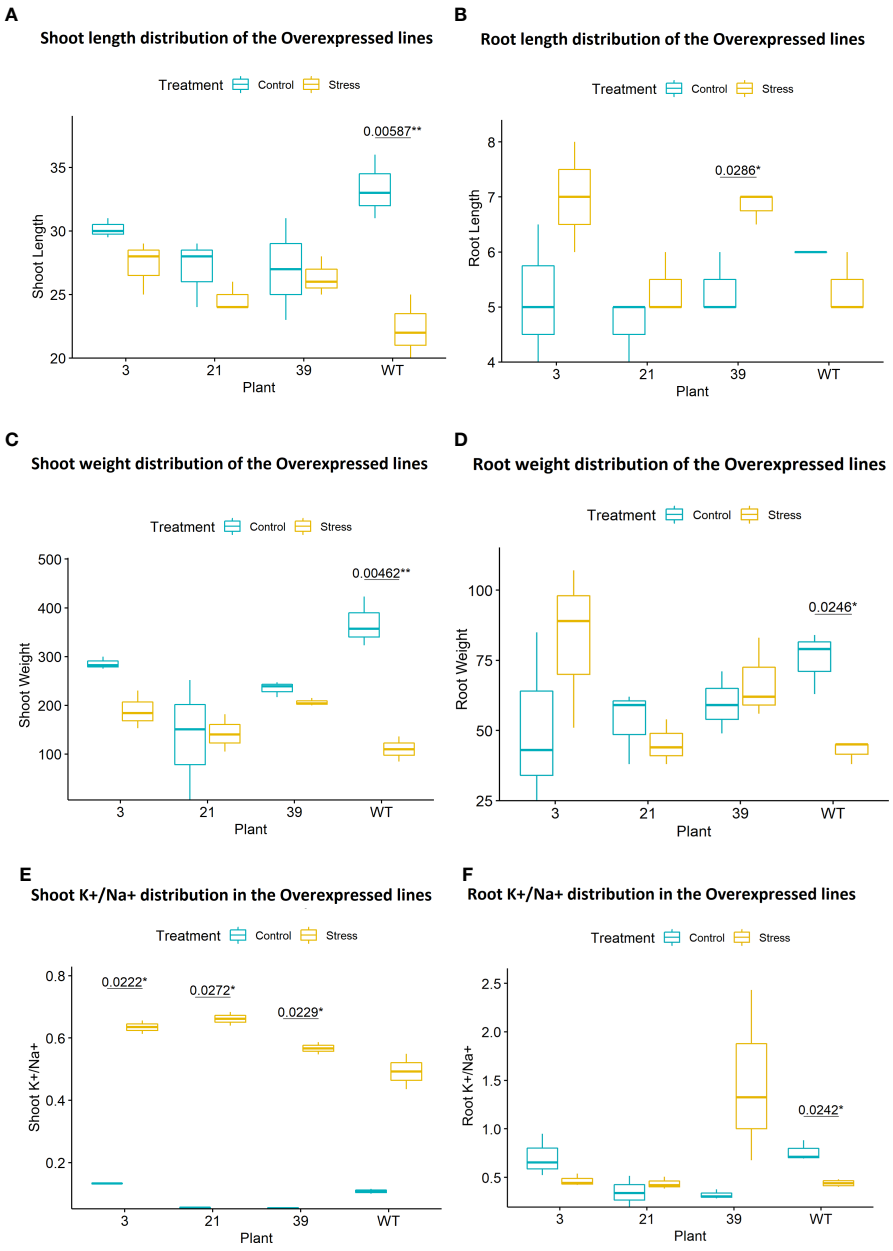


FIGURE 11
(A) Percent Reduction of shoot length was less in the transgenic lines and the wild type plants showed significantly high reduction ($p < 0.01$) under stress compared to no stress. (Transgenic lines = 3, 21, 39, WT = Wild Type BRRI dhan28). **(B)** Percent increase in root length was higher in the transgenic lines and significantly higher ($p < 0.05$) in line 39 under salt stress compared to no stress and the wild type plants showed reduction under stress. (Transgenic lines = 3, 21, 39, WT = Wild Type BRRI dhan28). **(C)** Percent Reduction of shoot weight was less in the transgenic lines and the wild type plants showed significantly high ($p<0.01$) reduction under stress compared to without stress condition. (Transgenic lines = 3, 21, 39, WT = Wild Type BRRI dhan28). **(D)** Percent reduction in root weight was less in the transgenic lines and the wild type plants under 120 mM salt stress showed significant reduction ($p < 0.05$) compared to without stress condition. (Transgenic lines = 3, 21, 39, WT = Wild Type BRRI dhan28). **(E)** Significant increase ($p < 0.05$) in K+/Na+ ratio was observed in the shoot tissues of transgenic lines 3, 21 and 39 under salt stress compared to their without stress condition which was not observed in the wildtype. **(F)** Significant decrease ($p < 0.05$) in K+/Na+ ratio was observed in root in the wild type compared to no stress. Transgenic line 39 showed higher increase under salt stress. (Transgenic lines = 3, 21, 39, WT = Wild Type BRRI dhan28). **P < 0.01, *P < 0.1.

vacuole through activation of TPK channels to replenish stress-induced cytosolic K^+ depletion (Latz et al., 2007). So, the ideal scenario is to increase K^+ selective TPK activity (Wu et al., 2018). Again, HAK is the only K^+ uptake transporter that is known to operate in K^+ uptake from soil at external concentrations well below 10 μM and this characteristic activity for high-affinity K^+ transport renders plants expressing the HAK transporter genes very tolerant to low-K^+ conditions (Assaha et al., 2017). Our hypothesis is strengthened by the fact that the separate downregulation of each of the genes, Os*TPKa* as well as *OsHAK_like* in the salt tolerant rice landrace led to poorer seed germination both without and with salt. The reduction in shoot length was statistically significant (Figures 6A-D). Therefore, upregulation of *OsHAK_like* in the tolerant varieties likely mediates the higher uptake of K^+ from soil during salt-stress.

4.3 *OsTPKa* and *OsHAK_like* might work in a coordinated manner to combat salinity stress induced K^+ depletion in cytoplasm

From the coordinated expression analysis study of *OsTPKa* and *OsHAK_like* (Figures 2, 3) in root of Horkuch under salt-stress it was found that there was a slow but gradual increase in expression of *OsHAK_like* whereas for *OsTPKa*, the expression was higher at the initial stage and decreased at the latter stages of salinity stress. It is possible that at the initial stage of salinity stress, the vacuolar TPKa is activated *via* the rapid increase in Ca^{2+} (Latz et al., 2007) and the rapid de-polarization of the membrane and K^+ efflux is countered. Meanwhile, the released Ca^{2+} reduces the Na^+ load by activating the SOS pathway, while the simultaneous H^+ pump activity further re-polarizes the membrane which in turn may subsequently activate the HAK_like K^+ transporter (Shabala and Cuin, 2008). Therefore, tolerant plants may rely on a rapid replenishment of K^+ supply from vacuolar storage *via* TPK mediated K^+ efflux into the cytoplasm to maintain osmotic balance and to normalize membrane potential. The activity of the TPK may allow the cells the time to complete the signaling processes and increase the number of transcripts for high affinity transporters (HAK) to further increase K^+ uptake and restore the cellular K^+ pool volume (Wu et al., 2018). Now, when HAK transporters take part in maintaining K^+ in cytoplasm, there is no need for vacuolar storage support and finally, other vacuolar inward transporters help replenish the lost K^+ in the vacuole. Therefore, from this study, it can be hypothesized that both genes may work in a coordinated manner to increase the chance of survival under salinity stress in Horkuch. However, this hypothesis needs to be proven with by evaluation of the role of both potassium and Ca^{2+} in this coordination which are being planned.

4.4 *OsTPKa* and *OsHAK_like* mutant transgenic lines showed a decrease in tolerance suggesting their role in survival under salinity

The expression analysis data of *OsTPKa* and *OsHAK_like* in the roots showed the overexpression of both genes in tolerant varieties like Horkuch and Pokkali compared to the sensitive one (IR29) under salt stress. However, the mechanism of how the upregulation of these two genes helps maintain a favorable ion gradient in tolerant varieties is not well characterized yet. So, to gain insight into the biophysical properties and physiological function of these genes, they were subjected to CRISPR/Cas9-based loss of function mutation in Horkuch.

It was observed that even at the T_0 generation, the mutant lines of Horkuch showed a remarkable sensitivity to 150 mM NaCl in leaf disc senescence assay indicating loss of tolerance in the transformed lines. Among selected lines that showed a sensitive response, 6 lines each for *OsTPKa* and 5 for *OsHAK_like* performed poorly with an extremely slower rate of seed germination, a significant reduction in shoot length and weight, and a very low percent increase in root length compared to wild type. Salinity-induced osmotic and ionic stresses are known to inhibit and delay seed germination along with interruption of nutrient uptake and growth retardation, especially and more prominently in sensitive varieties (Hakim et al., 2010). Therefore, our study indicates that due to mutation of the above-mentioned genes, tolerant Horkuch is behaving more like a sensitive variety which points to the importance of the two transporter genes in providing salinity tolerance.

To further verify the reduction in stress-tolerant characteristics of T_1 Horkuch *OsTPKa* and *OsHAK_like*, mutant lines, various tests including electrolyte leakage, H_2O_2 concentration and chlorophyll content were measured. Electrolyte leakage and H_2O_2 in shoots showed an increase with stress in the Horkuch mutant lines compared to the wild type indicating disturbed membrane integrity (Mishra et al., 2021). Percent increase in electrolyte leakage and peroxide (H_2O_2) content due to salt stress was lesser in the overexpressed lines under salinity compared to the wild type plants. For the mutant line, the content of the photosynthetic pigment in rice plant grown under salt condition declined significantly, being especially susceptible to salt. The chlorophyll pigment content in leaves of tolerant rice genotypes under stress is known to be maintained better than in sensitive varieties. In our study, the reduction of chlorophyll content was significantly higher in the mutant Horkuch lines which can lead to a decrease in photosynthetic efficiency and sugar content and a subsequently negative effect on plant survival under stress (Amirjani, 2011). There was striking increase in Na^+/K^+ ratio in *OsTPKa* and *OsHAK_like*

downregulated lines compared to corresponding wild type which indicates the importance of above-mentioned genes in maintaining the appropriate K^+ concentration under salt stress.

4.5 *OsTPKa* overexpressed transgenic lines showed increased tolerance to salinity

The overexpressed lines where the *OsTPKa* gene from Horkuch was cloned and transformed in a salt sensitive genotype BRRI dhan28, showed a significant increase in growth during germination as well as during initial growth at seedling stage. They looked healthy and taller than their wildtype counterpart even without any stress application (Supplementary Figure 4A). However, this significant difference with the wild type plants were observed only during the earlier growth of the plants. At 12 days old, wild type plants and putative transgenic plants were of similar height. Differences became evident between WT and transgenic plants, where the latter showed significantly lower reduction in plant height after salinity stress of 18 days (Supplementary Figure 4B). The putative transgenic plants' survival rate was much higher under salt stress at the seedling stage. This interesting observation indicates that the overexpressed lines may have some kind of growth advantage due to supply of extra potassium during the early days compared to the wild type even without salt stress. The plants without stress were subjected to generation advancement and a panicle wise selection using LDS assay could select the lines for phenotypic screening at seedling stage in the T_1 generation.

Under salt stress, in the over-expressed lines of *OsTPKa*, the root length and weight were significantly higher and there was significant increase of K^+/Na^+ under salt stress in the transgenic lines compared to the wild type plants. They showed less percent reduction of chlorophyll content under salt stress compared to the wildtype plant. However, this change was not that significant indicating that the overexpressed lines get growth advantage for their survival and maintenance but may not have extra benefit for enhanced photosynthesis or chlorophyll production in the leaf tissues.

The overexpressed T_1 transgenic lines showed significantly higher K^+/Na^+ ratio in their shoot and root tissues compared to no stress condition. In the wild type plants' root tissue, a significant reduction was observed under salt stress as opposed to the transgenic plants. The root tissues of transgenic line 39 showed a highly significant rise in K^+/Na^+ content compared to their control conditions, while the wild type plant showed a significant reduction. This trend was also supported by the higher gene expression level of OsTPKa in root tissue of line 39 compared to the other lines (Supplementary Figure 5). Potassium is one of the major nutrients required for growth and development of plants. Not only does early vigor give the plant better survivability under salt stress, but the higher potassium also protects against the toxic effects of the sodium. Therefore, the study supports the possible role of *OsTPKa* and *OsHAK_like* in providing salt tolerance.

5 Conclusion

CRISPR-mediated knockout of the two transporter genes, *OsTPKa* and *OsHAK_like* in salt-tolerant genotype (Horkuch) made it very sensitive to salt stress. The transformed plants showed poorer performance in several phenological tests. Moreover, *OsTPKa* from Horkuch when cloned and transformed into sensitive genotype, BRRI dhan28, conferred salt tolerance to the latter. The loss and gain of function study indicated the important role of the OsTPKa channel and OsHAK_like transporter in conferring salt tolerance in the rice plant. Collectively, this study helped us understand that the time-specific increased activity of these two transporters may help the rice landrace Horkuch defend against salt stress. It also is indicative of the likely mechanism by which this rice landrace has favorably adapted to coastal salinity over many years. The loci of these two transporters are close enough (Haque et al., 2022) for these to be simultaneously bred into commercial rice genotypes by use of DNA markers. Alternatively, the simultaneous activation of their respective promoters through gene-editing may lead to the quick development of salt tolerant high-yielding rice.

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

UH, SE and IJ performed the lab works and screening. UH, SE and ZS wrote the manuscript. ZS and SE designed the experiment. All authors contributed to the article and approved the submitted version.

Funding

The project was supported by The World Academy of Sciences (TWAS) research grant no 19-173 RG/BIO/AS_I. The University of Dhaka is gratefully acknowledged for making it possible to publish the open access article by arranging to pay for the reduced APC charges.

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.2022.1089109/full#supplementary-material

References

Ahmad, I., Devonshire, J., Mohamed, R., Schultze, M., and Maathuis, F. J. M. (2016). 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

Ahmadizadeh, M., Vispo, N. A., Calapit-Palao, C. D. O., Pangaan, I. D., Viña, C. D., and Singh, R. K. (2016). Reproductive stage salinity tolerance in rice: a complex trait to phenotype. *Indian J. Plant Physiol.* 21, 528–536. doi: 10.1007/s40502-016-0268-6

Ahmed, T., Biswas, S., Elias, S. M., Rahman, M. S., Tuteja, N., and Seraj, Z. I. (2018). In planta transformation for conferring salt tolerance to a tissue-culture unresponsive indica rice (Oryza sativa l.) cultivar. *In Vitro Cell.Dev.Biol.-Plant* 54, 154–165. doi: 10.1007/s11627-017-9870-1

Amin, M., Elias, S. M., Hossain, A., Ferdousi, A., Rahman, M. S., Tuteja, N., et al. (2012). Over-expression of a DEAD-box helicase, PDH45, confers both seedling and reproductive stage salinity tolerance to rice (Oryza sativa l.). *Mol. Breed.* 30, 345–354. doi: 10.1007/s11032-011-9625-3

Amirjani, M. R. (2011). Effect of salinity stress on growth, sugar content, pigments and enzyme activity of rice. *Int. J. Bot.* 7, 73–81. doi: 10.3923/ijb.2011.73.81

Amtmann, A., Troufflard, S., and Armengaud, P. (2008). The effect of potassium nutrition on pest and disease resistance in plants. *Physiol. Plant* 133, 682–691. doi: 10.1111/j.1399-3054.2008.01075.x

Assaha, D. V. M., Ueda, A., Saneoka, H., Al-Yahyai, R., and Yaish, M. W. (2017). The role of na(+) and k(+) transporters in salt stress adaptation in glycophytes. *Front. Physiol.* 8, 509. doi: 10.3389/fphys.2017.00509

Batelli, G., Verslues, P. E., Agius, F., Qiu, Q., Fujii, H., Pan, S., et al. (2007). SOS2 promotes salt tolerance in part by interacting with the vacuolar h^+ ATPase and upregulating its transport activity. *Mol. Cell. Biol.* 27, 7781–7790. doi: 10.1128/MCB.00430-07

Chen, G., Hu, Q., Luo, L., Yang, T., Zhang, S., Hu, Y., et al. (2015). Rice potassium transporter OsHAK1 is essential for maintaining potassium-mediated growth and functions in salt tolerance over low and high potassium concentration ranges. *Plant Cell Environ.* 38 2747–2765. doi: 10.1111/pce.12585

Chen, Z., Zhou, M., Newman, I. A., Mendham, N. J., Zhang, G., and Shabala, S. (2007). Potassium and sodium relations in salinised barley tissues as a basis of differential salt tolerance. *Funct. Plant Biol.* 34 (2), 150–162. doi: 10.1071/FP06237

Dabravolski, S. A., and Isayenkov, S. V. (2022). Recent updates on the physiology and evolution of plant TPK/KCO channels. *Funct. Plant Biol.* doi: 10.1071/FP22117

Gupta, M., Qiu, X., Wang, L., Xie, W., Zhang, C., Xiong, L., et al. (2008). KT/HAK/KUP potassium transporters gene family and their whole-life cycle expression profile in rice *(Oryza sativa)*. *Mol. Genet. Genomics* 280 (5), 437–452. doi: 10.1007/s00438-008-0377-7

Hakim, M., Juraimi, A., Begum, M., Hanafi, M., Ismail, M. R., and Selamat, A. (2010). Effect of salt stress on germination and early seedling growth of rice (Oryza sativa l.). *Afr. J. Biotechnol.* 9, 1911–1918. doi: 10.5897/AJB09.1526

Haque, T., Elias, S. M., Razzaque, S., Biswas, S., Khan, S. F., Jewel, G. M. N. A., et al. (2022). Salt tolerance QTLs of an endemic rice landrace, horkuch at seedling and reproductive stages. *Sci. Rep.* 12, 17306. doi: 10.1038/s41598-022-21737-9

Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.* 27, 297–300. doi: 10.1093/nar/27.1.297

Hussain, S., Zhang, J.-H., Zhong, C., Zhu, L.-F., Cao, X.-C., Yu, S.-M., et al. (2017). Effects of salt stress on rice growth, development characteristics, and the regulating ways: A review. *J. Integr. Agric.* 16, 2357–2374. doi: 10.1016/S2095-3119(16)61608-8

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

Isayenkov, S., Isner, J., and Maathuis, F. J. M. (2011b). Membrane localization diversity of TPK channels and their physiological role. *Plant Signaling Behav.* 6, 1201–1204. doi: 10.4161/psb.6.8.15808

Isayenkov, S. V., and Maathuis, F. J. M. (2019). Plant salinity stress: Many unanswered questions remain. *Front. Plant Sci.* 10, 80. doi: 10.3389/fpls.2019.00080

Jahiruddin, M., and Satter, M. (2010). Agricultural research priority: Vision 2030 and beyond. *Final Rep. Sub-sector: land Soil resource management*. Bangladesh: Bangladesh Agric. Res. Council Bangladesh Agric. Univ. p56.

Jayakannan, M., Bose, J., Babourina, O., Rengel, Z., and Shabala, S. (2013). Salicylic acid improves salinity tolerance in arabidopsis by restoring membrane potential and preventing salt-induced k^+ loss *via* a GORK channel. *J. Exp. Bot.* 64, 2255–2268. doi: 10.1093/jxb/ert085

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulia, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAY vectors for agrobacterium-mediated plant transformation. *Trends Plant Sci.* 7, 193–195. doi: 10.1016/S1360-1385(02)02251-3

Knight, H., Trewavas, A. J., and Knight, M. R. (1997). Calcium signaling in arabidopsis thaliana responding to drought and salinity. *Plant J.* 12, 1067–1078. doi: 10.1046/j.1365-313X.1997.12051067.x

Kumari, S., Chhillar, H., Chopra, P., Khanna, R. R., and Khan, M. I. R. (2021). Potassium: A track to develop salinity tolerant plants. *Plant Physiol. Biochem.* 167, 1011–1023. doi: 10.1016/j.plaphy.2021.09.031

Latz, A., Becker, D., Hekman, M., Muller, T., Beyhl, D., Marten, I., et al. (2007). TPK1, a Ca(2+)-regulated arabidopsis vacuole two-pore k(+) channel is activated by 14-3-3 proteins. *Plant J.* 52, 449–459. doi: 10.1111/j.1365-313X.2007.03255.x

Latz, A., Mehlmer, N., Zapf, S., Mueller, T. D., Wurzinger, B., Pfister, B., et al. (2013). Salt stress triggers phosphorylation of the arabidopsis vacuolar k^+ channel TPK1 by calcium-dependent protein kinases (CDPKs). *Mol. Plant* 6 (4), 1274–1289. doi: 10.1093/mp/sss158

Lisa, L. A., Elias, S. M., Rahman, M. S., Shahid, S., Iwasaki, T., Hasan, A., et al. (2011). Physiology and gene expression of the rice landrace horkuch under salt stress. *Funct. Plant Biol.* 38, 282–292. doi: 10.1071/FP10198

Maathuis, F. J. M. (2011). Vacuolar two-pore k^+ channels act as vacuolar osmosensors. *New Phytol.* 191, 84–91. doi: 10.1111/j.1469-8137.2011.03664.x

MacRobbie, E. A. C. (2006). Osmotic effects on vacuolar ion release in guard cells. *Proc. Natl. Acad. Sciences U.S.A.* 103, 1135–1140. doi: 10.1073/pnas.0510023103

Mishra, M., Wungrampha, S., Kumar, G., Singla-Pareek, S. L., and Pareek, A. (2021). How do rice seedlings of landrace pokkali survive in saline fields after transplantation? physiology, biochemistry, and photosynthesis. *Photosynth Res.* 150, 117–135. doi: 10.1007/s11120-020-00771-6

Munns, R., and Tester, M. (2008). Mechanisms of salinity tolerance. *Annu. Rev. Plant Biol.* 59, 651–681. doi: 10.1146/annurev.arplant.59.032607.092911

Okada, T., Sousuke Y., M. Y., Ko, K., Atsuhiko, S., Yuta, T., Kozo, I., et al. (2018). Characterization of rice KT/HAK/KUP potassium transporters and k^+ uptake by HAK1 from oryza sativa. *Plant Biotechnol.* 35 (2), 101–111. doi: 10.5511/plantbiotechnology.18.0308a

Parvin, S., Biswas, S., Razzaque, S., Haque, T., Elias, S. M., Tammi, R. S., et al. (2015). Salinity and drought tolerance conferred by in planta transformation of SNAC1 transcription factor into a high-yielding rice variety of Bangladesh. *Acta Physiologiae Plantarum* 37, 68. doi: 10.1007/s11738-015-1817-8

UNDP (United Nations Development Programme) (1997). *Human Development Report 1997*: Human Development to Eradicate Poverty. New York.

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

Rahman, M. A., Thomson, M. J., De Ocampo, M., Egdane, J. A., and Salam, M. A. (2019). Shah-E-Alam, m. ismail, AM assessing trait contribution and mapping novel QTL for salinity tolerance using the Bangladeshi rice landrace capsule. *Rice* 12, 63. doi: 10.1186/s12284-019-0319-5

Razzaque, S., Elias, S. M., Haque, T., Biswas, S., Jewel, G., Rahman, S., et al. (2019). Gene expression analysis associated with salt stress in a reciprocally crossed rice population. *Sci. Rep.* 9, 8249. doi: 10.1038/s41598-019-44757-4

Razzaque, S., Haque, T., Elias, S. M., Rahman, M. S., Biswas, S., Schwartz, S., et al. (2017). Reproductive stage physiological and transcriptional responses to salinity stress in reciprocal populations derived from tolerant (Horkuch) and susceptible (IR29) rice. *Sci. Rep.* 7, 46138. doi: 10.1038/srep46138

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual* (*2nd* ed.). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sanan-Mishra, N., Pham, X. H., Sopory, S. K., and Tuteja, N. (2005). Pea DNA helicase 45 overexpression in tobacco confers high salinity tolerance without affecting yield. *Proc. Natl. Acad. Sci.* 102 (2), 509–514. doi: 10.1073/pnas.0406485102

Shabala, S., and Cuin, T. A. (2008). Potassium transport and plant salt tolerance. *Physiologia Plantarum* 133, 651–669. doi: 10.1111/j.1399-3054.2007.01008.x

Shabala, S., Demidchik, V., Shabala, L., Cuin, T. A., Smith, S. J., Miller, A. J., et al. (2006). Extracellular Ca^{2+} ameliorates NaCl-induced k^+ loss from arabidopsis root and leaf cells by controlling plasma membrane K1-permeable channels. *Plant Physiol.* 141, 1653–1665. doi: 10.1104/pp.106.082388

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, 224. doi: 10.3389/fpls.2013.00224

Shen, Y., Shen, L., Shen, Z., Jing, W., Ge, H., Zhao, J., et al. (2015). The potassium transporter OsHAK21 functions in the maintenance of ion homeostasis and tolerance to salt stress in rice. *Plant Cell Environ.* 38 (12), 2766–2779. doi: 10.1111/pce.12586

Wu, H., Zhang, X., Giraldo, J. P., and Shabala, S. (2018). It is not all about sodium: revealing tissue specificity and signalling roles of potassium in plant responses to salt stress. *Plant Soil* 431, 1–17. doi: 10.1007/s11104-018-3770-y

Wu, H., Zhu, M., Shabala, L., Zhou, M., and Shabala, S. (2015). K^+ retention in leaf mesophyll, an overlooked component of salinity tolerance mechanism: a case study for barley. *J. Integr. Plant Biol.* 57, 171–185. doi: 10.1111/jipb.12238

Xie, K., Minkenberg, B., and Yang, Y. (2014). Targeted gene mutation in rice using a CRISPR-Cas9 system. *Bio-Protocol* 4. doi: 10.21769/BioProtoc.1225

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

Xie, J. H., Zapata-Arias, F. J., Shen, M., and Afza, R. (2000). Salinity tolerant performance and genetic diversity of four rice varieties. *Euphytica* 116, 105–110. doi: 10.1023/A:1004041900101

Yamaguchi-Shinozaki, K., and Shinozaki, K. (2006). Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu. Rev. Plant Biol.* 57, 781–803. doi: 10.1146/annurev.arplant.57.032905.105444

Yang, Z., Gao, Q., Sun, C., Li, W., Gu, S., and Xu, C. (2009). Molecular evolution and functional divergence of HAK potassium transporter gene family in rice (Oryza sativa l.). *J. Genet. Genomics* 36 (3), 161–172. doi: 10.1016/S1673-8527(08)60103-4

Yang, T., Zhang, S., Hu, Y., Wu, F., Hu, Q., Chen, G., et al. (2014). The role of a potassium transporter OsHAK5 in potassium acquisition and transport from roots to shoots in rice at low potassium supply levels. *Plant Physiol.* 166 (2), 945–959. doi: 10.1104/pp.114.246520

Yoshida, S.Institute, I.R.R (1976). *Laboratory manual for physiological studies of rice* (International Rice Research Institute).

Yu, Z., Duan, X., Luo, L., Dai, S., Ding, Z., and Xia, G. (2020). How plant hormones mediate salt stress responses. *Trends Plant Sci.* 25, 1117–1130. doi: 10.1016/j.tplants.2020.06.008

Zhang, H., Zhu, J., Gong, Z., and Zhu, J.-K. (2022). Abiotic stress responses in plants. *Nat. Rev. Genet.* 23, 104–119. doi: 10.1038/s41576-021-00413-0

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 08 May 2023
DOI 10.3389/fpls.2023.1124619

OPEN ACCESS

EDITED BY
Yusuf Khan,
Oslo University Hospital, Norway

REVIEWED BY
Claudia Teixeira Guimaraes,
Brazilian Agricultural Research Corporation (EMBRAPA), Brazil
Lohithaswa Hirenallur Chandappa,
University of Agricultural Sciences, Bangalore, India

*CORRESPONDENCE
Basudeb Sarkar
basudeb70@gmail.com

†These authors contributed equally to this work

RECEIVED 15 December 2022
ACCEPTED 27 March 2023
PUBLISHED 08 May 2023

CITATION
Sarkar B, Varalaxmi Y, Vanaja M, RaviKumar N, Prabhakar M, Yadav SK, Maheswari M and Singh VK (2023) Mapping of QTLs for morphophysiological and yield traits under water-deficit stress and well-watered conditions in maize. *Front. Plant Sci.* 14:1124619. doi: 10.3389/fpls.2023.1124619

Mapping of QTLs for morphophysiological and yield traits under water-deficit stress and well-watered conditions in maize

Basudeb Sarkar*†, Yellisetty Varalaxmi†, Maddi Vanaja, Nakka RaviKumar, Mathyam Prabhakar, Sushil Kumar Yadav, Mandapaka Maheswari and Vinod Kumar Singh

Division of Crop Sciences, Indian Council of Agricultural Research (ICAR)—Central Research Institute for Dryland Agriculture, Hyderabad, Telangana, India

Maize productivity is significantly impacted by drought; therefore, improvement of drought tolerance is a critical goal in maize breeding. To achieve this, a better understanding of the genetic basis of drought tolerance is necessary. Our study aimed to identify genomic regions associated with drought tolerance-related traits by phenotyping a mapping population of recombinant inbred lines (RILs) for two seasons under well-watered (WW) and water-deficit (WD) conditions. We also used single nucleotide polymorphism (SNP) genotyping through genotyping-by-sequencing to map these regions and attempted to identify candidate genes responsible for the observed phenotypic variation. Phenotyping of the RILs population revealed significant variability in most of the traits, with normal frequency distributions, indicating their polygenic nature. We generated a linkage map using 1,241 polymorphic SNPs distributed over 10 chromosomes (chrs), covering a total genetic distance of 5,471.55 cM. We identified 27 quantitative trait loci (QTLs) associated with various morphophysiological and yield-related traits, with 13 QTLs identified under WW conditions and 12 under WD conditions. We found one common major QTL (*qCW2–1*) for cob weight and a minor QTL (*qCH1–1*) for cob height that were consistently identified under both water regimes. We also detected one major and one minor QTL for the Normalized Difference Vegetation Index (NDVI) trait under WD conditions on chr 2, bin 2.10. Furthermore, we identified one major QTL (*qCH1–2*) and one minor QTL (*qCH1–1*) on chr 1 that were located at different genomic positions to those identified in earlier studies. We found co-localized QTLs for stomatal conductance and grain yield on chr 6 ($qg_s6–2$ and *qGY6–1*), while co-localized QTLs for stomatal conductance and transpiration rate were identified on chr 7 ($qg_s7–1$ and *qTR7–1*). We also attempted to identify the candidate genes responsible for the observed phenotypic variation; our analysis revealed that the major candidate genes associated with QTLs detected under water deficit conditions were related to growth and development, senescence, abscisic acid (ABA) signaling, signal transduction, and transporter activity in stress tolerance. The QTL regions identified in this study may be useful

in designing markers that can be utilized in marker-assisted selection breeding. In addition, the putative candidate genes can be isolated and functionally characterized so that their role in imparting drought tolerance can be more fully understood.

KEYWORDS

drought, physiological and yield traits, SNPs, QTLs, candidate genes

1 Introduction

Maize is a widely consumed staple food and is also used for feed and as an industrial material. However, drought stress has become a major challenge to its productivity, particularly during the anthesis-silking and grain-filling stages (Lobell et al., 2014; Liu and Qin, 2021). Improving drought tolerance in maize is a complex task due to the polygenic nature of this trait and a large amount of genotype × environment interaction (Shinozaki and Yamaguchi-Shinozaki, 2007; Xue et al., 2013; Thirunavukkarasu et al., 2014). Conventional breeding has mainly improved grain yield (GY, g/plant) in favorable environments, and has not done so in drought-prone areas. To address these challenges, a combination of different breeding strategies and the use of genomic tools is necessary. The identification of quantitative trait loci (QTLs) and candidate genes, along with the use of marker-assisted selection in breeding, is critical for this process (Lebreton et al., 1995; Simko et al., 1997; Collins et al., 2008). To improve plant tolerance to drought stress, it is essential to have access to genotypic and phenotypic data, which can be continuously analyzed to gain a better understanding of plant responses (Ribaut et al., 2009). Drought stress can lead to a range of morphophysiological and biochemical changes in plants, such as decreased leaf water content and photosynthesis levels, as well as altered metabolism. These changes can result in reduced plant height, cob weight, biomass, and grain yield (Tester and Langridge, 2010). Linkage and association mapping using next-generation sequencing (NGS) technologies is becoming increasingly popular in the identification of QTLs for complex traits, such as drought tolerance, which is essential for marker-assisted selection (MAS) in breeding. Biparental mapping populations are typically used in QTL identification, in which genotypes with contrasting traits are crossed to produce recombinant inbred lines (RILs), followed by multiple generations of selfing. Through QTL mapping, chromosomal fragments linked with the trait of interest can be identified.

Previously, genetic linkage maps were created using PCR-based markers, such as random amplified polymorphic DNA markers (RAPDs) and simple sequence repeats (SSRs), as well as non-PCR-based markers, such as restriction fragment length polymorphisms (RFLPs). However, rapid advancements in sequencing technology have led to the use of single nucleotide polymorphisms (SNPs) for the development of high-resolution linkage maps (Elshire et al., 2011). These developments in genomics have enabled the mapping of genomic regions associated with drought tolerance through QTLs and association mapping. Several major QTLs associated with drought stress tolerance in maize have been reported in studies by Sanguineti et al. (1999); Malosetti et al. (2008); Messmer et al. (2009); Messmer et al., (2011), Almeida et al. (2013); Liu and Qin (2021), and Leng et al. (2022). In a meta-analysis by Chen et al. (2017), in which 33 published studies of yield-related traits in maize were analyzed, 76 meta-QTLs were identified out of 999 QTLs across the maize genome, although these were reported for normal growth conditions. In a recent review, Liu and Qin (2021) highlighted the progress that has been made in the genetic dissection of drought tolerance in maize at different phenophases of the crop through linkage mapping and association mapping, using various molecular markers including RFLPs, SSRs, and SNPs. In QTL mapping studies, QTLs can be categorized as either 'constitutive' (Collins et al., 2008; that is, the same QTLs are detected in different environments) or 'adaptive' (that is, QTLs are detected only in specific environments; Almeida et al., 2013). Identification of constitutive or adaptive QTLs can provide valuable insights into ways of improving field-level stress tolerance. Co-localized QTLs for different traits under stress can help in determining whether a particular trait is constitutive or adaptive and in determining its role in improving field-level drought tolerance. This information is important for the identification and selection of appropriate breeding strategies to develop drought-tolerant maize varieties.

Maize is considered a model crop for research in plant genetics due to the availability of a vast amount of omics data (Wallace et al, 2014). The first release of the maize B73 reference genome (Schnable et al., 2009) led to the development of several omics datasets, including DNA resequencing, transcriptomic, metabolomic, and proteomic data (Gore et al., 2009; Chia et al., 2012; Jiao et al., 2012; Li et al., 2013; Wen et al., 2014; Walley et al., 2016; Wen et al., 2018; Jiang et al., 2019; Liu et al., 2020; Wang et al., 2020). Recently, pan-maize gene sets and a pan-*Zea* genome map have been constructed to aid in genetic improvement (Hufford et al., 2021; Gui et al., 2022), and population-level transcripts of diverse lines are also available (Hirsch et al., 2014; Jin et al., 2016). With the wealth of whole-genome sequence data available, mining for candidate genes responsible for phenotypic variation could provide valuable insights into the molecular mechanisms of drought tolerance in maize.

Against this background, the present study aimed to map, through linkage mapping analysis using a subset of maize RILs,

the genomic regions that are associated with drought tolerance-related morphophysiological and yield traits, by phenotyping under both well-watered (WW) and water-deficit (WD) conditions and genotyping *via* high-throughput SNP sequencing. An additional aim was the identification of both major and minor effect QTLs and associated candidate genes.

2 Materials and methods

2.1 Plant material

Contrasting genotypes for drought tolerance were identified based on multi-year evaluation of genotypes for various morphophysiological traits (Maheswari et al., 2016). The drought-tolerant genotype SNJ201126 and susceptible genotype HKI161 were used for development of a mapping population following the single cob method. Initial biparental crossing between tolerant and susceptible genotypes, was conducted during the rainy season of the year 2014. Subsequently, the F_1 generation was self-pollinated for nine generations to develop a mapping population consisting of 264 single-plant progenies.

2.2 Phenotyping for morphophysiological and yield-related traits

The RIL populations, consisting of 264 single-plant progenies and their parents, were planted in a single-row plot of 2.5 m, with 60 cm spacing between rows and 25 cm between plants. Separate experiments were conducted under both WW and WD conditions, following a randomized complete block design (RCBD) with three replications. In the experiment under WW conditions, populations were grown under normal growth conditions until maturity, with irrigation whenever required. However, in the experiment under WD conditions, irrigation was provided only up to the vegetative stage, i.e., 45 days after sowing (DAS); this was followed by imposition of a water deficit for a period of 10 days in order to expose plants to drought stress, which coincided with the anthesis-silking interval (ASI). These two sets of experiments with different treatments were repeated for two seasons: specifically, the rainy season of 2018 and the post-rainy season 2018–19. The experiments were carried out with appropriate plant protection measures in place, and in accordance with recommended practices for growing healthy crop. During the 2018 rainy season, the weekly average temperature varied between 17.9°C and 31.3°C, with relative humidity of 57.1%–84.5%; during the post-rainy season, the temperatures recorded fell between 11.4°C and 30.4°C, with relative humidity of 40.9%–83%. The total rainfall recorded was 377 mm and 16 mm during the rainy and post-rainy seasons, respectively (Figure S1).

Various morphophysiological parameters were recorded under both WW and WD conditions: these consisted of Normalized Difference Vegetation Index (NDVI), net CO_2 assimilation rate (A_{net}), stomatal conductance to water vapor (g_s), transpiration rate (TR), leaf temperature (LT), and anthesis-silking interval (ASI). Yield-contributing traits (i.e., cob height (CH), cob weight (CW), total biomass (TB), and grain yield (GY) were recorded for three representative plants of each genotype. NDVI was measured using a GreenSeeker® 505 device (Manuel NTech Industries Inc., Ukiah, CA, USA). This device measures the reflected light on the canopy of crops in the 660 nm (red) and 770 nm (near-infrared) bands. The NDVI value for any given point in the image, at a particular phenophase of the crop, is equal to the difference in the intensities of reflected light in the red and infrared range divided by the sum of these intensities. A_{net}, g_s, LT, and TR were measured using the LI-6400 portable photosynthesis system (LI-COR Instruments, Inc., Lincoln, NE, USA).

2.3 SNP genotyping

The mean phenotyping data for the mapping population (consisting of 264 single-plant progenies), as evaluated under both WW and WD conditions, were used for a cluster analysis, carried out by the average-linkage distance method using SAS® version 9.3 statistical software (SAS Institute Inc., Cary, NC, USA; Cary, 2011). The cluster analysis using combined mean data on morphophysiological and yield-related traits [i.e., relative water content, canopy temperature depression, quantum yield, Soil Plant Analysis Development (SPAD) chlorophyll meter readings, NDVI, proline content, net CO_2 assimilation rate, g_s, TR, LT, plant height, CH, cob length, cob girth, number of kernel rows per cob, number of kernels per row, CW, grain yield, TB per plant, and harvest index] under WD conditions was used to group the mapping population into diverse groups. Specifically, the population was grouped into eight clusters (Table S1) based on the average distances between all pairs of cluster members between the clusters. The RIL IDs (264 in total), their corresponding cluster IDs, and the corresponding distances are shown in Table S1. A subset of 79 RILs were selected from these eight clusters in such a way as to fully capture the genetic diversity of the mapping population (Table S2). The frequency distribution of this subset, when compared with the whole population for different traits, was found to represent the phenotypic variation of the population. Along with this subset of 79 RILs, the parents SNJ201126 and HKI161 (in triplicate) were subjected to SNP genotyping at Bionivid Technology Pvt. Ltd., Bengaluru, India. The Illumina NGS workflow for SNP genotyping was employed, as follows. First, the young leaves of 15-day-old seedlings of each genotype were used for DNA isolation using the DNAeasy Plant Mini Kit (Qiagen, Hilden, Germany). Next, the DNA quality and quantity were determined *via* agarose gel electrophoresis and a NanoDrop spectrophotometer, respectively. For library construction, DNA was fragmented randomly and adapters were ligated to the 5′ and 3′ ends. These fragments were then amplified by PCR and purified from the gel. Clusters were generated by loading the library into a flow cell, where fragments were captured on a lawn of surface-bound oligos complementary to the library adapters. After cluster generation, the templates were sequenced. Sequencing was carried out using Illumina SBS technology, a system that detects single bases as they are incorporated into template strands. As all four reversible terminator-bound dNTPs are present during each sequencing cycle, natural competition minimizes incorporation bias and reduces the raw error rate in comparison to other technologies. This enables highly accurate

base-by-base sequencing that virtually eliminates sequence-context-specific errors, even within repetitive sequence regions and homopolymers. Sequencing data were subsequently converted to raw data for analysis. The Illumina sequencer generates raw images utilizing sequencing control software for system control and base calling through an integrated primary analysis software tool called Real-Time Analysis (RTA). The base calls (BCL) binary was converted to FASTQ format using Illumina package bcl2fastq. The total numbers of bases and reads, along with values of GC (%), Q20 (%), and Q30 (%), were calculated for all samples. The raw sequences of genotyping data were deposited in the NCBI database (http://www.ncbi.nlm.nih.gov/sra/PRJNA913688) with accession number PRJNA913688.

2.4 Statistical analysis

The phenotyping data of the subset of 79 RILs, which were generated in two seasons under WW and WD conditions, were used for statistical analysis, as the sequencing data were generated for the same set. Individual and combined (i.e., over seasons and treatments) analyses of variance (ANOVAs) were carried out for morphophysiological traits; Pearson's simple correlation coefficients were also calculated and heritability estimates were made using SAS version 9.3. For combined analyses, the homogeneity of variance component was determined using Bartlett's test (Bartlett, 1937). Broad-sense heritability was calculated as per the following formula:

$$\text{Broad-sense heritability } (H^2) = \frac{\sigma^2 G}{\sigma^2 P}$$

where $\sigma^2 G$ is the total genotypic variance and $\sigma^2 P$ is the total phenotypic variance. Frequency distribution histograms for all traits were generated using Matplotlib tools (Hunter, 2007). Matplotlib is a cross-platform data visualization and graphical plotting library for Python. The Pyplot module was used to generate plots, and the Scipy.stats module (Virtanen et al., 2020) was used to compute and draw histograms of the WW and WD data; evenly spaced points over a specified interval on the *x*-axis were created using numpy.linspace, and the norm of the probability density function is displayed on the plots. The Kolmogorov–Smirnov method (Kolmogorov–Smirnov Test, 2008) was used to test for normality. Descriptive statistics were calculated using SAS, version 9.3.

2.5 Bioinformatics analyses

2.5.1 SNP calling and filtering

The raw reads of the sequencing data of 79 RILs and their parents were generated in FASTQ format for all samples and imported into a TASSEL GBS pipeline (Glaubitz et al., 2014; Zhang et al., 2015), implemented in TASSEL version 5.0. Maize genotype B73, Zm-B73-Reference-NAM-5.0-Genome-Assembly–NCBI (https://www.ncbi.nlm.nih.gov/assembly/GCF_902167145.1/) was used as the reference genome. The qualifying filtering steps of the sequence reads were mapped onto the genome using the Burrows–Wheeler Alignment (BWA) tool (Li and Durbin, 2009). The mapped reads were then exported as a sequence alignment map (SAM) file for SNP calling and genotyping (Bradbury et al., 2007). A total of 176 Gb of data was generated for all sequenced samples. To filter the parent's call, replicates were first merged by ensuring that at least two replicates had observed calls, and the most common allele was taken as the parent call, with the alternate call within replicates of parents taken to reflect genotyping errors. Functional annotation of SNPs was carried out using the Ensembl Plants variant effect predictor (VEP) tool (http://plants.ensembl.org/Zea_mays/Tools/VEP?db=core) with maize reference assembly Zm-B73-REFERENCE-NAM-5.0. The SNPs already reported in the database were clustered into a non-redundant reference SNP cluster and assigned a unique rsID; for SNPs that did not clearly correspond to a clear rsID, an internal ID was given. In addition, SNPs with a minor allele frequency (MAF), i.e., a frequency of< 5%, were filtered out before analysis.

2.5.2 Linkage map construction and QTL analysis

Linkage map construction and QTL analysis were carried out using QTL IciMapping software, version 4.2 (Meng et al., 2015). The SNP data consisting of DNA bases (i.e., A, T, G, or C) were converted into the format recognized by the QTL IciMapping software using the SNP conversion functionality. SNPs showing non-polymorphism in parents or progenies, or missing in one or more of the parents, were removed by this functionality. The datasets thus generated for the RILs population therefore consisted of either one of the parental types (A or B) or missing data. Linkage map construction was carried out using the MAP functionality and comprised three steps: grouping, ordering, and rippling. The grouping of markers was based on anchored marker information and a logarithm of the odds (LOD) threshold score of 2.5 for unanchored markers. The ordering algorithm used was K-optimality by recombination, using the random nearest neighbor (NN) count route (10). The criteria used in rippling were a window size of five and recombination frequency (REC). Finally, the output was used to generate the linkage map. The anchoring and genotypic data generated along with the phenotypic data were used for QTL identification. A total of 1,241 SNPs were finally selected for analysis. The Biparental Populations (BIP) functionality of the software was used to study the association of these SNPs with morphophysiological and yield-related traits. The mapping method used was the inclusive composite interval mapping (ICIM) method for QTL with additive effects (i.e., ICIM-ADD). The mapping parameters set were stepwise scanning by 1 cM, deletion of missing phenotypic data, phenotype on marker variables (PIN) of 0.001, and a LOD threshold of 2.5. The QTL effects were estimated based on LOD, additive effect of identified loci, and percentage of phenotypic variation explained (PVE%). QTLs with PVE% of ≥ 15% and< 15% were regarded as major and minor effect QTLs, respectively. The standard procedure was followed for QTL nomenclature (McCouch, 2008).

The linkage map view package of R (The R Foundation for Statistical Computing, Vienna, Austria; Ouellette et al., 2018) was used to display the QTLs identified under WW and WD conditions, capturing only the regions where QTLs were located. An epistatic analysis was carried out using the IciMappingVer.4.1 EPI epistatic

module, with default parameter settings (LOD = 5, step = 1 cM, and stepwise regression probability< 0.0001). The combined phenotyping data from two seasons (i.e., the rainy and post-rainy seasons) and two different environments (i.e., the WW and WD conditions) were used to carry out a joint QTL analysis with additive-by-environment (A-by-E) interactive effects in a multi-environment trial (MET) module of the ICIM method (Meng et al., 2015). The parameters of the QTL analysis were set as follows: LOD threshold: 2.5; 1,000 permutations; step: 1 cM; and PIN: 0.001. The confidence interval (CI) of each QTL was determined by using LOD > 3.

2.5.3 Functional annotation of SNPs

Functional annotation of the selected SNPs was carried out using the VEP tool in Ensembl Plants (https://plants.ensembl.org/Zea_mays/Tools/VEP). The VEP tool analyses the variants and predicts the functional consequences of both known and unknown variations. The reference genome assembly was Zm-B73-REFERENCE-NAM-5.0 for each SNP to identify rs-ID, location, allele, consequence, gene, feature type, biotype, exon/intron, and TREMBL protein IDs were identified.

2.5.4 Identification of candidate genes in the genomic region spanning QTLs for water-deficit stress conditions

The QTL intervals obtained in the linkage map were further studied for the prediction of candidate gene(s) associated with the respective QTLs, using the maize genome sequence available at Ensembl Plants (https://ensembl.gramene.org/Zea_mays/). Marker intervals were mapped for their physical locations and the sequences between the intervals were retrieved using NCBI-BLAST. The numbers and types of genes present in the sequences were also identified. Functional annotation of the genes present within the QTL regions was carried out using the Database for Annotation, Visualization and Integrated Discovery (DAVID) tool for gene functional annotation (https://david.ncifcrf.gov/summary.jsp) and the maize genetics and genomics database (GDB) (https://www.maizegdb.org). The genomic region covering each QTL was further searched for the presence of QTLs reported by genome-wide association studies using the maize genome database (https://jbrowse.maizegdb.org).

3 Results

3.1 Phenotypic variation in the mapping population

The population consisted of 79 RILs, which demonstrated wide variation in their morphophysiological and yield-related traits (i.e., NDVI, A_{net}, g_s, TR, LT, ASI, CH, CW, TB, and GY) under WW and WD conditions (Table S3). All traits were affected by WD. Histograms of the frequency distributions of these traits in the WW and WD conditions are presented in Figure 1. All traits followed a near-normal distribution, with the exception of stomatal conductance, which is positively skewed in the WW condition. Based on the significance values obtained using the Kolmogorov–Smirnov test, the traits A_{net}, LT, CH, CW, GY, and TB followed a normal distribution under both under WW and WD conditions, whereas the traits g_s and TR followed a normal distribution under the WD condition (Table S4). This indicates that the selected RILs captured the genetic variability of the entire population to be utilized for QTL identification. The descriptive statistics of the mapping population under WW and WD conditions for morphophysiological and yield-related traits, including coefficients of variation (CV%), skewness, kurtosis, and heritability, are provided in Table S5. Moderate broad-sense heritability (H^2), ranging between 0.45 and 0.71, was observed for all traits, and a wide range of coefficients of variation was also observed among these traits.

A combined ANOVA over trials across seasons (i.e., rainy vs. post-rainy) and conditions (i.e., WW vs. WD) indicated significant interactions among season, treatment type, and genotype for most traits. For total biomass (TB), all interaction effects were significant except seasons × treatments (Table S6). For ASI, the interaction effects for treatments × genotypes and seasons × treatments × genotypes were non-significant. The significant variation by genotype, environment, and their interaction for a number of traits indicated that these traits were influenced by both genetic and environmental factors. A simple correlation coefficient analysis revealed significant positive correlations of NDVI with CH, CW, GY, and TB; A_{net} with g_s and TR; g_s with TR; CH with TB, GY, and TB; CW with GY and TB; and GY with TB under both WW and WD conditions. In addition, a significant positive correlation of NDVI with LT, and of CH with CW and GY, was observed under WD conditions (Table S7).

3.2 QTL mapping

The numbers of raw SNPs and polymorphic SNPs between the parents after filtering with MAF< 5%, along with their distribution in various chromosomes, the number of mapped SNPs in the linkage map, and the average marker interval, are presented in Table 1. The largest number of markers (219) was detected on chr 2, and the smallest (65) on chr 10. In the present study, 27 QTLs were identified; of these, 13 were detected under WW conditions, 12 under WD conditions, and two (*qCH1–1* and *qCW2–1*) under both WW and WD conditions (Table 2; Figure 2). Major and minor QTLs were detected only in chromosomes 1, 2, 3, 5, 6, and 7 (out of the 10 maize chromosomes studied under WW and WD conditions), and no QTLs were detected in chromosomes 4, 8, 9, or 10. Under WD conditions, QTLs for CH (*qCH1–1* and *qCH1–2*) and ASI (*qASI1–1*) and for NDVI, CW, TB, and GY were detected in chromosomes 1 and 2. Similarly, three (*qA$_{net}$3–1, qg$_s$3–1, qCH3–1*), two (*qg$_s$6–1, qGY6–1*) and one (*qg$_s$7–2*) QTLs were detected in chromosomes 3, 6, and 7, respectively, under WD conditions (Figure 2). QTLs with PVE% greater than 15% were regarded as major QTLs. Among the 13 QTLs detected under WW conditions, three major QTLs were identified for traits A_{net}, g_s, and TR, with LOD scores ranging from 2.54 to 6.07 and PVE% ranging from 15.86% to 21.47%. Ten minor QTLs were detected for traits LT, TR, g_s, TB, CH, and ASI, with LOD scores ranging from 2.54 to 3.2 and

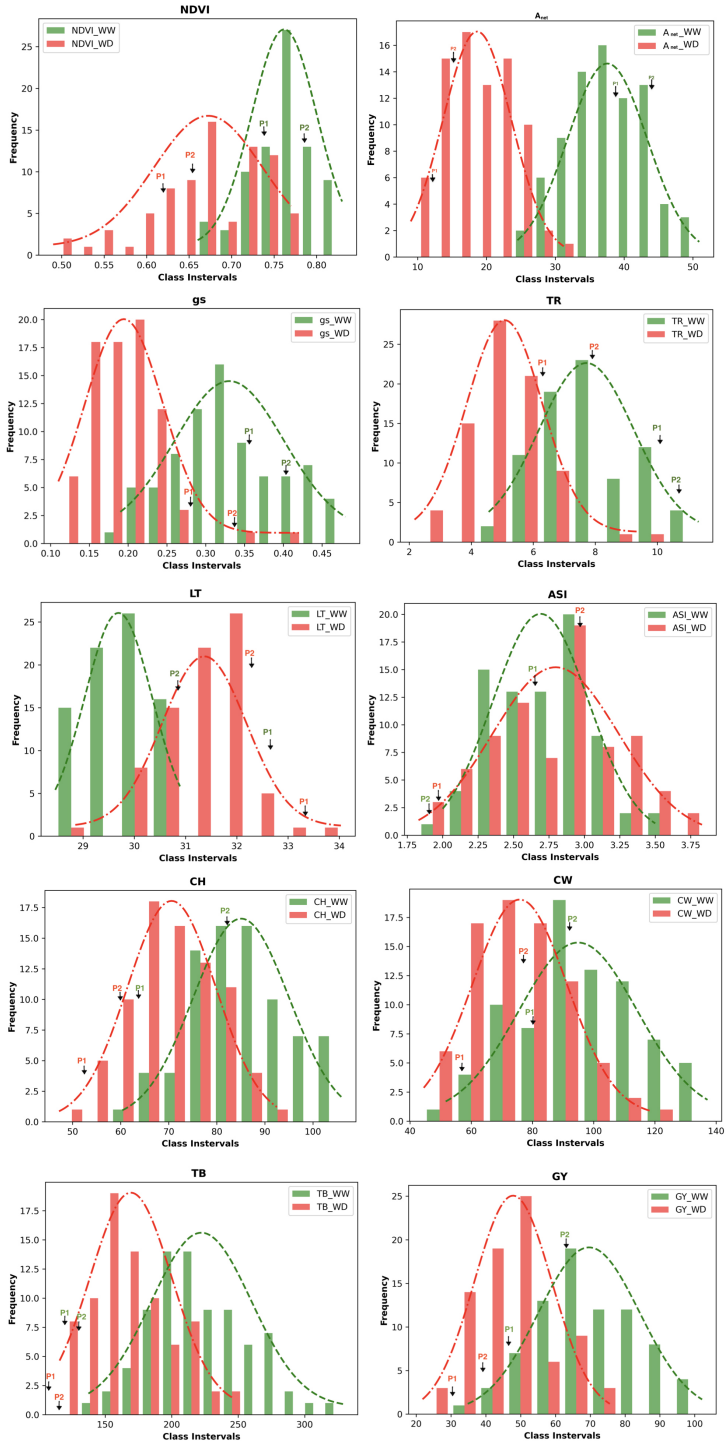


FIGURE 1
Frequency distribution of the phenotypic data of the RILs population of the various morphophysiological and yield related traits. The values of the parents (P_1 - HKII61 and P_2 SNJ201126 are indicated by arrows); NDVI, Normalized difference vegetation index; Anet, Net CO_2 assimilation rate; g_s, Stomatal conductance to water vapor; TR, Transpiration rate; LT, Leaf temperature; CH, Cob height; ASI, Anthesis-silking interval; CW, Cob weight; TB, Total biomass; GY,Grain yield/plant; WW, Well-watered; WD, Water-deficit stress.

PVE% ranging from 6.75% to 14.91% (Table S8). Among the 12 QTLs detected under WD conditions, three were major QTLs identified for traits NDVI, g_s, and CH, with LOD scores ranging from 3.49 to 4.45 and PVE% ranging from 15.05% to 18.59%. Nine minor QTLs were detected for traits A_{net}, NDVI, ASI, CH, g_s, TB, and GY, with LOD scores ranging from 2.52 to 3.37 and PVE% ranging from 9.69% to14.25% (Table 2). Among these, two QTLs were detected under both WW and WD conditions: these were a minor QTL for the CH trait and a major QTL for the CW trait, located on chromosomes 1 and 2, respectively.

TABLE 1 Numbers of single nucleotide polymorphisms (SNPs) on the 10 chromosomes of maize used for quantitative trait locus (QTL) mapping.

SNPs	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5	Chr 6	Chr 7	Chr 8	Chr 9	Chr 10	Total
Raw SNPs (*n*)	7,145	5,539	4,997	4,853	5,246	3,879	3,846	4,306	3,438	3,247	46,496
Polymorphic SNPs between the parents after filtering with MAF< 5% (*n*)	1,878	1,409	1,419	1,171	1,302	1,030	1,188	1,171	842	851	12,261
Mapped SNPs in linkage map (*n*)	144	219	182	85	96	95	163	119	73	65	1,241
Average marker interval (cM)	5.61	3.14	3.86	5.84	5.77	4.15	3.49	4.12	4.97	6.23	

MAF, minor allele frequency; cM, centimorgan.

For the trait NDVI, one major and one minor QTL were detected on chr 2, with LOD scores of 3.93 and 3.19 and capturing 18.59% and 14.02% PVE, respectively. For the A_{net} trait, major and minor QTLs were detected on chr 3, with LOD scores of 3.24 and 2.52 and capturing 17.24% and 14.25% PVE, respectively. For the g_s trait, two major QTLs on chr 6 and three minor QTLs, one on chr 3 and two on chr 7, were detected. The LOD scores for these ranged from 2.54 to 4.45, and PVE% ranged from 9.69% to 18.21%. For the TR trait, one major QTL was detected on chr 7 (with a LOD score of 6.07 and PVE% of 21.47%), and two minor QTLs were detected, one on chr 1 (with a LOD score of 2.87 and PVE% of 10.81%) and one on chr 3 (with a LOD score of 2.72 and PVE% of 10.11%).

For the ASI trait, three minor QTLs were detected, two on chr 1 and one on chr 6, with LOD scores of 2.61, 3.02, and 3.2 and PVE% of 13.92%, 14.91%, and 13.56%, respectively. For the CH trait, one major QTL, with a LOD score of 3.49 and PVE% of 15.05%, was detected on chr 1. In addition, three minor QTLs were detected on chromosomes 1, 3, and 8, with LOD scores of 2.52, 2.8, and 2.77 and PVE% of 13.84%, 11.18%, and 14.9%, respectively. For the TB trait, three minor QTLs were detected, two on chr 2 and one on chr 5, with LOD scores of 2.53, 2.58, and 2.55 and PVE% of 13.51%,

TABLE 2 Quantitative trait loci (QTLs) identified for various morphophysiological and yield-related traits under water-deficit (WD) conditions.

QTL name	Conditions	QTL type	Chromosome	Position of QTL	Left marker	Right marker	LOD	PVE (%)	Add	Interval map distance (cM)	Map distance (cM)
qNDVI2–1	WD	Minor	2	174	rs812099243	rs822182360	3.19	14.02	–0.02	172.5–175.5	3
qNDVI2–2	WD	Major	2	388	rs131350195	S2_66658066	3.93	18.59	0.03	378.5–388.5	10
qA$_{net}$3–1	WD	Minor	3	314	S3_169283017	S3_173528165	2.52	14.25	–1.94	312.5–317.5	5
qgs3–1	WD	Minor	3	168	S3_5950551	S3_5721251	3.37	12.87	–0.02	164.5–171.5	7
qgs6–1	WD	Major	6	109	S6_126753475	rs836167502	4.45	18.21	–0.02	105.5–112.5	7
qgs7–2	WD	Minor	7	372	S7_139259301	S7_139259336	2.89	9.69	0.02	370.5–374.5	4
qASI1–1	WD	Minor	1	129	rs128441140	rs128842621	2.61	13.92	0.22	120.5–139.5	19
qCH1–1	WW, WD	Minor	1	266	S1_38965222	S1_38965211	2.52	13.84	3.53	263.5–266.5	3
qCH1–2	WD	Major	1	740	S1_6365045	rs818095140	3.49	15.05	4.16	738.5–741.5	3
qCH3–1	WD	Minor	3	235	S3_194753667	rs277236564	2.8	11.18	–3.35	232.5–239.5	7
qCW2–1	WW, WD	Major	2	442	S2_213060766	S2_18995588	3.03	17.95	8.59	434.5–445.5	11
qTB2–2	WD	Minor	2	455	S2_18995588	S2_14679066	2.58	11.5	17.91	444.5–464.5	20
qGY2–1	WD	Minor	2	457	S2_18995588	S2_14679066	2.78	10.54	6.2	445.5–467.5	22
qGY6–1	WD	Minor	6	275	S6_89546385	S6_179562549	2.54	11.33	6.41	250.5–286.5	36

cM, centimorgan; NDVI, Normalized Difference Vegetation Index; A_{net}, net CO_2 assimilation rate (μmol CO_2 m^{-2} s^{-1}); g_s, stomatal conductance to water vapor (mol m^{-2} s^{-1}); ASI, anthesis-silking interval (days); CH, cob height (cm); CW, cob weight (g/cob); TB, total biomass (g/plant); GY, grain yield (g/plant); WW, well-watered conditions; WD, water-deficit conditions; Chr, chromosome; LOD, logarithm of odds ratio; PVE (%), percentage of total phenotypic variance explained by the QTL; Add, additive effect; QTL, quantitative trait locus.

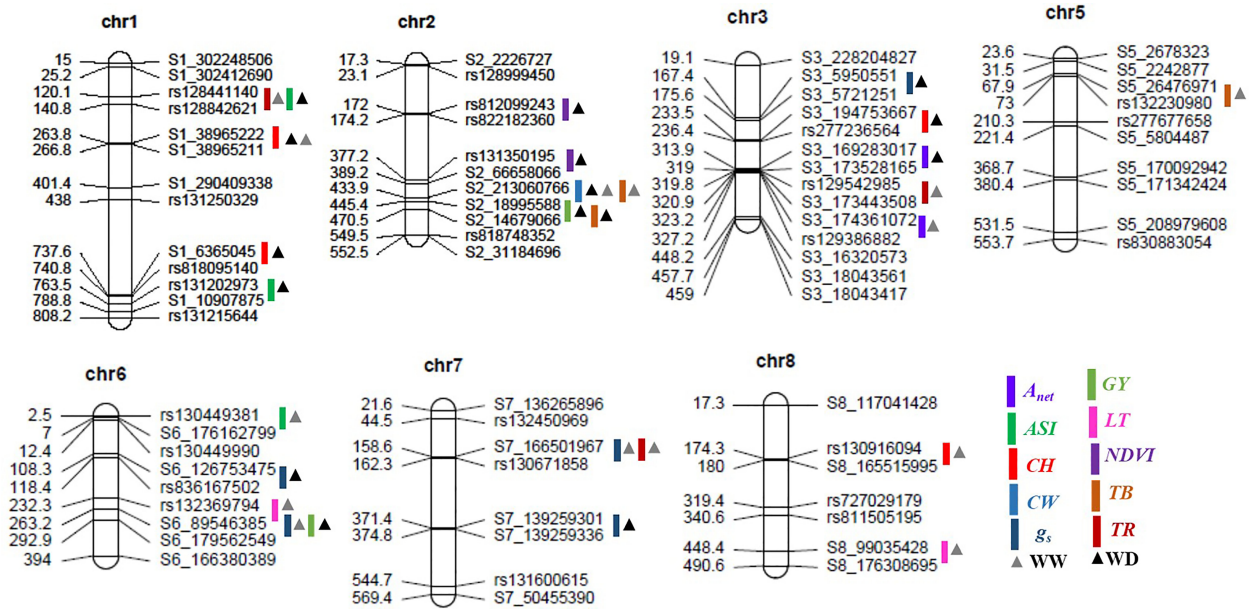


FIGURE 2
Positions of the quantitative trait loci (QTLs) for various morphophysiological and yield-related traits on seven chromosomes in the recombinant inbred line (RIL) population grown under well-watered and water-deficit conditions. The scaled numbers on the left side of the chromosomes indicate genetic length (cM), with the corresponding markers on the right side. The colored bars and triangles represent the QTLs identified for the various traits and the condition, respectively. A_{net}, net CO_2 assimilation rate; ASI, anthesis-silking interval; CH, cob height; CW, cob weight; g_s, stomatal conductance to water vapor; GY, grain yield; LT, leaf temperature; NDVI, Normalized Difference Vegetation Index; TB, total biomass; TR, transpiration rate; chr, chromosome; WW, well-watered conditions; WD, water-deficit conditions.

11.50%, and 13.04%, respectively. For the GY trait, two minor QTLs were detected, one on chr 2 and one on chr 6, with LOD scores of 2.78 and 2.54 and PVE% of 10.54% and 11.33%, respectively. Finally, for the CW trait, one major QTL was detected on chr 2, with a LOD score of 3.03 and PVE% of 17.91%.

3.3 Identification of co-localized QTLs

Of the 27 QTLs detected under WW and WD conditions, three pairs of co-localized QTLs were identified. For traits TR and ASI, two QTLs (*qTR1–1* and *qASI1–1*) were co-localized at the marker interval rs128441140–rs128842621 on chr 1, with LOD scores of 2.87 and 2.61 and PVE% of 10.81% and 13.92%, respectively. For traits g_s and GY, two QTLs (*qg$_s$6–2* and *qGY-6–1*) were co-localized at the marker interval S6_89546385–S6_179562549 on chr 6, with a LOD score of 2.54 and 2.54 and PVE% of 15.86% and 11.77%, respectively. Finally, for the g_s and TR traits, two QTLs (*qg$_s$7–1* and *qTR7–1*) were co-localized at the marker interval S7_166501967–rs130671858 on chr 7, with LOD scores of 2.73 and 6.07 and PVE% of 11.3% and 21.07%, respectively.

3.4 Epistatic interaction among QTLs

Eleven significant digenic epistatic QTLs for traits A_{net}, g_s, TR, LT, ASI, CW, and TB were detected (Table 3; Figures S2A, B). The corresponding LOD scores ranged between 5.0 and 5.91, and PVE% ranged from 16.60% to 37.43%. The negative epistatic values (add-by-add) indicated that there was a stronger epistatic effect of recombinant genotype than of parental genotype. The epistatic effect of QTLs was negative for traits TB, A_{net}, and LT, but positive for traits g_s, ASI, TR, A_{net}, and CW. The genomic region on chr 2 between the markers rs129243511 and S2_229825946 showed epistatic interaction for the CW trait on chr 8, located between the markers S8_145945904 and S8_148216407. This interaction contributed 37.43% of the PVE. Two epistatic interactions were identified for A_{net}. The first involved the region on chr 1 between the markers S1_3798004 and rs128441140, which showed epistatic interaction with the region on chr 3 located between the markers rs129555629 and S3_183844216. This interaction contributed 17.60% of the PVE. For the second, a region on chr 2 between the markers rs72722896 and rs276685886 showed epistatic interaction for A_{net} with a region on chr 3, located between the markers S3_2276048 and S3_89193637. This interaction contributed 16.71% of the PVE. Finally, the genomic region on chr 2 between the markers rs131971876 and rs129196105 showed epistatic interaction for the TR trait with the region on chr 7 located between the markers S7_131791490 and S7_136261108. This interaction contributed 29.55% of the PVE.

3.5 QTL–environment interaction

Using the MET (multi-environmental trials) module of ICIM, a total of 53 QTLs were identified for various traits (Table S9). Of these, 21 QTLs were common to those identified by the ICIM-ADD

TABLE 3 Epistatic interactions observed for morphophysiological and yield-related traits.

Trait	Condition	First chromosome	Position1 (cM)	Left marker 1	Right marker 1	Second chromosome	Position2 (cM)	Left marker 2	Right marker 2	LOD	PVE (%)	Add-by-add[a]
A_{net}	WW	1	115	S1_3798004	rs128441140	3	350	rs129555629	S3_183844216	5.08	17.60	3.56
A_{net}	WW	2	365	rs727228961	rs276685886	3	635	S3_2276048	S3_89193637	5.91	16.71	-3.90
g_s	WW	6	40	S6_177038002	S6_148279548	7	120	rs130642102	rs836234066	5.32	16.60	0.05
g_s	WW	7	270	S7_11679195	rs130488903	8	250	rs814260212	S8_73973624	5.48	18.05	0.05
TR	WD	2	265	rs131971876	rs129196105	7	365	S7_131791490	S7_136261108	5.79	29.55	0.63
LT	WD	4	60	S4_166565742	S4_182462050	6	170	rs55624911	rs130311785	5.22	34.85	-0.60
ASI	WD	2	640	S2_221752549	rs833320055	3	270	rs839843727	rs277263572	5.86	24.89	0.31
ASI	WD	1	25	S1_302248506	S1_302412690	3	210	rs131368069	rs132076316	5.00	20.18	0.27
CW	WD	2	60	rs129243511	S2_229825946	8	220	S8_145945904	S8_148216407	5.13	37.43	7.58
TB	WD	5	270	rs129997839	rs130004128	5	295	S5_24426031	S5_29004727	5.10	22. 50	-24.04
TB	WD	2	55	rs129234757	rs129243510	5	255	S5_9853336	S5_10571036	5.19	18.89	-18.16

cM, centimorgan; A_{net}, net CO_2 assimilation rate (μmol CO_2 m^{-2} s^{-1}); g_s, stomatal conductance to water vapor (mol m^{-2} s^{-1}); TR, transpiration rate (mmol H_2O m^{-2} s^{-1}); LT, leaf temperature ($^{\circ}C$); ASI, anthesis-silking interval (days); CW, cob weight (g/cob); TB, total biomass (g/plant); WW, well-watered conditions; WD, water-deficit conditions; LOD, logarithm of the odds ratio, where the threshold value was ≥ 5; PVE(%), total phenotypic variance in percentage explained by the QTL; add-by-add[a], additive-by-additive effect; negative additive effect value indicates the direction of favorable allele from donor parent; QTL, quantitative trait locus.

10.3389/fpls.2023.1124619

method. One QTL for each of the traits g_s, ASI, and GY and two QTLs for the LT trait were identified using the ICIM-ADD method but were not detected by MET analysis. Conversely, the MET analysis identified 32 QTLs for traits A_{net}, CH, ASI, TB, GY, NDVI, LT, and CW that had not been detected using the ICIM-ADD method. The proportion of phenotypic variation captured by additive and dominance effects [PVE(A)] ranged from 0.13% to 14.36%, and the proportion captured by additive- and dominance-by-environment effects [PVE(A by E)] for corresponding QTLs ranged from 0% to 7.92% (Table S9). Thus, PVE(A by E) was significantly lower than PVE(A). Most QTLs detected through the MET module of ICIM were non-significant. However, traits A_{net}, ASI, g_s, and TR were found under MET to make greater contributions to phenotypic variation (with contributions ranging from 5.24% to 21.7%), they were not found to contribute to the significant QTL × E interaction effect (Table 4).

3.6 Functional annotation of SNPs

The functional annotation of genes associated with the major and minor QTLs for morphophysiological and yield-related traits and their biological/molecular functions under WD and WW conditions are shown in Tables 5, S10, respectively. The QTL regions with important genes imparting stress tolerance for different traits belonged to the categories of signal transduction (GY: *Zm00001eb297570*, protein serine/threonine phosphatase); transcription factors [NDVI: *Zm00001eb118010*, G2-like transcription factor 27; ASI: *Zm00001eb295810*, NAC-type transcription factor (NAC87)]; transporter activity (A_{net}: *Zm00001eb146040*, chloride channel protein; g_s: *Zm00001eb324180*, sugar carrier protein C; TR: *Zm00001eb015510*, phospholipid-transporting ATPase; CH: *Zm00001eb363270*, calcium-transporting ATPase); cell wall biosynthesis and its organization (TR: *Zm00001eb144960*, lipoxygenase; TR: *Zm00001eb145080*, pectin acetyl esterase); photosynthesis (g_s and TR: *Zm00001eb324240*, chlorophyll a/b-binding protein, chloroplast); and carbon utilization (CH: *Zm00001eb002270*, glyceraldehyde phosphate dehydrogenase B1).

3.7 Identification of candidate genes in the genomic region spanning QTLs under WD conditions

Table S11 shows the names of the QTLs, their chromosome locations, left markers, right markers, QTL intervals (cM), the physical location of each QTL region (i.e., start and end), its size, and the number of genes within the QTL region. Of the 14 QTLs identified under WD conditions, two QTLs, *qg$_s$7–2* and *qCH1–1*, encompassed a single gene, namely the Nicotinamide adenine dinucleotide phosphate-binding Rossmann-fold superfamily protein (*Zm00001eb316940*) and LIM homeodomain proteins transcription factor 2 (*Zm00001eb011970*), respectively. Although the physical distances between the markers of the QTLs *qg$_s$6–1*, *qASI1–1*, *qCW2–1*, and *qGY6–1* were large, linkage with the trait could not be ascertained, meaning that these were not used. The physical sizes of the QTLs *qNDVI2–1*, *qNDVI2–2*, *qA$_{net}$3–1*, *qg$_s$3–1*, *qCH1–2*, *qCH3–1*, *qTB2–2*, and *qGY2–1* were below< 5 Mbp; these encompassed 22, 108, 264, 35, 31, 64, 372, and 372 protein-coding genes, respectively. Genes that played a significant role in WD tolerance in these QTL intervals were also identified. The details of QTLs detected in earlier studies in the QTL regions, i.e., *qA$_{net}$3–1*, *qg$_s$3–1*, *qASI1–1*, *qCH3–1*, and *qGY6–1*, are listed in Table S12.

4 Discussion

High variability was observed in various morphophysiological and yield-related traits under WW and WD conditions in a field phenotyping study of the RIL population. In particular, significant variation was observed in the NDVI, A_{net}, g_s, TR, LT, ASI, CH, CW, TB, and GY traits. The RIL population displayed transgressive segregation for traits A_{net}, CH, TB, and GY under both

TABLE 4 Quantitative trait locus (QTL) × environment (E) interactions in the recombinant inbred line (RIL) population over two seasons (rainy and post-rainy).

Trait	Chromosome	Position	Left marker	Right marker	LOD	LOD (AbyE)	PVE	PVE (AbyE)	Add	AbyE_01	AbyE_02
A_{net}	3	314	S3_169283017	S3_173528165	2.55	1.1	9.51	3.55	–1.09	0.84	–0.84
A_{net}	3	324	S3_174361072	rs129386882	3.23	1.35	15.53	7.92	–1.22	–1.24	1.24
ASI	1	786	rs131202973	S1_10907875	3.18	1.06	5.24	0.37	–0.09	–0.02	0.02
g_s	3	168	S3_5950551	S3_5721251	3.33	2.17	6.14	2.64	–0.01	0.01	-0.01
g_s	6	109	S6_126753475	rs836167502	4.66	2.15	10.17	1.81	–0.02	0.01	-0.01
g_s	7	372	S7_139259301	S7_139259336	2.89	2.14	5.71	3.06	0.01	-0.01	0.01
g_s	8	232	S8_152991912	S8_123369436	2.65	2.65	5.59	5.46	0.00	-0.01	0.01
TR	7	162	S7_166501967	rs130671858	6.29	1.48	21.7	7.34	0.45	0.32	–0.32

cM, centimorgan; A_{net}, net CO_2 assimilation rate (μmol CO_2 m^{-2} s^{-1}); g_s, stomatal conductance to water vapor (mol m^{-2} s^{-1}); NDVI, Normalized Difference Vegetation Index; TR, transpiration rate (mmol H_2O m^{-2} s^{-1}); LOD, logarithm of the odds ratio; LOD (AbyE), LOD score for additive- and dominance-by-environment effects; PVE (AbyE), phenotypic variation explained by additive- and dominance-by-environment effects at the current scanning position; AbyE_01, additive- and dominance-by-environment effect 1 at the current scanning position; AbyE_02, additive- and dominance-by-environment effect 2 at the current scanning position.

10.3389/fpls.2023.1124619

TABLE 5 List of annotated genes present within quantitative trait loci (QTLs) identified under water-deficit (WD) conditions for various morphophysiological and yield traits.

QTL name	Chromosome	Position (start–end bp)	Position of SNP	SNP	Gene size (bp)	Locus ID	Annotation	Biological process
qNDVI2–1	2	241,929,343–241,931,437	241,930,638	G	2094	*Zm00001eb118010*	G2-like transcription factor 27 (glk27)	DNA binding
$qA_{net}3–1$	3	169,275,199–169,279,402	169,283,017	C	4203	*Zm00001eb144000*	E3 ubiquitin–protein ligase RGLG1	Metal ion binding
	3	169,278,633–169,284,255	169,283,017	C	3002	*Zm00001eb144010*	Amino acid/auxin permease 20	Amino acid transport
	3	173,533,121–173,537,574	173,528,165	A	4318	*Zm00001eb145030*	BTB/POZ and TAZ domain-containing protein 3	Ubiquitin conjugation pathway
$qg_s3–1$	3	5,718,725–5,721,785	5,721,251	G	3060	*Zm00001eb120960*	Putative transcription factor bHLH041	Protein dimerization activity
$qg_s6–1$	6	126,751,036–126,754,755	126,753,475	A	3719	*Zm00001eb280280*	Brassinosteroid-insensitive 1-associated receptor kinase 1	Protein serine/threonine kinase activity
	6	120,726,441–120,737,326	120,736,548	G	10,885	*Zm00001eb278890*	Transcription elongation factor SPT5	mRNA binding/ transcription regulation
$qg_s7–2$	7	139,256,603–139,259,621	139,259,301	C	3018	*Zm00001eb316940*	NAD(P)-binding Rossmann-fold superfamily protein	Oxidoreductase activity
qASI1–1	1	53,207,544–53,208,672	53,209,688	C	1128	*Zm00001eb015510*	Phospholipid-transporting ATPase	Phospholipid translocation
qCH1–1	1	38,963,680–389,674,58	38,965,222	A	3565	*Zm00001eb011970*	LIM zinc-binding domain-containing protein	Cell cycle related
qCH1–2	1	6,360,197–6,362,989	6,365,045	A	2792	*Zm00001eb002270*	Glyceraldehyde phosphate dehydrogenase B1	Glucose metabolic process
qCH3–1	3	194,748,309–194,750,865	194,753,667	A	2556	*Zm00001eb151120*	Pentatricopeptide repeat-containing protein	Organellar biogenesis
	3	195,624,083–195,628,950	195,627,269	G	3257	*Zm00001eb151330*	Receptor-like serine/threonine protein kinase	Protein phosphorylation/ assimilate partitioning
qCW2–1	2	241,494,911–241,497,812	241,498,576	G	2651	*Zm00001eb117750*	Proline-rich receptor-like protein kinase PERK4	Protein serine/threonine kinase activity
qTB2–2	2	14,676,667–14,679,369	14,679,066	T	2562	*Zm00001eb072580*	Osjnbb0016d16.16-like protein	–
qGY2–1	2	14,676,667–14,679,369	14,679,066	T	2562	*Zm00001eb072580*	Osjnbb0016d16.16-like protein	–
qGY6–1	6	179,555,644–179,566,050	179,562,549	C	7615	*Zm00001eb297570*	Protein serine/threonine phosphatase	Protein dephosphorylation
	6	179,563,176–179,566,051	179,562,549	C	2875	*Zm00001eb297580*	Pentatricopeptide repeat-containing protein mitochondrial	Zinc ion binding

NDVI, Normalized Difference Vegetation Index; A_{net}, net CO_2 assimilation rate (μmol CO_2 m^{-2} s^{-1}); g_s, stomatal conductance to water vapor (mol m^{-2} s^{-1}); ASI, anthesis-silking interval (days); CH, cob height (cm); CW, cob weight (g/cob); TB, total biomass (g/plant); GY, grain yield (g/plant); SNP, single nucleotide polymorphism.

conditions, suggesting the presence of new combinations of multiple genes with more positive or negative effects on quantitative traits than were present in either parent. The identification of progeny plants with favorable gene combinations means that they can serve as donors for further genetic improvement. The frequency distributions for the RIL subset showed near-normal distributions for most traits, indicating their polygenic nature. Positive correlations were observed between the A_{net} trait and its related traits, such as g_s and TR, and between yield and its related traits, such as CW, TB, and GY, under both conditions. NDVI was positively correlated with CH, CW, TB, and GY under both conditions. TR also showed a positive correlation with g_s as did A_{net} with g_s and TR, under WD conditions. Previous studies have also reported positive correlations between NDVI and GY (Trachsel et al., 2016). Interestingly, significant positive correlations were observed

between morphophysiological and yield-related traits in the present study (Table S7), whereas only non-significant correlations were reported in an earlier study (Nikolić et al., 2012).

4.1 QTL mapping

The genetic map in this study consisted of 1,241 SNP markers spread across 10 linkage groups, covering a total genetic distance of 5,471.55 cM, with an average marker density of 4.4 cM per marker. RIL mapping populations are widely used for QTL identification (Li et al., 2013; Yang et al., 2016), and in this study, 27 QTLs were identified for various morphophysiological and yield-related traits under both WW conditions (13 QTLs) and WD conditions (12 QTLs). Notably, for the NDVI trait, both a major and a minor QTL were detected on chr 2, with bin position 2.10 under WD conditions. Previous studies have also reported QTLs for NDVI and plant height in two $BC_1F_{2:3}$ backcross populations (LPSpop and DTPpop), with 18 QTLs identified in total (Trachsel et al., 2016). In the DTPpop population, QTLs for NDVI influencing stay-green habit (*SEN6*) were also detected in bins 8.01 and 2.07. The present study identified several QTLs associated with various morphophysiological and yield-related traits under both WW and WD conditions. For the A_{net} trait, a major QTL (under WW conditions) and a minor QTL (under WD conditions) were identified on chr 3 (bp 3.05), which also showed the presence of a minor QTL for the TR trait under WW conditions. QTLs for the TR trait were also detected on chrs 1 and 7. Two major QTLs for the g_s trait were identified on chr 6 (under WD conditions) and two minor QTLs, one on chr 3 (under WD conditions) and one on chr 7 (under WW conditions), were also identified. A previous study by Yu et al. (2015) identified 32 QTLs associated with chlorophyll a, chlorophyll b, total chlorophyll content, net CO_2 photosynthetic rate, stomatal conductance, intercellular CO_2 concentration, and TR. Another study reported on the photosynthetic performance of maize grown in drought environments (Zhao and Zhong, 2021). In a previous study conducted by Pelleschi et al. (2006), 19 major QTLs were identified for various physiological traits under drought-stressed and WW regimes. QTLs for the stomatal conductance trait have been found on all chromosomes in maize except for chr 5 (Quarrie et al., 1994; Sanguineti et al., 1999). In the present study, two minor QTLs for the LT trait (under WW conditions) were identified on chrs 6 and 8. Three minor QTLs for the ASI trait were also identified: two on chr 1 (one under WW and the other under WD conditions) and one on chr 6 (under WW conditions). A total of 33 QTLs for the ASI trait under WD stress have been reported in earlier studies, distributed across all chromosomes (Semagn et al., 2013). Additionally, QTLs for ASI under both WW and WD conditions were also identified in another study (Hu et al., 2021). In this study, we identified one major QTL under WD stress and two minor QTLs (one under WW and one under WD conditions) for cob height on chr 1, as well as one minor QTL on each of chrs 3 (under WD conditions) and 8 (under WW conditions). Previous studies have also reported QTLs for cob height on all 10 chromosomes (Beavis et al., 1994; Lubberstedt et al., 1997; Li et al., 2016), with 21 QTLs identified for both PH and CH in three common genomic regions in two biparental populations (Li et al., 2016). However, our study identified major and minor QTLs (*qCH1–2* and *qCH1–1*) at different positions on chr 1 compared with those identified in previous studies. QTLs for cob height have also been reported on chrs 3, 4, 5, 6, and 8 in previous studies, but at different bs (Yang et al., 2008; Zhao et al., 2018; Tadesse et al., 2020). Grain yield is a highly complex trait that is influenced by many genes, each exerting only a small effect (Hallauer and Miranda, 2010). Earlier studies have reported on a QTL atlas that includes the major genes for GY and its associated traits (Zhou et al., 2020). In our study, we identified QTLs for the CW and GY traits on chr 2, which were physically located at bps 241,494,911–241,497,812 and bps 14,676,667–14,679,369, respectively. These QTLs are different from the previously reported meta-QTLs (Zhou et al., 2020). We also identified another QTL for grain yield on chr 6, located at bps 179,555,644–179,566,050, which is also different from the meta-QTLs that have been previously reported on this chromosome. Additionally, we identified three minor QTLs for the TB trait: two on chr 2 (one under WW and the other under WD conditions) and one on chr 5 (under WW conditions). Similar studies have also identified QTLs for biomass production and leaf area. For example, Chen et al. (2011) identified seven QTLs for biomass production and leaf area in the marker interval bnlg1832–P2M8/j (bp 1.05) on chr 1. Rahman et al. (2011) identified a QTL for yield on chr 1 that was co-located with the QTLs for root traits, total biomass, and osmotic potential in a region of about 15 cM. These studies suggest that genetic control of biomass and yield is complex and involves multiple QTLs distributed across different chromosomes and genomic regions.

The missing proportion of phenotypic variance for a trait may be attributed to epistasis (Carlborg and Haley, 2004), a term that refers to non-allelic interaction that can modify the degree of phenotypic expression by suppressing or enhancing the expression of interacting genes (Mackay et al., 2014; Rahman et al., 2017). Although our study was limited by a smaller population size, it is possible that epistatic effects, in addition to a few major and minor loci, contributed to the variation observed in the morphophysiological and yield-related traits studied, such as A_{net}, g_s, TR, LT, ASI, CW, and TB. However, further validation using a larger mapping population is necessary to confirm these findings. To better understand the stability of QTLs across different environments, it would be beneficial to evaluate this population in multiple locations, across multiple years, and under different treatments, such as variable temperature and watering regimes (Boer et al., 2007). Although our study used a smaller population size, a joint analysis over multiple years was conducted to establish the stability of QTLs and to evaluate the interactions between QTLs and the environment. Our findings suggest that, with the exception of g_s with respect to the QTL located at the marker interval S8_152991912–S8_123369436 on chr 8, most of the traits were associated with non-significant effects of environment in terms of the expression of QTLs, which captured only a small proportion of the phenotypic variation explained by additive-by-environment effects.

4.2 *In silico* analysis of the genomic region spanning QTLs and identification of candidate genes for stress tolerance

In this study, analysis of the genomic region of the QTLs using *in silico* methods led to the identification of candidate genes

associated with stress tolerance. The QTL *qNDVI2–1* contained two important genes: Golden 2-like (GLK) transcription factor 27 (*Zm00001d007962*, *GRMZM2G173882*, and *Zm00001eb118010*) and cold-regulated 413 plasma membrane protein 2 (*Zm00001d007968*). The Golden 2-like (GLK) transcription factor 27 gene plays a crucial role in regulating chloroplast growth and development, and also contributes to the maintenance of stay-green traits (Chen et al., 2016; Qin et al., 2021), whereas the cold-regulated 413 plasma membrane protein 2 gene enhances osmotic stress tolerance through enhanced expression of *AtCor78*/*AtRD29A* (Zhou et al., 2018). In the QTL interval of *qNDVI2–2*, three genes were identified: barren stalk 2 (*Zm00001eb084940*), AP2-EREBP transcription factor 131 (*Zm00001eb084810*, *GRMZM2G087059*, and *Zm00001d003884*), and potassium high-affinity transporter 1 (*Zm00001eb084630*, *GRMZM2G093826*, and *Zm00001d00386*). These genes contribute to stress tolerance through their involvement in signaling (Yao et al., 2019), plant growth and development (Xie et al., 2019), and facilitation of potassium (K^+) ion distribution in shoots (Qin et al., 2019). Overall, these genes in the QTL region for the NDVI trait play roles in maintaining stress tolerance through signaling and transporter activity.

In our study, we also found that the QTL interval of $qA_{net}3–2$ contained several genes that encode various transcription factors, such as ABI3/VP1 transcription factor 31 (*Zm00001eb144270*, *Zm00001d042460*), AP2-like ethylene-responsive transcription factor (*GRMZM2G141638*, *ZM00001EB144510*), ARR-B-transcription factor 6 (*Zm00001eb144290*, *Zm00001d042463*), bHLH transcription factor 132 (*Zm00001d042482 GRMZM2G114873*), AP2/EREBP transcription factor 53 (*Zm00001d042492*), and GRAS transcription factor 7 (*GRMZM2G013016*). These transcription factors play important roles in the regulation of downstream genes through binding to DNA elements in the promoter regions of the target genes. These genes also play vital roles in plant growth, development, hormone signaling, and stress responses (Kimotho et al., 2019). Additionally, we identified transporters, such as the magnesium (*GRMZM2G139822*, *ZM00001EB144080*) and proline transporters 1 and 2 (*ZM00001EB144010*, *GRMZM2G078024*), that are essential for maintenance of membrane homeostasis. Magnesium is a vital component of chlorophyll, which plays a crucial role in absorbance of sunlight during photosynthesis. It also acts as a phosphorus carrier in plants and is essential for phosphate metabolism (Hermans et al., 2013). Overall, our findings suggest that the *qAnet3–2* QTL region plays a vital role in the regulation of transcription factors and transporters, which are crucial for various physiological processes in plants.

The QTL $qg_s3–1$ contains three genes that are critical for plant stress tolerance and development. The bHLH transcription factor 70 (*GRMZM2G397755* and *Zm00001d039459*) is a key regulator of stress-related genes, enabling the plant to activate adaptive responses under various abiotic stresses. This transcription factor also plays a vital role in synthesis of flavonoids, which are essential for stress tolerance (Qian et al., 2021). The epidermal patterning factor-like protein (*GRMZM2G077219*, *Zm00001d039470*, and *Zm00001eb121050*) is involved in the development of stomatal cells in the upper epidermis of plant leaves (Lee et al., 2012). This protein is crucial for gaseous exchange in plants, which is vital for photosynthesis and respiration. The third gene, glutaredoxin 12 (*Grx12*) (*Zm00001eb121030*, *GRMZM2G303044*, *Zm00001d039468*, and *Zm00001eb121020*), is a stress-related redox sensor that plays a significant role in signaling through glutathione. *Grx12* enables glutathione to play a signaling role through glutathionylation of target proteins (Zaffagnini et al., 2012). The minor QTL $qg_s7–2$ includes the gene *Zm00001eb316940*, which encodes for the NAD (P)-binding Rossmann-fold superfamily protein. Although the exact function of this protein is not well understood, it is known to bind to NAD and NADP and is predicted to play a role in metabolic processes.

The QTL *qCH1–1* was expressed under both WW and WD conditions, and it is associated with the gene *Zm00001eb011970*, which encodes the LIM zinc-binding domain-containing protein DA1–2. This protein is related to ubiquitin binding and the expression of cell cycle genes, which contribute to long-distance phloem transport (Park et al., 2014). The LIM domain is a protein interaction domain that is involved in many cellular processes, including signal transduction, transcriptional regulation, and cytoskeletal organization. The QTL *qCH1–2* encompasses the gene encoding calcium-dependent protein kinase 36 (*CDPK 36*), which translates elevated calcium concentration into enhanced protein kinase activity and subsequent downstream signaling events (Singh et al., 2018). Calcium is an important secondary messenger in plants that plays a crucial role in a wide range of signaling pathways, including stress responses, development, and growth.

The QTL *qCH3–1* encompasses two important genes: receptor-like protein kinase (RLK) G-type lectin S-receptor-like serine/threonine-protein kinase (*ZM00001EB151330*), and homeobox transcription factors (*Zm00001d043231* and *ZM00001EB151130*). The first gene, RLK, is a member of the cell-surface receptor-like protein kinase family, which is critical in perception of signals. RLKs have active functions in various physiological processes such as plant growth and development, and in responses to both biotic and abiotic stresses (Ye et al., 2017). The G-type lectin S-receptor-like serine/threonine-protein kinase is involved in signaling during plant reproduction and defense against pathogens. Homeobox genes play a crucial role in specifying cell identity and positioning during embryonic development (Khan et al., 2018). They regulate various developmental processes, such as organogenesis, morphogenesis, and differentiation. The QTL *qCH3–1* includes the genes encoding RLK and homeobox-transcription factors, which play critical roles in plant growth, development, and responses to both biotic and abiotic stresses.

The QTLs *qTB2–1* and *qGY2–1* were found to be located within the same genomic region and contained several genes encoding for various transcription factors. ABI3/VP1 transcription factor 12 is involved in seed maturation and germination, and the AP2/EREBP transcription factor is known to regulate gene expression in response to various environmental cues. Calcium-dependent protein kinase 5 (*CDPK 5*) is a key component of the calcium signaling pathway and plays a crucial role in stress responses. The GRAS transcription factor is involved in various developmental processes, including root and shoot development, while MYB-

related transcription factor 20 is involved in the regulation of secondary metabolism. The RING zinc finger protein-like RING/U-box superfamily protein and zinc finger C3HC4-type family of proteins are involved in protein degradation and the regulation of gene expression, respectively (Kimotho et al., 2019). In addition to these transcription factors, the QTL region also contained various stress-inducible genes, such as abscisic acid receptor *PYL9*, which plays a critical role in drought and salt stress responses, and the dehydration-responsive element-binding protein 1D, which is involved in the regulation of water stress-responsive genes. The guard cell S-type anion channel SLAC1 is involved in stomatal regulation, while the WD40 repeat-like superfamily of proteins are involved in various developmental processes, including cell division and differentiation. The naked endosperm, sucrose synthase, and xyloglucan galactosyltransferase genes are all involved in seed development and are essential for proper seed maturation.

Overall, this study has identified several candidate genes that play crucial roles in various physiological processes, including the perception of external signaling, expression of functional proteins involved in stress tolerance, and the regulation of gene expression in response to environmental cues. These findings have important implications for crop improvement, as they provide valuable insights into the molecular mechanisms underlying stress tolerance and growth and development in plants.

5 Conclusion

The investigation conducted in this study revealed a substantial amount of variation in a range of morphophysiological and yield-related traits in the mapping population under both WW and WD conditions. Both major and minor QTLs were identified for these traits. Interestingly, one major QTL (*qCH1–2*) and one minor QTL (*qCH1–1*) for cob height were identified at different genomic positions than in earlier studies. Co-localized QTLs were also detected for traits g_s and GY on chr 6 (*qg$_s$6–2* and *qGY6–1*), and for traits g_s and TR on chr 7 (*qg$_s$7–1* and *qTR7–1*). The major candidate genes associated with QTLs were also detected under WD stress conditions and were found to be involved in growth and development, senescence, ABA signaling, signal transduction, and transporter activity processes contributing to WD tolerance. To facilitate marker-assisted selection in breeding, the QTL regions identified in this study could be fine-mapped and converted into SSR markers. In addition, the putative candidate genes could be isolated and functionally characterized, and the high-yielding and better-performing RILs could be used for genetic improvement of maize.

Data availability statement

The original contributions presented in the study are publicly available. The data can be found here: https://www.ncbi.nlm.nih.gov/sra/PRJNA913688.

Author contributions

BS carried out the parent selection, crossing, and development of the RILs. Statistical analysis, manuscript writing, editing, and critical review of the manuscript: YV. Field phenotyping of the parents, crossing program, RIL evaluation, QTL mapping, and drafting of the manuscript: MV. Physiological characterization of the population: NRK. Assistance with data analysis: MP. Infrastructure, facilities, and manuscript editing: SKY. Field experiments and review of the manuscript: MM. Manuscript editing and review: VKS. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by The Indian Council of Agricultural Research (ICAR) under the national flagship project entitled National Innovations in Climate Resilient Agriculture (NICRA).

Acknowledgments

This research was carried out under the National Innovations in Climate Resilient Agriculture (NICRA) project. The authors are thankful to the ICAR-Indian Institute of Maize Research, CIMMYT, Hyderabad Office, and to the ICAR - National Bureau of Plant Genetic Resources Regional Station, Hyderabad, for providing maize seed.

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.2023.1124619/full#supplementary-material

References

Almeida, G. D., Makumbi, D., Magorokosho, C., Nair, S., Borém, A., Ribaut, J. M., et al. (2013). QTL mapping in three tropical maize populations reveals a set of constitutive and adaptive genomic regions for drought tolerance. *Theor. Appl. Genet.* 126, 583–600. doi: 10.1007/s00122-012-2003-7

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. R. Stat. Soc. Ser. A.* 160, 268–282. doi: 10.1098/rspa.1937.0109

Beavis, W. D., Smith, O. S., Grant, D., and Fincher, R. (1994). Identification of quantitative trait loci using a small sample of top crossed and F4 progeny from maize. *Crop Sci.* 34, 882–896. doi: 10.2135/cropsci1994.0011183X003400040010x

Boer, M. P., Wright, D., Feng, L. Z., Podlich, D. W., Luo, L., Cooper, M., et al. (2007). A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. *Genetics.* 177, 1801–1813. doi: 10.1534/genetics.107.071068

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

Carlborg, Ö., and Haley, C. (2004). Epistasis: too often neglected in complex trait studies? *Nat. Rev. Genet.* 5, 618–625. doi: 10.1038/nrg1407

Cary, N. C. (2011). *The SAS system for windows. SAS version 9.3. procedure guide* (SAS Inc).

Chen, L., An, Y., Li, Y. X., Li, C., Shi, Y., Song, Y., et al. (2017). Candidate loci for yield-related traits in maize revealed by a combination of metaQTL analysis and regional association mapping. *Front. Pl Sci.* 8. doi: 10.3389/fpls.2017.02190

Chen, J., Cai, Y., Xu, L., Wang, J., Zhang, W., Wang, G., et al. (2011). Identification of QTLs for biomass production in maize (*Zea mays* l.) under different phosphorus levels at two sites. *Front. Agric. China.* 5, 152–161. doi: 10.1007/s11703-011-1077-3

Chen, M., Ji, M., Wen, B., Liu, L., Li, S., Chen, X., et al. (2016). GOLDEN 2-LIKE transcription factors of plants. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01509

Chia, J. M., Song, C., Bradbury, P. J., Costich, D., de Leon, N., Doebley, J., et al. (2012). Maize HapMap2 identifies extant variation from a genome in flux. *Nat. Genet.* 44 (7), 803–807. doi: 10.1038/ng.2313

Collins, N. C., Tardieu, F., and Tuberosa, R. (2008). Quantitative trait loci and crop performance under abiotic stress: Where do we stand? *Plant Physiol.* 147, 469–486. doi: 10.1104/pp.108.118117

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

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Qui, S., et al. (2014). TASSEL-GBS: A high-capacity genotyping by sequencing analysis pipeline. *PloS One* 9, e90346. doi: 10.1371/journal.pone.0090346

Gore, M. A., Chia, J. M., Elshire, R. J., Sun, Q., Ersoz, E. S., Hurwitz, B. L., et al. (2009). A first-generation haplotype map of maize. *Sci. (New York N.Y.)* 326 (5956), 1115–1117. doi: 10.1126/science.1177837

Gui, S., Wei, W., Jiang, C., Luo, J., Chen, L., Wu, S., et al. (2022). A pan-zea genome map for enhancing maize improvement. *Genome Biol.* 23 (1), 1–22. doi: 10.1186/s13059-022-02742-7

Hallauer, A. R., and Miranda, J. B. (2010). *Quantitative genetics in maize breeding* (New York: Springer).

Hermans, C., Conn, S. J., Chen, J., Xiao, Q., and Verbruggen, N. (2013). An update on magnesium homeostasis mechanisms in plants. *Metallomics* 5, 1170–1183. doi: 10.1039/c3mt20223b

Hirsch, C. N., Foerster, J. M., Johnson, J. M., Sekhon, R. S., Muttoni, G., Vaillancourt, B., et al. (2014). Insights into the maize pan-genome and pan-transcriptome. *Plant Cell* 26 (1), 121–135. doi: 10.1105/tpc.113.119982

Hu, X. M., Wang, G., Du, X. M., Zhang, H. W., Xu, Z. X., Wang, J., et al. (2021). QTL analysis across multiple environments reveals promising chromosome regions associated with yield-related traits in maize under drought conditions. *Crop J.* 9, 759–766. doi: 10.1016/j.cj.2020.10.004

Hufford, M. B., Seetharam, A. S., Woodhouse, M. R., Chougule, K. M., Ou, S., Liu, J., et al. (2021). *De novo* assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Sci. (New York N.Y.)* 373 (6555), 655–662. doi: 10.1126/science.abg5289

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. doi: 10.1109/MCSE.2007.55

Jiang, L. G., Li, B., Liu, S. X., Wang, H. W., Li, C. P., Song, S. H., et al. (2019). Characterization of proteome variation during modern maize breeding. *Mol. Cell. proteomics: MCP* 18 (2), 263–276. doi: 10.1074/mcp.RA118.001021

Jiao, Y., Zhao, H., Ren, L., Song, W., Zeng, B., Guo, J., et al. (2012). Genome-wide genetic changes during modern breeding of maize. *Nat. Genet.* 44 (7), 812–815. doi: 10.1038/ng.2312

Jin, M., Liu, H., He, C., Fu, J., Xiao, Y., Wang, Y., et al. (2016). Maize pan-transcriptome provides novel insights into genome complexity and quantitative trait variation. *Sci. Rep.* 6, 18936. doi: 10.1038/srep18936

Khan, N., Hu, C. M., Khan, W. A., Wang, W., Ke, H., Huijie, D., et al. (2018). Genome-wide identification, classification, and expression pattern of homeobox gene family in *brassica rapa* under various stresses. *Sci. Rep.* 8, 16265. doi: 10.1038/s41598-018-34448-x

Kimotho, R. N., Baillo, E. H., and Zhang, Z. (2019). Transcription factors involved in abiotic stress responses in maize (*Zea mays* l.) and their roles in enhanced productivity in the post genomics era. *Peer J.* 7, e7211. doi: 10.7717/peerj.7211

Kolmogorov–Smirnov Test (2008). *The concise encyclopedia of statistics*. Ed. Y. Dodge (New York, NY: Springer), 283–287. doi: 10.1007/978-0-387-32833-1_214

Lebreton, C., Lazic-Jancic, V., Steed, A., Pekic, S., and Quarrie, S. A. (1995). Identification of QTL for drought responses in maize and their use in testing causal relationships between traits. *J. Exp. Bot.* 46, 853–865. doi: 10.1093/jxb/46.7.853

Lee, J. S., Kuroha, T., Hnilova, M., Khatayevich, D., Kanaoka, M. M., McAbee, J. M., et al. (2012). Direct interaction of ligand-receptor pairs specifying stomatal patterning. *Genes Dev.* 26, 126–136. doi: 10.1101/gad.179895.111

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, 8410. doi: 10.3390/ijms23158410

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics.* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, Y. X., Li, C., Bradbury, P. J., Liu, X., Lu, F., Romay, C., 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, C., Li, Y., Sun, B., Peng, B., Liu, C., Liu, Z., et al. (2013). Quantitative trait loci mapping for yield components and kernel-related traits in multiple connected RIL populations in maize. *Euphytica.* 193, 303–316. doi: 10.1007/s10681-013-0901-7

Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., et al. (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* 45 (1), 43–50. doi: 10.1038/ng.2484

Li, X., Zhou, Z., Ding, J., Wu, Y., Zhou, B., Wang, R., et al. (2016). Combined linkage and association mapping reveals QTL and candidate genes for plant and ear height in maize. *Front. Plant Sci.* 7, 833. doi: 10.3389/fpls.2016.00833

Liu, S., Li, C., Wang, H., Wang, S., Yang, S., Liu, X., et al. (2020). Mapping regulatory variants controlling gene expression in drought response and tolerance in maize. *Genome Biol.* 21 (1), 163. doi: 10.1186/s13059-020-02069-1

Liu, S., and Qin, F. (2021). Genetic dissection of maize drought tolerance for trait improvement. *Mol. Breed.* 41, 8. doi: 10.1007/s11032-020-01194-w

Lobell, D. B., Roberts, M. J., Schlenker, W., Braun, N., Little, B. B., Rejesus, R. M., et al. (2014). Greater sensitivity to drought accompanies maize yield increase in the U.S. Midwest. *Sci. (New York N.Y.)* 344 (6183), 516–519. doi: 10.1126/science.1251423

Lubberstedt, T., Melchinger, A. E., Schon, C. C., Utz, H. F., and Klein, D. (1997). QTL mapping in test crosses of European flint lines of maize: 1. comparison of different testers for forage yield traits. *Crop Sci.* 37, 921–931. doi: 10.2135/cropsci1997.0011183X003700030037x

Mackay, I. J., Bansept-Basler, P., Barber, T., Bentley, A. R., Cockram, J., Gosman, N., et al. (2014). An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: Creation, properties, and validation. *G3 (Bethesda)* 4, 1603–1610. doi: 10.1534/g3.114.012963

Maheswari, M., Vijayalakshmi, T., Varalaxmi, Y., Sarkar, B., Yadav, S. K., Singh, J., et al. (2016). Functional mechanisms of drought tolerance in maize through phenotyping and genotyping under well-watered and water stressed conditions. *Eur. J. Agron.* 79, 43–57. doi: 10.1016/j.eja.2016.05.008

Malosetti, M., Ribaut, J. M., Vargas, M., Crossa, J., and van Eeuwijk, F. A. (2008). A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* l.). *Euphytica.* 161, 241–257. doi: 10.1007/s10681-007-9594-0

McCouch, S. R. (2008). Gene nomenclature system for rice. *Rice* 1, 7284. doi: 10.1007/s12284-008-9004-9

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

Messmer, R., Fracheboud, Y., Banziger, M., Stamp, P., and Ribaut, J. M. (2011). Drought stress and tropical maize: QTL for leaf greenness, plant senescence, and root capacitance. *Field Crop Res.* 124, 93–103. doi: 10.1016/j.fcr.2011.06.010

Messmer, R., Fracheboud, Y., Banziger, M., Vargas, M., Stamp, P., and Ribaut, J. M. (2009). Drought stress and tropical maize: QTL-by-environment interactions and stability of QTL across environments for yield components and secondary traits. *Theor. Appl. Genet.* 119, 913–930. doi: 10.1007/s00122-009-1099-x

Nikolić, A., Ignjatović-Micić, D., Dodig, D., Andelković, V., and Lazić-Jančić, V. (2012). Identification of QTLs for yield and drought-related traits in maize: assessment of their causal relationships. *Biotechnol. Equip.* 26, 2952–2960. doi: 10.5504/bbeq.2012.0016

Ouellette, L. A., Reid, R. W., Blanchard, S. G., and Brouwer, C. R. (2018). LinkageMapView-rendering high-resolution linkage and QTL maps. *Bioinf. (Oxford England)* 34 (2), 306–307. doi: 10.1093/bioinformatics/btx576

Park, J. I., Ahmed, N. U., Jung, H. J., Arasan, S. K., Chung, M. Y., Cho, Y. G., et al. (2014). Identification and characterization of LIM gene family in *Brassica tianrapa*. *BMC Genomics* 15, 641. doi: 10.1186/1471-2164-15-641

Pelleschi, S., Leonardi, A., Rocher, J., Cornic, G., de Vienne, D., Thévenot, C., et al. (2006). Analysis of the relationships between growth, photosynthesis and carbohydrate metabolism using quantitative trait loci (QTLs) in young maize plants subjected to water deprivation. *Mol. Breed.* 17 (1), 21–39. doi: 10.1007/s11032-005-1031-2

Qian, Y., Zhang, T., Yu, Y., Gou, L., Yang, J., Xu, J., et al. (2021). Regulatory mechanisms of bHLH transcription factors in plant adaptive responses to various abiotic stresses. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.677611

Qin, Y., Wu, W., and Wang, Y. (2019). *ZmHAK5* and *ZmHAK1* function in k^+ uptake and distribution in maize under low k^+ conditions. *J. Intergr Plant Biol.* 61, 691–705. doi: 10.1111/jipb.12756

Qin, M., Zhang, B., Gu, G., Yuan, J., Yang, X., Yang, J., et al. (2021). Genome-wide analysis of the G2-like transcription factor genes and their expression in different senescence stages of tobacco (*Nicotiana tabacum* l.). *Front. Genet.* 12. doi: 10.3389/fgene.2021.626352

Quarrie, S. A., Steed, A., Lebreton, C., Gulli, M., Calestani, C., and Marmiroli, N. (1994). QTL analysis of ABA production in wheat and maize and associated physiological traits. *Fiziol Rast.* 41, 565–571.

Rahman, M. A., Bimpong, I. K., Bizimana, J. B., Pascual, E. D., Arceta, M., Mallikarjuna Swamy, B. P., et al. (2017). Mapping QTLs using a novel source of salinity tolerance from hasawi and their interaction with environments in rice. *Rice.* 10, 47–63. doi: 10.1186/s12284-017-0186-x

Rahman, H., Pekic, S., Lazic-Jancic, V., Quarrie, S. A., Shah, S. M., Pervez, A., et al. (2011). Molecular mapping of quantitative trait loci for drought tolerance in maize plants. *Genet. Mol. Res.* 10 (2), 889–901. doi: 10.4238/vol10-2gmr1139

Ribaut, J. M., Betran, J., Monneveux, P., and Setter, T. (2009). "Drought tolerance in maize," in *Handbook of maize: Its biology*. Eds. J. L. Bennetzen and S. C. Hake (New York, NY: Springer). doi: 10.1007/978-0-387-79418-1_16

Sanguineti, R. T., Landi, P., Salvi, S., Maccaferri, M., and Casarini, E. S. (1999). QTL analysis of drought-related traits and grain yield in relation to genetic variation for leaf abscisic acid concentration in field-grown maize. *J. Exp. Bot.* 50, 1289–1297. doi: 10.1093/jxb/50.337.1289

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: Complexity, diversity, and dynamics. *Sci. (New York N.Y.)* 326 (5956), 1112–1115. doi: 10.1126/science.1178534

Semagn, K., Beyene, Y., Warburton, M. L., Tarekegne, A., Mugo, S., Barbara Meisel, B., et al. (2013). Meta-analyses of QTL for grain yield and anthesis silking interval in 18 maize populations evaluated under water-stressed and well-watered environments. *BMC Genom.* 14, 313. doi: 10.1186/1471-2164-14-313

Shinozaki, K., and Yamaguchi-Shinozaki, K. (2007). Gene networks involved in drought stress response and tolerance. *J. Exp. Bot.* 58, 221-227. doi: 10.1093/jxb/erl164

Simko, I., McMurry, S., Yang, H. M., Manschot, A., Davies, P. J., and Ewing, E. E. (1997). Evidence from polygene mapping for a causal relationship between potato tuber dormancy and abscisic acid content. *Plant Physiol.* 115, 1453–1459. doi: 10.1104/pp.115.4.1453

Singh, A., Sagar, S., and Biswas, D. K. (2018). Calcium dependent protein kinase, a versatile player in plant stress management and development. *Crit. Rev. Plant Sci.* 36, 336–352. doi: 10.1080/07352689.2018.1428438

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

Tester, M., and Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science.* 327, 818–822. doi: 10.1126/science.1183700

Thirunavukkarasu, N., Hossain, F., Arora, K., Sharma, R., Shiriga, K., Mittal, S., et al. (2014). Functional mechanisms of drought tolerance in subtropical maize (*Zea mays* l.) identified using genome-wide association mapping. *BMC Genom.* 15, 1182. doi: 10.1186/1471-2164-15-1182

Trachsel, S., Sun, D., SanVicente, F. M., Zheng, H., Atlin, G. N., Suarez, E. A., et al. (2016). Identification of QTL for early vigor and stay-green conferring tolerance to drought in two connected advanced backcross populations in tropical maize (*Zea mays* l.). *PloS One* 11 (3), e0149636. doi: 10.1371/journal.pone.0149636

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, R., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* 17 (3), 261–272. doi: 10.1038/s41592-019-0686-2

Wallace, J. G., Larsson, S. J., and Buckler, E. S. (2014). Entering the second century of maize quantitative genetics. *Heredity* 112 (1), 30–38. doi: 10.1038/hdy.2013.6

Walley, J. W., Sartor, R. C., Shen, Z., Schmitz, R. J., Wu, K. J., Urich, M. A., et al. (2016). Integration of omic networks in a developmental atlas of maize. *Sci. (New York N.Y.)* 353 (6301), 814–818. doi: 10.1126/science.aag1125

Wang, B., Lin, Z., Li, X., Zhao, Y., Zhao, B., Wu, G., et al. (2020). Genome-wide selection and genetic improvement during modern maize breeding. *Nat. Genet.* 52 (6), 565–571. doi: 10.1038/s41588-020-0616-3

Wen, W., Jin, M., Li, K., Liu, H., Xiao, Y., Zhao, M., et al. (2018). An integrated multi-layered analysis of the metabolic networks of different tissues uncovers key genetic components of primary metabolism in maize. *Plant journal: Cell Mol. Biol.* 93 (6), 1116–1128. doi: 10.1111/tpj.13835

Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., et al. (2014). Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat. Commun.* 5, 3438. doi: 10.1038/ncomms4438

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

Xue, Y., Warburton, M. L., Sawkins, M., Zhang, X., Setter, T., Xu, Y., et al. (2013). Genome-wide association analysis for nine agronomic traits in maize under well-watered and water-stressed conditions. *Theor. Appl. Genet.* 126, 2587–2596. doi: 10.1007/s00122-013-2158-x

Yang, X. J., Lu, M., Zhang, S. H., Zhou, F., Qu, Y. Y., Xie, C., et al. (2008). QTL mapping of plant height and ear position in maize (Zea mays L.). *Heriditas* 30, 1477–1486. doi: 10.3724/SP.J.1005.2008.01477

Yang, X. J., Lu, M., Zhang, S. H., Zhou, F., Qu, Y. Y., Xie, C., et al. (2018). Genome-wide identification, classification, and expression pattern of homeobox gene family in *brassica rapa* under various stresses. *Sci. Rep.* 8, 16265. doi: 10.1038/s41598-018-34448-x

Yang, C., Zhang, L., Jia, A., and Rong, T. (2016). Identification of QTL for maize grain yield and kernel-related traits. *J. Genet.* 95, 239–247. doi: 10.1007/s12041-016-0628-zb

Yao, H., Skirpan, A., Wardell, B., Matthes, M. S., Best, N. B., McCubbin, T., et al. (2019). The barren stalk2 gene is required for axillary meristem development in maize. *Mol. Plant* 12, 374–389. doi: 10.1016/j.molp.2018.12.024

Ye, Y., Ding, Y., Jiang, Q., Wang, F., Sun, J., and Zhu, C. C. (2017). The role of receptor-like protein kinases (RLKs) in abiotic stress response in plants. *Plant Cell Rep.* 36, 235–242. doi: 10.1007/s00299-016-2084-x

Yu, T. T., Liu, C. X., Mei, X. P., Wang, J. G., Wang, G. Q., and Cai, Y. L. (2015). Correlation and QTL analyses for photosynthetic traits in maize. *J. Southwest Uni.* 37, 1–10.

Zaffagnini, M., Bedhomme, M., Marchand, C. H., Couturier, J. R., Gao, X. H., Rouhier, N., et al. (2012). Glutaredoxin s12: Unique properties for redox signaling. *Antioxid Redox Signal.* 16, 17–32. doi: 10.1089/ars.2011.3933

Zhang, X., Pérez-Rodríguez, P., Semagn, K., Beyene, Y., Babu, R., Lopez-Cruz, M. A., et al. (2015). Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. *Heredity.* 114, 291–299. doi: 10.1038/hdy.2014.99

Zhao, X., Fang, P., Zhang, J., and Peng, Y. (2018). QTL mapping for six ear leaf architecture traits under water-stressed and well-watered conditions in maize (Zea mays l.). *Plant Breed.* 137 (1), 60–72. doi: 10.1111/pbr.12559

Zhao, X. Q., and Zhong, Y. (2021). Genetic dissection of the photosynthetic parameters of maize (*Zea mays* l.) in drought-stressed and well-watered environments. *Russ J. Plant Physiol.* 68, 1125–1134. doi: 10.1134/S1021443721060236

Zhou, Z., Li, G., Tan, S., Li, D., Weiß, T. M., Wang, X., et al. (2020). A QTL atlas for grain yield and its component traits in maize (*Zea mays*). *Plant Breed.* 139, 562–574. doi: 10.1111/pbr.12809

Zhou, A., Liu, E., Li, H., Li, Y., Feng, S., Gong, S., et al. (2018). PsCor413pm2, a plasma membrane-localized, cold-regulated protein from phlox subulata, confers low temperature tolerance in arabidopsis. *Int. J. Mol. Sci.* 19, 2579. doi: 10.3390/ijms19092579

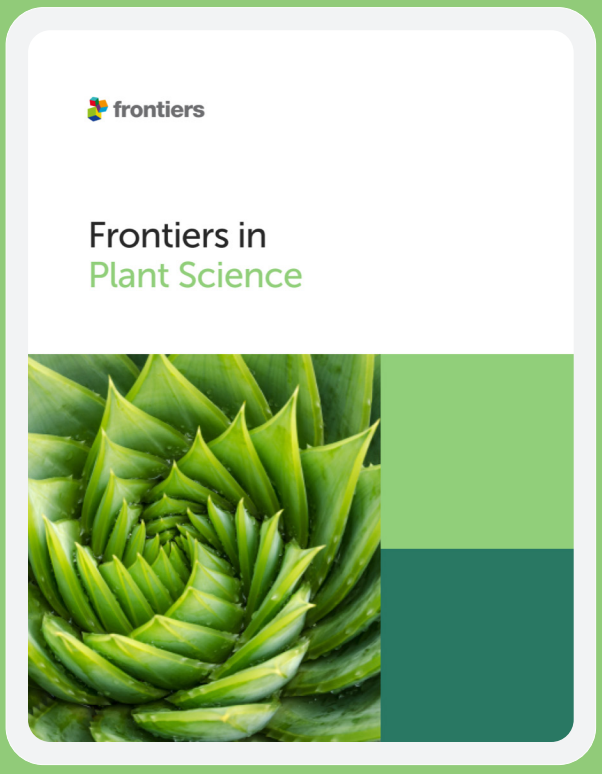
Frontiers in
Plant Science

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