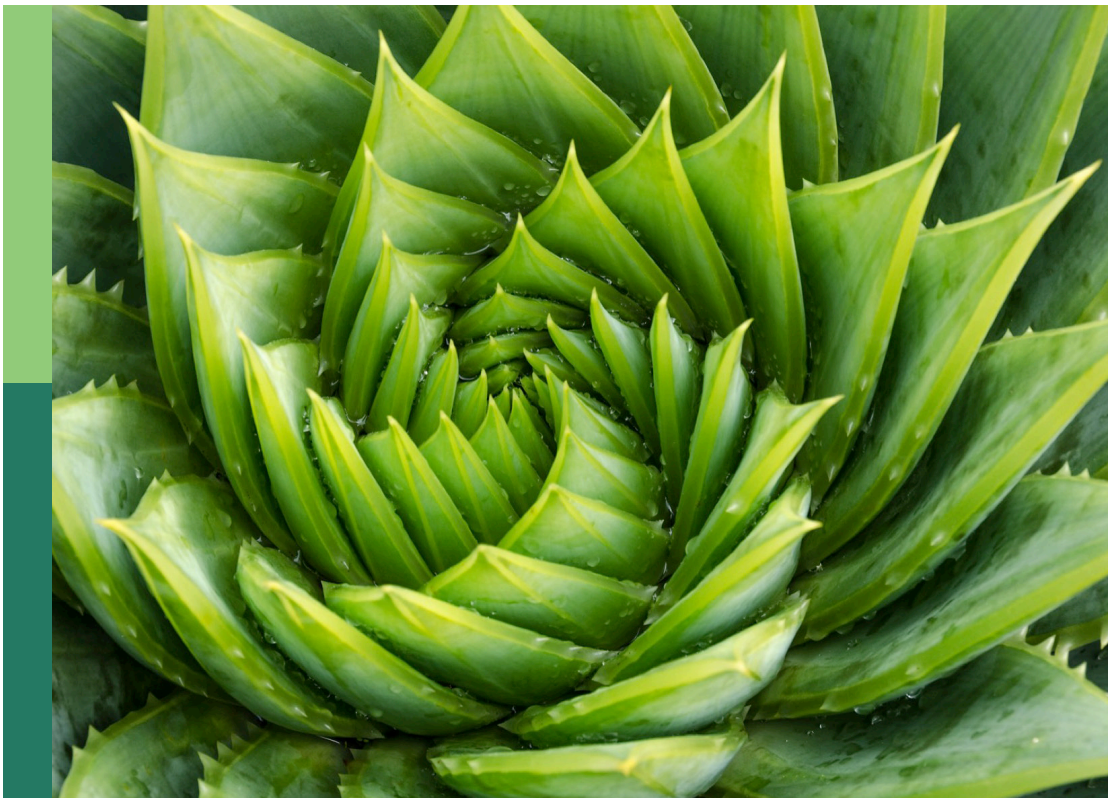
frontiers | Research Topics

Dissecting complex gene families to understand their roles in climate-resilience

Edited by
Mehanathan Muthamilarasan, Venkateswara Rao and Bhanu Prakash Petla

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-3570-7
DOI 10.3389/978-2-8325-3570-7

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

Dissecting complex gene families to understand their roles in climate-resilience

Topic editors

Mehanathan Muthamilarasan — University of Hyderabad, India
Venkateswara Rao — University of Hyderabad, India
Bhanu Prakash Petla — International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

Citation

Muthamilarasan, M., Rao, V., Petla, B. P., eds. (2023). *Dissecting complex gene families to understand their roles in climate-resilience*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3570-7

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 12 September 2023
DOI 10.3389/fpls.2023.1284812

OPEN ACCESS

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

*CORRESPONDENCE
Venkateswara Rao
venkateswararao@uohyd.ac.in
Bhanu Prakash Petla
p.bhanuprakash@cgiar.org
Mehanathan Muthamilarasan
muthu@uohyd.ac.in

RECEIVED 29 August 2023
ACCEPTED 06 September 2023
PUBLISHED 12 September 2023

CITATION
Rao V, Petla BP and Muthamilarasan M (2023) Editorial: Dissecting complex gene families to understand their roles in climate-resilience. *Front. Plant Sci.* 14:1284812. doi: 10.3389/fpls.2023.1284812

Editorial: Dissecting complex gene families to understand their roles in climate-resilience

Venkateswara Rao[1]*, Bhanu Prakash Petla[2]* and Mehanathan Muthamilarasan[1]*

[1]Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad, Telangana, India, [2]International Crops Research Institute for the Semi-arid Tropics, Hyderabad, Telangana, India

KEYWORDS

genome wide analysis, climate resilience, gene family analysis, stress response, functional characterization

Editorial on the Research Topic
Dissecting complex gene families to understand their roles in climate-resilience

Climate resilience is an important trait that the global plant research community focuses on, for ensuring food and nutritional security. Understanding the genetic determinants underlying this complex trait is key for tweaking the candidate genes, alleles, or QTLs, to enhance climate resilience. Genomics plays a major role in decoding the plant genomes to identify the coding and non-coding elements present within the DNA. Single-copy genes are present scarcely among the coding elements, whereas the remaining genes are encoded as gene families. More than one member of a particular gene family allows the genes to perform a multitude of functions, and their expression is tightly regulated at different levels. Whole genome sequencing reveals the information of genes and their corresponding gene families in the given genome. Characterization of these gene families provides insights into their structural and functional diversity. Thus, recent studies focus on dissecting the structural and functional aspects of gene families, which further identify the candidate genes for downstream characterization. Given the importance of genome-wide analyses in delineating the functional roles of different gene families in imparting climate resilience, the Research Topic invited articles on this aspect of gene family analyses.

Interestingly, the Research Topic received ten original research articles. Suri et al. studied the peptidyl-prolyl cis-trans isomerases (PPIases) in wheat (*Triticum aestivum*) to delineate their role in heat stress response. The authors mainly focused on FKBP (FK506-binding protein), Par (parvulin) and PTPA (phospho-tyrosyl phosphatase activators) gene families that belong to the PPIase superfamily. Through computational approaches, 71 FKBP, 12 Par, and 3 PTPA encoding genes were identified and characterized. Expression analysis of TaFKBP, TaPar and TaPTPA genes in response to heat stress at 50°C showed the putative role of candidate genes in conferring thermotolerance. As wheat is highly sensitive to temperature, Suri et al.have provided leads to understand the PPIase-mediated thermotolerance in this crop. Similarly, Chandra et al. identified two novel transcription factors in wheat, namely, TaZHD1 and TaZHD10, using a comparative genomic approach.

These genes belong to the zinc finger homeodomain class of transcription factors and play a major role in leaf rolling during moisture stress conditions. During drought stress, a significant upregulation of TaZHD10 was observed compared to TaZHD1, indicating their involvement in stress-induced leaf rolling. Further, the authors have performed structural and functional characterization of these two genes using *in silico* tools to gain further insights. Altogether, the study provided novel insights into the role of TaZHD1 and TaZHD10 in promoting leaf rolling during moisture stress in wheat, which can further be manipulated for enhanced tolerance to this stress.

Protein kinases play a prominent role in regulating plants' growth, development, and stress response. Given their importance, studies on identifying and characterizing kinases using whole genome sequencing data are rising. In this context, Santos et al. have identified the kinase-encoding genes of rubber (*Hevea brasiliensis*) to delineate their role in abiotic stress responses. The study identified 2842 protein kinase-encoding genes, which were classified into 20 different groups and 122 families. These genes and their corresponding proteins were characterized using different *in silico* tools to study their chromosomal localization, physicochemical properties, subcellular localization, duplication and divergence patterns, and gene ontology annotation. Further, *in silico* expression profiling was performed using the publicly available RNA-seq data and a co-expression network was constructed. Thus, the study has identified candidate protein kinases that can further be functionally characterized to study their involvement in abiotic stress response. Similarly, Feng et al. studied a class of protein kinases named sucrose non-fermenting-1 (SNF1)-related kinases (SnRKs) in maize (*Zea mays*) for their response to abiotic stresses. A total of 60 SnRK-encoding genes were identified by the authors, which were further studied for their evolutionary relationships, chromosome locations, gene structures, conserved motifs, and *cis*-elements in promoter regions. Also, homology modelling of SnRK proteins was performed, followed by *in silico* expression profiling using the data retrieved from the MaizeGDB database. To validate the findings, the authors treated a maize inbred line (B73) with dehydration and salt stress and the expression pattern of candidate SnRK genes was analyzed. The study, overall, provided insights into the structure and organization of the SnRK gene family in the maize genome and pinpointed candidate genes that can further be studied to understand their role in abiotic stress response.

Metal transporters are reportedly involved in regulating abiotic stresses besides their role in controlling growth and development. Magnesium transporter proteins (MGTs) are on such transporter known for their involvement in response to adverse environmental conditions. In this direction, Tang et al. studied the MGT genes in wheat. Their study identified 24 genes, which were analyzed for phylogenetic relationships, gene and protein structures, chromosomal localization, and organization of *cis*-elements in the promoter region of each gene. In addition, the authors have performed *in silico* expression profiling using publicly available RNA-seq data, and to further validate the findings, qRT-PCR was performed for candidate genes in wheat subjected to different stresses and hormone treatment. Wheat plants subjected to magnesium deficiency, aluminium chloride stress, and abscisic acid treatment were used for qRT-PCR-based expression profiling. The expression pattern was studied in root and shoot tissues of plants subjected to different treatments, and the expression values were compared with that of control (untreated) plants. The authors have observed a differential expression of candidate genes during the stress and hormone treatments compared to control plants, and their study showed the putative involvement of *MGT1B* in maintaining magnesium homeostasis.

Bcl-2-associated athanogene (BAG) proteins regulate growth and development, senescence, and response to environmental stimuli. Given this, Gu et al. studied the structure and organization of BAG proteins and their encoding genes in tobacco (*Nicotiana tabacum*). The study identified 19 BAG proteins in the tobacco genome, which were further characterized for their three-dimensional structures and phylogenetic relationships. Additionally, the authors used qRT-PCR to identify the genes regulating leaf senescence. From this data, *BAG5c* was chosen for further characterization. To confirm the role of BAG5c in regulating senescence, *bag5c* knock-down lines were generated using virus-induced gene silencing approach. The authors have observed a downregulated expression of *NtCP1*, *NtSEN4*, and *NtSAG12* in BAG5c-silenced plants, indicating the positive role of this gene in regulating leaf senescence. Analyzing the subcellular localization revealed the presence of this protein in the cell wall and nucleus. Further, yeast two-hybrid assay indicated the interaction of NtBAG5c and HSP70 in yeast cells. Similarly, in poplar (*Populus trichocarpa*), Liao et al. studied the structure and organization of the laccase gene family. Laccases play roles in lignin biosynthesis, thus gaining importance in biofuel research. The study revealed the presence of 53 laccase-encoding genes in the poplar genome, which are further analyzed using different computational approaches. This includes evolutionary studies, physical mapping, phylogenetic classification, gene structure analysis and motif distribution, domain architecture analysis, and comparative mapping. Also, RNA-seq-based expression profiling was performed to identify the candidate genes, which were further analyzed using qRT-PCR in 84K poplar trees (*P. alba* × *P. glandulosa*). The results suggested the putative involvement of *LAC23* in the lignification process, and to confirm this, the authors generated *proLAC23::LAC23-eGFP* lines. Expression and localization of LAC23 and lignin content showed a strong correlation between each other. The study provided clues on the role of LAC23 in lignin synthesis, thus having a role in realizing the biofuel potential of this important tree species.

Two studies used high-throughput transcriptomics and proteomics tools to identify stress-specific genes and proteins in the marine alga (*Dunaliella salina*) and rice (*Oryza sativa*), respectively. Ramachandran et al. subjected *D. salina* to different salt concentrations to understand the physiological, biochemical, and molecular changes underlying hypersaline tolerance. Similarly, Kumar et al. subjected three cultivars of rice, *viz*., Moroberekan (heat sensitive), IR64 (moderately heat tolerant), and Nagina22 (heat tolerant), to heat stress (short- and long-term), followed by protein isolation from anthers and mass spectrometry (LC-MS/MS). The study identified a repertoire of differentially expressed

proteins in anthers during heat stress, which can further be used for imparting durable heat stress tolerance in rice using breeding and transgenic approaches. In another microalga, *Chlamydomonas reinhardtii*, Yadav et al. have used two cyclic electron transport mutants, namely PGRL1 (Proton Gradient Regulation) and PGR5, to understand their role in the assembly and regulation of photosynthetic complexes under high light stress. In addition to biophysical and biochemical analysis of the mutants, the authors performed total proteome analysis using the nLC-MS/MS approach. The data suggested the involvement of cyclic electron transport in maintaining photosynthesis and photoprotection during high light stress. Altogether, the study provides insights into the role of cyclic electron transport in maintaining thylakoid protein abundance and super-complex organization under high light conditions.

In conclusion, the Research Topic has received insightful research contributions showing the role of novel genes in conferring climate resilience. Interestingly, the articles ranged from microalgae to land plants, covering different aspects of dissecting the stress tolerance trait in these organisms. With the advent of next-generation and third-generation sequencing platforms, sequencing complex genomes is now possible, resulting in the release of genome information in public databases. The availability of genome sequencing information enables the researchers to dissect candidate gene families to understand their role in conferring climate resilience. However, besides relying on fundamental bioinformatic analyses and expression profiling, the researchers should focus on the functional characterization of candidate genes to delineate their precise roles during stress incidence and endurance. Also, focus shall be given to studying the involvement of candidate genes in conferring tolerance to multiple (combined or consequential) stresses that impact crop yield and productivity in field conditions. Such information can further be translated to enhance tolerance to individual and multiple stresses in crop plants using advanced approaches such as genomics-assisted breeding, transgene technology, and genome editing.

Author contributions

VR: Writing – original draft. BP: Writing – original draft. MM: Writing – original draft, Writing – review & editing.

Acknowledgments

We thank all the authors and reviewers who have participated in this Research Topic.

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

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 22 November 2022
DOI 10.3389/fpls.2022.1038881

OPEN ACCESS

EDITED BY
Mehanathan Muthamilarasan,
University of Hyderabad, India

REVIEWED BY
Xiujun Zhang,
Wuhan Botanical Garden, (CAS), China
Venura Herath,
University of Peradeniya, Sri Lanka

*CORRESPONDENCE
Vinod
vinod.genetics@gmail.com
Shailendra Kumar Jha
jhashail78@gmail.com

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

RECEIVED 07 September 2022
ACCEPTED 25 October 2022
PUBLISHED 22 November 2022

CITATION
Chandra AK, Jha SK, Agarwal P, Mallick N, Niranjana M and Vinod (2022) Leaf rolling in bread wheat (*Triticum aestivum* L.) is controlled by the upregulation of a pair of closely linked/duplicate zinc finger homeodomain class transcription factors during moisture stress conditions.
Front. Plant Sci. 13:1038881.
doi: 10.3389/fpls.2022.1038881

Leaf rolling in bread wheat (*Triticum aestivum* L.) is controlled by the upregulation of a pair of closely linked/ duplicate zinc finger homeodomain class transcription factors during moisture stress conditions

Ajay Kumar Chandra, Shailendra Kumar Jha*, Priyanka Agarwal, Niharika Mallick, M. Niranjana and Vinod*

Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi, India

Zinc finger-homeodomain (ZF-HDs) class IV transcriptional factors (TFs) is a plant-specific transcription factor and play a key role in stress responses, plant growth, development, and hormonal signaling. In this study, two new leaf rolling TFs genes, namely *TaZHD1* and *TaZHD10*, were identified in wheat using comparative genomic analysis of the target region that carried a major QTL for leaf rolling identified through multi-environment phenotyping and high throughput genotyping of a RIL population. Structural and functional annotation of the candidate *ZHD* genes with its closest rice orthologs reflects the species-specific evolution and, undoubtedly, validates the notions of remote-distance homology concept. Meanwhile, the morphological analysis resulted in contrasting difference for leaf rolling in extreme RILs between parental lines HD2012 and NI5439 at booting and heading stages. Transcriptome-wide expression profiling revealed that *TaZHD10* transcripts showed significantly higher expression levels than *TaZHD1* in all leaf tissues upon drought stress. The relative expression of these genes was further validated by qRT-PCR analysis, which also showed consistent results across the studied genotypes at the booting and anthesis stage. The contrasting modulation of these genes under drought conditions and the available evidenced for its epigenetic behavior that might involve the regulation of metabolic and gene regulatory networks. Prediction of miRNAs resulted in five Tae-miRs that could be associated with RNAi mediated control of *TaZHD1* and *TaZHD10* putatively involved in the metabolic pathway controlling rolled leaf phenotype. Gene interaction network analysis indicated that *TaZHD1* and *TaZHD10* showed pleiotropic effects and might

also involve other functions in wheat in addition to leaf rolling. Overall, the results increase our understanding of *TaZHD* genes and provide valuable information as robust candidate genes for future functional genomics research aiming for the breeding of wheat varieties tolerant to leaf rolling.

KEYWORDS

leaf rolling, wheat, candidate genes, ZHD TFs, upregulation, moisture stress

Introduction

In recent decades, the improvement of yield and yield-related traits has remained one of the key objectives for the breeding of important crops. The enhancement of photosynthesis is an imperative approach to effusively exploit efficient crop yields (Gao et al., 2019). The leaf blade is the primary organ of photosynthesis and home of dry matter accumulation in crops. Main leaf functions, such as respiration, transpiration, and photosynthesis, rely on three-dimensional architecture of leaf (Zhang et al., 2009). Leaf rolling (LR) can transform the crops' photosynthetic efficiency and light condition (Gao et al., 2019). For example, low to moderate leaf rolling in *Oryza sativa* (rice) and *Zea mays* (maize) can optimize light transmission and higher canopy photosynthetic efficiency, which can significantly improve drought stress responses by radiant heat absorption and reduction in water loss *via* transpiration, thereby hoarding grain yield (Zhang et al., 2009). In contrast, severe leaf rolling could hinder the leaf's function and plant development and significantly reduce the yield enactment of the crops.

Zinc finger-homeodomain (ZHD) proteins, a benevolent of plant-specific transcription factors, influence various important biological processes in plants (Khatun et al., 2017; Abdullah et al., 2018; Liu et al., 2021). ZHD proteins contain a C_2H_2-type zinc finger (ZF) and homeodomain (HD) domains, which can regulate the expression of target regulatory genes by specific binding to *Cis*-acting regulatory elements (Wang et al., 2016). The first ZHD protein/genes clusters were identified in Flaveria, a C_4 plant encoding *PEPCase* genes (Windhovel et al., 2001). Subsequently, several plant-specific *ZHD* family genes were identified in various plant species such as *Glycine max* (Deng et al., 2002), *Arabidopsis thaliana* (Hu et al., 2008), *Triticum aestivum* (Bhattacharjee and Jain, 2013), *Oryza sativa* (Xu et al., 2014; Yoon et al., 2020), *Brassica rapa* (Wang et al., 2016), *Solanum lycopersicum* (Khatun et al., 2017), *Gossypium hirsutum* (Abdullah et al., 2018), and *Fagopyrum tataricum* (Liu et al., 2019). Many *ZHD* family genes are involved in vascular development, light signaling, phytohormonal signaling, and biogenesis of plant organs. At the same time, some *ZHD* genes are reported to have a regulatory role in the floral organ development and abiotic stress (salt and dehydration) responses (Wang et al., 2016; Abdullah et al., 2018). For example, the expression level of *AtZHD1* is modulated by abscisic acid, salt, and dehydration stressors (Wang et al., 2016). In addition, NAC proteins can interact with HD proteins that can simultaneously overexpress *NAC* and *ZHD* as a potential regulator of drought tolerance in Arabidopsis (Hu et al., 2008). Further, Hu et al. (2008) also reported that overexpression of *MIF1* interferes with the normal functioning of ZHD proteins. Additionally, *GmZHD1* and *GmZHD2*, the two soybean (*G. max*) ZHD proteins overexpressed calmodulin isoform 4 (*GmCaM4*) encoding gene against the pathogen infestation (Wang et al., 2016). Studies suggesting *ZHD* gene may play significant roles while defining plant growth, development, stress responses, and leaf rolling in plants.

The epidermis-specific expression of *Outer cell layer 1-5* (*OCL1-5*) encodes a protein containing homeodomain-leucine zipper (HD-Zip) TFs modulate rolled leaf, suggesting an increase in the physiological activity of bulliform cells at adaxial surface of leaf blade (Gao et al., 2019). Moreover, the overexpression of *rice outermost cell-specific gene 5* (*OsRoc5*) (Zou et al., 2011) and *OsRoc8* (Fang et al., 2021), a class IV gene homologous to Arabidopsis *GLABRA2* causes adaxially rolled leaves *via* increased size and number of bulliform cells. In contrast, *OsZHD10* and its closest homolog *AtZHD10*, a homeodomain-leucine zipper Arabidopsis class I gene overexpressed in the adaxial region, induce leaf rolling by increased bulliform cells activity in rice (Kim et al., 2007). Similarly, the overexpression of *OsZHD1* and its closest homolog *OsZHD2*, a zinc finger homeodomain class

Abbreviations: ZHD, Zinc finger-homeodomain; TFs, Transcriptional factors; RILs, Recombinant inbred lines; OCL, Outer cell layer; Roc, Rice outermost cell-specific; RAP-DB, Rice Annotation Project Database; QTL, Quantitative trait loci; NCBI, National Center for Biotechnology Information; BLAST, Basic Local Alignment Search Tool; ICAR, Indian Council of Agricultural Research; FYM, Farmyard manure; qRT-PCR, Quantitative real-time polymerase chain reaction.

homeobox TFs induced leaf rolling by increased number and size of the bulliform cells at their abaxial surface (Xu et al., 2014; Zhu et al., 2017; Yoon et al., 2020). A similar role was often displayed by *Zea mays* zinc-finger homeodomain protein 10 (*ZmZHD10*) while defining bulliform cell's shrinkage during heat, salt, and drought tolerance (Abdullah et al., 2018). Recently, 37 zinc-finger homeodomain proteins (TaZHDs) have been identified in wheat. Gene expression analysis of six core groups of *TaZHD* genes suggested that these may be involved in the regulation of critical biological processes, including cold, salt, and drought stress tolerance as a component mediator of leaf rolling trait (Liu et al., 2021), as reported in other crops like maize and rice (Duan et al., 2018; Gao et al., 2019). The information suggested that these genes/transcription factors (TFs) might be putatively involved in adaxial or abaxial leaf rolling by modulation of bulliform cells. However, while the potential regulatory role of *ZHD* genes/TFs has been extensively studied in Arabidopsis, rice, maize, and other model crops, the molecular basis of *ZHD* genes in leaf rolling and its component trait (drought tolerance) in wheat have not yet been identified.

The allohexaploid nature of cultivated wheat is vital in making bread wheat a superior crop to its ancestors because allelic/loci interactions between sub-genomes might contribute to its higher flexibility for gene expression, which may trigger its wide adaptability under drought stress conditions (Guan et al., 2020). IWGSC's annotation of an exceedingly complex whole genome sequence of wheat has made it possible to identify the candidate genomic loci (genes) on a genome-wide scale. In this study, we report the identification and characterization of *ZHD* genes, a novel candidate leaf rolling gene(s) containing zinc finger homeodomain class homeobox TFs that might be involved in leaf rolling modulation under drought stress conditions. The information generated through the present study could bridge our knowledge gaps concerning ZHD proteins as a critical regulator of bulliform cells' activity and their possible molecular networks underlying leaf rolling in these important cereal crops.

Material and methods

Plant materials, drought treatment, and growth statistics

In this study, four contrasting recombinant inbred lines (RILs) having extreme contrast for leaf rolling index (LRI) viz., high LR (RIL D-13 and RIL D-65) and low LR (RIL D-7 and RIL D-9) along with their parents (NI5439; drought resilient and HD2012; drought susceptible) (Verma et al., 2020), was used for phenotypic analysis, histological study and gene expression profiling. Each RIL representing 25-30 plants was grown in a polyhouse of the Division of Genetics, ICAR-Indian Agriculture Research Institute, New Delhi (28.6377° N, 77.1571° E, and 228.61 m altitude over mean sea level) under controlled conditions during the natural growing season of 2021-22. The experiment was set up in plastic pots filled with a 3:1 ratio of soil and farmyard manure (FYM), with each test RILs replicated thrice in a completely randomized design. During the experimentation period, the plants were allowed to grow under proper conditions (relative humidity: 75-80%; temperature: 25 ± 5 °C; light and dark cycles: 13 + 11 hours; water up to 60-80% soil capacity) (Dong et al., 2019). Moreover, standard agricultural practices were followed to raise a healthy crop instead of any disease management.

Three independent biological replicates were used for each drought treatment and irrigated. A total of 12-15 plants representing each test RILs were subjected to drought treatment as the plants attained booting (stage 10) and heading (stage 10.5) (Lindsey et al., 2017). At each stage, drought treatment was given for ten days without water, while the irrigated (control) pots were allowed to water at regular intervals. For histological and gene expression study, flag leaves of test RILs were harvested as plants exhibiting rolled leaves. The histological study was done with fresh leaf, while for gene expression studies, harvested samples were rapidly frozen in liquid N_2 and stored at -80°C until use.

Histological observation of leaves for rolling

Considering the significance of bulliform cells in defining leaf rolling in plants (Yang et al., 2016), free-hand and semi-thin sections of fully expanded flag leaves were used for histological observation in the present study. Harvested leaves were thoroughly washed with running water for free-hand sections and were sectioned about 5 cm in size from their middle portion. Similarly, semi-thin sections of about 50-100 µm thick were done by using a scalpel blade from its middle portion. Furthermore, thin slices were stained with 0.2% Acetocarmine (Sigma) for 2 minutes at 37°C followed by observation and digital photographs were captured with a light microscope (Nikon Y-TV55, Japan) fitted with Nikon Eclipse H600L camera (Nikon, Japan). At least three independent biological replicates were used for each experiment (Yang et al., 2016).

Identification of candidate gene

In the previous study, we identified a stable QTL putatively associated with leaf rolling Qlr.nhv-5D.2 on the 5D chromosome, flanked with markers AX-94892575 and AX-95124447 at 338665336 and 410953022 intervals, respectively (Verma et al., 2020). Marker intervals were analyzed for the number and type of genes (transcripts), physical locations,

protein domains and families, and other gene (protein) specific features with modified parameters using the EnsemblPlants tool, BioMart (https://plants.ensembl.org/biomart/martview/) (Kinsella et al., 2011). The complete nucleotide and coding sequences of identified genes (transcripts) were retrieved from the IWGSC database, readily available at EnsemblPlants (https://plants.ensembl.org/Triticum_aestivum/Info/Index). Homology (BLASTn) search was performed against the recently released wheat genome assembly using BLAST (https://plants.ensembl.org/Triticum_aestivum/Tools/Blast) (Bolser et al., 2016). Furthermore, the complete nucleotide and coding sequences of candidate genes reportedly involved in leaf rolling in model cereals (*O. sativa* and *Z. mays*) were retrieved from Rice Annotation Project Database, RAP-DB (https://rapdb.dna.affrc.go.jp/) and Maize Genetics & Genomics database, MaizeGDB (https://www.maizegdb.org/), respectively (Sakai et al., 2013; Portwood et al., 2019). Finally, putative wheat genes (transcripts) identified within the QTL intervals were used as the query was cross-checked by NCBI blast (https://blast.ncbi.nlm.nih.gov/Blast.cgi) for data reliability based on higher query coverage, lower E-value, and highest % identity (Johnson et al., 2008).

Chromosomal distribution of candidate gene

Physical mapping exhibiting specified positions of the identified candidate genes corresponding to wheat chromosome was constructed based on the information generated through EnsemblPlants (https://plants.ensembl.org/Triticum_aestivum/Info/Index) and IWGSC-URGI (https://wheat-urgi.versailles.inra.fr/), wheat genome databases (Bolser et al., 2016; Alaux et al., 2018).

Structural and functional characterization of candidate gene

The DNA sequence of selected wheat genes (transcripts) was subjected to an ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/) for the prediction of coding DNA sequence (CDS) within the gene (Irene et al., 2002). The structural information such as 5′-UTR, exon, intron, and 3′-UTR of candidate genes and their orthologs was extracted using Gene Structure Display Server 2.0 (http://gsds.gao-lab.org/) (Hu et al., 2015). To discover the structural variation in the genes (proteins), conserved domains and motifs were analyzed using the Conserved Domain Architecture Retrieval Tool (https://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi) and MEME suite version 5.0.5 (http://meme-suite.org/tools/meme), respectively (Bailey et al., 2015; Khatun et al., 2017). We utilized MEME suite with slightly modified parameters, i.e., a) motif width: 6-250 amino acids, b) site distribution per sequence: zero or one, and c) maximum no. of motifs: 50 were selected for motif discovery on candidate ZHD proteins. Putative *Cis*-acting regulatory elements in the promoter regions of the identified genes were predicted using the PlantCARE program (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/), a database search engine (Xu F. et al., 2012; Baldoni et al., 2015). Furthermore, the physicochemical characteristics of identified leaf rolling wheat proteins were anticipated by an ExPASy database server, ProtParam (https://web.expasy.org/cgi-bin/protparam/protparam) (Chandra et al., 2020). Sub-cellular and functional localization of leaf rolling wheat proteins was annotated using an integrative web server, BUSCA (http://busca.biocomp.unibo.it/) (Verma et al., 2020).

Protein structure analysis

Homology modeling is a powerful tool for chimeric proteins' structural and functional assignment. To cross-check the significant structural similarity between the proteins and their orthologs, the predicted proteins were superimposed based on available homologous proteins in the SWISS-MODEL database server (https://swissmodel.expasy.org/interactive/4QhJDj/templates/) (Gasteiger et al., 2005). Since, in earlier study we had already reported homology modeling for 3D molecular structure and superposition of predicted TaZHD1 protein (Verma et al., 2020), so in the present investigation we only considered second ZHD protein i.e., TaZHD10 for further comparative protein structure analysis using the homology modeling approach (Supplementary Figure 1A). The 3D structure homology models at higher complexity levels of the predicted leaf rolling wheat protein TaZHD10 was modeled using SWISS-MODEL (https://swissmodel.expasy.org/interactive) (Gasteiger et al., 2005). Furthermore, to refine and validate the structure of the predicted models, the Ramachandran plot of each protein model was plotted by analyzing psi (Ψ) and phi (Φ) torsion angels using SWISS-MODEL structure assessment (https://swissmodel.expasy.org/assess/), against amino acid residues in the predicted proteins (Waterhouse et al., 2018).

Sequence alignment and comparative phylogenetic analysis

Protein sequences of predicted leaf rolling wheat proteins and their orthologs were aligned using T-Coffee (https://www.ebi.ac.uk/Tools/msa/tcoffee/) (Di Tommaso et al., 2011). To explore the evolutionary relationship and functional similarity among annotated proteins, phylogenetic tree topology was constructed using PhyML v3.1 (http://www.phylogeny.fr/simple_phylogeny.cgi), a freely available software package (Guindon et al., 2010). To demonstrate the phylogenetic tree,

the Maximum-Likelihood method coupled with 1000 iterations of bootstrapping values was used (Wei et al., 2020).

Prediction of miRNAs targeting candidate genes

Full-length CDS and genomic DNA sequences of identified wheat leaf rolling genes were subjected to a query against an updated version of the wheat miRNAs library, freely available on psRNATarget v2.0 (https://www.zhaolab.org/psRNATarget/analysis?function=2), a plant small RNA target analysis server (Dai et al., 2018). Following default parameters rooted in the psRNATarget algorithm were used likely no. of top targets: 200, expectation: 5, the penalty for G: U pair: 0.5, the penalty for mismatches: 1, seed region: 2-13 nt, extra weight in seed regions: 1.5, no. of mismatch allowed in seed regions: 2, HSP (complementarity scoring) size: 19 and translation inhibition range: 10-11 nt (Liebsch and Palatnik, 2020).

Differential expression of candidate ZHD gene

Transcriptome-wide expression profiling

For estimation of transcript abundance and expression profile of the identified leaf rolling genes, the freely available RNA-seq expression data files (RefSeq v2.1) of wheat (Chinese Spring) were used (Zhu et al., 2021). The relative expression of each gene under drought stress in various tissues at different developmental stages was presented as a heat map. The heat map was generated from the relative abundance of transcripts of each gene, calculated as FPKM (Fragments Per Kilobase of transcript per 10 million reads) using the wheat expression browser (http://www.wheat-expression.com), powered by expVIP (Borrill et al., 2016).

RNA isolation, cDNA preparation, and quantitative real-time PCR (qRT-PCR)

The total RNA was extracted from flag leaf tissues of test samples at two different developmental stages, viz., booting (stage 10) and heading (stage 10.5) stage, using the Trizol solution (Invitrogen). Total RNA was treated with RNase-free DNase I (Fermentas) to remove residual DNA contamination. The first-strand cDNA was generated from DNase-treated total RNA (2.5 µg/µl) with oligo (dT) primers using High-capacity cDNA reverse transcription kit (Applied Biosystems). qRT-PCR was performed using GoTaq qPCR master mix (Promega) on a Bio-Rad CFX96 Real-Time system (CA, USA). A set of gene-specific primers were designed using the NCBI Primer-BLAST tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) (Untergasser et al., 2012) and commercially synthesized by Eurofins Genomics, India (Supplementary Table 1). All test samples were analyzed in three independent biological replications and three technical replicates. The *TaActin1* (LOC123179078), a typical constitutively expressed wheat housekeeping gene was used as an internal control to normalize the expression of candidate genes. The relative expression levels of *ZHD* genes were analysed using irrigated NI5439 as a reference. The relative fold in expression of each gene concerning cycle threshold was calculated using following formulae (Schmittgen and Livak, 2008), represents the cycle numbers required for the detection of fluorescent signal to minimize the background noise.

For the relative expression of candidate *ZHD* gene in any sample,

a) ΔCt_{goi} = Mean (Ct_{goi} - Ct_{actin}), here, goi = Gene of interest

b) $\Delta\Delta Ct = \Delta Ct_{goi\ (soi)} - \Delta Ct_{goi\ (irrigated\ NI5439)}$ soi = Sample of interest

c) Exp (2^ - ΔΔCt) = POWER (2, - ΔΔCt) Ct- value = Cycle threshold

Gene interaction network analysis

Gene co-expression, co-localization, genetic interaction, as well as physical and shared protein domain analysis, are helpful tools for predicting the biological network integration of predicted genes (Yun et al., 2010; Perrella et al., 2018). In the case of wheat, there is no specific database server for protein-protein, gene-gene, and molecular interaction network study. As a result, biological network and interactome data from *A. thaliana* were used as a reference for real-time gene interaction network analysis of wheat leaf rolling genes on the GeneMANIA web server (https://genemania.org/) (Warde-Farley et al., 2010).

Epigenetic aspects of leaf rolling

Epigenetic changes script significant impact to plant genomes as transcriptional imprints to adapt and respond against severe stress conditions by reprogramming gene expression (Mohanty et al., 2012; Yong-Villalobos et al., 2015). To elucidate the epigenetic regulation, we analyzed the genome-wide DNA methylation patterns of leaf rolling *ZHD* genes under the differential stress condition (irrigated, drought, and 24 HAW) at the different genic regions. In the case of wheat, there is no specific epigenome platform available for methylome study. As a result, methylome dataset from *O. sativa* were used as a reference for integrated epigenetic regulation and comprehensive analysis of wheat *ZHD* genes using the Plant Methylome database web server (https://epigenome.genetics.uga.edu/PlantMethylome/), powered by JBrowse 1.12.5 (Buels et al., 2016).

Results

Histological observation for leaf rolling

Morphological observation of leaves for rolling resulted in variable architectural differences in drought-treated plants than in the irrigated (control). The gross morphology of wheat genotypes studied under irrigated and drought conditions at the late heading stage is presented in Figure 1. At booting (stage 10), all drought-treated plants exhibited semi-roll to completely rolled leaves with significant erectness, owing to reduced leaf lamina joint angle while irrigated plants showed drooping type and flat leaves with no rolling. Similarly, the pattern of the architecture of leaves was almost similar as observed during the booting stage, but the relative intensity of flatness and rolling was significantly higher at the early to late heading (stage 10.5) stage. The relative flatness and rolling of RILs with parents were observed as consistent with our earlier reports. The calculated leaf rolling index (LRI) substantially varied from higher rolling (RIL D-13 and RIL D-65) to low rolling (RIL D-7 and RIL D-9) RILs compared to parents (reported in our earlier studies and documented by Verma et al., 2020). Considering the key role of bulliform cells in defining leaf rolling in wheat, histological analysis was done in present study. We observed that the bulliform cells in irrigated RIL D-65 leaf blades are arranged in a "U" shape and were fully flaccid, while the bulliform cells of treated RIL D-65 are arranged in a "V" shape and were tabid (Figures 2A, B).

Identification of candidate gene

In the previous study, we had identified a stable QTL putatively associated with leaf rolling *Qlr.nhv-5D.2* on the 5D chromosome, flanked with markers *AX-94892575* and *AX-95124447* at 338665336 and 410953022 intervals, respectively (Verma et al., 2020). A total of 5388 transcripts were predicted using a genome-wide analysis approach in this marker interval. Out of 185 unique transcripts, only three viz., *TraesCS5D02G253100.1* and *TraesCS5D02G234600.1*, and *TraesCS5D02G234700.1* speculated to be involved in leaf rolling in *T. aestivum* were identified (Table 1). The homology search performed against the already available databases led to the identification of multiple hits. The most significant hits were further shortlisted based on various filters. Genes associated with leaf rolling (*OsZHD1* and *OsZHD10*) from rice showed significant homology with all three transcripts. The transcript *TraesCS5D02G253100.1* showed 96.9% homology and E-value (1.8e-34) with *OsZHD1* (*LOC4347315*) while *TraesCS5D02G234600.1* resulted in 100% homology and E-value (7.8e-149) with *OsZHD10* (*LOC4345673*). In contrast, *TraesCS5D02G234700.1* showed only 40.5% homology and E-value (9.2e-09) with *OsZHD10*, whereas no significant hits were found with *OsZHD1*. Because of this, we excluded our *TraesCS5D02G234700.1* transcript and were not considered for further investigation. Finally, two genes, viz., *TraesCS5D02G253100.1* (*TaZHD1*) and *TraesCS5D02G234600.1* (*TaZHD10*), were considered to be putatively involved in leaf rolling in bread wheat. All the identified genes were annotated according to the name designated to their orthologs. To specify the name of annotated sequences, the first initial of the studied genus (*Triticum*) and species name (*aestivum*) were used in the alias (Ta), followed by the short form (abbreviation) of the protein encoded with their homologs (Table 1).

Chromosomal distribution of candidate *ZHD* genes

The physical positions of identified candidate genes, *TaZHD1* and *TaZHD10*, to corresponding 5D chromosome is depicted in Figure 3. A comprehensive study of each gene using the BioMart database server revealed that *TraesCS5D02G253100.1* (*TaZHD1*) is localized on 5D:359474816-359475706 whereas *TraesCS5D02G234600.1* (*TaZHD10*) is confined on 5D:342174542-342176173, at the long arm of the corresponding chromosome. Further, the genetic distance between both genes was calculated using IWGSC-URGI wheat genome databases and was found up to 17.30 Mb, which was considered merely closer to each other.

Structural and functional annotation of identified ZHD genes

The Open Reading Frame (ORF) or coding DNA sequence (CDS) of each candidate *ZHD* gene was explored by subjecting genomic sequences to an ORF finder. The total number of anticipated ORFs per gene ranged from 5 to 10 ORFs, with an average of eight (8) ORFs. Comprehensive analysis of *TaZHD1* resulted in 9 ORFs, while *TaZHD10* resulted in 5 ORFs with variable lengths. Finally, nucleotide and translated protein sequences of the longest ORFs of each predicted gene were retrieved and used for further structural and functional analysis. The detailed information on best-predicted ORFs, including start, stop, and length statistics, are presented in Supplementary Table 2.

The structural organization of genes, including 5'-UTR, exon-introns pattern, and 3'- UTR contain essential information to understand gene evolution and was predicted by subjecting genomic sequences and CDS to Gene Structure Display Server (Figure 4A). The results showed that both *TaZHD1* and *TaZHD10* genes contain a single exon flanked

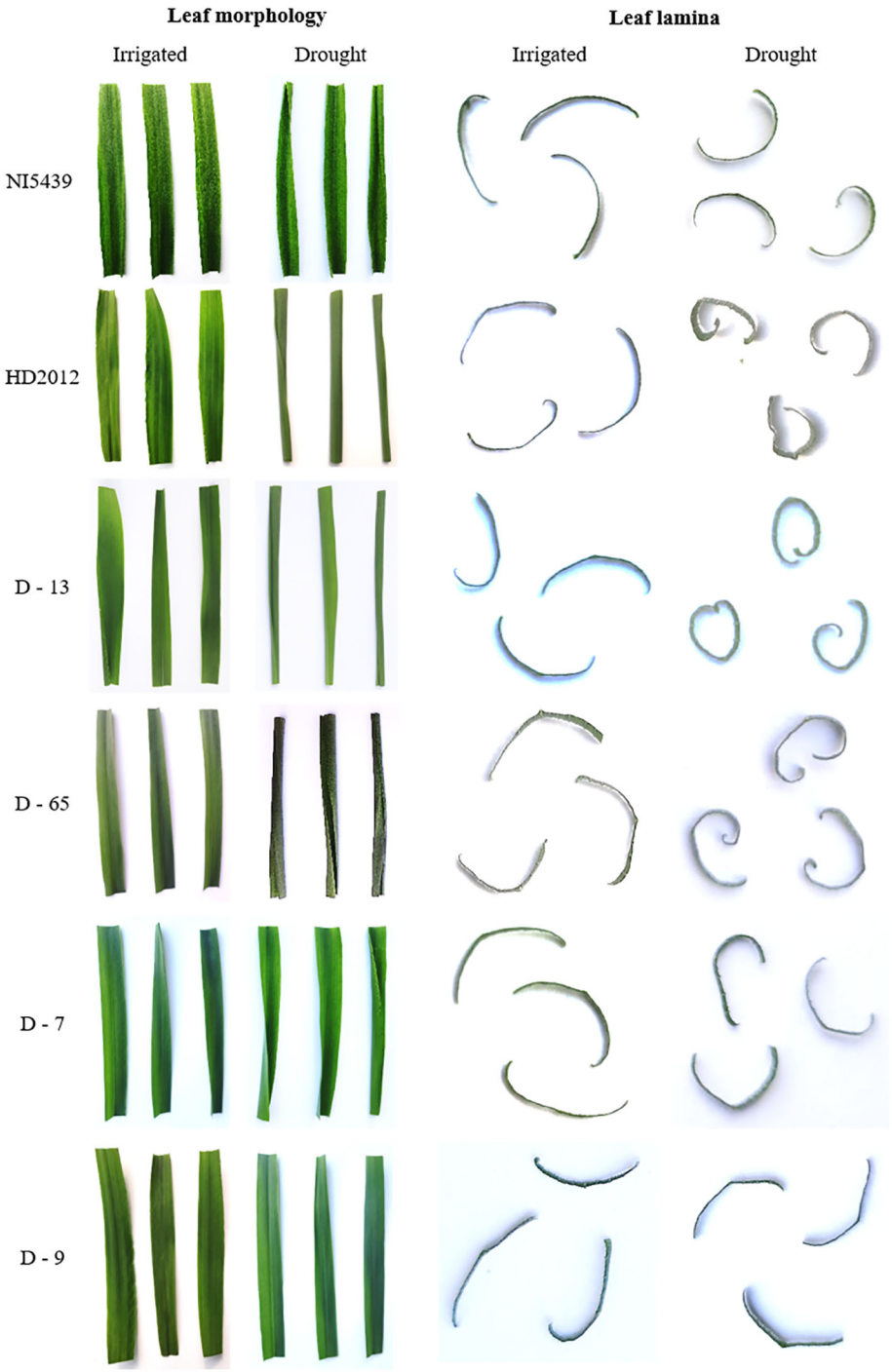


FIGURE 1
Morphological observations of flag leaf of wheat RILs with parents for rolling (From top to bottom: NI5439; drought resilient, HD2012; drought susceptible, high LR RILs = D-13 and D-65, and low LR RILs = D-7 and D-9).

by upstream and downstream sequences, as displayed in their respective homologs. The length of upstream and downstream sequences was almost equal, but exons' size was highly variable in both genes. The number and lengths of exons were relatively more conserved, suggesting that both genes govern potentially conserved biological functions.

During the comprehensive analysis of physicochemical properties of translated ZHD proteins, the chimeric proteins

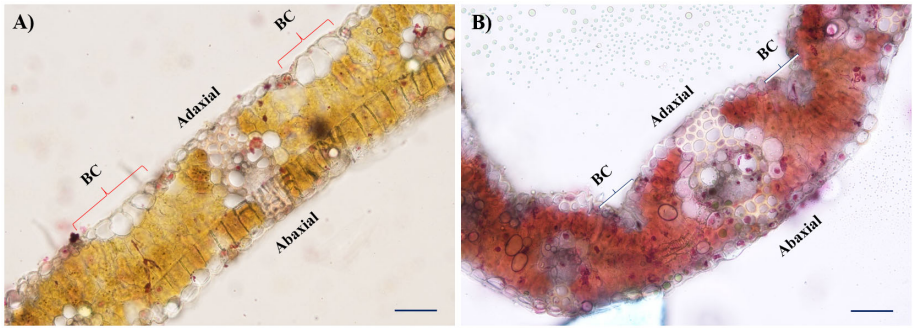


FIGURE 2
Histological observation of flag leaves contrasting in their propensity for bulliform cells (BC) activity in leaf rolling under moisture stress. **(A)** Cross-sections of RIL D-65 under irrigated condition, and **(B)** Cross-sections of drought treated RIL (D-65) leaves under 20X resolution. Scale bars: 50 μm.

resulted in a wide array of variability in terms of molecular weight, a total number of amino acids, grand average of hydropathicity (GRAVY), aliphatic index, instability index, and isoelectric point (pI). The characteristic properties revealed that the identified chimeric proteins are physiochemically identical to their respective homologs (Supplementary Table 2).

Proteins belonging to the ZF-HD IV family are potential plant-specific transcription factors such as *TaZHD1* and *TaZHD10*. To determine whether *TaZHD1* and *TaZHD10* are found in the nucleus, the subcellular location of these proteins was investigated. *TaZHD1* and *TaZHD10* were found solely in the nucleus, indicating that they are nuclear proteins, according to the results (Supplementary Table 2).

Wheat *ZHD* genes possess zinc finger homeodomain (ZF-HD) IV family transcription factors

Wheat genome annotation analysis using the *EnsemblPlants* database showed that *TaZHD1* and *TaZHD10* are located on the 5D chromosome and encode a 387 and 316 amino acid residue protein, respectively. Multilevel sequence analysis of candidate proteins revealed the presence of two functional domains, ZF-HD dimer superfamily and homeo_ZF_HD superfamily domain designated by cl04737 and cl22771 accession numbers, respectively (Figure 4B). Both superfamily domains belong to plant-specific transcription factors and are found to be ubiquitously present in *A. thaliana*, C_4, and other C_3 plants

TABLE 1 Identification of candidate genes putatively involved in leaf rolling in wheat and their orthologs.

S. No.	Gene Stable ID*	Gene annotation	Source	Protein type	Genomic Location	Query Length (bp)	E-value	Identity (%)
1.	*LOC4347315* (*LOC_Os09g29130*)	*OsZHD1*	*O. sativa japonica group*	Zinc-finger homeodomain protein 1-like	Chr 9: NC_029264.1 (17703982.17705654)	1673	0.0	87.1
2.	*LOC4345673* (*LOC_Os08g34010*)	*OsZHD10*	*O. sativa japonica group*	Zinc-finger homeodomain protein 10-like	Chr 8: NC_029263.1 (21307444:21309247)	1741	2.00e-80	89.0
3.	*LOC103653493*	*ZmZHD10*	*Z. mays*	Zinc-finger homeodomain protein 10	Chr 4: NC_050099.1 (88731765.88733255)	1491	9.00e-64	84.6
4.	*LOC119309747*	*TdZFH9a*	*T. dicoccoides*	Zinc-finger homeodomain protein 9-like	Chr 5B: NC_041389.1 (418045587.418047238)	1652	0.0	95.2
5.	*LOC119301727*	*TdZFH9b*	*T. dicoccoides*	Zinc-finger homeodomain protein 9-like	Chr 5A: NC_041388.1 (389777977.389779664)	1688	0.0	95.5
6.	*TraesCS5D02G253100.1*	*TaZHD1*	*T. aestivum*	Uncharacterized proteins	5D:359474816-359475706	891	1.8e-34	96.9
7.	*TraesCS5D02G234600.1*	*TaZHD10*	*T. aestivum*	Uncharacterized proteins	5D:342174542-342176173	1632	7.8e-149	100.0

*Gene stable ID obtained from the NCBI or ensemble plants (wheat) database.

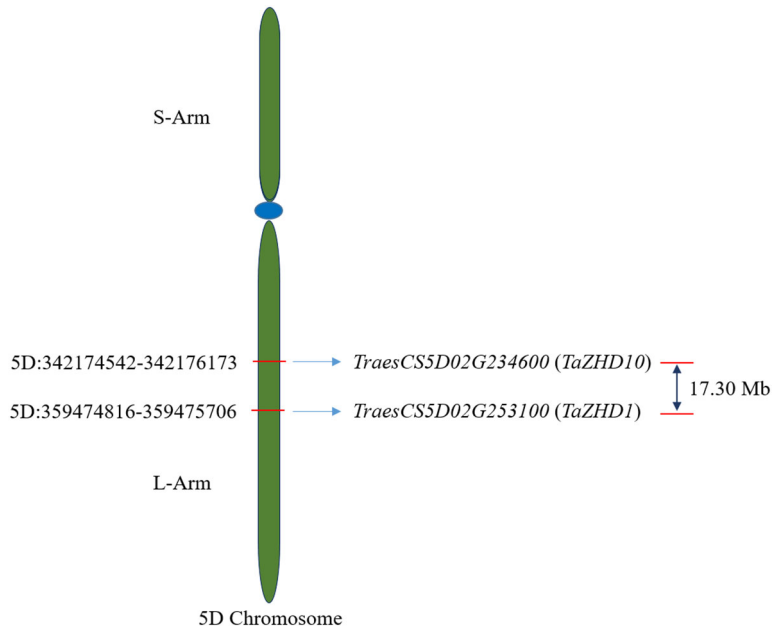


FIGURE 3
Chromosomal distribution of identified candidate *ZHD* genes in wheat.

while regulating the expression of key genes. Further, based on the similarity index to proteins of known structure, these domains are believed to be involved in forming hetero- and homodimers. They may form zinc fingers, an essential constituent of DNA binding domains.

In addition to this, motif analysis identified a set of 11 conserved motifs distributed across the putative chimeric proteins. The relative distribution of conserved domains and motifs across the putative chimeric proteins are depicted in Figures 4B–D, and presented in Supplementary Table 3. Out of these, the motif-I, IV, and V are considered to be the longest (50 aa), followed by motif-VII (49 aa) and motif-II (34 aa), whereas motif-VI (09 aa) is designated as the shortest one based on their length of consensus sequences. Further, to localize the specific positions of identified domains and motifs, we aligned protein sequences of identified candidate proteins with their homologs. The results revealed that the ZF-HD dimer superfamily is estimated to be embedded between 40-125 peptides, while the homeo_ZF_HD superfamily domain lies between 220 to 300 peptides as designated in their respective homologs (Figure 5). The ZF-HD dimer superfamily was found to share the consensus sequences of motif-VIII (40%), motif-II (100%), and motif-III (80%) while embedded with some undefined regions. Similarly, the homeo_ZF_HD superfamily domain was formed by sharing consensus sequences of motif-IV (10%) and motif-VII (100%) and some undefined regions. Besides this, multilevel sequences analysis revealed that the motif-XI and VIII share some overlapped regions of about four amino acid long consensus sequences while motif-X and IV share 14 amino acid long overlapped regions.

Finally, to explore the evolutionary significance of identified genes (proteins), a phylogenetic tree was constructed using the Maximum-Likelihood method. Based on the higher bootstrap values, the rooted phylogenetic tree topology grouped annotated genes into two phylogenetic clusters viz., clusters A and B (Figure 6A). Cluster A includes genes like *OsZHD1* and *TaZHD1*, while cluster B includes genes like *ZmZHD10, OsZHD10, TaZHD10, TdZFH9a*, and *TdZFH9b*. The study showed that *TaZHD1* shares high similarities with other homologs in rice (*OsZHD1*). Likewise, the *TaZHD10* is more closely related to homologs in rice (*OsZHD10*) and emmer wheat (*TdZFH9a* and *TdZFH9b*) but distantly with homologs of maize (*ZmZHD10*). Furthermore, multiple sequence alignment of amino acids showed that *TaZHD1* shares 96.9% identity with *OsZHD1*. Similarly, *TaZHD10* shares a significantly higher identity with rice *OsZHD10* (89.0%) and emmer wheat homologs, including, *TdZFH9a* (95.2%) and *TdZFH9b* (95.5%), while least with maize (84.6%) homologs. Altogether, this information suggested that *TaZHD1* and *TaZHD10* share high homology, a common origin from similar orthologs, and possess the typical features of a class zinc finger homeodomain (ZF-HD) IV family transcription factor/gene.

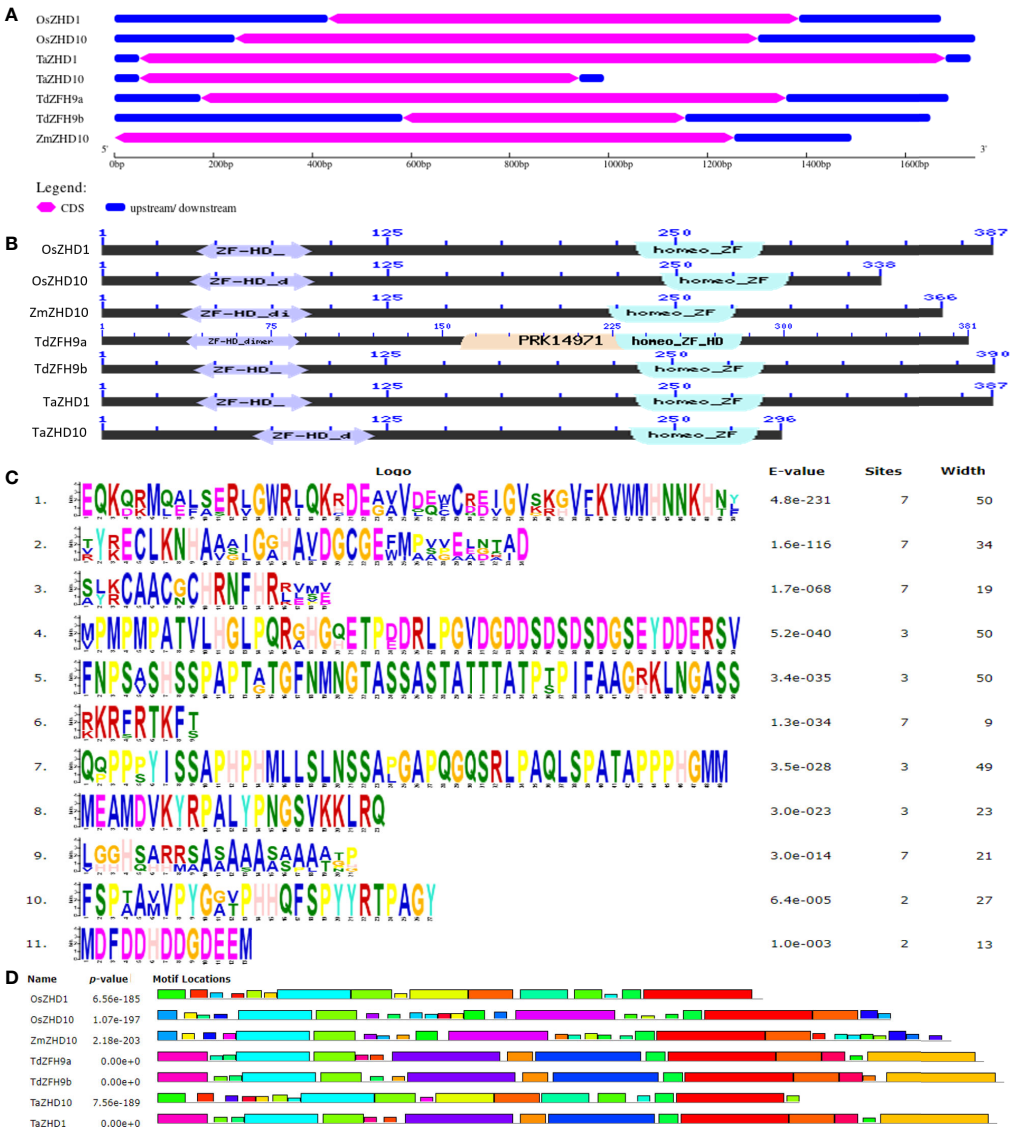


FIGURE 4
Comparative analysis of **(A)** structural organization of *ZHD* genes, **(B)** conserved domains, **(C)** conserved motifs, and **(D)** locations of motifs, of candidate leaf rolling ZHD protein of orthologs and their best possible match with wheat.

Structural analysis of candidate ZHD proteins

Modeled 3D structure of TaZHD10 shared up to 100% similarity with respective homolog. Using an automated Swiss-Model server, the modeled 3D structure was further examined for superposition with acceptable template structures. The TaZHD10 protein was found superposed (<1 Å RMSD) with more than seven ZF-HD homeobox family proteins of *A. thaliana* sharing up to 55.74-60.34% similarity. Among them, the homeobox domain of *A. thaliana* hypothetical protein (F22K18.140) displayed up to 59.65% similarity upon superposition (Supplementary Table 4). The similarity of proteins up to this level was believed to be adequate for annotation of the 3D molecular structure of predicted proteins. As per homology modeling and superposition rule offered by automated Swiss-Model server, modeled protein exhibiting more than 30% similarity with template structure is considered an excellent model (Supplementary Figure 1A).

To evaluate the quality of modeled 3D structure of TaZHD1 and TaZHD10 protein, we further analyzed the Ramachandran plot of each protein in contrast with template structure (homologs). As depicted in Supplementary Figure 1B, phi (Φ) and psi (Ψ) torsion angles of Ramachandran plots revealed an

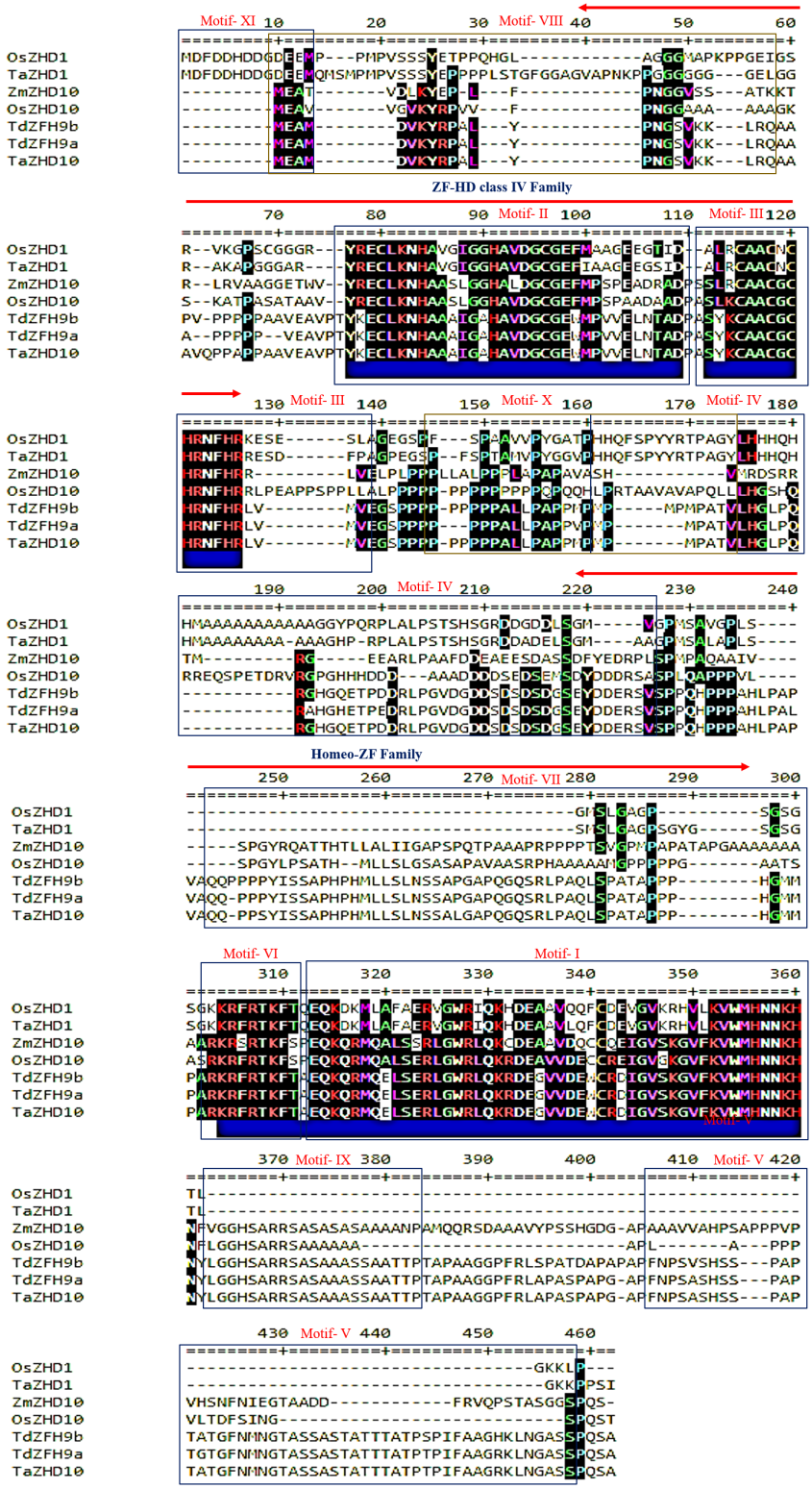
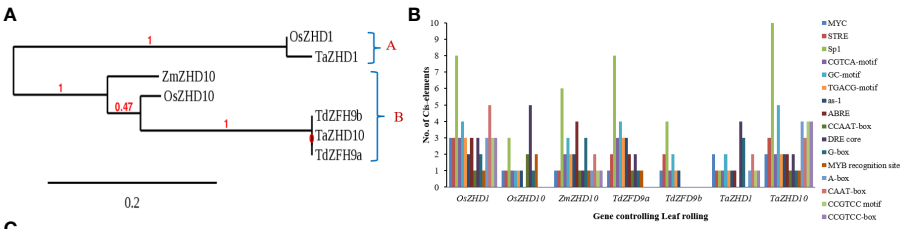


FIGURE 5
Multiple sequence alignment of the conserved domain and motifs distributed across the candidate leaf rolling *ZHD* genes in wheat and its orthologs.



C

	Core promoter	Abiotic-stress responsive							Light-responsive elements		Enhancer-like element	Unknown function				
	CAAT-box	MYC	CGTCA-motif	TGACG-motif	ABRE	CCAAT-box	DRE core	MYB recognition site	Sp1	G-box	GC-motif	STRE	as-1	A-box	CCGTCC motif	CCGTCC-box
OsZHD1	5	3	3	3	3	1	3	1	8	2	4	3	2	3	3	3
OsZHD10		1	1	1		2	5	2	3	1	1	1	1			
ZmZHD10	2	1	2	2	4	1	1	1	6	3	3	1	2	1	1	1
TdZFD9a		1	3	3	2	1	2	1	8	1	4	2	3			
TdZFD9b		1	1	1					4		2	2	1			
TaZHD1	2	2	1	1	1		4		1	3	2	1	1	1	1	1
TaZHD10	3	2	2	2	2	1	2	1	10	1	5	3	2	4	4	4

FIGURE 6
Comparative analysis of, **(A)** phylogenetic tree topology, **(B)** *Cis*- acting regulatory elements in promoters, and **(C)** functional classes of annotated *Cis*-acting regulatory elements of candidate leaf rolling *ZHD* genes of wheat and their orthologs (*Cis*-acting regulatory elements identified in genes were represented in the form of bars).

excellent geometry of modeled 3D structure of TaZHD1 and TaZHD10 proteins. Ramachandran plot calculation showed up to 2.53% residues in generous outlier regions and 12.86% residues in rotamer outlier regions. In comparison, up to 86.56% of residues were present in the most favored regions, commonly called Ramachandran regions. The results suggested the excellent quality of modeled proteins as displayed through PROCHECK algorithms, a widely used method to estimate the stereochemical quality of modeled proteins. Comprehensive descriptions of Ramachandran plot structure assessment of identified leaf rolling gene (protein) in wheat with homologs are presented in Supplementary Table 5.

Analysis of *Cis*-acting regulatory elements in *ZHD* genes

In the promoters' region, the frequency of *Cis*-acting regulatory elements plays a vital role in defining the stress-responsive or tissue-specific expression of a key gene under variable environmental conditions (Khatun et al., 2017). A web-based search was carried out from available databases to identify possible hormones and stress-responsive *Cis*-acting regulatory elements in the promoter regions of wheat *ZHD* genes. The PlantCARE database predicted a set of 34 *Cis*-acting regulatory elements scattered across promoter regions of the putative candidate genes. The result revealed that both candidate genes possess CAAT-box (CAAT) as basal or core promoter and enhancer regions. In contrast, other 33 *Cis*-acting regulatory elements were also found, directly or indirectly involved in biotic & stress-responsiveness, phytohormonal signaling, light-responsiveness, and other developmental processes. During comprehensive analysis, the study revealed that both candidate genes possess large numbers of *Cis*- acting regulatory elements in promoter regions as recorded in their homologs. Because of a large set of identified *Cis*- acting regulatory elements, the study was finally focused on the 16 most frequent *Cis*-acting regulatory elements in promoters. Out of 16, we found seven abiotic-stress responsive *Cis*-acting regulatory elements, including MYC (CATGTG/CAATTG), CGTCA-motif (CGTCA), TGACG-motif (TGACG), ABRE (ACGTG/GCCGCGTGGC), CCAAT-box (CAACGG), DRE core (GCCGAC) and MYB recognition site (CCGTTG), involved in the regulation of drought stress-inducible genes. In addition, CGTCA-motif and TGACG-motif are engaged in the MeJA-responsive expression of genes. *Cis*-acting regulatory elements like Sp1 (GGGCGG) and G-box (TACGTG/CACGAC) play a specific role as light-responsive elements, whereas GC-motif (AGCGCGCCG/CCCCCG) is an enhancer-like element involved in anoxic specific inducibility of target genes. Furthermore, the particular role of other *Cis*-acting regulatory elements like STRE (AGGGG), as-1 (TGACG), A-box (CCGTCC), CCGTCC motif (CCGTCC), and CCGTCC-box (CCGTCC) are yet to be defined.

The *TaZHD1* possessed all 14 *Cis*-acting regulatory elements except CCAAT-box and MYB recognition site, while the *TaZHD10* had all 16 types *Cis*-acting regulatory elements in their promoter regions (Figure 6B). Besides, *Cis*-acting regulatory elements like DRE core (4) and G-box (3) exist in many copies in *TaZHD1*, correspondingly the STRE (3), Sp1 (10), GC-motif (5), A-box (4), CAAT-box (3), CCGTCC motif (4) and CCGTCC-box (4) are present in many copies in *TaZHD10* (Figure 6C). Consequently, because of a huge set of stress-responsive *Cis*-acting regulatory elements in promoter regions of *TaZHD1* and *TaZHD10*, the present analysis can conclude that *TaZHD1* and

TaZHD10 are a type of abiotic (drought) stress-inducible genes that might involve in the expression of rolled leaf phenotype under a drought-induced environment in wheat.

Prediction of miRNAs targeting candidate *ZHD* genes

Earlier studies revealed that the miRNAs could play defining role in leaf differentiation by negative regulation of target gene expression at the post-transcriptional level (Moon and Hake, 2011; Xu et al., 2018). So, to unveil the possible role of miRNAs in leaf rolling identified wheat, *TaZHD1*, and *TaZHD10* were searched against the psRNATarget v2.0 server. The study resulted in 5 wheat miRNAs (Tae-miRs) targeting both candidate genes. Identified 5 wheat Tae-miRs belong to three different miR families (Supplementary Table 6). Three miRNAs viz., tae-miR1130b-3p, tae-miR531, and tae-miR9666a-3p were targeting *TaZHD1* representing two different miR families. Correspondingly, two miRNAs viz., tae-miR9664-3p and tae-miR9672b were found to be targeting the *TaZHD10*, also representing two different miR families. The expected (E) value of all miRNAs ranged from 4.5 to 5.0 suggesting greater functional specificity of identified Tae-miRs. Furthermore, all miRNAs also exhibited significant pairwise alignment with their complementary query sequence advocating that the identified genes belong to protein-encoding Tae-miRs, which might be involved in generating functional proteins. Besides this, all predicted Tae-miRs showed cleavage inhibition properties with a multiplicity ratio (1). Thus, the results further demonstrated that identified Tae-miRs might associate with RNA-mediated silencing of *TaZHD1* and *TaZHD10* putatively involved in the metabolic pathway controlling rolled leaf phenotype.

Differential expression profiling of Wheat *ZHD* genes

Transcriptome-wide expression profiling

To gain further molecular insights into the relative expression patterns of *TaZHD1* and *TaZHD10* transcripts, we used available RNA-Seq transcriptome datasets for leaf differentiation, plant growth, tissues and organ development, and biotic and abiotic stress responses. Expression profiling of *TaZHD1* and *TaZHD10* transcript resulted in differential expression patterns at three different developmental viz., seedling, vegetative, and reproductive stages. Both transcripts were displayed lower to a higher level of transcript expression in all studied tissues (roots, leaves, shoots, grains, and developmental spikes). Furthermore, *TaZHD1* and *TaZHD10* transcripts showed significantly higher expression levels ranging from 0-10 tpm, in the tissues at seedling, three-leaf stage, tillering, fifth leaf stage, seven leaf stage, booting, flag leaf stage, early-to-late heading, grain filling, and ripening stage. As demonstrated in a heat map (Figure 7), both transcripts also showed dramatically higher expression upon biotic (stripe rust, powdery mildew, and fusarium wilt) and abiotic (drought, heat, cold, and phosphorus starvation) stress-induced conditioning wheat, especially in Chinese spring. *In-silico* expression profiling of *TaZHD1* and *TaZHD10* transcripts was further predicted under drought stress compared to control. Expression of *TaZHD1* was significantly reduced under drought stress in tissues of root, stem, 3-7 leaf stage, and developmental spikes compared to irrigated. In contrast, *TaZHD1* was expressed significantly higher in flag leaf and leaf sheath tissues under drought stress at the seedling and reproductive stages. Besides, this *TaZHD10* transcript was predicted to be 2-5 fold up-regulations under drought stress with reference to *TaZHD1*. Similarly, *TaZHD10* was expressed significantly higher in flag leaf, leaf sheath, and developmental spike tissues while reduced in roots at seedling and reproductive stages. Both transcripts showed low to no expression in the roots and 3-leaf stage, while expressed almost equally in flag leaf, leaf-sheath, and spikes throughout all developmental stages (Figure 7). Altogether, results indicated that *TaZHD10* showed many fold up-regulation and was ubiquitously expressed in drought-induced leaf tissues, which might also be considered a major component trait of leaf rolling. Henceforth, out of these two transcripts, *TaZHD10* might be the best candidate gene controlling rolled leaf trait under a drought-induced environment in this important cereal crop.

Quantitative real-time PCR profiling

To further elucidate the functional aspects of identified candidate *ZHD* transcripts we conducted qRT-PCR analysis at two different developmental stages. Likewise, differential expression patterns of these transcripts were also recorded at three different time scales, viz., irrigated, drought (10 days after water withdrawal), and after 24 hours of watering (24 HAW). Both transcripts were variably expressed in flag leaf tissues of all genotypes at both booting (stage 10), and heading (stage 10.5) stages. Under drought conditions, transcripts were initially up-regulated and reached the highest expression level (2-4 folds' vs irrigated) and then down-regulated (1-3 folds vs drought) upon 24 HAW. Consistent with this, we observed a notable increase/decrease in expression of the *TaZHD1* at a particular growth stage often caused by the up-or down-regulation of *TaZHD10* (Figures 8A, B). For example, over-expression of *TaZHD10* was accompanied by the decreased expression of *TaZHD1* at the booting stage and down-regulation of *TaZHD1* caused an increased expression of *TaZHD10* at the heading stage. In contrast, upon drought treatment, the transcript abundance of *TaZHD10* had showed 2-3 folds up-regulation in expression than *TaZHD1* in all studied genotypes.

Overall, the expression pattern we observed for these candidate genes by qRT-PCR analysis was similar to their expression in the obtained RNA-Seq transcriptome data. In addition, differential expression profiles of genes in studied genotypes are highly congruent with the results as observed in

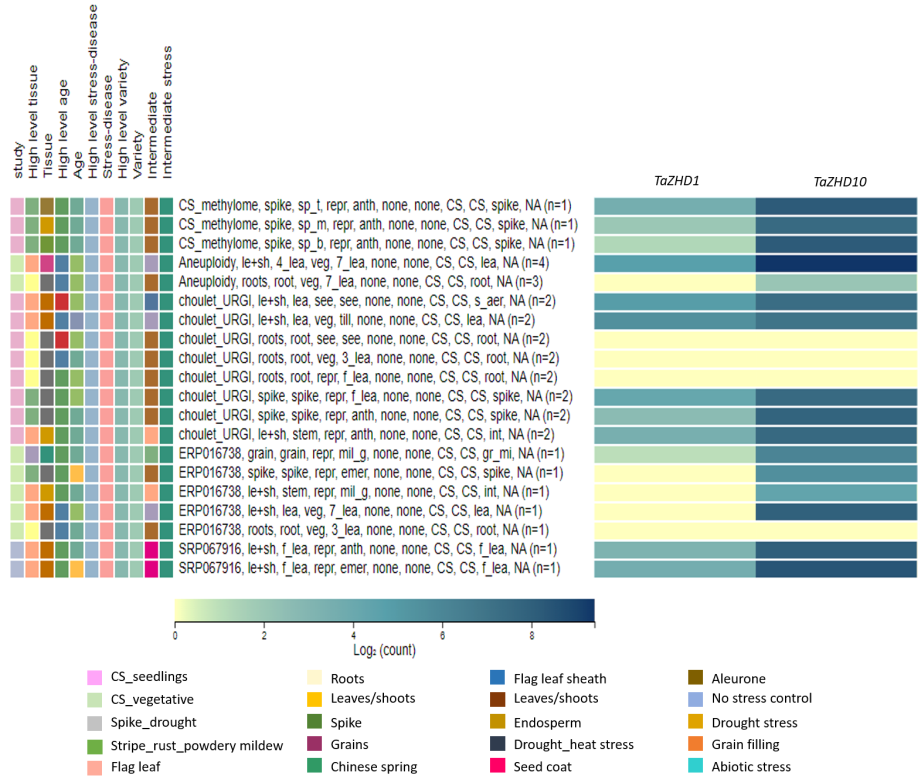


FIGURE 7
Transcriptome-wide expressions profiling of candidate leaf rolling *ZHD* genes in wheat tissues at different developmental stages and under studied biotic and abiotic stress conditions.

histological observations (Figures 1 and 2) and leaf rolling index (for more details please follow Verma et al., 2020). Finally, based on the differential expression patterns we hypothesized that there might exist a complex regulatory loop between *TaZHD1* and *TaZHD10* which might control rolled-leaf phenotype. Furthermore, our investigations suggested that the zinc finger homeodomain (ZF-HD) IV family rich proteins might also play an important and complementary role in the transcriptional activator of *TaZHD1* and *TaZHD10* -controlling bulliform cell differentiation. More intriguingly, the study further advocated that the relative expression of candidate transcripts in wheat is highly genotype and treatment-specific, while also varied upon developmental stages. Relative expression profiles of transcripts in tissues of different wheat genotypes studied under irrigated, drought and 24 HAW are presented in Figures 8A, B.

Gene interaction network analysis of *ZHD* genes and other regulatory partners

The predicted molecular interactome network revealed a set of zinc finger homeodomain (ZF-HD) IV family rich proteins having different regulatory partners based on the various parameters such as physical interactions, shared protein domains, genetic interactions, co-localization, and co-expression analysis (Figure 9). The interaction network clearly states that both *ZHD1* and *ZHD10* with their regulatory partners are involved in various metabolic and gene regulatory networks. Furthermore, up to 52.06% are engaged in close physical interactions, 21.72% are designated as candidate gene/transcription factors (TFs), 18.12% regulate the co-expression of particular genes, and 4.90% share their functional protein domains. Only 2.44% are involved in genetic interactions, while 0.76% is co-localized within similar cellular or sub-cellular compartments. The interaction network of both *ZHD1* and *ZHD10* share some common regulatory partners. The majority of them are dominated by the Zinc-finger homeodomain class proteins (ZHD3-9 and ZHD11-14), mini zinc finger proteins (MIF1-3), and *A. thaliana* homeobox protein 22 (ATHB22).

Apart from these, both *ZHD1* and *ZHD10* also exhibited significant interactions, suggesting that both might involve a common metabolic and gene regulatory network. Comprehensive analysis of the *ZHD1* interaction network further revealed that regulatory partners like GYF domain-containing protein (AT5G42950), axial regulatory YABBY1

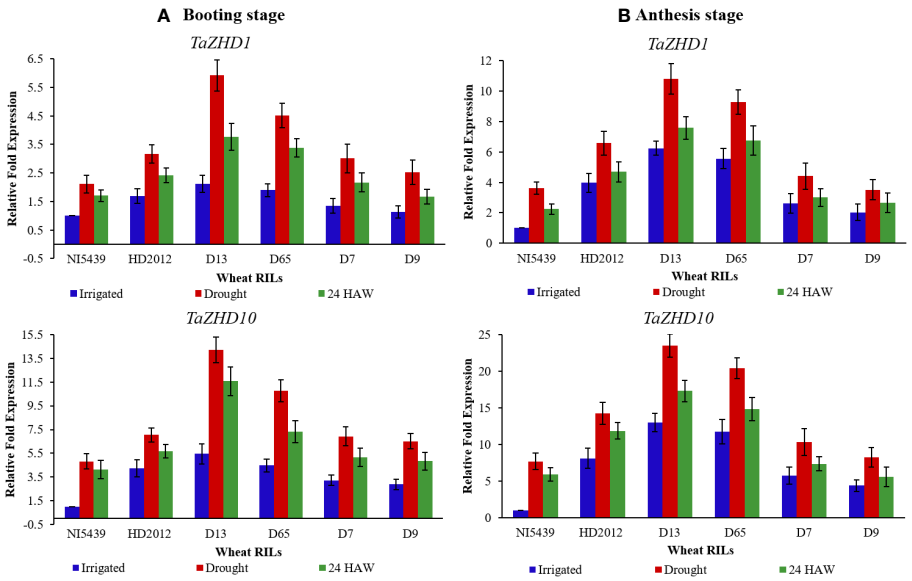


FIGURE 8
Relative expression profiling of candidate leaf rolling *ZHD* genes in tissues of wheat RILs at **(A)** booting stage, and **(B)** anthesis (heading) stage (From left to right: NI5439; drought resilient, HD2012; drought susceptible, high LR RILs = D-13 and D-65, and low LR RILs = D-7 and D-9; HAW, hours after watering). Each test sample represents three independent biological and three technical replicates.

(YAB1), SKU5 are similar 3 (sks3), an unknown protein (AT1G10990) does not exhibit significant interactions with *ZHD10* protein/genes. Similarly, the regulatory partners of *ZHD10* like NAC domain-containing protein 20 (ANAC020), Profilin-2 (PRO2), Arginine-glutamic acid dipeptide repeat protein (AT1G09050), and unknown protein (AT1G09040) are not showing any significant interactions with *ZHD1* protein/genes. Collectively, the resultant gene regulatory network integrates many genes controlling leaf rolling by coherent development of bulliform cells as presented in Figure 9.

Epigenetic aspects of leaf rolling

We investigated the genomic DNA methylation levels in context of mCG, mCHG and mCHH, and found the wide coverage across the genic structure of both target genes. Consistent with the findings of qRT-PCR profiling, the DNA methylation profiles revealed heavily methylated and demethylated 5' UTRs, introns, exons, and 3'UTRs regions for the mCG, mCHG and mCHH sequences which varied upon growth (stress) conditions (Supplementary Table 7). Under irrigated condition, there is no or little methylated and demethylated regions of target genes were observed. Unlike this, the mCG and mCHG sequence revealed significant methylated regions whereas the mCHH methylation of target genes was observed under drought condition. Congruent to this, variable demethylation levels in mCG, mCHG and mCHH sequences were recorded against the 24 hours of watering. We also observed significant differences in methylation patterns, specifically in the mCG context. *TaZHD10* revealed methylation while demethylation of *TaZHD1* at target exons was significantly more effective than *TaZHD10*. The results also confirmed findings of qRT-PCR profiling that the higher expression levels of *TaZHD10* than *TaZHD1* is a result of variable methylation-demethylation events. Taken together, methylome study suggested a critical role of DNA methylation in the regulation of *ZHD* genes, and also provides molecular insights into the epigenetic potential for controlling *TaZHD1* and *TaZHD10* mediated leaf rolling modulated by the drought-stress conditions.

Discussion

Leaf architecture, such as shape and size, are critical morphological traits that affect the yield performance of the plant. Grass species such as maize and rice require the integration of critical developmental events to develop a single leaf from the founder cells, including differentiation, development, and elongation of leaf structures, leaf sheath, and leaf blade (Sakaguchi et al., 2010). Very young and tender leaf rolls inherently in the rice and maize when it is still enclosed with the sheath, is mainly *via* an unknown mechanism. Congruent to this, the mature leaves generally roll adaxially under stress conditions, probably through a mechanism

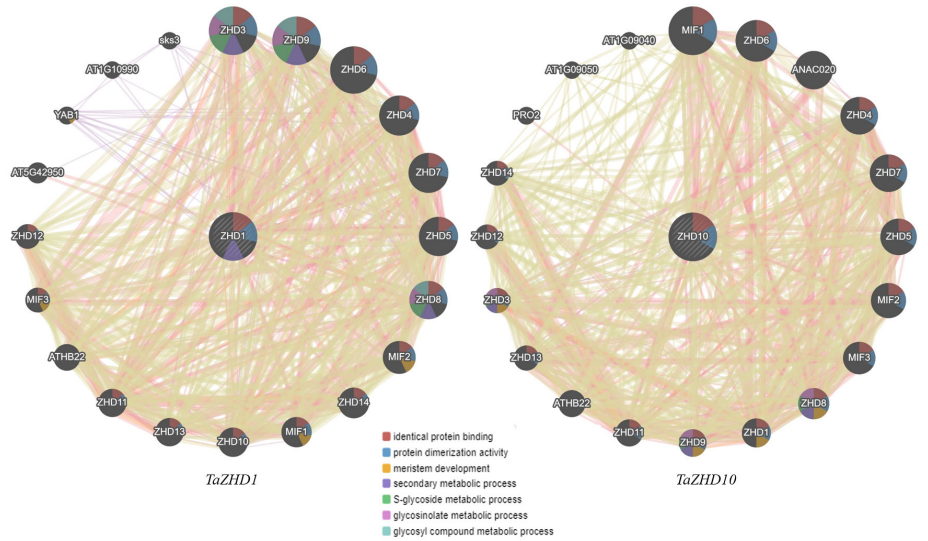


FIGURE 9
Gene interaction network of candidate leaf rolling *ZHD* genes with their regulatory partners taking *A. thaliana* as reference.

controlled by the bulliform cells. Bulliform cells are thin-walled, large, highly vacuolated, and bulky cells present on the adaxial epidermis between the vascular bundles of the leaf blade (Fujino et al., 2008). It is proven that the adaxial bulliform cells lose turgor pressure under water stress, which results in leaf rolling. In contrast, when stress is relieved, bulliform cells maintain water potential and revive the surface tension. However, the actual physiology of bulliform cells in leaf rolling underwater deficiency is still to be proven (Li et al., 2016). Earlier studies advocate that bulliform cells might control the development of the young leaves, which are still rolled up in leaf sheath (Alvarez et al., 2008). It has also been proposed that bulliform cells could cause the erection of the leaf blade, reduce water loss, and increase stomatal resistance (Sakamoto et al., 2006). Apart from these, there are several factors which defines leaf rolling in plants. In general, leaf rolling caused by bulliform cells on adaxial leaf blade are regulated by the architecture (shape and size), number of bulliform cells and the different environmental factors (Alvarez et al., 2008).

Since the leaf architecture is a prominent agronomic trait in wheat breeding (Verma et al., 2020), the physio-bio-molecular mechanism of leaf rolling was of great interest, especially under the current environmental stress and nutrient limiting condition. Under drought stress, plant species like Arabidopsis, rice, wheat, maize, and sorghum exhibited characteristic leaf rolling as a tolerance response. In addition, leaf rolling has polygenic control with additive effects at target loci, especially in wheat (Sirault et al., 2008). There is scarce information about the genetic basis of leaf rolling in bread wheat (hexaploid), except for a report regarding stable QTL (*Qlr.nhv-5D.2*), associated with leaf rolling under moisture stress, located on the 5D chromosome (Verma et al., 2020). However, in emmer wheat (tetraploid) 11 QTLs related to leaf rolling were reported, significantly found to co-localized on 1A, 2A, 2B, 4B, 5A, 5B, 6A, 6B, 7A, and 7B chromosomes (Peleg et al., 2009).

Leaf rolling is a complex trait controlled by a myriad of genes. More than 70 QTLs/genes have been extensively studied in crops like rice and maize to regulate leaf rolling, but the complete molecular prospects of these complex traits are still not yet fully explored. The HD-Zip class IV family proteins are plant-specific transcription factors (TFs) classified based on sequence similarity into HD-Zip class I-IV family proteins (Gao et al., 2019). HD-Zip class IV family proteins (genes) occur in monocots (rice and maize), dicots (arabidopsis), and other plants (Bhattacharjee and Jain, 2013; Xu et al., 2014; Wang et al., 2016; Khatun et al., 2017; Abdullah et al., 2018; Yoon et al., 2020; Liu et al., 2021). These *ZHD* family genes are reported to involve in various morpho-physiological roles in plants, including vascular bundle development, light signaling, phytohormonal signaling, biogenesis of plant organs and stress responses (Tran et al., 2007, Liu et al., 2021). Arabidopsis contains 16 HD-Zip class IV family genes that play important roles during trichome development, root development, differentiation of epidermal cells, and anthocyanin accumulation (Xu F. et al., 2012; Gao et al., 2019). Similarly, rice contains 11 HD-Zip class IV family genes among which *OsRoc1* to *OsRoc8*, are specifically over-expressed in epidermal tissues of rice resulting in abaxial and/or adaxial leaf rolling *via* increased size and number of bulliform cells (Yun et al., 2010; Zou et al., 2011; Fang et al., 2021). More often, maize contains

17 HD-Zip class IV family genes among which *ZmOCL1, ZmOCL3, ZmOCL4*, and *ZmOCL5* are exclusively over-expressed in epidermal tissues whereas expression of *ZmOCL2* is restricted to the sub-epidermal tissues led to leaf rolling.

In addition, some *ZHD* genes are reported to have a significant modulatory role in defining organ identity under salt and dehydration stress (Wang et al., 2016; Abdullah et al., 2018). In arabidopsis, *AtZHD10* interacts with TZP proteins and regulates root development by hypocotyl elongation (Perrella et al., 2018). Relative expression of Arabidopsis ZHD protein (*AtZHD1*) is induced under salt and dehydration stress (Wang et al., 2016). It was later evidenced that ZHD TFs of *AtZHD1* can interact with promoters of *Early response to dehydration stress 1* (*ERD1*) at the zinc finger homeodomain recognition site (CACTAAATTGTCAC) and subsequently result in co-expression of both *ERD1* and *AtZHD1* under ABA responsiveness (Tran et al., 2007; Singh and Laxmi, 2015; Abdullah et al., 2018). Furthermore, it was also revealed that the NAC TF can also interact with ZHD-rich proteins that can simultaneously overexpress *NAC* and *ZHD* as a potential regulator of salt and drought tolerance in Arabidopsis (Hu et al., 2008). Finally, as reported in the present investigation, the results advocated that the identified *ZHD* genes encode a ZHD protein modulated by a HD-Zip class IV TFs (Ariel et al., 2007). As the potential regulator of drought tolerance (via leaf rolling) in various plant species, the HD-Zip class IV TFs can modulate leaf rolling *via* increased size and number of bulliform cells (Zou et al., 2011; Dong et al., 2019; Fang et al., 2021), particularly in the wheat crop.

In the present study, we reported for the first time two new wheat, *TaZHD1* and *TaZHD10* genes containing a zinc finger homeodomain class homeobox transcription factors (TFs). We found that *TaZHD1* and *TaZHD10* are the closest orthologs of rice *OsZHD1* and *OsZHD10* and may play a key role in leaf rolling. Gene structure prediction revealed that the identified *TaZHD1* and *TaZHD10* possess the same exon-intron patterns, and the positions of the exons were highly conserved, which was consistent with earlier studies (Khatun et al., 2017; Abdullah et al., 2018). Besides, due to the consistent pattern of genes, the relative proportion of the length of exons in wheat was revealed to be highly coincident with earlier reports in rice, maize, tomato, cotton, and Arabidopsis (Xu G. et al., 2012; Khatun et al., 2017; Abdullah et al., 2018; Liu et al., 2021). The unique grouping of wheat and rice reflects the species-specific evolution of the *ZHD* family genes or could be either due to the domain or motifs-based phylogeny analysis. Undoubtedly, this is not surprising to consider the concept of remote-distance homology between the two co-evolving crop species (Moore and Purugganan, 2003). This result further demonstrates the reliability of evolutionary significance analysis for the independent classification of *ZHD* family genes. The *TaZHD1* and *TaZHD10* comprise ZF-HD dimer superfamily domains as functional domains presumably requisite for DNA-binding, suggesting that *TaZHD* genes act as potential transcription factors (Abu-Romman, 2014; Li et al., 2016). Moreover, potential domains or motifs and their gain or loss during gene duplication events might have contributed to the development of paralogous pairs of genes (Cannon et al., 2004; Xu G. et al., 2012).

The sequence analysis of identified chimeric ZHD proteins revealed wide variability in molecular weight, number of amino acids, GRAVY, aliphatic index, instability index, and isoelectric point (pI). The studied physicochemical properties might directly influence the stability of the chimeric protein (Chandra et al., 2021). Stability variations could be one of the driving factors for reduced leaf lamina joint angle with erect and rolled leaf behavior of the ZHD proteins while accumulation in sub-cellular compartments (i.e., nucleus) of the plants (Liu et al., 2021).

Cis-acting regulatory elements are a regulatory component of promoter regions of many genes that provide clues for stress responsiveness, light-responsiveness, hormonal responsiveness, growth and development, and other gene-specific expression patterns under specific environmental and growth conditions (Yun et al., 2010; Xu F. et al., 2012; Abdullah et al., 2018; Chandra et al., 2021). As reported in earlier studies, we predicted many *Cis*-regulatory elements in *TaZHD1* and *TaZHD10* promoter regions. We categorized these *Cis*-acting regulatory elements into three major groups based on their involvement in metabolic and gene regulation in the plant, including a) light-responsive, b) stress-responsive, c) hormonal responsive, d) involved in growth and developmental processes, and e) unknown function (Mohanty et al., 2012). CGTCA-motif and TGACG-motif are involved in (Methyl-Jasmonate) MeJA-responsive expression of genes (Abdullah et al., 2018). *Cis*-acting regulatory elements like Sp1 and G-box play a specific role as light-responsive elements, whereas GC-motif is an enhancer-like element involved in anoxic specific inducibility of target genes (Ezer et al., 2017; Liu et al., 2021). Besides zinc finger, the presence of numerous *Cis*-acting regulatory elements, including MYC (myelocytomatosis oncogene), CGTCA-motif, TGACG-motif, ABRE (ABA-responsive element), CCAAT-box, DRE core (dehydration-responsive element) and MYB (myeloblastosis oncogene) recognition site, suggesting the existence of multiple regulatory mechanisms in response to drought stress (Singh and Laxmi, 2015; Baldoni et al., 2015; Khatun et al., 2017; Dong et al., 2019; Chandra et al., 2021; Liu et al., 2021), that resulted in leaf rolling in plants (Dong et al., 2019). Additionally, the CCAAT-box contains MYBH1 binding sites that regulate various abiotic stresses and developmental processes (Laloum et al., 2013). Correspondingly, MYC controls plant growth and jasmonate (JA) induced defence gene activation (Yun et al., 2010; Singh and Laxmi, 2015). In contrast, the specific role of other *Cis*-acting regulatory elements like STRE (stress-responsive element), as-1, A-box, CCGTCC motif, and CCGTCC-box are yet to be defined

(Xu F. et al., 2012; Gubert et al., 2014; Yoon et al., 2020; Liu et al., 2021). The drought responsive *Cis*-acting regulatory elements in the genes are consistent with their drought-responsive up-regulation during course of leaf rolling in plants.

Moreover, some reports elucidated that miRNAs can also play an important role while defining leaf differentiation by post-transcriptional silencing of target genes (Moon and Hake, 2011; Xu et al., 2018; Liebsch and Palatnik, 2020). For example, overexpression of *OsAGO7* resulted in adaxial leaf rolling in rice, which is knocked down by RNAi- the silencing mechanism. The transgenic plant exhibited significant pleiotropic defects resulting in phenotypic development like low plant height and narrow and erect rolled leaves (Shi et al., 2007; Wu et al., 2009). Additionally, miRNA160 and miRNA166 have played an important role in leaf rolling *via* xylem differentiation and reduced stomatal conductance by targeting *OsARF18* and *OsHB4*, respectively (Huang et al., 2016; Zhang et al., 2018). Thus, as reported in earlier studies, some other miRNAs like tae-miR1130b-3p, tae-miR531, tae-miR9666a-3p, tae-miR9664-3p, and tae-miR9672b can indirectly affect the leaf rolling by targeting *TaZHD1* and *TaZHD10*.

In the era of modern biological sciences, gene interaction network analysis through system biology has become an appealing research topic. Interactome networks render the functioning and offer a possible interaction of particular proteins within a cell to define any metabolic process (Yun et al., 2010; Borrill et al., 2016; Liu et al., 2021). Ever since, because of the unavailability of molecular interactome data of wheat in the public domain, we have used Arabidopsis interactome databases to predict the possible interaction of *TaZHD1* and *TaZHD10*. The results showed that *TaZHD1* might interact with the *TaZHD10* (Figure 8), which was congruent with our other results. The ZHD class IV transcription factors (TFs) can activate or suppress the expression of downstream target genes, thus regulating the important metabolic and gene regulatory networks (Tran et al., 2007; Abu-Romman, 2014; Abdullah et al., 2018). Therefore, we used available RNA sequence datasets to quantify the relative expression of *TaZHD1* and *TaZHD10* to explore the functional aspects. The result revealed the transcriptional expression of the identified gene, coinciding with earlier reports in rice confirming the paralogous pair of genes evolved through common gene duplication events (Moore and Purugganan, 2003).

qRT-PCR expression profiling of candidate genes in flag leaf tissue further revealed significant up-regulation or down-regulation consistent with other reported genes, *OsZHD1, OsZHD2, OsZHD10, YABBY1-7, SRL1*, and *ROC5* in rice and maize (Ohmori et al., 2011; Zou et al., 2011; Xiang et al., 2012; Xu et al., 2014). In present study, differential expression profiling under the drought-induced condition in the broadest section of young flag leaves was found significantly variable in the extreme RILs compared to parents (NI5439 and HD2012). The D-13 and D-65 RILs having higher leaf rolling than parents have highest expression of both the transcripts at both the stage and after adding water got down-regulated. Similarly, both transcripts in low leaf rolling (RIL D-7 and RIL D-9) were initially up-regulated under drought conditions (2-4 folds' vs irrigated), and then down-regulated upon 24 HAW (1-3 folds vs drought). This result clearly states the role of *TaZHD1* and *TaZHD10* in leaf rolling in wheat. Unlike this, the study also reports co-expression of *TaZHD1* with *TaZHD10* under drought-induced conditions. It could be because of co-evolution of multiple or duplicate copy of gene as a phenomenon of stress tolerance mechanism. Our result was consistent with Hoffmann and Hercus (2000), reported environmental stressors as a key evolutionary forces for co-evolution of multiple or duplicate copy of gene. Not much earlier, significance of gene-duplication events and cross-talk between co-evolved genes was extensively studied in plants, specifically for salinity and drought tolerance in Arabidopsis (Panchy et al., 2016).

Overexpression of *TaZHD1* and *TaZHD10* can induce abaxial or adaxial curled leaf. *TaZHD1* and *TaZHD10* were the only homologs of *OsZHD1* and *OsZHD10* in rice, and its overexpression can also tempt abaxial or adaxial curled leaf. Similarly, the overexpression of *OsZHD1* with the closest homolog *OsZHD2*, results in induced leaf rolling by increasing the number and size of the bulliform cells at their abaxial surface (Xu et al., 2014; Zhu et al., 2017; Yoon et al., 2020). More often, a similar role was displayed by *ZmZHD10*, which exhibits characteristic bulliform cells shrinkage during drought, heat, and salt tolerance (Abdullah et al., 2018). These genes were stated to affect the differentiation of bulliform cells in leaf blades that result in leaf rolling in plants (Zou et al., 2011; Xu et al., 2014; Yang et al., 2016; Li et al., 2019). Together, the morphological pattern of the rolled leaf was perfectly matched with the expression patterns of studied genes, further suggesting that the *ZHD* class IV transcription factors (TFs) genes might act as an upstream regulator of bulliform cell differentiation necessary for leaf rolling. The study also indicated the pleiotropic role of annotated *ZHD* genes in wheat. Since, it was already reported in arabidopsis, rice, maize, tomato, and soybean that the *ZHD* family genes are involved in multiple plant developmental processes and stress tolerance (Xu et al., 2014; Wang et al., 2016; Liu et al., 2019; Yoon et al., 2020). So, likely the reported *TaZHD1* and *TaZHD10* could also be involved in other unknown functions in wheat *via* independent mechanisms or interaction in addition to leaf rolling.

Plant responses to different biotic and abiotic stress include various metabolic and genic reprogramming which are highly important for priming regulation of gene expression. More recently, evidence suggested the epigenetic regulation of gene expression at transcriptional and post-transcriptional level in response to abiotic stress (Kou et al., 2011; Yong-Villalobos et al., 2015). Different biotic and abiotic stresses can act as a stimulus

and lead to alteration in genic expression by various epigenetic mechanisms like DNA methylation, histone modifications, nucleosome positioning, and miRNA (Saraswat et al., 2017; Guo et al., 2019). Among these epigenomics marks DNA methylation and histone modifications, are the most studied events in plants for exploration of epigenetic regulation of target genes (Hudson and Buck-Koehntop, 2018). For example, *A. thaliana ZHD* like i.e, *ASYMMETRIC LEAVES (AS2)* and Leaf Polarity gene *ETTIN/AUXIN RESPONSIVE FACTORS* (*ETT/ARF3*) are repressed due to CpG methylation resulted in altered leaf adaxial-abaxial development (Vial-Pradel et al., 2018). Similarly, histone modifications are a vital component of epigenetic mechanisms and are reported to have a pivotal role in controlling leaf differentiation (Xu et al., 2018). Since *TaZHD1* and *TaZHD10* are members of the zinc finger homeodomain (ZF-HD) IV family gene, they can play a pivotal role in gene expression *via* DNA methylation. Thus, differential expression patterns of *TaZHD1* and *TaZHD10* in flag leaf tissues under a drought-induced environment finally indicated its epigenetic aspects concerning the regulation of metabolic and gene regulatory networks involved in leaf rolling in wheat crops.

Conversely, the study further calls for systematic investigation would be necessary to extrapolate the functions of wheat *TaZHD1* and *TaZHD10* in leaf rolling by modern multi-omics technologies. In addition, we cannot exclude the prospect that the variation in the intergenic regions of the mapped QTL (*Qlr.nhv-5D.2*) interval may lead to a foundation for leaf rolling. A hypothetical metabolic and gene regulatory model that describes how identified *TaZHD1* and *TaZHD10* are involved in the leaf rolling under severe drought stress in wheat is presented in Figure 10. Besides all, the study further suggested that a complex metabolic and gene regulatory network might be involved in the leaf rolling in this important food crop. Finally, the present investigation can offer a valuable foundation for future functional research of other *TaZHD* genes or transcription factors for breeding wheat varieties tolerant to leaf rolling.

Conclusion

This is the first comprehensive study that resulted in identification and characterization of two new leaf rolling genes, namely *TaZHD1* and *TaZHD10* in wheat, using a comparative genomics approach. This gene belongs to the zinc finger homeodomain (ZF-HD) class IV family and encodes a ZF-HD dimer superfamily domain-containing protein. The transcriptome-wide differential expression profiling resulted in many fold up-regulations of *TaZHD1* and *TaZHD10* in drought-induced leaf tissues. The result was further validated by qRT-PCR analysis which also demonstrated their significant

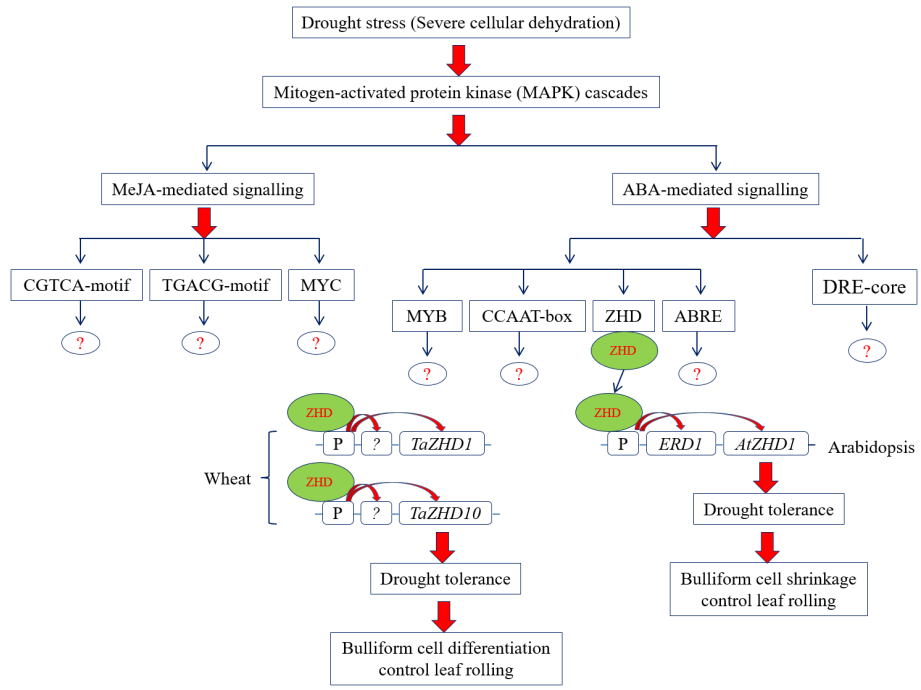


FIGURE 10
Proposed metabolic and gene regulatory model of leaf rolling under severe drought stress condition in wheat.

upregulation upon drought while down-regulated upon 24 HAW. The contrasting modulation of these genes under a drought-induced environment and the available reports of its epigenetic behavior might provoke erect and rolled leaves. Overall, the substantiate validation of *TaZHD1* and *TaZHD10* suggested their functional redundancy and pinpoint importance in leaf rolling by regulating bulliform cells differentiation. However, the study also calls for a further comprehensive investigation to decipher the knowledge gaps concerning the ZHD proteins as a key regulator of bulliform cells activity and their possible molecular networks underlying leaf rolling in this important cereal crop.

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

V and SJ conceptualized the research. SJ, V, and AC designed the experiments. V, MN, NM and SJ contributed experimental materials; AC and PA executed lab experiments and data collection; AC, SJ and V analyzed and interpreted data; AC and SJ wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by ICAR, Govt. of India sponsored project entitled "Incentivizing research in agriculture; Project IV: Molecular genetic analysis of resistance/tolerance to different stresses in rice, wheat, chickpea and mustard including sheath blight complex genomics - wheat component" (Project no.- 15613160025).

Acknowledgments

The authors acknowledge ICAR sponsored project "Incentivizing research in agriculture" for funding and ICAR-Indian Agricultural Research Institute, New Delhi for providing all 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.2022.1038881/full#supplementary-material

References

Abdullah, M., Cheng, X., Cao, Y., Su, X., Manzoor, M. A., Gao, J., et al. (2018). Zinc finger-homeodomain transcriptional factors (ZHDs) in upland cotton (*Gossypium hirsutum*): Genome-wide identification and expression analysis in fiber development. *Front. Genet.* 9. doi: 10.3389/fgene.2018.00357

Abu-Romman, S. (2014). Molecular cloning and expression analysis of zinc finger-homeodomain transcription factor *TaZFHD1* in wheat. *South. Afr. J. Bot.* 91, 32–36. doi: 10.1016/j.sajb.2013.11.014

Alaux, M., Rogers, J., Letellier, T., Flores, R., Alfama, F., Pommier, C., et al. (2018). Linking the international wheat genome sequencing consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome. Biol.* 19, 111. doi: 10.1186/s13059-018-1491-4

Alvarez, J. M., Rocha, J. F., and Machado, S. R. (2008). Bulliform cells in *Loudetiopsis chrysothrix* (Nees) conert and *Tristachya leiostachya* nees (Poaceae): Structure in relation to function. *Braz. Arch. Biol. Technol.* 51 (1), 113–119. doi: 10.1590/S1516-89132008000100014

Ariel, F. D., Manavella, P. A., Dezar, C. A., and Chan, R. L. (2007). The true story of the HD-zip family. *Trends Plant Sci.* 12 (9), 419–426. doi: 10.1016/j.tplants.2007.08.003

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* 43 (W1), W39–W49. doi: 10.1093/nar/gkv416

Baldoni, E., Genga, A., and Cominelli, E. (2015). Plant MYB transcription factors: Their role in drought response mechanisms. *Int. J. Mol. Sci.* 16 (7), 15811–15851. doi: 10.3390/ijms160715811

Bhattacharjee, A., and Jain, M. (2013). "Homeobox genes as potential candidates for crop improvement under abiotic stress," in *Plant acclimation to environmental stress*. Eds. N. Tuteja and S. S. Gill (New York: Springer), 163–176. doi: 10.1007/978-1-4614-5001-6_7

Bolser, D., Staines, D. M., Pritchard, E., and Kersey, P. (2016). Ensembl plants: Integrating tools for visualizing mining and analyzing plant genomics data. *Methods Mol. Biol.* 1374, 115–140. doi: 10.1007/978-1-4939-3167-5_6

Borrill, P., Ramirez-Gonzalez, R., and Uauy, C. (2016). expVIP: A customizable RNA-seq data analysis and visualization platform. *Plant Physiol.* 170 (4), 2172–2186. doi: 10.1104/pp.15.01667

Buels, R., Yao, E., Diesh, C. M., Hayes, R.D., Munoz-Torres, M., Helt, G., et al. (2016). JBrowse: A dynamic web platform for genome visualization and analysis. *Genome Biol.* 17, 66. doi: 10.1186/s13059-016-0924-1

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

Chandra, A. K., Pandey, D., Tiwari, A., Sharma, D., Agarwal, A., Sood, S., et al. (2020). An omics study of iron and zinc homeostasis in finger millet: Biofortified foods for micronutrient deficiency in an era of climate change? *OMICS: A J. Int. Biol.* 24 (12), 1–18. doi: 10.1089/omi.2020.0095

Chandra, A. K., Pandey, D., Tiwari, A., Gururani, K., Agarwal, A., Dhasmana, A., et al. (2021). Metal based nanoparticles trigger the differential expression of key regulatory genes which regulate iron and zinc homeostasis mechanism in finger millet. *J. cer. Sci.* 100, 103235. doi: 10.1016/j.jcs.2021.103235

Dai, X., Zhaohong, Z., and Patrick, X. Z. (2018). psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res* 46, W1, W49–W54. doi: 10.1093/nar/gky316

Deng, X., Phillips, J., Meijer, A. H., Salamini, F., and Bartels, D. (2002). Characterization of five novel dehydration-responsive homeodomain leucine zipper genes from the resurrection plant *C. plantagineum*. *Plant Mol. Biol.* 49, 601–610. doi: 10.1023/A:1015501205303

Di Tommaso, P., Sebastien, M., Ioannis, X., Orobitg, M., Montanyola, A., Chang, J.M., et al. (2011). T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* 39, 2, W13–W17. doi: 10.1093/nar/gkr245

Dong, J., Zhao, J., Zhang, S., Yang, T., Liu, Q., Mao, X., et al. (2019). Physiological and genome-wide gene expression analyses of cold-induced leaf rolling at the seedling stage in rice (*Oryza sativa* l). *Crop J.* 7 (4), 431–443. doi: 10.1016/j.cj.2019.01.003

Duan, L., Han, J., Guo, Z., Tu, H., Yang, P., Zhang, D, et al. (2018). Novel digital features discriminate between drought resistant and drought sensitive rice under controlled and field conditions. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00492

Ezer, D., Shepherd, S.J., Brestovitsky, A., Dickinson, P., Cortijo, S., Charoensawan, V., et al. (2017). The G-box transcriptional regulatory code in arabidopsis. *Plant Physiol.* 175, 2, 628–640. doi: 10.1104/pp.17.01086

Fang, J., Guo, T., Xie, Z., Chun, Y., Zhao, J., Peng, L., et al. (2021). The URL1-ROC5-TPL2 transcriptional repressor complex represses the *ACL1* gene to modulate leaf rolling in rice. *Plant Physiol.* 185 (4), 1722–1744. doi: 10.1093/plphys/kiaa121

Fujino, K., Matsuda, Y., Ozawa, K., Nishimura, T., Koshiba, T., Fraaije,, et al. (2008). *NARROW LEAF 7* controls leaf shape mediated by auxin in rice. *Mol. Genet. Genomics* 279 (5), 499–507. doi: 10.1007/s00438-008-0328-3

Gao, L., Yang, G., Li, Y., Fan, N., Li, H., Zhang, M., et al. (2019). Fine mapping and candidate gene analysis of a QTL associated with leaf rolling index on chromosome 4 of maize (*Zea mays* l). *Theor. Appl. Genet.* 132 (11), 3047–3062. doi: 10.1007/s00122-019-03405-1

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). "Protein identification and analysis tools on the ExPASy server," in *The proteomics protocols handbook springer protocols handbooks*. Ed. J. M. Walker (New Jersey, USA: Humana Press). doi: 10.1385/1-59259-890-0:571

Guan, J., Garcia, D., Zhou, Y., Appels, R., Li Mao, A. L., et al. (2020). The battle to sequence the bread wheat genome: A tale of the three kingdoms. *Gen. Prot. Bio.* 18 (3), 221–229. doi: 10.1016/j.gpb.2019.09.005

Gubert, C. M., Christy, M. E., Ward, D. L., Groner, W.D., Liljegren, S.J., et al. (2014). *ASYMMETRIC LEAVES1* regulates abscission zone placement in *Arabidopsis* flowers. *BMC Plant Biol.* 14, 195. doi: 10.1186/s12870-014-0195-5

Guindon, S., Dufayard, J., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., et al. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biol.* 59, 3, 307–321. doi: 10.1093/sysbio/syq010

Guo, T., Wang, D., Fang, J., Zhao, J., Yuan, S., Xiao, L., et al. (2019). Mutations in the rice *OsCHR4* gene encoding a CHD3 family chromatin remodeler induce narrow and rolled leaves with increased cuticular wax. *Int. J. Mol. Sci.* 20 (10), 2567. doi: 10.3390/ijms20102567

Hoffmann, A. A., and Hercus, M. J. (2000). Environmental stress as an evolutionary force. *BioScience* 50, 3, 217–226. doi: 10.1641/0006-3568(2000)050[0217:esaaef]2.3.co;2

Huang, J., Li, Z., and Zhao, D. (2016). Deregulation of the *OsmiR160* target gene *OsARF18* causes growth and developmental defects with an alteration of auxin signaling in rice. *Sci. Rep.* 6, 29938. doi: 10.1038/srep29938

Hu, W., dePamphilis, C. W., and Ma, H. (2008). Phylogenetic analysis of the plant-specific zinc finger-homeobox and mini zinc finger gene families. *J. Integr. Plant Biol.* 50 (8), 1031–1045. doi: 10.1111/j.1744-7909.2008.00681.x

Hudson, N. O., and Buck-Koehntop, B. A. (2018). Zinc finger readers of methylated DNA. *Molecules* 23 (10), 2555. doi: 10.3390/molecules23102555

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., Gao, G., et al. (2015). GSDS 20: An upgraded gene feature visualization server. *Bioinformatics.* 31 (8), 1296–1297. doi: 10.1093/bioinformatics/btu817

Irene, T. R., Sykes, K. F., Rayner, S., and Johnston, S. A. (2002). ORF-FINDER: A vector for high-throughput gene identification. *Gene* 282 (1-2), 33–41. doi: 10.1016/s0378-1119(01)00819-8

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

Khatun, K., Nath, U. K., Robin, A. H. K., Park, J.I., Lee, D.J., Kim, M.B., et al. (2017). Genome-wide analysis and expression profiling of zinc finger homeodomain (ZHD) family genes reveal likely roles in organ development and stress responses in tomato. *BMC Genomics* 18, 695. doi: 10.1186/s12864-017-4082-y

Kim, S. G., Kim, S. Y., and Park, C. M. (2007). A membrane-associated NAC transcription factor regulates salt-responsive flowering *via FLOWERING LOCUS t* in arabidopsis. *Planta.* 226, 647–654. doi: 10.1007/s00425-007-0513-3

Kinsella, R. J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., et al. (2011). Ensembl BioMarts: A hub for data retrieval across taxonomic space. *Database: J. Bio. Data. Curat* 2011, bar030. doi: 10.1093/database/bar030

Kou, H. P., Li, Y., Song, X. X., Ou, X. F., Xing, S. C., Ma, J., et al. (2011). Heritable alteration in DNA methylation induced by nitrogen-deficiency stress accompanies enhanced tolerance by progenies to the stress in rice (*Oryza sativa* l). *J. Plant Physiol.* 168 (14), 1685–1693. doi: 10.1016/j.jplph.2011.03.017

Laloum, T., De Mita, S., Gamas, P., Baudin, M., and Niebel, A. (2013). CCAAT-box binding transcription factors in plants: Y so many? *Trends Plant Sci.* 18 (3), 157–166. doi: 10.1016/j.tplants.2012.07.004

Liebsch, D., and Palatnik, J. F. (2020). MicroRNA miR396 GRF transcription factors and GIF co-regulators: a conserved plant growth regulatory module with potential for breeding and biotechnology. *Curr. Opin. Plant Biol.* 53, 31–42. doi: 10.1016/j.pbi.2019.09.008

Lindsey, L. E., Paul, P., and Lentz, E. (2017). *Wheat growth stages and associated management* (Columbus, Ohio, USA: Ohio State University Fact Sheet). Available at: https://ohioline.osu.edu/factsheet/agf-126. AGF-126.

Liu, M., Wang, X., Sun, W., Ma, Z., Zheng, T., Huang, L., et al. (2019). Genome-wide investigation of the ZF-HD gene family in tartary buckwheat (*Fagopyrum tataricum*). *BMC Plant Biol.* 19, 248. doi: 10.1186/s12870-019-1834-7

Liu, H., Yang, Y., and Zhang, L. (2021). Zinc finger-homeodomain transcriptional factors (ZF-HDs) in wheat (*Triticum aestivum* l): Identification evolution expression analysis and response to abiotic stresses. *Plants.* 10 (3), 593. doi: 10.3390/plants10030593

Li, Y., Yang, Y., Liu, Y., Li, D., Zhao, Y., Li, Z., et al. (2019). Overexpression of *OsAGO1b* induces adaxially rolled leaves by affecting leaf abaxial sclerenchymatous cell development in rice. *Rice.* 12, 60. doi: 10.1186/s12284-019-0323-9

Li, C., Zou, X., Zhang, C., Shao, Q., Liu, J., Liu, B., et al. (2016). *OsLBD3-7* overexpression induced adaxially rolled leaves in rice. *PloS One* 11 (6), e0156413. doi: 10.1371/journal.pone.0156413

Mohanty, B., Herath, V., Wijaya, E., Yeo, H.C., Benildo, G., Lee, D.Y., et al. (2012). Patterns of cis-element enrichment reveal potential regulatory modules involved in the transcriptional regulation of anoxia response of japonica rice. *Gene.* 511 (2), 235–242. doi: 10.1016/j.gene.2012.09.048

Moon, J., and Hake, S. (2011). How a leaf gets its shape. *Curr. Opin. Plant Biol.* 14 (1), 24–30. doi: 10.1016/j.pbi.2010.08.012

Moore, R. C., and Purugganan, M. D. (2003). The early stages of duplicate gene evolution. *PNAS* 100 (26), 15682–15687. doi: 10.1073/pnas.2535513100

Ohmori, Y., Toriba, T., Nakamura, H., Ichikawa, H., and Hirano, H.Y.. (2011). Temporal and spatial regulation of *DROOPING LEAF* gene expression that promotes midrib formation in rice. *Plant J.* 65 (1), 77–86. doi: 10.1111/j.1365-313X.2010.04404.x

Panchy, N., Lehti-Shiu, M., and Shiu, S. H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171 (4), 2294–2316. doi: 10.1104/pp.16.00523

Peleg, Z., Fahima, T., Krugman, T., Abbo, S., Yakir, D.A.N., Korol, A.B., et al. (2009). Genomic dissection of drought resistance in durum wheat x wild emmer wheat recombinant inbreed line population. *Plant Cell Environ.* 32 (7), 758–779. doi: 10.1111/j.1365-3040.2009.01956.x

Perrella, G., Davidson, M., O'Donnell, L., Nastase, A. M., Herzyk, P., Breton, G., et al. (2018). ZINC-FINGER interactions mediate transcriptional regulation of hypocotyl growth in arabidopsis. *PNAS* 115 (19), E4503–E4511. doi: 10.1073/pnas.1718099115

Portwood, J. L., Woodhouse, M. R., Cannon, E. K., Gardiner, J. M., Harper, L.C., Schaeffer, M.L., et al. (2019). MaizeGDB 2018: The maize multi-genome genetics and genomics database. *Nucleic Acids Res.* 47 (D1), D1146–D1154. doi: 10.1093/nar/gky1046

Sakaguchi, J., Jun-Ichi, I., Yukihiro, I., Nakamura, A., Fukuda, H., and Sawa, S.. (2010). *COE1* an LRR–RLK responsible for commissural vein pattern formation in rice. *Plant J.* 63 (3), 405–416. doi: 10.1111/j.1365-313X.2010.04250.x

Sakai, H., Lee, S. S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., et al. (2013). Rice annotation project database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol.* 54 (2), e6. doi: 10.1093/pcp/pcs183

Sakamoto, T., Morinaka, Y., Ohnishi, T., Sunohara, H., Fujioka, S., Ueguchi-Tanaka, M., et al. (2006). Erect leaves caused by brassinosteroid deficiency increase biomass production and grain yield in rice. *Nat. Biotechnol.* 24, 105–109. doi: 10.1038/nbt1173

Saraswat, S., Yadav, A. K., Sirohi, P., and Singh, N. K. (2017). Role of epigenetics in crop improvement: Water and heat stress. *J. Plant Biol.* 60, 231–240. doi: 10.1007/s12374-017-0053-8

Schmittgen, T., and Livak, K. (2008). Analyzing real-time PCR data by the comparative C_T method. *Nat. Protoc.* 3, 1101–1108. doi: 10.1038/nprot.2008.73

Shi, Z., Wang, J., Wan, X., Shen, G., Wang, X., Zhang, J., et al. (2007). Over-expression of rice *OsAGO7* gene induces upward curling of the leaf blade that enhanced erect-leaf habit. *Planta.* 226, 99–108. doi: 10.2307/23389653

Singh, D., and Laxmi, A. (2015). Transcriptional regulation of drought response: A tortuous network of transcriptional factors. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00895

Sirault, X. R. R., Condon, A. G., Rebetzke, G. J., and Farquhar, G. D. (2008) *Genetic analysis of leaf rolling in wheat*. Available at: https://seslibraryusydeduau/handle/2123/3213.

Tran, L. S., Nakashima, K., Sakuma, Y., Osakabe, Y., Qin, F., and Simpson, S.D.. (2007). Co-Expression of the stress-inducible zinc finger homeodomain ZFHD1 and NAC transcription factors enhances expression of the *ERD1* gene in arabidopsis. *Plant J.* 49 (1), 46–63. doi: 10.1111/j.1365-313X.2006.02932.x

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, e115–e115. doi: 10.1093/nar/gks596

Verma, A., Niranjana, M., Jha, S. K., Mallick, N., Agarwal, P., and Vinod,. (2020). QTL detection and putative candidate gene prediction for leaf rolling under moisture stress condition in wheat. *Sci. Rep.* 10, 18696. doi: 10.1038/s41598-020-75703-4

Vial-Pradel, S., Keta, S., Nomoto, M., Luo, L., Takahashi, H., Suzuki, M., et al. (2018). Arabidopsis zinc-Finger-Like protein ASYMMETRIC LEAVES2 (AS2) and two nucleolar proteins maintain gene body DNA methylation in the leaf polarity gene *ETTIN* (*ARF*3). *Plant Cell Physiol.* 59, 7, 1385–1397. doi: 10.1093/pcp/pcy031

Wang, W., Wu, P., Li, Y., and Hou, X. (2016). Genome-wide analysis and expression patterns of ZF-HD transcription factors under different developmental tissues and abiotic stresses in Chinese cabbage. *Mol. Genet. Genomics* 291 (3), 1451–1464. doi: 10.1007/s00438-015-1136-1

Warde-Farley, D., Sylva, L., Donaldson, O., Comes, O., Zuberi, K., Badrawi, R., et al. (2010). The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 38, 2, W214–W220. doi: 10.1093/nar/gkq537

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res* 46, W1, W296–W303. doi: 10.1093/nar/gky427

Wei, Q., Chen, R., Wei, X., Liu, Y., Zhao, S., Yin, X., et al. (2020). Genome-wide identification of R2R3-MYB family in wheat and functional characteristics of the abiotic stress responsive gene TaMYB344. *BMC Genomics* 21, 792. doi: 10.1186/s12864-020-07175-9

Windhovel, A., Hein, I., Dabrowa, R., and Stockhaus, J. (2001). Characterization of a novel class of plant homeodomain proteins that bind to the C_4 phosphoenolpyruvate carboxylase gene of *Flaveria trinervia*. *Plant Mol. Biol.* 45, 201–214. doi: 10.1023/A:1006450005648

Wu, L., Zhang, Q., Zhou, H., Ni, F., Wu, X., and Qi, Y. (2009). Rice microRNA effector complexes and targets. *Plant Cell.* 21, 3421–3435. doi: 10.1105/tpc.109.070938

Xiang, J., Zhang, G., Qian, Q., and Xue, H. (2012). *SEMI-ROLLED LEAF1* encodes a putative glycosylphosphatidylinositol-anchored protein and modulates rice leaf rolling by regulating the formation of bulliform cells. *Plant Physiol.* 159, 1488–1500. doi: 10.1104/pp.112.199968

Xu, P., Ali, A., Han, B., and Wu, X. (2018). Current advances in molecular basis and mechanisms regulating leaf morphology in rice. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01528

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

Xu, F., Park, M. R., Kitazumi, A., Herath, V., Mohanty, B., Yun, S.J., et al. (2012). Cis-regulatory signatures of orthologous stress-associated bZIP transcription factors from rice, sorghum and arabidopsis based on phylogenetic footprints. *BMC Genomics* 13, 497. doi: 10.1186/1471-2164-13-497

Xu, Y., Wang, Y., Long, Q., Huang, J., Wang, Y., and Zhou, K.. (2014). Overexpression of *OsZHD1* a zinc finger homeodomain class homeobox transcription factor induces abaxially curled and drooping leaf in rice. *Planta.* 239 (4), 803–816. doi: 10.1007/s00425-013-2009-7

Yang, S. Q., Li, W. Q., Miao, H., Gan, P. F., Qiao, L., and Chang, Y. L.. (2016). REL2, a gene encoding an unknown function protein which contains DUF630 and DUF632 domains controls leaf rolling in rice. *Rice* 9 (1), 37. doi: 10.1186/s12284-016-0105-6

Yong-Villalobos, L., González-Morales, S. I., Wrobel, K., Gutiérrez-Alanis, D., Cervantes-Peréz, S. A., Hayano-Kanashiro, C., et al. (2015). Methylome analysis reveals an important role for epigenetic changes in the regulation of the arabidopsis response to phosphate starvation. *PNAS.* 112 (52), E7293–E7302. doi: 10.1073/pnas.1522301112

Yoon, J., Cho, L. H., Yang, W., Pasriga, R., Wu, Y., Hong, W. J., et al. (2020). Homeobox transcription factor *OsZHD2* promotes root meristem activity in rice by inducing ethylene biosynthesis. *J. Exp. Bot.* 71 (18), 5348–5364. doi: 10.1093/jxb/eraa209

Yun, K. Y., Park, M. R., Mohanty, B., Herath, V., Xu, F., Mauleon, R., et al. (2010). Transcriptional regulatory network triggered by oxidative signals configures the early response mechanisms of japonica rice to chilling stress. *BMC Plant Biol.* 10, 16. doi: 10.1186/1471-2229-10-16

Zhang, G. H., Xu, Q., Zhu, X. D., Qian, Q., and Xue, H. W. (2009). *SHALLOT-LIKE1* is a KANADI transcription factor that modulates rice leaf rolling by regulating leaf abaxial cell development. *Plant Cell* 21 (3), 719–735. doi: 10.1105/tpc.108.061457

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

Zhu, T., Wang, L., Rimbert, H., Rodriguez, J. C., Deal, K.R., De Oliveira, R., et al. (2021). Optical maps refine the bread wheat *Triticum aestivum* cv Chinese spring genome assembly. *Plant J. Cell Mol. Boil.* 107 (1), 303–314. doi: 10.1111/tpj.15289

Zhu, Q., Yu, S., Chen, G., Ke, L., and Pan, D. (2017). Analysis of the differential gene and protein expression profile of the rolled leaf mutant of transgenic rice (*Oryza sativa* l). *PloS One* 12 (7), e0181378. doi: 10.1371/journal.pone.0181378

Zou, L., Sun, X., Zhang, Z., Liu, P., Wu, J.X., Tian, C.J., et al. (2011). Leaf rolling controlled by the homeodomain leucine zipper class IV gene *Roc5* in rice. *Plant Physiol.* 156, 1589–1602. doi: 10.1104/pp.111.176016

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 December 2022
DOI 10.3389/fpls.2022.1053524

OPEN ACCESS

EDITED BY
Bhanu Prakash Petla,
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

REVIEWED BY
Venura Herath,
University of Peradeniya, Sri Lanka
Harmeet Kaur,
National Institute for Plant Biotechnology, Indian Council of Agricultural Research, India
Bharati Pandey,
Panjab University, India

*CORRESPONDENCE
Prabhjeet Singh
singhprabhjeet62@gmail.com

†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 25 September 2022
ACCEPTED 23 November 2022
PUBLISHED 15 December 2022

CITATION
Suri A, Singh H, Kaur K, Kaachra A and Singh P (2022) Genome-wide characterization of FK506-binding proteins, parvulins and phospho-tyrosyl phosphatase activators in wheat and their regulation by heat stress. *Front. Plant Sci.* 13:1053524. doi: 10.3389/fpls.2022.1053524

Genome-wide characterization of FK506-binding proteins, parvulins and phospho-tyrosyl phosphatase activators in wheat and their regulation by heat stress

Anantika Suri[1], Harpreet Singh[2], Kirandeep Kaur[1†], Anish Kaachra[3†] and Prabhjeet Singh[1*]

[1]Department of Biotechnology, Guru Nanak Dev University, Amritsar, India, [2]Department of Bioinformatics, Hans Raj Mahila Maha Vidyalaya, Jalandhar, India, [3]Biotechnology Division, Institute of Himalayan Bioresource Technology, Council of Scientific and Industrial Research, Palampur, HP, India

Peptidyl-prolyl *cis-trans* isomerases (PPIases) are ubiquitous proteins which are essential for *cis-trans* isomerisation of peptide bonds preceding the proline residue. PPIases are categorized into four sub-families *viz.*, cyclophilins, FK506-binding proteins (FKBPs), parvulins and protein phosphatase 2A phosphatase activators (PTPAs). Apart from catalysing the *cis-trans* isomerization, these proteins have also been implicated in diverse cellular functions. Though PPIases have been identified in several important crop plants, information on these proteins, except cyclophilins, is scanty in wheat. In order to understand the role of these genes in wheat, we carried out genome-wide identification using computational approaches. The present study resulted in identification of 71 FKBP (*TaFKBP*) 12 parvulin (*TaPar*) and 3 PTPA (*TaPTPA*) genes in hexaploid wheat genome, which are distributed on different chromosomes with uneven gene densities. The TaFKBP and TaPar proteins, besides PPIase domain, also contain additional domains, indicating functional diversification. *In silico* prediction also revealed that TaFKBPs are localized to ER, nucleus, chloroplast and cytoplasm, while the TaPars are confined to cytoplasm and nucleus. The TaPTPAs, on the contrary, appear to be present only in the cytoplasm. Evolutionary studies predicted that most of the *TaFKBP, TaPar* and *TaPTPA* genes in hexaploid wheat have been derived from their progenitor species, with some events of loss or gain. Syntenic analysis revealed the presence of many collinear blocks of *TaFKBP* genes in wheat and its sub-genome donors. qRT-PCR analysis demonstrated that expression of *TaFKBP* and *TaPar* genes is regulated differentially by heat stress, suggesting their likely involvement in thermotolerance. The findings of this study will provide basis for further functional characterization of these genes and their likely applications in crop improvement.

KEYWORDS

FK506-binding proteins, parvulins, phospho-tyrosyl phosphatase activators, heat stress, wheat

Introduction

The plants cope up with different biotic and abiotic stresses through adjustments at physiological, biochemical and molecular levels, that include the production of antioxidative enzymes, accumulation of compatible solutes, signaling molecules, chaperones, etc. (Wang et al., 2003). The peptidyl-prolyl *cis-trans* isomerases (PPIases) are yet another class of proteins that have also been reported to play an important role in stress adaptation of plants (Sharma and Singh, 2003; Wang et al., 2003; Singh et al., 2019; Singh H et al., 2020; Thirumalaikumar et al., 2021). The PPIases are the only enzymes known that can catalyze the *cis-trans* transition of peptidyl-prolyl bonds. Though the peptide bonds in proteins occur in *trans* conformation, about 6.5% of the peptidyl-prolyl bonds occur in *cis* state due to the five-membered ring structure of prolyl residue (Galat and Rivière, 1998). The *cis-trans* transition of the peptidyl-prolyl bond is a rate-limiting step in correct folding of proteins and, hence, requires intervention of PPIases (Fischer et al., 1989) which are typical enzymes and follow the Michaelis-Menton kinetics. Depending upon their sensitivity towards immunosuppressive drugs, the PPIases are categorised into two sub-families, immunophilins and non-immunophilins. While the immunophilins comprise of cyclosporin A (CsA)-binding proteins, cyclophilins, and FK506-binding proteins, FKBPs, the non-immunophilin PPIases include structurally distinct parvulins and phospho-tyrosyl phosphatase activators (PTPAs), that are not sensitive to any of these immunosuppressive agents (Rahfeld et al., 1994). The cyclophilins and FKBPs, characterized by the presence of cyclophilin-like domain (CLD) and FK506-binding domain (FKBD), respectively, are ubiquitous proteins and have been reported in almost all organisms ranging from microbes to plants and animals (Göthel and Marahiel, 1999; Singh H et al., 2020; Yadav et al., 2022; Zhou et al., 2022). These proteins are localized in different sub-cellular organelles, *viz*., mitochondria, endoplasmic reticulum, cytosol and nucleus (Breiman et al., 1992; Luan et al., 1996; Xu et al., 1998; Carol et al., 2001). The cyclophilins and FKBPs are encoded by large gene families in plants. Genome-wide computational analysis revealed the presence of 21, 23, 24, 29, 30 and 38 different FKBPs in *Prunus persica*, *Arabidopsis thaliana*, *Solanum lycopersicum*, *Oryza sativa*, *Zea mays*, and *Malus domestica*, respectively (He et al., 2004; Gollan and Bhave, 2010; Wang W et al., 2012; Zhang Y et al., 2014; Dong et al., 2018; Waseem et al., 2018). Members of FKBPs have also been characterized in different fungal species such as *Aspergillus* spp. and *Penicillium* spp. (Joseph et al., 1999; Singh et al., 2021). Likewise, 28, 29, 30, 33, 38-78, 62, 83 and 91 cyclophilins have been reported in *O. sativa*, *A. thaliana*, *M. domestica*, *Medicago truncatula*, *Gossypium* spp., soybean, *Triticum aestivum* and *Brassica napus*, respectively (Gasser et al., 1990; Romano et al., 2004; Trivedi et al., 2012; Mainali et al., 2014; Hanhart et al., 2017; Chen et al., 2019; Singh et al., 2019; Ge et al., 2020; Wang et al., 2022). Contrary to the cyclophilins and FKBPs, the PTPAs and parvulins are less abundant and only one or two members have been reported in human, yeast and *Penicillium* spp. (Rahfeld et al., 1994; Lu et al., 1996; Jordens et al., 2006; Magnusdottir et al., 2006; Mueller et al., 2006; Singh et al., 2021).

In mammals, the cyclophilins and FKBPs are involved in immunosuppression, as interaction of these proteins with CsA and FK506, respectively, prevent transcription of genes encoding interleukin-2 (Liu et al., 1991; O'Keefe et al., 1992), leading to suppression of immune response. The immunophilins also regulate important cellular processes in plants, with the roles of cyclophilins discussed extensively in a recent review (Singh H et al., 2020 and references therein). Recently, an *Arabidopsis* cyclophilin, AtCYP18-1, was reported to play an important role in splicing of introns that are retained in response to heat stress during germination (Jo et al., 2022). Cyclophilin *ROC3* was implicated in regulation of ABA-induced stomatal closure and drought stress response in *Arabidopsis* (Liu et al., 2021). Similarly, ectopic expression of a pigeonpea cyclophilin (*CcCYP*) was shown to confer tolerance against multiple abiotic stresses in transgenic rice (Juturu et al., 2021). The FKBPs too have been shown to perform diverse cellular functions *viz*., signal transduction (Gopalan et al., 2004), transcription (Kang et al., 2008), assembly of multiprotein complexes (Geisler and Bailly, 2007; Nigam et al., 2008; Boudko et al., 2014). protein trafficking (Luan et al., 1994), apoptosis (Meiri et al., 2010), fertility (Yu et al., 2012), biosynthesis of long chain fatty acids (Harrar et al., 2003), regulation of gene expression (Li and Luan, 2010), photosynthetic membrane assembly in plants (Lima et al., 2006) and as histone chaperones (Singh A et al., 2020). Several FKBPs also exhibit chaperonic activity that is independent of PPIase activity. For example, one of the wheat FKBPs, FKBP73, exihbited chaperonic function despite abrogation of its PPIase activity (Kurek et al., 2002). Expression of *FKBP* genes, *FKBP62* and *FKBP65*, was enhanced in response to wounding, NaCl and malondialdehyde treatment in *Arabidopsis* (Vucich and Gasser, 1996; Weber et al., 2004; Aviezer-Hagai et al., 2007; Meiri and Breiman, 2009). Yu et al. (2017) demonstrated that constitutive expression of a *Z. mays FKBP* gene, *ZmFKBP20-1*, resulted in higher tolerance to drought and salt stress in the transgenic *Arabidopsis* plants. The role of a rice FKBP, OsFKBP20-1b, in stress adaptation was attributed to regulation of RNA processing through its interaction with a splicing factor OsSR45 (Park et al., 2020).

It is, thus, evident that different PPIases play important roles in growth and development processes of plants as well as in adaptation to different abiotic stresses. Earlier studies by our group demonstrated that the cyclophilin family in wheat consists of 83 genes, several of which are modulated under heat stress (Singh et al., 2019). However, except for cyclophilins (Singh et al., 2019), information on PPIases belonging to FKBPs,

parvulins and PTPAs is scanty in this important crop plant. Therefore, the present study was undertaken to carry out *in silico* identification and characterization of FKBPs, parvulins and PTPAs in wheat and to study their regulation by thermal stress. The results of this study revealed that the wheat genome encodes for 71 FKBPs, 12 parvulins, and 3 PTPAs, and expression of several of these genes is modulated differentially by heat stress. These findings will provide impetus for further functional characterization of these gene families, and their likely applications in crop improvement through biotechnological and breeding strategies.

Materials and methods

Identification and retrieval of *FKBP*, *parvulin* and *PTPA* gene sequences

The different FKBPs, parvulins, and PTPAs encoded by the wheat genome were identified by using the coding (CDS) and protein sequences of different organisms for homology-based search. For FKBPs, sequences from *O. sativa*, *A. thaliana*, *P. persica* and *M. domestica*; for PTPAs, sequences from *Drosophila melanogaster*, *Homo sapiens*, *Saccharomyces cerevisiae* and *Saccharomyces pombe*; and for parvulins, sequences of *A. thaliana*, *H. sapiens*, *S. cerevisiae* and *Escherichia coli* were employed for this study. The CDS and protein sequences for different organisms were obtained from their respective databases (UniProt; Universal Protein knowledgebase; https://www.uniprot.org/, NCBI; National Centre for Biotechnology Information; https://www.ncbi.nlm.nih.gov/, RGAP; Rice Genome Annotation Project; http://rice.plantbiology.msu.edu/analyses_search_locus.shtml, TAIR; The Arabidopsis Information Resource; https://www.arabidopsis.org/, GDR; Genome Database for Rosaceae; https://www.rosaceae.org/retrieve/sequences) and from other publicly available sources (Supplementary Table S1). BLASTN and BLASTP (Ensembl Plants; https://plants.ensembl.org) programs were used for homology search with an e-value cut off of 10 and 1, respectively. The redundant sequences having the same transcript IDs were removed from the dataset. The matches based on score, coverage and percent identity were retrieved and selected for further analysis. For FKBPs, the retrieved protein sequences were used to build a Hidden Markov model (HMM; HMMER software package; Eddy, 2011). This model was used as a query for running HMM search to obtain more distantly homologous FKBPs from the wheat proteome (IWGSC RefSeq v2.0). Ensembl plants (https://plants.ensembl.org/Triticum_aestivum/Info/Index) was used to download the genomic sequences, CDS, untranslated regions (UTRs), introns, exons, cDNA as well as protein sequences of the corresponding matches.

Conserved domain search and subcellular localization prediction

Domain analysis was performed using PFAM (http://pfam.xfam.org; Finn et al., 2016), Prosite (https://prosite.expasy.org/; Sigrist et al., 2012) and conserved domain database of NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Only those members which displayed the presence of typical functional domain of their respective PPIase family were subjected to additional characterization and investigation. Graphical representation of the domains and motifs was performed using IBS1.0 stand alone program (Liu W et al., 2015). The subcellular localization was predicted by using LocTree3 protein subcellular localization prediction system (https://rostlab.org/services/loctree3/; Goldberg et al., 2014).

Sequence analysis

The deduced amino acid sequences were aligned with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). The molecular weights and pIs of the putative PPIases were predicted by biosynthesis's peptide property calculator web server Version 3.1 (https://www.biosyn.com/peptidepropertycalculator/peptidepropertycalculator.aspx). Amino acid sequences of the respective domains were extracted using EMBOSS: extractseq web interface (bioinformatics.nl/cgi-bin/emboss/extractseq; Olson, 2002). Active site residues of the aligned domain sequences were analysed by employing ESPript 3.0 (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi; Robert and Gouet, 2014).

Gene structure, motif analysis and phylogenetic analysis of FKBPs, parvulins, and PTPAs

Gene structures were analyzed using GSDS 2.0 server (http://gsds.cbi.pku.edu.cn/; Hu et al., 2015). Conserved motifs of the protein sequences were identified with MEME Suite (https://meme-suite.org/meme/; Bailey et al., 2015), and edited using TBtools (https://github.com/CJ-Chen/TBtools; Chen et al., 2020). Mega X was employed to construct phylogenetic trees by Neighbor Joining (NJ) method using 1000 bootstrap values. Respective protein sequences of FKBPs, parvulins and PTPAs from *T. aestivum*, *A. thaliana*, *O. sativa*, *Aegilops tauschii*, *Triticum dicoccoides* and *Triticum urartu* were used for the construction of phylogenetic trees. Constructed phylogenetic trees were subsequently edited using Interactive Tree of Life web server (https://itol.embl.de/upload.cgi).

Chromosome mapping and syntenic analysis

Predicted wheat *FKBP*, *parvulin* and *PTPA* genes were mapped to the corresponding chromosomes using MapChart2.32 (Voorrips, 2002). The map obtained was further processed with InKscape (https://inkscape.org/) to highlight various features. Gene duplication events between *T. aestivum*, its progenitors (*T. urartu*, *T. dicoccoides* and *A. tauschii*), *O. sativa* and *A. thaliana* were analyzed with MCScanX (Multiple Collinearity Scan toolkit, https://github.com/wyp1125/MCScanX; Wang Y et al., 2012) keeping threshold values of <1e-4 and 5 for e-value and match size, respectively. The tandem gene duplications and collinear blocks identified with MCScanX were plotted using shinyCircos online application (https://venyao.shinyapps.io/shinyCircos/; Yu et al., 2017). Dual synteny plot was generated by TBtools software (https://github.com/CJ-Chen/TBtools; Chen et al., 2020).

Phylogenetic analysis

Phylogenetic tree was generated to understand the evolutionary relationship among different FKBP, parvulin and PTPA proteins. Alignment of the amino acid sequences of all the proteins was carried out by Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/; Sievers et al., 2011). The phylogenetic tree was generated with MEGA X (Molecular Evolutionary Genetics Analysis) across computing platforms (Kumar et al., 2018) using Neighbour-Joining method with 1000 bootstrap replicates. The tree was exported to the Interactive Tree of Life (https://itol.embl.de/; Letunic and Bork, 2016) for subsequent processing and annotation.

Identification of cis-regulatory elements

The upstream 2 kb promoter regions of the gene sequences were retrieved from Ensembl plants database and used as inputs to search cis-regulatory elements in the database (HSEAT, Heat Shock Element Analysis Tool; https://sourceforge.net/projects/heast/).

Effect of heat stress on expression of *FKBP*, *parvulin* and *PTPA* genes

Seeds of wheat (*T. aestivum* L.) cultivar HD2967 were surface sterilized using Tween 20 and subsequently sown in soil, and grown in the culture room at 25 ± 2 °C. For acclimation to heat stress, the seven days old seedlings were exposed to 37°C for 2 h. Heat stress was imposed by incubating the seedlings at lethal temperature of 50°C for 4.5 h with and without acclimation at 37°C for 2 h. Seedlings kept at 25 °C served as experimental control. Healthy seedlings after the treatments were collected at different time intervals and snap frozen in liquid nitrogen until further analysis.

Total RNA from the leaves was isolated using TRIZOL solution (Invitrogen, USA), and reverse transcription was performed with random hexamer primers using superscript III First-strand synthesis system kit following the manufacturer's instructions (Invitrogen, USA). For expression analysis, primers binding to common regions of different homoeologs were designed with Primer-BLAST tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome). The quantitative real-time PCR (qRT-PCR) was performed on Ariamx real-time PCR system (Agilent Technologies, USA) with Brillant III ultra-Fast SYBR green QPCR master mix (Agilent Technologies, USA). Expression analysis was carried out in three independent biological replicates. *Actin* was used as an internal control for normalization and the fold change was determined using $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). The data obtained were subjected to one-way analysis of variance (ANOVA; graphpad prism).

For digital expression analysis (DGE), expression pattern of *TaFKBP* and *TaPar* genes was determined using WheatExp containing homoelogue-dependent gene expression profiles for wheat (https://wheat.pw.usda.gov/WheatExp/; Pearce et al., 2015). Protein sequences were searched against WheatExp database using tblastn and the matched hits were retrieved. RNAseq expression data from a previous work (Liu Z et al., 2015) was used as a reference for DGE analysis of the selected PPIases. The heat map of normalized (log10) expression values was generated using heatmap function from the NMF CRAN library (https://cran.r-project.org/web/packages/NMF/index.html) in RStudio version 4.0.2 (http://www.rstudio.com/; Myers et al., 2020). Hierarchical clustering of rows was performed with Euclidean distance and complete linkage.

Protein-protein interaction network

The protein sequences of TaFKBPs, TaPars and TaPTPAs were submitted to STRING (v11.5) to predict the relationships between themselves and with other proteins (Szklarczyk et al., 2019) based on orthologs from *A. thaliana*. The parameter "max number of interactors to show" was set to "no more than 10 interactors" while the interacting partners were subjected to k-means clustering to obtain 3 clusters. The query protein (*T. aestivum*) names were used to label the nodes.

Results

Identification of gene families encoding wheat FKBPs, Parvulins and PTPAs

Bioinformatics analysis of wheat genome using Ensembl plants resulted in identification of 71 *TaFKBP*, 12 *TaPar* and 3

TaPTPA genes (Supplementary Tables S1, S2). The nomenclature adopted in this study is as used earlier for wheat cyclophilins (Singh et al., 2019). The wheat FKBP (TaFKBP), parvulin (TaPar) and PTPA (TaPTPA) proteins were suffixed with their predicted molecular weight and chromosomal localization. Localization analysis revealed that the *TaFKBP* genes are present on all the 21 wheat chromosomes, with the maximum number (8) on chromosome 2B and the lowest, one each, on chromosomes 3B, 4A, 4B and 4D (Figure 1). In contrast, only eight wheat chromosomes showed the presence of *TaPar* genes, with the chromosome 2D harbouring the maximum three members. The chromosomes 3B, 4A, 5A, 5B and 5D were predicted to contain only a single *TaPar* gene (Figure 1). Only a single homeologue triplet of *PTPA* genes was observed in wheat, which were localized on chromosome 7A, 7B and 7D, respectively. *In silico* analysis of one of the *FKBPs, TaFKBP23-6-U*, indicated the localization of this gene in an uncharacterized part of wheat genome, denoted by '*U*' (Figure 1). Though majority of the identified genes were present as triplets due to the three progenitor genomes AA, BB and DD, contributed by the diploid species *T. urartu, Aegilops speltoides* and *A. tauschii*, respectively (Feldman and Levy, 2005), absence or existence of additional copies was also observed for some *TaFKBP* and *TaPar* genes. For example, *TaFKBP29-1-2A, TaFKBP24-2-2B and TaFKBP29-2-2D*, a homeologue triplet localized on chromosome 2A, 2B and 2D, depicted the presence of two additional copies on chromosome 2B (*TaFKBP7-1-2B; TaFKBP16-7-2B*) and one on 2D (*TaFKBP20-6-2D*). The gene encoding TaFKBP7-1-2B appears to be a paralogue containing partial FKBD sequence (missing N-terminal region) having more than 97% identities with TaFKBP29-1-2A, TaFKBP24-2-2B and TaFKBP29-2-2D proteins. Similarly, the *TaPar* genes, *TaPar13-2-2A, TaPar13-3-2B* and *TaPar13-1-2D* also exhibited additional copies on chromosomes 2D (*TaPar12-1-2D*), 3B (*TaPar11-1-3B*) and 4A (*TaPar11-2-4A*).

The number of amino acid residues in TaFKBPs and TaPars varies from 68 to 649, and 99 to 301, respectively, whereas the molecular weights of these sub-families range between 7.59 to 77.10 kDa, and 11.06 to 32.70 kDa, respectively (Supplementary Table S1). Genome-wide analysis revealed that only three PTPAs are encoded by the wheat genome, with TaPTPA41-1-7D* and TaPTPA42-1-7B consisting of 393 amino acid residues each, compared to 394 in TaPTPA42-2-7A. Since the CDS and amino acid sequences of TaPTPA41-1-7D* contain some unannotated residues, its exact molecular weight and pI could not be predicted. The molecular weights of TaPTPA42-1-7B and TaPTPA42-2-7A were computed as 42.46 and 42.57 kDa, respectively. Isoelectric points (pIs) of different TaFKBPs, TaPars and TaPTPAs range between 4.32 to 10.68, 6.15 to 9.92, and 6.9 to 7.1, respectively (Supplementary Table S1). While the TaFKBPs were predicted to localize to different subcellular compartments *viz.*, ER (1), nucleus (6), chloroplast (32) and cytoplasm (32), the TaPars are likely to localize only in cytoplasm (4) and nucleus (8). On the contrary, all the three TaPTPA proteins were predicted to be present only in the cytoplasm.

Chromosome localization and synteny analysis of wheat *TaFKBP*, *TaPar* and *TaPTPA* genes

Gene duplication events are vital for the amplification of gene families and their subsequent evolution in wheat. A tandem duplication event occurs when two or more similar genes are found located within a 200 kb of chromosomal region (Holub, 2001). Duplication events that gave rise to *TaFKBP*, *TaPar* and *TaPTPA* genes were studied in *O. sativa*, *A. thaliana*, *T. aestivum* (AABBDD) and its progenitors, *T. urartu* (AA), *T. dicoccoides* (AABB) and *A. tauschii* (DD) using MCScanX (Supplementary Table S3) (Wang Y et al., 2012). Due to absence of genome sequence of the B sub-genome donor *A. speltoides*, wild emmer wheat (*T. dicoccoides*) was used as the source of this sub-genome. Neither *T. aestivum* nor any of its progenitors showed syntenic relationship for parvulin and the PTPA genes. Syntenic relationships were also not observed for FKBP genes between wheat, *O. sativa*, *A. thaliana* and *T. urartu*. Several tandem duplicated pairs of *FKBP* genes were identified in *T. urartu* (2), *T. dicoccoides* (2), *A. tauschii* (3) and *T. aestivum* (29) (Figure 2A; Supplementary Table S4). Microcollinearity analyses of the genome segments across different progenitor species revealed the presence of several orthologous segments sharing collinear blocks of *FKBP* genes (Figures 2A, B; Supplementary Table S5). *T. aestivum* contains four blocks of *FKBP* genes collinear with each of its two progenitor species, *A. tauschii* and *T. dicoccoides*. In addition, six segmentally duplicated blocks of *FKBP* genes were also identified in *T. aestivum*, compared to a single block in *T. dicoccoides*. These results indicate that tandem as well as segmental gene duplication events might have played an important role in the evolution of *FKBP* gene family in wheat. Furthermore, our results also suggest that the wheat *FKBP* genes are entirely derived from its progenitors since most of the wheat chromosomes, except 1B, 2A, 2B, 2D and 3D, harbor the same number of genes as are present on the corresponding chromosomes of the sub-genome donors (Supplementary Table S6). However, the chromosomes 1B, 2A, 2B, 2D and 3D depicted the presence of higher number of *FKBP* genes relative to the sub-genome donors, indicating enrichment due to duplication events. Another reason for this difference could be the loss of these genes in *T. dicoccoides* and *A. tauschii* during evolution. Relative to *T. dicoccoides* and *T. aestivum*, the difference in *FKBP* genes in *T. urartu* may also be ascribed to similar events. In contrast, the wheat chromosomes 1D, 4B and 5D harbor fewer number of *FKBP* genes than its sub-genome

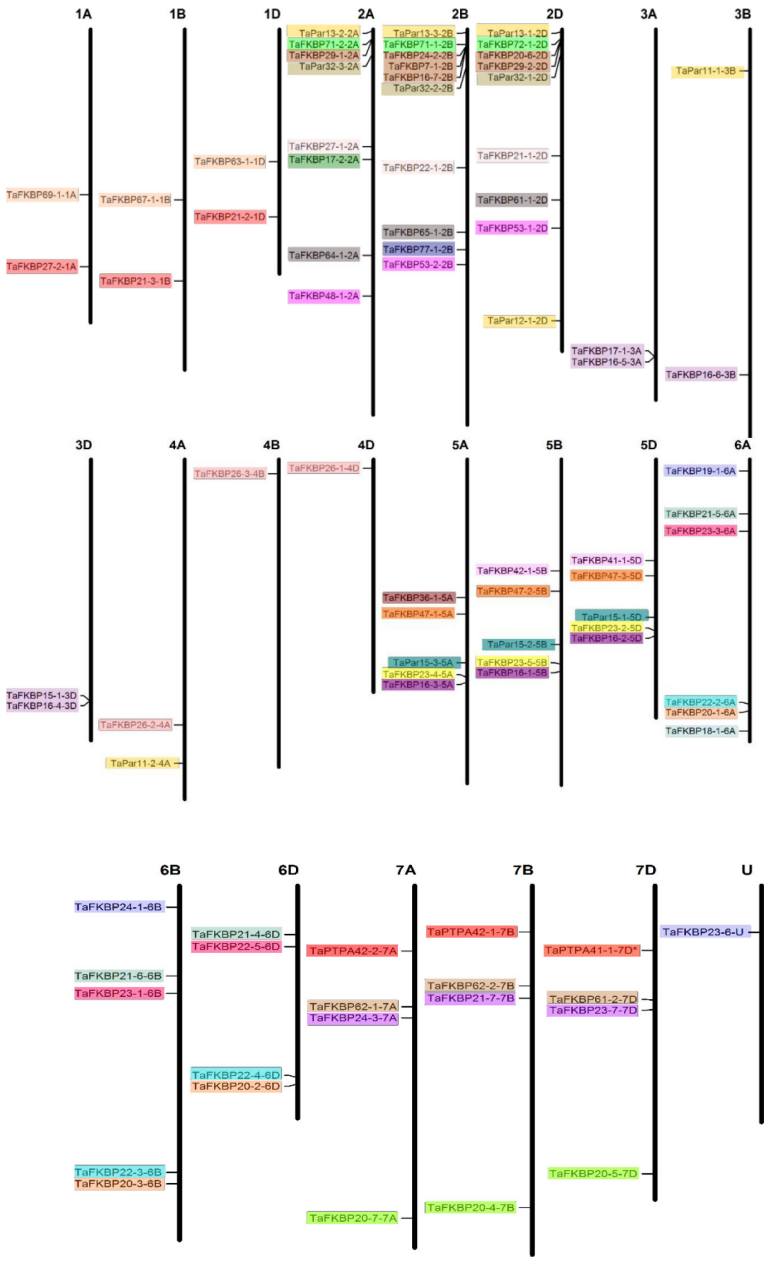


FIGURE 1
Chromosomal distribution of the genes encoding FKBPs, parvulins and PTPAs in wheat. The chromosome size was determined from the IWGSC1.0 assembly and is drawn to scale. Homeologous groups of genes are represented in same colors (U represents uncharacterized part of genome).

donors, suggesting loss during the course of evolution. Similar trend was also observed for *TaPar* genes on chromosomes 2A, 2B and 2D (Supplementary Table S7). The *TaFKBP* genes (71) showed homology with 14, 40, 20, 16, 23, 16 and 18 counterparts in *T. urartu, T. dicoccoides, A. tauschii*, *Arabidopsis*, *O. sativa*, *P. persica* and *M. domestica*, respectively. Similarly, different homologues of *TaPars* (12) were also observed in *A. thaliana* (3), *E. coli* (1), human (2), *S. cerevisiae* (1), *T. urartu* (3), *T. dicoccoides* (9) and *A. tauschii* (5). The *A. thaliana*, *O. sativa*, *T. urartu* and *A. tauschii* also depicted the presence of a single homologue of *TaPTPA* gene. Absence of *PTPA* genes in one of the wheat progenitors, *T. dicoccoides* suggests, that the gene encoding TaPTPA42-1-7B might have been derived from the wheat B sub-genome donor *A. speltoides.* Alternately, it is also likely that these genes might have been lost during evolution, as observed for *14-3-3* gene family in wheat (Shao

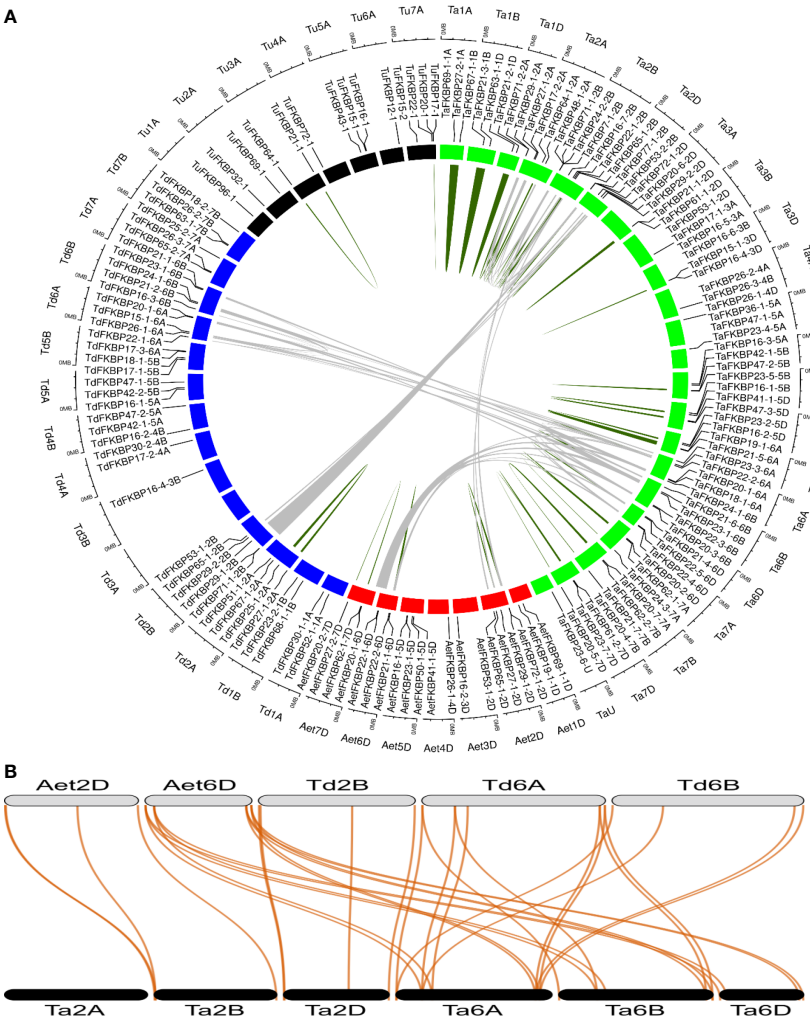


FIGURE 2
Collinearity analysis of *TaFKBP* genes using genome sequences of *T. aestivum* (Ta, AABBDD, green box) and progenitor species *A. tauschii* (AeT, DD, red box), *T. urartu* (Tu, AA, black box) and *T. dicoccoides* (Td, AABB, blue box). **(A)** A circular layout of *TaFKBP* orthologues with collinear blocks joined across the corresponding chromosomes in *T. aestivum* and progenitor species by grey lines. The tandemly duplicated genes are marked with dark green lines. The scale on the circle is in Megabases. **(B)** The specific collinear blocks between *Ta* (black color), *Aet* (grey color) and *Td* (grey color) are shown as a dual synteny plot. The collinear genes are connected by red lines.

et al., 2021). This study provides further evidence that the genes encoding FKBPs, parvulins and PTPAs are conserved across major plant species and their abundance in the wheat genome may be attributed to various duplication events during the course of evolution.

Gene structure organization

Analysis of gene organization revealed significant variability in the exon-intron structure of different *TaFKBP*, *TaPar* and *TaPTPA* genes (Figure 3). Genes of all the three PPIase sub-families showed the presence of introns in their open reading frames (ORFs), with the number of introns varying significantly within each of the sub-families. For instance, the number of introns in the ORF of these genes ranges between 1-19 in *TaFKBPs*, 1-7 in *TaPars* and 1-2 in *TaPTPAs*. The genes encoding PTPAs and FKBPs exhibited homoeolog-dependent variability in the number of introns, as also observed earlier for the wheat cyclophilin genes (Singh et al., 2019). For example, *TaFKBP16-7-2B* and *TaFKBP20-6-2D* consist of six introns each in their ORFs compared to 2-10 in their other homeologues *TaFKBP7-1-2B*, *TaFKBP24-2-2B*, *TaFKBP29-1-2A* and *TaFKBP29-2-2D*. Similarly, *TaPTPA42-1-7B* and *TaPTPA42-2-7A* depicted a single intron in their ORFs, relative to two in the third homeolog, *TaPTPA41-1-7D**. On the contrary, the *TaPar*

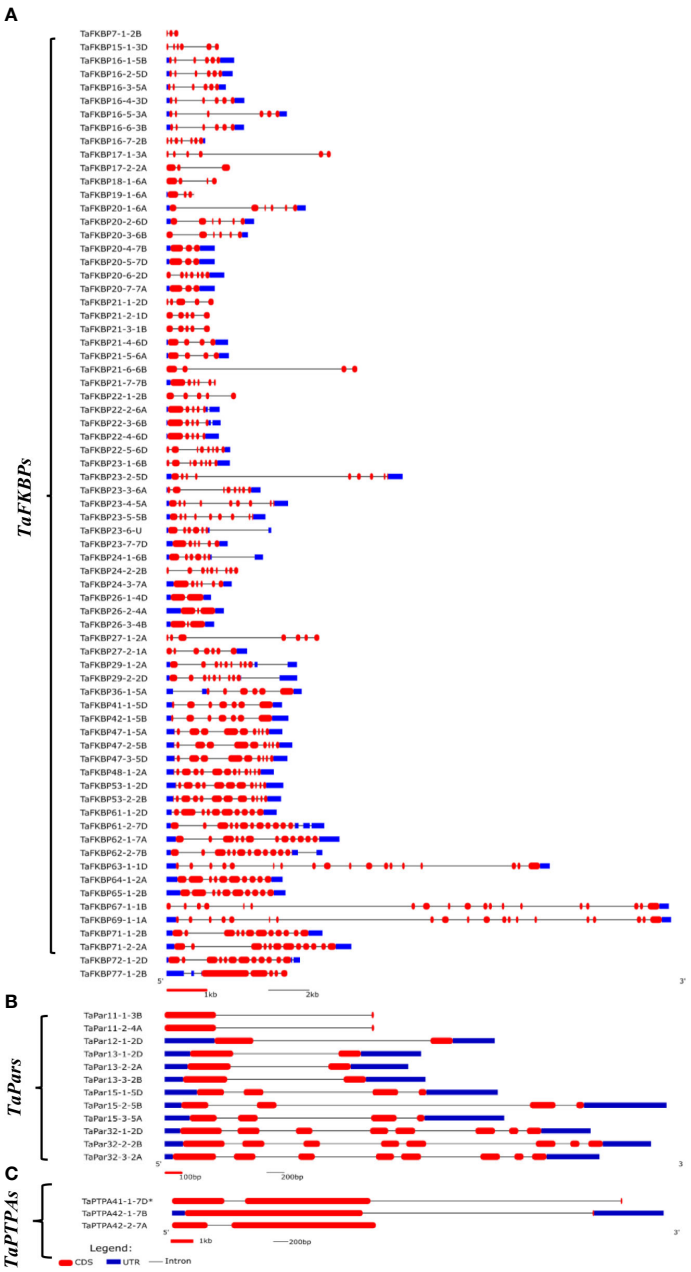


FIGURE 3
Structure of genes encoding FKBPs **(A)**, Parvulins **(B)** and PTPAs **(C)** in T. aestivum. * represents exact molecular weight and pI not determined.

genes showed conservation in the abundance of introns within members of the same homeologous groups. The length of introns also differs among different PPIase genes, with both the smallest (53 bp) and the largest (7738 bp) being observed in *TaFKBPs*, *TaFKBP15-1-3D* and *TaFKBP67-1-1B*, respectively. Within the *TaPTPA* gene family, the smallest (272 bp) and the largest (3302 bp) introns were observed in *TaPTPA41-1-7D**. The homeologous group of *TaPar* genes, comprising of *TaPar32-1-2D*, *TaPar32-2-2B* and *TaPar32-3-2A*, depicted the smallest length (80 bp) for intron 4, while the largest (2833 bp) was observed in *TaPar15-2-5B*. While majority of the *TaFKBPs* (50) contain untranslated regions (UTRs) in both the 5' and 3' regions, the same could not be observed for 12 of these genes. Nine *TaFKBPs* exhibited introns only in their 3' UTR and two in only 5'UTR. Absence of 5' UTR or 3' UTR was also observed among *TaFKBPs*. While 5' UTR is lacking in six *TaFKBPs* (*TaFKBP16-7-2B*, *TaFKBP20-3-6B*, *TaFKBP20-6-2D*, *TaFKBP22-5-6D*, *TaFKBP27-2-1A* and *TaFKBP67-1-1B*), three

TaFKBPs (*TaFKBP19-1-6A*, *TaFKBP21-7-7B*, *TaFKBP77-1-2B*) are devoid of 3' UTR. Though UTR of several *TaFKBPs* depicted the presence of introns, none of the *TaPar* and *TaPTPA* genes exhibited this feature. The genes *TaPar11-1-3B*, *TaPar11-2-4A*, *TaPTPA41-1-7D** and *TaPTPA42-2-7A* showed absence of both 5' and 3'UTRs. Unlike the cyclophilin and FKBP gene families in lower organisms *viz., Leptosphaeria maculans* (Singh et al., 2014), that contain shorter introns and compact genomes, the length of introns is larger than exons in most of the *TaFKBP, TaPar* and *TaPTPA* genes.

Domain architecture and active site residue analysis

The FKBP, parvulin and PTPA proteins in hexaploid wheat contain the characteristic signature domains, with several members of former two families also containing additional domains (Figure 4). Unlike the TaFKBPs and TaPars, several of which are multi-domain proteins, the TaPTPAs consist of only the PTPA domain. Besides FKBD, 21 of the 71 TaFKBPs also demonstrated the presence of other domains, such as nucleoplasmin-like (NPL) and tetratricopeptide repeat (TPR) domain, implying diversification in their roles (Supplementary Table S1). Except for NPL domain-containing FKBPs, such as TaFKBP47-1-5A, TaFKBP47-2-5B, TaFKBP47-3-5D, TaFKBP48-1-2A, TaFKBP53-1-2D and TaFKBP53-2-2B, that exhibit FKBD at C-terminus, this domain in other multi-domain proteins is localized at the N-terminus. One of the single domain proteins, TaFKBP19-1-6A, contains a partial N-terminus FKBD. Twelve of the multi-domain FKBPs also contain TPR domain in addition to three FKBDs, indicating their likely role in protein-protein interactions. Since the NPL domain-containing FKBPs have been demonstrated to play an important role in chromatin remodelling and regulation of transcription in *A. thaliana* (Li and Luan, 2010), the presence of this domain in the wheat FKBPs (TaFKBP47-1-5A, TaFKBP47-2-5B, TaFKBP47-3-5D and TaFKBP48-1-2A, TaFKBP53-1-2D, TaFKBP53-2-2B) implies the role of these proteins in regulation of gene expression.

Comparison of active site residues with the human FKBP, hFKBP12, revealed the absence of His (H-87) residue in all members of FKBP family in wheat (Supplementary Table S8; Figure S1A). Similarly, compared to their human counterpart hPar14, all wheat parvulins also lack the critical residue Asp (D-74; Supplementary Table S9 and Figure S1B). Variability in other active site residues was also observed in TaFKBPs and TaPars. The parvulins TaPar11-1-3B (11.06 kDa) and TaPar11-2-4A (11.16 kDa) possess only two of the conserved active site residues i.e., His (H-42) and Ser (S-72), indicating that these proteins might be enzymatically inactive. This speculation, however, needs experimental validation. The higher molecular weight TaFKBPs, such as TaFKBP61-1-2D (61.35 kDa), TaFKBP61-2-7D (61.97 kDa), TaFKBP62-1-7A (62.01 kDa), TaFKBP62-2-7B (62.02kDa), TaFKBP63-1-1D (63.97 kDa), TaFKBP64-1-2A (64.89 kDa), TaFKBP65-1-2B (65.60 kDa), TaFKBP67-1-1B (67.96 kDa), TaFKBP69-1-1A (69.90 kDa), TaFKBP71-1-2B (71.30 kDa), TaFKBP71-2-2A (71.80 kDa) and TaFKBP72-1-2D (72.07 kDa), which contain multiple FKBDs, also depicted difference in the occurrence of active site residues in different FKBDs. This observation suggests that the FKBDs with the maximum number of conserved active site residues might be enzymatically active while others may not be functional. This speculation, though warrants experimental confirmation, is also supported by a previous study which showed that of the three FKBDs in the wheat FKBP, wFKBP73, only one of the domains was enzymatically active (Unger et al., 2010). Our analysis also revealed that the different FKBDs in TaFKBPs are conserved among different homeologues. High identity of more than 94.31% was also observed in PTPA domains of TaPTPA41-1-7D*, TaPTPA42-1-7B and TaPTPA42-2-7A (Figure 5), signifying conservation of these proteins.

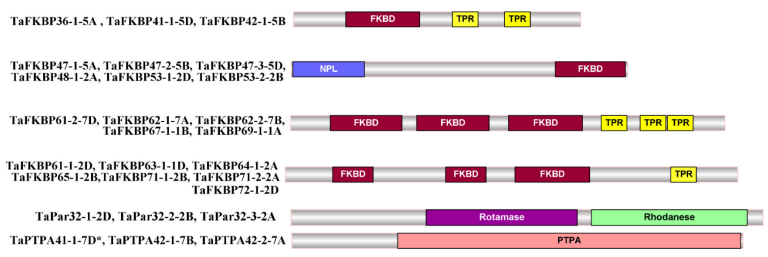


FIGURE 4
Schematic representation of multi-domain FKBP, parvulin and PTPA proteins of *T. aestivum*.

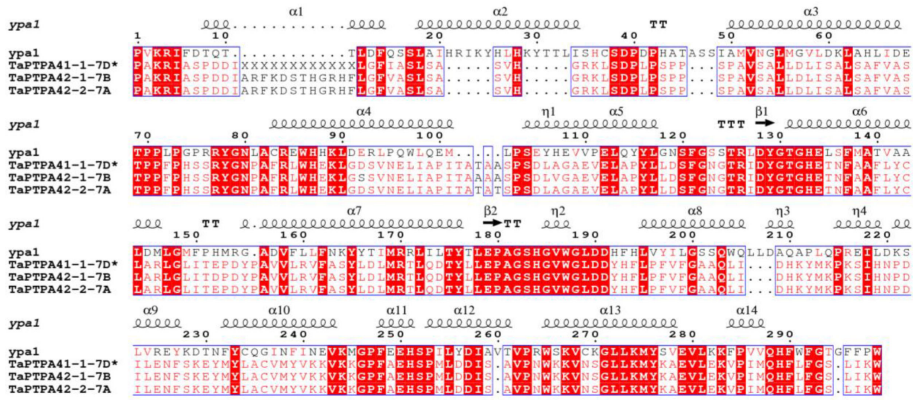


FIGURE 5
Multiple sequence alignment of PTPA domains of wheat PTPA proteins and yeast PTPA ypa1. * represents exact molecular weight and pI not determined.

Motif analysis

Protein sequences of TaFKBPs, TaPars and TaPTPAs were also analysed for the presence of conserved motifs. A total of 15 motifs, varying in length from 15 to 150 amino acids, were identified in TaFKBPs (Table 1; Figure 6A). The FKBD comprises of motifs 1, 3, 4, 5, 6 and 7, with the motifs 1, 3, 4 and 5 being observed in maximum number of proteins (67, 65, 56 and 59, respectively). The motifs 1, 3 and 4 consist of 21 amino acid residues each, compared to 15 residues in motif 5, with all constituting a component of FKBD. The absence of motifs 6 and 7 in TaFKBPs might be due to deletion of these sequences in their C-termini. Presence of only two motifs (1 and 7) in the highest molecular weight protein TaFKBP77-1-2B (77.10 kDa), might be due to the presence of partial FKBD in this protein. Apart from the conserved FKBDs, *in silico* analysis also predicted the presence of motifs in TPR (motifs 2 and 15) and NPL domains (motifs 11 and 12) of the multi-domain FKBPs. The motifs 1, 3, 4, 5 (FKBP_C) and 8 (unknown function) were observed most commonly, with 36 TaFKBPs exhibiting the presence of this combination. However, exclusive presence of motif 13 (ASP_Protease) and motif 14 (Domain of unknown function) in two different homeologue triplets i.e., TaFKBP48-1-2A, TaFKBP53-1-2D and TaFKBP53-2-2B; and TaFKBP47-1-5A, TaFKBP47-2-5B and TaFKBP47-3-5D, respectively, might be associated with some discrete functions of these proteins, that are yet to be elucidated. In another homeologue triplet, TaFKBP24-1-6B, TaFKBP23-6-U and TaFKBP19-1-6A, the former two proteins harbour motifs 1, 3, 4 (FKBP_C) and 8 (unidentified motif), while TaFKBP19-1-6A consists of only motifs 4 and 8. The presence of other motifs such as LBR tudor, cytidylate kin 2, ASP_Protease, domain of unknown function and CDC45-like protein in few of the wheat FKBPs suggests novel cellular roles that are yet to be studied. Akin to TaFKBPs, homeolog-specific variations in motifs were also observed in TaPars (Table 1; Figure 6B). On the contrary, all the three PTPAs, TaPTPA41-1-7D*, TaPTPA42-1-7B and TaPTPA42-2-7A, exhibited the same motif architecture (Table 1; Figure 6C).

Phylogenetic analysis of wheat FKBP, parvulin and PTPA proteins

To understand evolutionary relationships among FKBPs, parvulins and PTPAs of different plant species, we constructed separate phylogenetic tree for each of these protein families using alignments of amino acid sequences from wheat, its progenitors, and other plant species. For FKBPs, the phylogenetic tree constructed using NJ method in MEGA X was based on alignment of 196 FKBD-containing proteins from *T. aestivum* (70), *T. urartu* (14), *T. dicoccoides* (40), *A. tauschii* (20), *A. thaliana* (23) and *O. sativa* (29) (Figure 7A). Similarly, for parvulins also, the phylogenetic tree included 40 different parvulin protein sequences from *T. aestivum* (12), *T. urartu* (3), *T. dicoccoides* (9), *A. tauschii* (4), *A. thaliana* (3), *O. sativa* (4), *H. sapiens* (2), *S. cerevesiae* (1) and *E. coli* (2) (Figure 7B). For PTPAs, a total of seven proteins from *T. aestivum* (3), *T. urartu* (1), *A. tauschii* (1), *A. thaliana* (1) and *O. sativa* (1) were included for constructing the phylogenetic tree (Figure 7C).

Phylogenetic analysis classified the TaFKBPs into two different clades (Clades-I and II), which were further divided into various sub-groups consisting of homologous FKBPs from different plant species. Clade-I consisted of 21 members compared to 175 in Clade-II. The ultimate sub-groups (labeled a to u) comprised all the TaFKBP homeologue triads that were clustered together with their progenitor species and homologues from other plant species. The close clustering of homeologue triad was also observed for cyclophilins and 14-3-3 gene families in common wheat (Singh et al., 2019; Shao

TABLE 1 Comparative analysis of different conserved motifs in FKBP, parvulin and PTPA proteins in *T. aestivum*.

Motif	Motif sequence	Amino acids	Putative function (Pfam/Prosite/CDD)	Proteins having this motif
TaFKBPs				
1	IKGMKVGEKRRLTIPPELGYG	21	FKBP_C	67
2	NNAEKIEAAAKKKDEGNAWFKMGKYARASKRYEKAAKFIEYDSSFSEDEKKQSKPLKISCKLNNAACKLKLKDYKEAEKLCTKVLELDSTNVKALYRRAQAYTELV	106	TPR1/TPR2	9
3	GSPPEIPPNATLIFDVELLSV	21	FKBP_C	65
4	GDKVEVHYTGTLADGTVFDSS	21	FKBP_C	56
5	PFKFRLGSGQVIKGW	15	FKBP_C	59
6	HFCPALSKAVKTMKKGEKVLLTVKPQYGFGEQGRPASEDEGAVPPNATLHIDLZLVSWKTVTEIGDDKKILKKVLKEGEGYERPNDGAVVKVKLIGKLDDGTVFVKKGHDGQEPFEFKTDEEQVIEGLDRAVLTMKKGEVALVTIPPEHA	150	FKBP_C	9
7	SVKDICKDGGIFKKILKEGEKWENPKDPDEVTVKYEARLEDGTVVSKSEGVEFTVKD	57	FKBP_C	12
8	TKSGLKYKDLKVGEG	15	–	55
9	DLAVVPPNSTVYYEVELVSFDKEKESWDL	29	LBR tudor	12
10	KKLKEKVKEYNKKDAKFYKNMFNKKPKPENE	31	cytidylate kin 2	17
11	SAFWGVEVKPGKPYTHSHNPRHGRLRJTQATLGA	34	NPL	6
12	VVGNKEPVLLCALAPKLADVCHLQIELEEKPEVFS	35	NPL	14
13	RSVHLAGYYVGDVYEDIGDSDTGSESLQGSDDDFLASDDDDVVIPVSHGQMNTDSEDDSDYDEDYDSEDDEDLMYNQGRGKSSVVIEEIQEDEKPVDDNSRLQLAVRTPPAESVESEDEDGFPVSESKKSSKGSSKKDKNLNNGTSTED	149	ASP_Protease	3
14	SVLGQSSVHLSGYYLRPGSRGNAGEEDSESYGEDVGESDTDQDYEGSEDSYESDFIDDGDNEVPEDSDVSDSMDDGDVCSTPDHRKQDSEKHARKVKRQRRLKKKQQVDSSADKIADSPSKPAARRKRGSIFDSASEDEDFLAQSEEEN	149	Domain of unknown function/CDC45 like protein	3
15	HLNVAACLJKQPRFDLAIAACAKVLTENPVNVKALYRRGKAYAEAGRAEDAKEDFLK	57	TPR/domain of unknown function	9
TaPars				
1	CPSKENGGMLGWVRRGQMVPEFEEAAFSAPLNKVVRCKTKFGWHLVQVLSERDQCLLQDIQPEELHEKMQDPSFIEEAQLIDVREPDEVERASLPGFKVLPLRQFGTWGPVMTDEFNPQKDTYVLCHHGMRSMQVAKWLQSQGFQKIYNV	150	PPIC-type PPIase domain, Rhodanese domain	3
2	EFADVAQZHSDCPSAKRGGDLGTFPRGKMQKPFZEAAYALKVGEISDIIDTESGVHIILRTG	62	PPIC-type PPIase domain	7
3	ASHILIKHEGSRREASWKDPEGRVISATT	29	PPIC-type PPIase domain	9
4	SLSGLARRAPLLAVASPAPSSPAALSLLASARPVSAAWGSAMRPAGEHPRPGTRVLCTAAS	61	–	3
5	DGEGGAKGGKGKGKGGKGGDDLGTLAVEK	29	–	6
6	RADAAARLGELRPQI	15	–	9
7	QHSARPPAAGLHQRADHARQIQDCLVMDKINPCGQQYWYQC	41	–	2
8	IHAYSVKADSSIPTY	15	–	3
9	GKPGEGKEVKGKGKL	15	–	6
10	TGEETVR	7	–	3
11	DCASFADL	8	–	2
12	CRQFGA	6	–	2

(Continued)

10.3389/fpls.2022.1053524

TABLE 1 Continued

Motif	Motif sequence	Amino acids	Putative function (Pfam/Prosite/CDD)	Proteins having this motif
TaPTPAs				
1	NGTRIDYGTGHETNFAAFLYCLARLGLITEPDYPAVVLRVFASYLDLMRTLQDTYLLEPAGSHGVWGLDDYHFLPFVFGA AQLIDHKYMKPKSIHNPDILENFSKEYMYLACVMYVKKVKKGPFAEHSPMLDDISAVPNWKKVNSGLLKM	150	PTPA	3
2	WTGDSPPPAYRPIRMPAINAPTNTAAIVLSPVPQPLPVPPASPPFAFQAPAKRIASPDDIARFKDSTHGRHFLGFVASLSAS VHGRKLSDPLPSPPSPAVSALLDLISALSAFVASTPPFPHSSRYGNPAFRLWHEKLGDSVNELIAPIT	150	PTPA	3
3	MSNPESSPQAAASSTSPPSHAGHIHTPLCRSCG	33	–	3
4	EVLEKVPIMQHFLFGSLIKWE	21	PTPA	3
5	SPSDLAGAEVELAPYLLDSFG	21	PTPA	3
6	APAPAP	6	–	3

The different motifs were identified using MEME tool.

et al., 2021). While the sub-groups a, c, d, e, g, i and o lacked FKBP homologues from *T. urartu* (A genome), the sub-groups h, q and u showed the absence of FKBP homologues from *T. dicoccoides* (A genome). The clustering pattern depicted in the phylogenetic tree prompts us to hypothesise that the *FKBP* genes on chromosomes 1A, 2A, 4A, 5A, 6A and 7A of the common wheat either evolved from *T. dicoccoides*, or the corresponding genes present in *T. urartu* were later lost during the course of evolution. Conservation of most of the homeologue triplets implies that members of the *TaFKBP* gene family were not subjected to significant changes during the course of evolution and these genes in hexaploid wheat have been acquired from its progenitors due to polyploidy. The presence of FKBP homologues from *O. sativa* (monocot) and *A. thaliana* (dicot) in each of the sub-groups implies existence of these genes even before divergence of monocots and dicots. Furthermore, the grouping of FKBPs appears to be domain specific and also influenced by the localization pattern, and the arrangement of introns and exons in the corresponding genes. For example, Clade-I consisted of multi-domain FKBPs possessing TPR and FKBDs. One of its two sub-clades contained proteins with three FKBDs and one TPR domain, with their localization predicted to be in the nucleus. However, all other members with similar domain architecture, but localized to cytosol, formed a separate sub-clade in Clade-II. Members containing multiple domains formed different sub-groups in Clade-II (for example sub-groups m, n, o, p, q). On the other hand, most of the FKBPs consisting of only a single FKBD were clustered together based on their localization pattern. This implies that the localization signal (Supplementary Table S1) and conservation of the underlying structure (Figure 3A) might have played an important role in the functional divergence of FKBPs in common wheat.

The phylogenetic tree of TaPars depicted three major clades *viz.*, Clades-I, II and III. The Clades I and II comprised of 19 and 20 members, respectively, while the third clade contained only a single parvulin, surA, from *E. coli*, and was differentiated from other two clades that consisted of various eukaryotic parvulins. Both Clade-I and Clade-II were further divided into sub-clades (labeled a-d), with all the sub-groups, except sub-group a that consisted of only a single member TaPar12-1-2D from D genome of *T. aestivum*, represented by triplet homeologues. The sub-group b consisted of two additional proteins from *T. aestivum* (TaPar11-1-3B and TaPar11-2-4A), encoded by a gene each on chromosomes 3B and 4A. Except for sub-group d, none of the other sub-groups depicted the presence of members representing each of the three wheat sub-genomes. For example, sub-groups a and b lacked representatives from *T. urartu*, while sub-group c was devoid of corresponding homologue from *A. tauschii*. Computational analysis categorised the PTPA proteins into two groups, Clades-I and II, with the former comprising of closely clustered PTPA proteins from different species (*T. aestivum*, *T. urartu*, *A. tauschii* and *O. sativa*), while Clade-II included PTPA from dicot *A. thaliana*. These observations suggest that in contrast to *FKBP* genes, the *PTPA* genes might have diverged after differentiation of monocots from dicots.

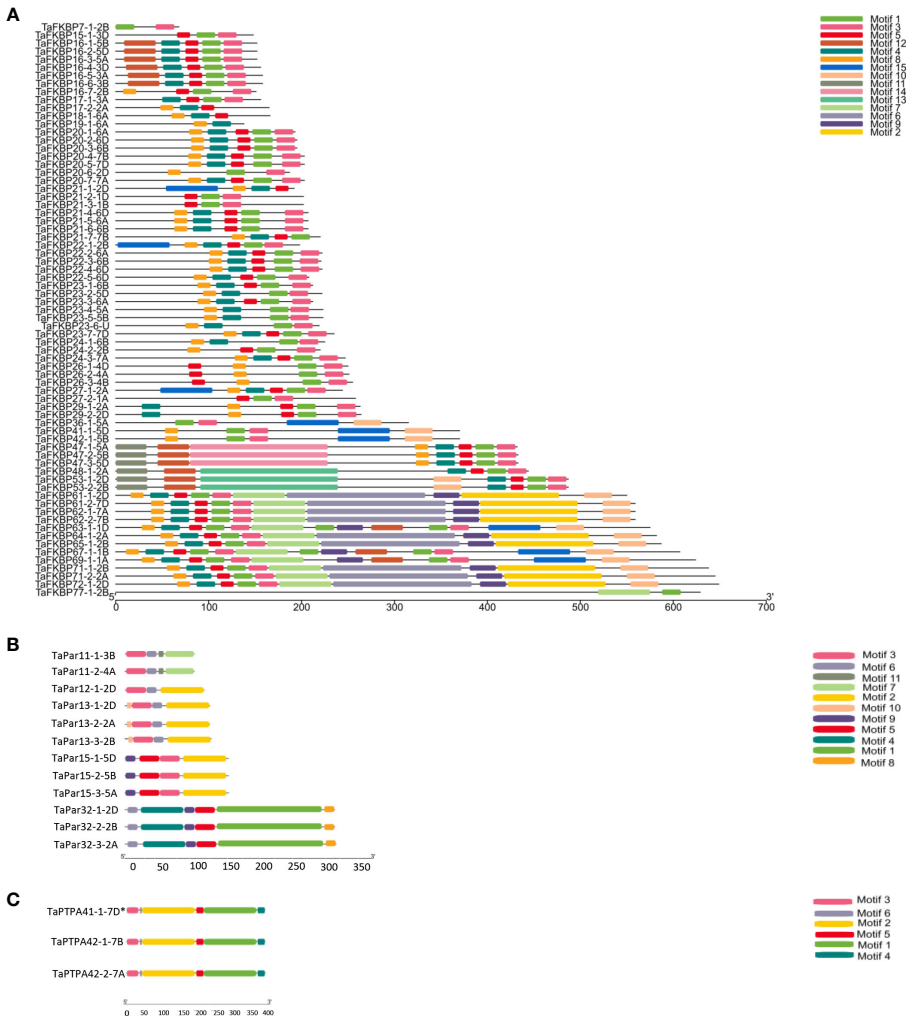


FIGURE 6
Organization of different conserved motifs in FKBP **(A)**, parvulin **(B)** and PTPA **(C)** proteins of *T. aestivum*.

Expression analysis of *TaFKBP*, *TaPar* and *TaPTPA* genes in response to heat stress

The effect of heat stress on expression of *TaFKBP*, *TaPar* and *TaPTPA* genes was investigated in response to various heat stress treatments using qRT-PCR (Figures 8–10). The survivability of the wheat seedlings was followed, as described earlier, after exposure to lethal temperature (50 °C) without and with acclimation of seedlings at 37 °C for 2 h (Singh et al., 2019). The seedlings that were acclimated at 37°C for 2 h before heat stress at 50°C for 4.5 h showed higher survival (80%) as compared to the seedlings that were exposed to lethal heat stress at 50°C without acclimation (19%), indicating acquisition of thermotolerance at sub-lethal temperature (Singh et al., 2019). For expression analysis, common primers for 28 different homeolog *TaFKBP* groups (71 genes), five for *TaPars* (12 genes) and a single primer pair for the *TaPTPAs* (3 genes) were designed and validated by conventional PCR. Of these, 14 and nine primer pairs, representing 43 and nine different *TaFKBPs* and *TaPars*, respectively, from different homeolog groups depicted amplification and were used further for qRT-PCR analysis (Supplementary Table S10). The *TaPTPA* genes, however, did not show amplification.

On the basis of expression profile, the *TaFKBP* genes were classified into five different groups (Figures 8, 9). The expression of Group 1 genes, except *TaFKBP53-1-2D/TaFKBP48-1-2A/TaFKBP53-2-2B* and *TaFKBP61-2-7D/TaFKBP62-1-7A/TaFKBP62-2-7B*, after increasing during acclimation (37 °C for 2 h) was also maintained under subsequent lethal heat stress (37 °C-50 °C-4.5 h), indicating the likely role of these genes in thermotolerance (Figure 8A). Among the Group 1 genes, highest expression after acclimation (27.7-fold) and in response to

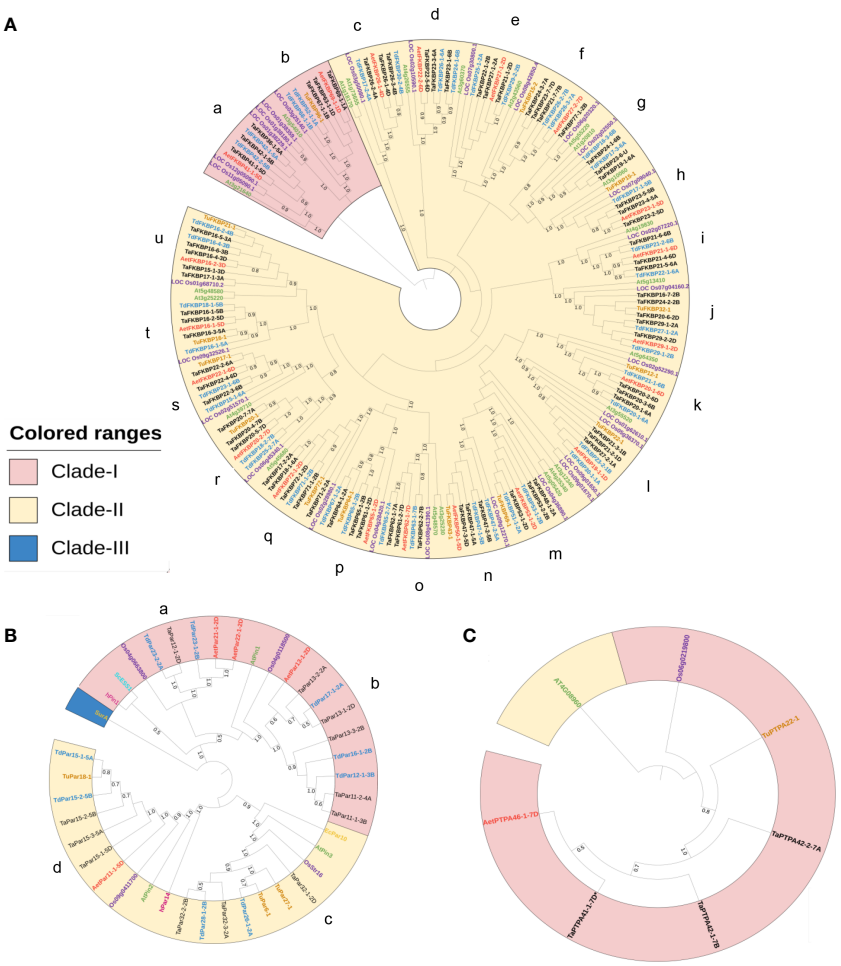


FIGURE 7
Phylogenetic analysis of FKBP **(A)**, Parvulin **(B)** and PTPA **(C)** proteins of *T. aestivum* (black), *T. urartu* (orange), *T. dicoccoides* (blue), *A. tauschii* (red), *A. thaliana* (green) and *O. sativa* (purple), *H. sapiens* (pink), *S. cerevisiae* (blue) and *E. coli* (yellow). Clustal Omega was used to perform the multiple sequence alignment. The phylogenetic tree was constructed by Neighbor-Joining (NJ) method of MEGA X standalone application with 1000 bootstrap replicates. The generated tree was annotated using iTOL online server. Different clades are highlighted with different colors.

subsequent lethal heat stress (40.8-fold) was observed for *TaFKBP61-2-7D/TaFKBP62-1-7A* and *TaFKBP62-2-7B*, respectively. On the contrary, *TaFKBP21-7-7B/TaFKBP23-7-7D* and *TaFKBP24-3-7A* registered the lowest increase in transcript level after acclimation (2.5-fold) and at 37 °C-50 °C-4.5 h (2.7-fold). However, these genes were the only members of Group 1 which depicted significantly higher expression compared to control 25 °C when exposed to direct heat stress at 50 °C for 4.5 h (Figure 8A).

The expression of Group 2 genes, compared to control, was enhanced significantly only when the thermal stress (50 °C-4.5 h) was imposed after acclimation (Figure 8B). Exposure to acclimation temperature (37 °C-2 h) and direct heat stress had no appreciable effect on the expression of these genes (Figure 8B). The genes that depicted an increase in expression in response to both direct heat stress and after acclimation were categorised in Group 3 (Figure 9A). Among the Group 3 genes, the increase in transcript abundance by heat stress at 50 °C ranged from 2.9-fold (*TaFKBP16-1-5B/TaFKBP16-2-5D/TaFKBP16-3-5A*) to 9.7-fold (*TaFKBP47-1-5A/TaFKBP47-2-5B/TaFKBP47-3-5D*) without acclimation, and from 1.9-fold (*TaFKBP16-1-5B/TaFKBP16-2-5D/TaFKBP16-3-5A*) to 19.9-fold (*TaFKBP71-1-2B/TaFKBP71-2-2A/TaFKBP72-1-2D*) after acclimation (37 °C-50 °C-4.5 h). The expression of these genes was not affected by sub-lethal temperature (37 °C for 2 h).

Contrary to Group 1, 2 and 3, the Group 4 comprised a single *TaFKBP* triplet, *TaFKBP20-1-6A/TaFKBP20-2-6D/TaFKBP20-3-6B*, that showed increase in transcript level (3.7-fold) only in response to direct exposure to 50 °C but was unaffected by any other temperature regime (Figure 9B). The genes *TaFKBP16-4-3D/TaFKBP16-5-3A/TaFKBP16-6-3B* and *TaFKBP36-1-5A*, whose expression was not responsive to any

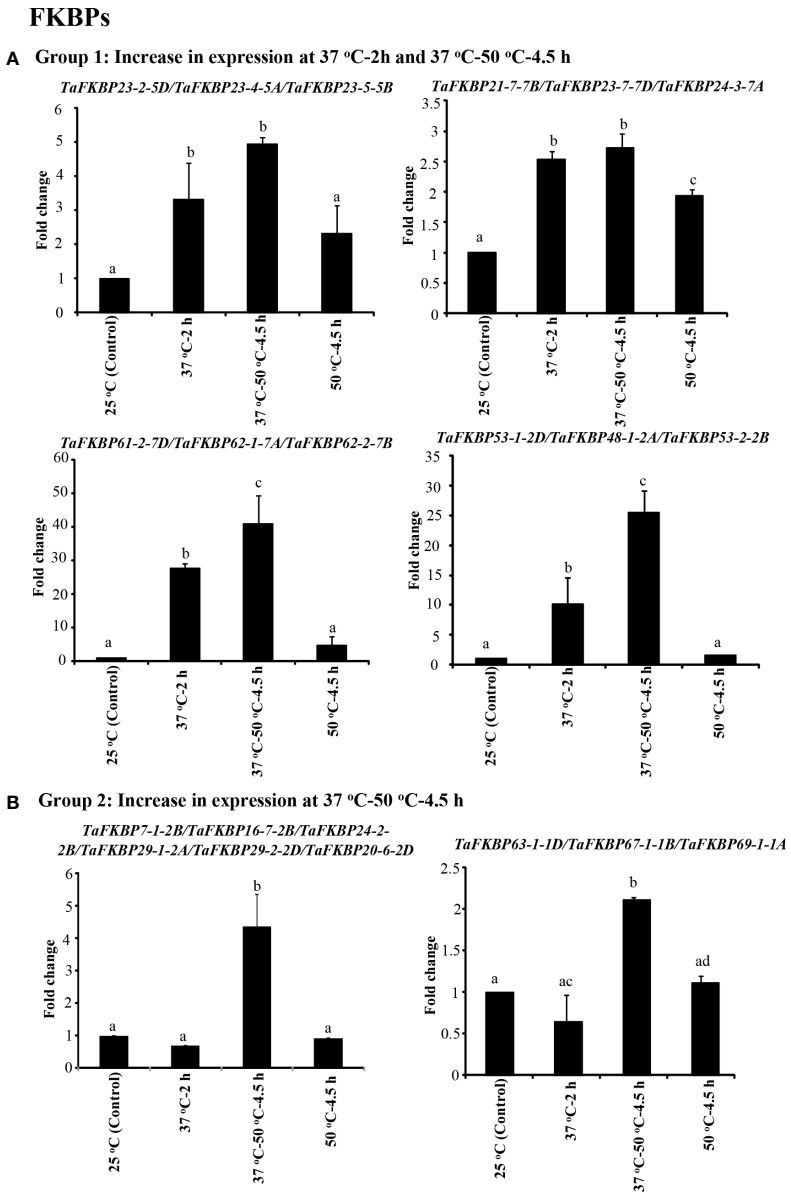


FIGURE 8
qRT-PCR analysis of *FKBP* genes in the seven days old wheat seedlings. **(A)** Genes whose expression was induced during acclimation phase (37° C-2 h) and was maintained for further 4.5 h on subsequent exposure to 50 °C. **(B)** Genes that registered an increase in expression only at 37 °C-50 °C-4.5 h. *Actin* was used as reference. The values depict the mean of three biological replicates ± standard error and the different lowercase letters denote significant difference between the treatments at $P \leq 0.001$ (Tukey-HSD test; α=0.05) (/represents homeologs).

of the temperature treatments, were grouped into Group 5 (Figure 9C). Expression analysis of *TaPar* genes revealed that compared to 25 °C control, the transcript levels of *TaPar32-1-2D/TaPar32-2-2B/TaPar32-3-2A* and *TaPar15-1-5D/TaPar15-2-5B/TaPar15-3-5A* were upregulated by heat stress after acclimation (37 °C-50 °C-4.5 h), but were not responsive to direct heat stress (50 °C for 4.5 h) (Figure 10). On the contrary, the expression of *TaPar13-1-2D/TaPar13-2-2A/TaPar13-3-2B* was enhanced only during acclimation (37 °C for 2 h), but was not affected by direct thermal stress.

Expression analysis of *TaFKBPs* and *TaPars* was also performed digitally under heat (40 °C for 1 h and 6 h) and drought stress conditions using RNAseq expression data (Figure 11; Liu Z et al., 2015). Heatmap analysis revealed differential modulation of several members of the *TaFKBP* gene family under heat stress. Of the 71 different *TaFKBP*

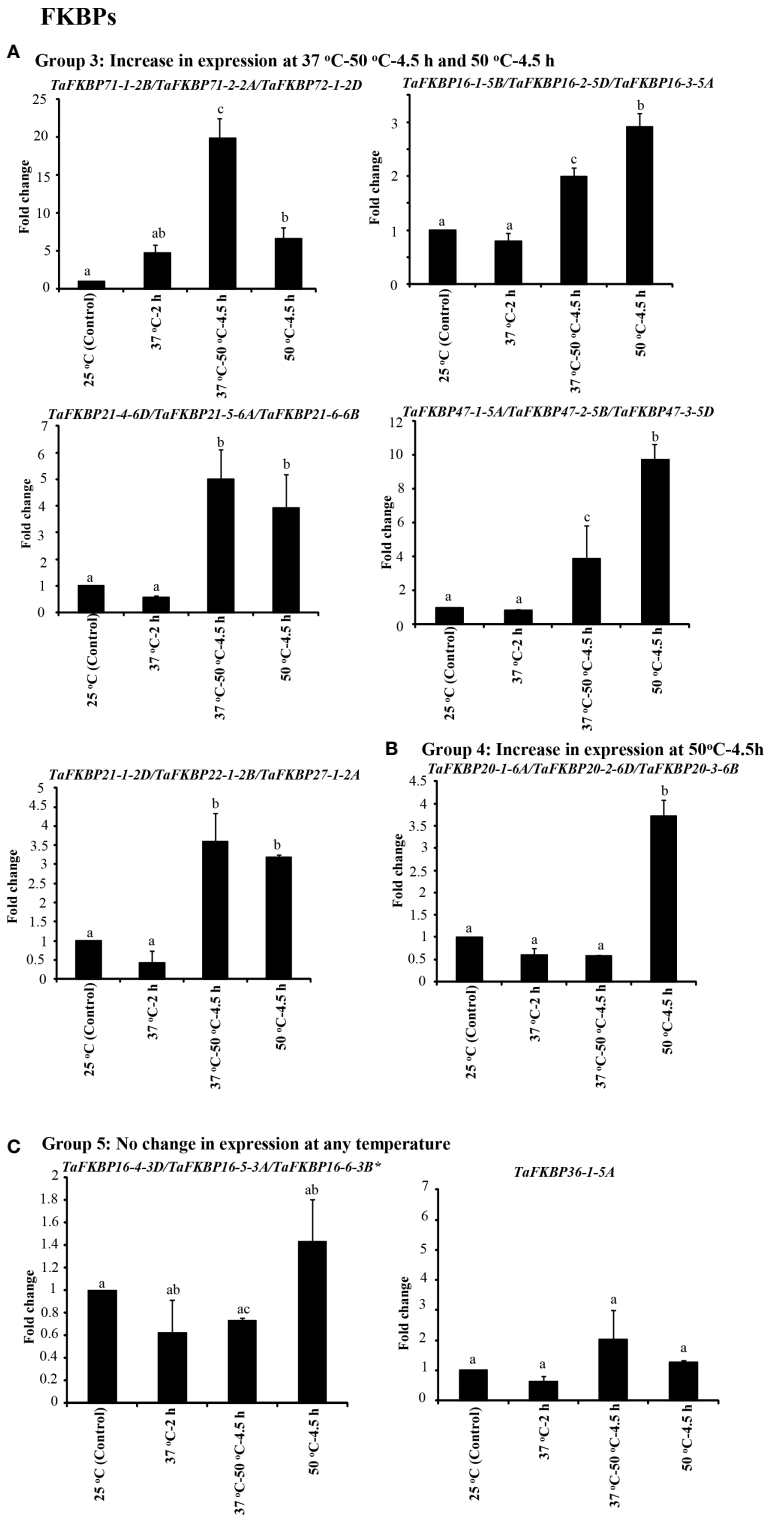


FIGURE 9
qRT-PCR analysis for studying the effect of heat stress on expression of *TaFKBP* genes in the seven days old wheat seedlings. **(A)** Genes showing upregulation in response to both direct heat stress (50 °C) and after acclimation (37 °C-50 °C-4.5 h), **(B)** only direct heat stress (50 °C) and **(C)** unaffected by heat stress. *Actin* was used as reference. The values depict the mean of three biological replicates ± standard error and the different lowercase letters denote significant difference between the treatments at $P \leq 0.001$ and $P \leq 0.05$* (Tukey-HSD test; α=0.05) (/represents homeologs).

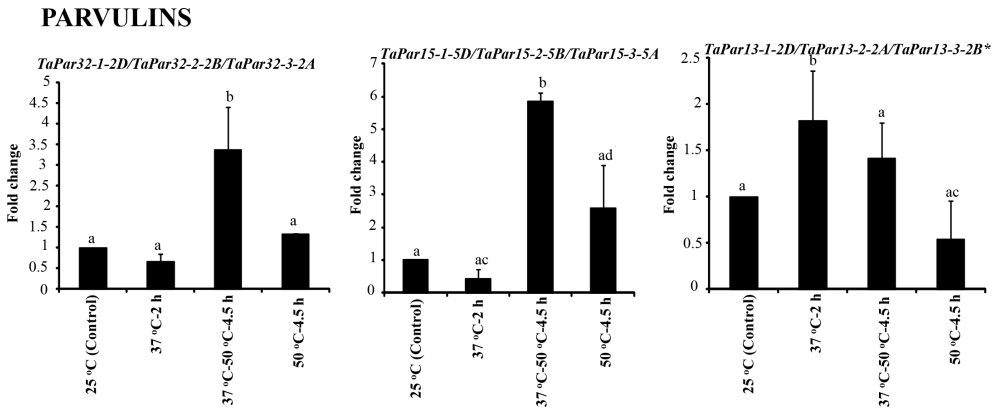


FIGURE 10
Effect of heat stress on expression of genes encoding parvulins in seven days old wheat seedlings using *Actin* as reference. The values depict the mean of three biological replicates ± standard error and the different lowercase letters denote significant difference between the treatments at P ≤ 0.001 and P ≤ 0.05* (Tukey-HSD test; α=0.05) (/represents homeologs).

genes, *TaFKBP64-1-2A* and *TaFKBP65-1-2B* showed substantial upregulation at 40 °C over control (Figure 11A). The significant increase in expression of *TaFKBP62-1-7A/TaFKBP62-2-7B* during acclimation phase (37 °C-50 °C-4.5 h) and their subsequent downregulation at higher temperature (50 °C-4.5 h), as observed in our study, is also supported by heatmap analysis which showed upregulation of these genes only when exposed to 40 °C for only 1 h, with prolonged stress for 6 h resulting in a decrease (Figure 11A). Induction of Group 3 genes *TaFKBP71-2-2A/TaFKBP72-1-2D* (Figure 9A) during acclimation is also supported by digital expression analysis, which showed higher expression after heat stress. Similarly, decrease in expression of *TaPar13-2-2A* and *TaPar13-3-2B* genes in response to direct heat stress (50 °C) is also consistent with the results of digital analysis that revealed downregulation of these genes at 40 °C (Figure 11B). Difference in the expression profile of several other genes, such as *TaFKBP20-1-6A/TaFKBP20-2-6D/TaFKBP20-3-6B*, observed between the present study and digital analysis, may be ascribed to difference in the temperature regime and the genotype used.

Heat shock elements have been implicated in regulation of genes by thermal stress (Schöffl et al., 1998). Therefore, to understand the role of cis-regulatory elements in the stress-responsiveness of different genes, 2.0 kb upstream regions of these genes were examined for the identification of heat stress responsive elements (Supplementary Table S11). These analyses resulted in identification of different heat shock elements (HSEs), such as stress response element (STRE), TTC rich type 1, TTC rich type 2, TTC rich type 3, TTC rich type 4, Gap type 1, Gap type 2, Gap type 3 and Perfect HSE (Qazi et al., 2020), which with the exception of few members, were observed in varying number (0-23) in the upstream regions of all genes. The perfect HSEs (Shakeel et al., 2011) were identified in the upstream regions of only seven *TaFKBP* genes (*TaFKBP20-1-6A, TaFKBP22-1-2B, TaFKBP23-5-5B, TaFKBP61-1-2D, TaFKBP77-1-2B, TaPTPA41-1-7D** and *TaPTPA42-2-7A*), out of which significant heat-induced increase in expression was observed only for *TaFKBP20-1-6A* (Group 4) on exposure to direct heat stress at 50 °C, and for *TaFKBP22-1-2B* (Group 3) at both 37 °C-50 °C-4.5 h and 50 °C-4.5 h (Figure 8). Similarly, the increase in expression observed at 37 °C-50 °C-4.5 h for the *TaPar* homeolog triplet *TaPar15-1-5D*/*TaPar15-2-5B* and *TaPar15-3-5A* could be due to the presence of HSEs present as STREs. Though the Group 5 genes did not show any change in transcript levels under any of the heat stress conditions, the upstream region of these genes also showed the presence of stress responsive cis elements, suggesting their likely role under different stress conditions.

Protein-protein interaction network

The PPI networks of TaFKPBs, TaPars and TaPTPAs (Figure 12) were constructed using the STRING database (Szklarczyk et al., 2019). In total, TaFKBPs matched to 19 distinct *A. thaliana* orthologs, with identities ranging from 34.9 to 91.1% (Supplementary Table S12). As observed, the PPI network of FKBPs consisted of 29 nodes and 170 edges (Figure 12A), with the former showing varied degree of interaction ranging from 2 to 23 (Supplementary Table S13). Among TaFKBPs, the highest degree of interaction (19) was exhibited by TaFKBP20-3-6B and TaFKBP20-7-7A, while the lowest (6) was shown by TaFKBP47-3-5D and TaFKBP53-2-2B, suggesting that the FKBP proteins were highly linked with other proteins and are likely to play important roles in different biological processes. The k-means clustering grouped all the

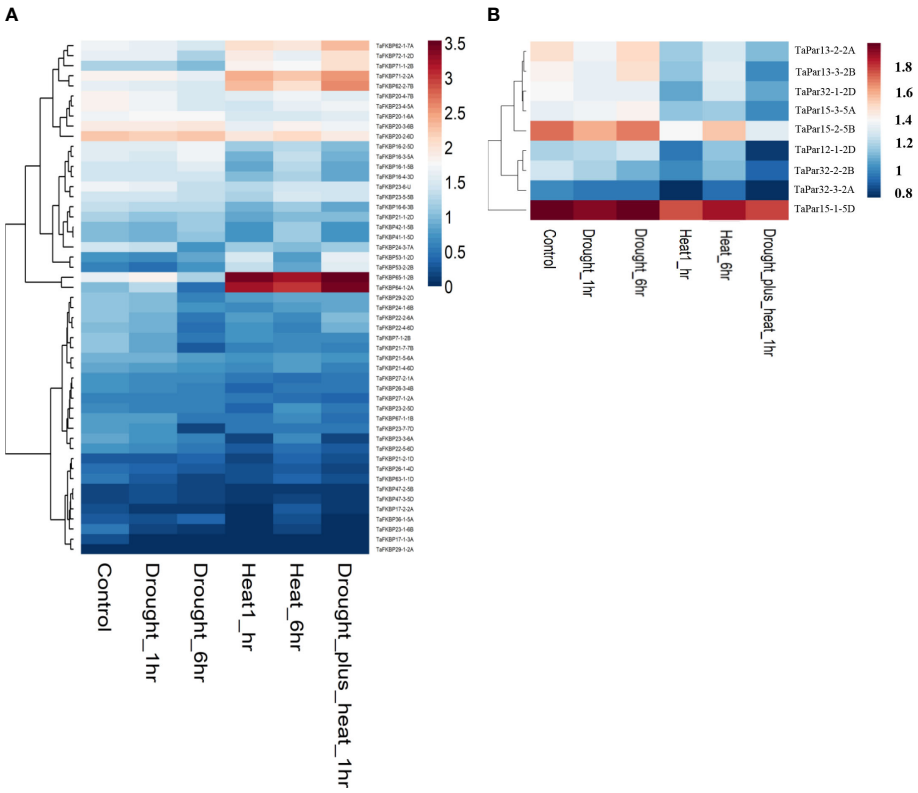


FIGURE 11
Heatmap showing expression of wheat FKBP **(A)** and parvulin **(B)** encoding genes under heat and drought stress conditions. Expression values were derived from the data available at WheatExp database (https://wheat.pw.usda.gov/WheatExp/) pertaining to treatment of one week old seedlings with heat stress at 40°C, drought or both for 1 h and 6 h duration (Liu Z et al., 2015). The scaled color gradient varies from blue (downregulation) to red (upregulation).

interacting partners into three clusters of size 13, 12 and 4 (Supplementary Table S14). The 12 TaPars matched to 3 distinct *A. thaliana* orthologs, with identities of 47.5 to 90.6% (Supplementary Table S12). The PPI network of TaPars (Figure 12B) consisted of 13 nodes and 24 edges, and the degree of interaction of nodes varied from 1 to 6 (Supplementary Table S13). The highest degree of interaction (7) among these proteins was exhibited by TaPar15-3-5A, and the lowest (4) by TaPar32-3-2A. The k-means clustering grouped all the interacting partners into three clusters of size 13, 12 and 4 (Supplementary Table S14). On the other hand, the three TaPTPA proteins matched a single *A. thaliana*, homologue with identity ranging from 57.8 to 59.1% (Supplementary Table S12). The PPI network of PTPAs (Figure 12C) consisted of 11 nodes and 47 edges with varied degree of interaction (Supplementary Table S13). TaPTPAs were clustered into three groups containing 3 to 4 members. As observed for TaFKBPs and TaPars, the PPI network of TaPTPAs also formed clusters with interacting partners, indicating their probable functions in the biological systems (Supplementary Table S14).

Discussion

The FKBPs, parvulins and PTPAs are imperative for *cis-trans* isomerisation of the peptidyl-prolyl bonds, and several studies have revealed their essential role in diverse cellular processes (Hanes et al., 1989; Luan et al., 1994; Lu et al., 1996; Janssens and Goris, 2001; Van Hoof and Goris, 2003; Gopalan et al., 2004; Meiri et al., 2010; Yu et al., 2012; Singh et al., 2021). The availability of whole genome sequences has enabled detailed *in silico* analysis and characterization of these gene families in different plant species. Though previous studies characterized FKBPs in several plants (He et al., 2004; Gollan and Bhave, 2010; Wang W et al., 2012; Zhang X et al., 2014; Dong et al., 2018; Waseem et al., 2018), information on this important family of genes, until this study, was lacking in wheat. The present analysis revealed that the hexaploid wheat genome encodes for 71 different *TaFKBP* genes, which are unevenly distributed on the 21 different chromosomes. On the contrary, the *TaPTPA* and *TaPar* genes are confined to a single or only few chromosomes, respectively. The expansion of *FKBP* gene family in wheat appears to be the result of different duplication events, such as

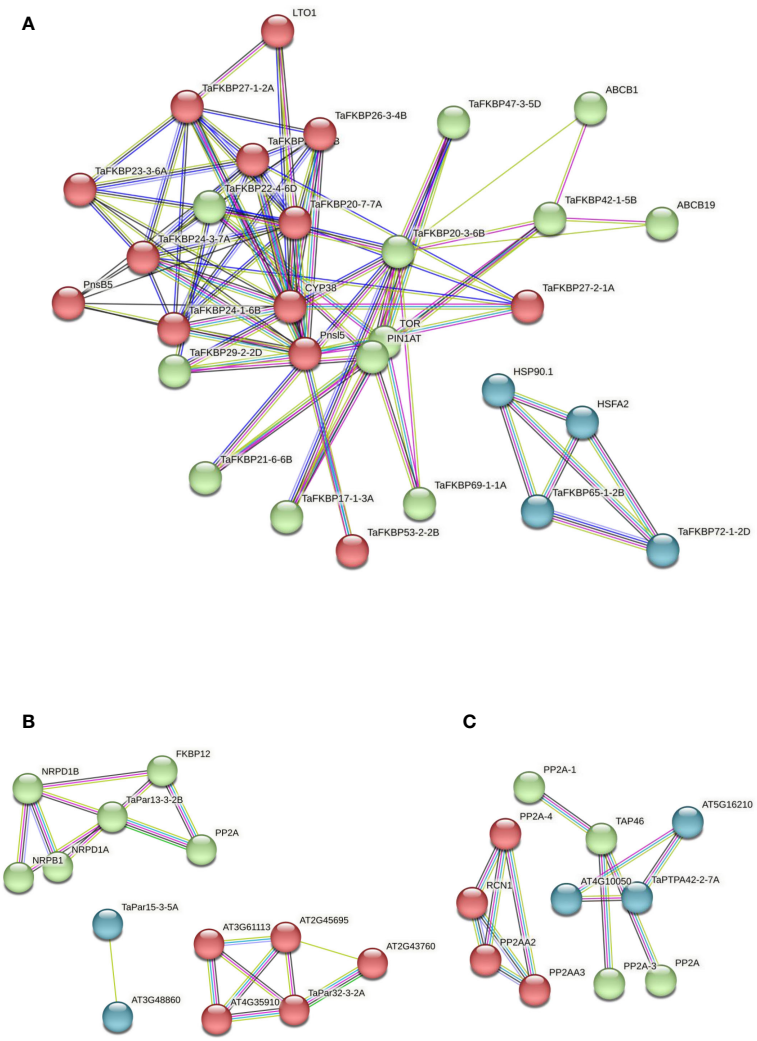


FIGURE 12
Protein-protein interaction (PPI) networks of wheat FKPBs **(A)**, parvulins **(B)**, and PTPAs **(C)**, constructed using the STRING database version 11.5 (https://string-db.org/; Szklarczyk et al., 2019). The clusters obtained using k-means clustering in each of the protein family are colored in red, green and blue. The network nodes represent proteins and the edges represent associations. The color of the edge line indicates the type of evidence i.e., violet: fusion, green: neighborhood, blue: cooccurrence, purple: experimental, yellow: text mining, light blue: database, black: coexpression.

whole genome, segmental and tandem duplications. Synteny analysis indicated the presence of many collinear blocks of *TaFKBP* genes in wheat and its sub-genome donors. In addition, the presence of triplet homeologues for most of the *TaFKBPs* suggest that whole genome duplication due to polyploidization is likely the major force behind evolution of *FKBP* gene family in wheat, which was also reported for cyclophilin, 14-3-3 and invertase genes (Singh et al., 2019; Shao et al., 2021; Wang et al., 2021). However, compared to their corresponding sub-genome donors, the difference in prevalance of *FKBP* genes on some hexaploid wheat chromosomes indicates either a gain due to duplication or loss during the course of evolution (Shao et al., 2021).

Phylogenetic analysis showed strong clustering of homoeologue triplets of TaFKBP, TaPar and TaPTPA proteins, with overall grouping influenced by localization patterns and arrangement of sequence motifs. The monocot and dicot homologues of FKBPs and parvulins were clustered together in each of the sub-groups, indicating the existence of these genes in plants even before divergence of the two (Figures 7A, B). Similar observation was also reported for the invertase gene family in common wheat (Wang et al., 2021). On the contrary, the monocot and dicot PTPAs were clustered separately, indicating the acquisition of these genes by wheat genome after differentiation of dicots and monocots. Structural analysis of the genes revealed that the length of exons, which is variable in *TaFKBP* genes, could be related to

their orthologs. This is evident, since the length of exon 3 (34 bp), 4 (62 bp) and 5 (53 bp) in homeologues *TaFKBP20-1-6A*, *TaFKBP20-2-6D*, and *TaFKBP20-3-6B* is similar to the length of exon 2, 3 and 4 in their *A. thaliana* ortholog *AtFKBP12* (Dong et al., 2018). Further, the length of these exons is also consistent with their other orthologs in *B. rapa, O. sativa, P. persica, S. lycopersicum* and *M. domestica* (Dong et al., 2018). The uniform pattern of exons was also observed for all the *TaPar* genes, and it corresponds with their *A. thaliana* orthologs *AtPIN1*, *AtPIN2* and *AtPIN3* (He et al., 2004). As reported for *FKBP* genes of *S. lycopersicum* (Waseem et al., 2018), considerable diversity was also noticed in the 5' and 3' UTR regions of *TaFKBPs*.

As observed earlier for cyclophilin proteins in wheat, the FKBPs also show variability in their domain architecture and localization, implying functional diversification (Singh et al., 2019). The presence of TPR domains in majority of the multi-domain FKBPs suggests their likely role in protein-protein interactions (Kurek et al., 2002). Two of the homeologous groups of FKBPs, TaFKBP47-1-5A/TaFKBP47-2-5B/TaFKBP47-3-5D, and TaFKBP48-1-2A/TaFKBP53-1-2D/TaFKBP53-2-2B also contain an NPL-domain along with the FKBD. It is speculated that as reported for AtFKBP53, which contains an NPL-domain (Li and Luan, 2010), these wheat FKBPs might also be acting as histone chaperones, which however, requires experimental validation. Different types of structural motifs observed in TaFKBPs indicate towards their functional specificities. For instance, the presence of motifs 2 and 15 suggests their likely participation in protein-protein interactions, since other multi-domain FKBPs *viz.*, wFKBP73 and wFKBP77 in wheat (Reddy et al., 1998), AtFKBP62 in *A. thaliana* (Meiri and Breiman, 2009) and human FKBP52 (Pirkl and Buchner, 2001), which contain these sequences, were implicated in interaction with the molecular chaperone HSP90 *via* their TPR domain. Similarly, the existence of the motifs 11 and 12 in the N-terminus NPL-domain of two groups of homeologues (TaFKBP47-1-5A/TaFKBP47-2-5B/TaFKBP47-3-5D, and TaFKBP48-1-2A/TaFKBP53-1-2D/TaFKBP53-2-2B) suggests that these proteins may be acting as histone chaperons, as observed for their orthologous proteins AtFKBP53 in *A. thaliana* (Li and Luan, 2010), and FPR3 and FPR4 in *S. cerevisiae* (Benton et al., 1994; Kuzuhara and Horikoshi, 2004). The motif 10 (cytidylate kinase 2) was observed in several (17) FKBPs implying their likely role in pyrimidine biosynthesis, as reported for the cytidylate kinase protein Cmk (Wojtkiewicz et al., 2020). The Aspartic_Protease (ASP_Protease) motif predicted in a homeolog triplet TaFKBP48-1-2A/TaFKBP53-1-2D/TaFKBP53-2-2B indicate the role of these proteins in biotic and abiotic stress responses in plants, since these proteases are known to regulate the pathogenesis-related proteins (Rodrigo et al., 1989; Rodrigo et al., 1991), and nitrogen remobilization under water stress conditions (de Carvalho et al., 2001). The homeologues TaFKBP47-1-5A/TaFKBP47-2-5B/TaFKBP47-3-5D, predicted to harbour motif 14 (Domain of unknown function/CDC45 like protein), may be involved in cell cycle regulation, as demonstrated for the AtCDC45 in *A. thaliana* (Stevens et al., 2004).

All wheat parvulins showed the presence of the characteristic rotamase domain, but the multi-domain parvulins TaPar32-1-2D/TaPar32-2-2B/TaPar32-3-2A also contain an additional rhodanese-like domain towards C-terminus, akin to their *A. thaliana* ortholog AtPin3 (He et al., 2004). The different TaPars exhibited homeolog-dependent differences with respect to the different motifs. For instance, the occurrence of motif 12 was observed only in TaPar32-1-2D and TaPar32-3-2A, but not in their third homeolog TaPar32-2-2B. Similarly, motifs 3 (PPIC-type PPIase domain) and 6 (function not known) were observed most commonly (TaPar11-1-3B, TaPar11-2-4A, TaPar12-1-2D, TaPar13-1-2D, TaPar13-2-2A, TaPar13-3-2B), while Motifs 7 and 11 (functions not known) were noted only in few of these proteins (TaPar11-1-3B, TaPar11-2-4A). It is likely that the differential presence of the motifs might signify functional divergence in the roles of different parvulin homeologues.

Expression analysis of *TaFKBP*, *TaPar* and *TaPTPA* genes

The expression of *TaFKBP, TaPar* and *TaPTPA* genes was analysed in response to heat stress to understand their likely role in thermotolerance of wheat seedlings. These results revealed that the expression of different *TaFKBP* and *TaPar* genes in wheat is modulated differentially by heat stress, which is in accordance with the earlier studies in *O. sativa*, *Z. mays*, *M. domestica*, and *Penicillium* (Nigam et al., 2008; Yu et al., 2012; Dong et al., 2018; Waseem et al., 2018; Singh et al., 2021). For instance, the heat stress-induced increase in *TaFKBP61-2-7D/TaFKBP62-1-7A/TaFKBP62-2-7B* (Group 1) was also observed for their maize ortholog, *ZmFKBP62a* (Yu et al., 2012). Similarly, the increase in expression of *TaFKBP63-1-1D/TaFKBP67-1-1B/TaFKBP69-1-1A* (Group 2) after exposure to lethal heat stress following acclimation (37 °C-50 °C-4.5 h) is consistent with the temperature-induced increase in expression of *ZmFKBP72* in maize (Yu et al., 2012). Apart from the heat-upregulated members, the *TaFKBP* and *TaPar* families in wheat also possess genes that are downregulated under thermal stress, indicating functional divergence in their roles, which might be contributing to maintenance of homeostasis under stress conditions. The identification of commonly found heat shock elements in the promoter regions of heat-stress inducible genes, such as *TaFKBP20-1-6A* (Group 4) and *TaFKBP21-1-2D/TaFKBP27-1-2A* (Group 3), suggests their putative role in stress response, that, however, needs to be explored further by overexpression and knockdown approaches. Of the different *FKBP* genes analyzed, *TaFKBP61-2-7D/TaFKBP62-1-7A/TaFKBP62-2-7B* (Group 1) showed highest increase in transcript levels during acclimation

(27.7-fold) and lethal heat stress following acclimation (40.8-fold). These proteins are approximately 98.03% identical with the wFKBP73, that in addition to its PPIase activity was also shown to have chaperonic function (Supplementary Figure S2A; Kurek et al., 2002). Therefore, it is likely that as observed for the wheat cyclophilin *TaCYPA-1* (Kaur et al., 2016), the enhanced levels of *TaFKBP61-2-7D/TaFKBP62-1-7A/TaFKBP62-2-7B* under heat stress might be providing protection against heat-induced damage by preventing protein aggregation and mediating protein folding through their chaperone/PPIase activity. As observed for *TaFKBPs*, expression of *TaPars* was also affected differentially by heat stress, prompting us to speculate that besides developmental regulation, these proteins might have an important role in stress response of plants. Though expression of none of the three *TaPTPA* homeologues was detected in the present study, role of these genes can't be ruled out and, hence, requires further analysis.

Protein-Protein interaction network

The PPI networks of TaFKPBs, TaPars and TaPTPAs demonstrated that these proteins interact among themselves as well as with other proteins, suggesting their involvement in diverse biological functions (Figure 12). The PPI network of FKBPs showed that the TaFKBP65-1-2B and TaFKBP72-1-2D might be playing essential roles in response to abiotic stress (Figure 12A), since both of them share the network cluster with heat shock protein 90-1 (HSP90.1) and heat shock transcription factor A-2 (HSFA2), indicating their likely role in regulation of thermotolerance in plants, as reported for ROF1 in *A. thaliana* (AtFKBP62) (Meiri and Breiman, 2009). The role of TaFKBP65-1-2B and TaFKBP72-1-2D in thermotolerance is also substantiated by their enhanced expression in heat stress, as revealed by heatmap analysis (Figure 11A). Similarly, the putative interaction of TaFKBP42-1-5B with ABC transporter B family members ABCB1 and ABCB19 is consistent with the previous study reporting interaction of *Arabidopsis* ortholog AtFKBP42 (TWD1) with vacuolar ABC transporters, which is important for auxin transport (Geisler et al., 2003; Geisler et al., 2004).

Computational analysis also predicted the interaction between TaPar13-3-2B and serine/threonine- protein phosphatase 2α (PP2A), implying its role in regulation of cell cycle in plants (Yeh and Means, 2007; Figure 12B). The putative interaction between TaPTPA42-2-7A and HEAT (huntingtin-elongation-A subunit-TOR) repeat-containing protein (AT5G16210) indicates its association with the PP2A scaffolding subunit that comprises of 15 tandem repeats of HEAT motif (Walter et al., 1989; Hemmings et al., 1990; Figure 12C). Though the PPI network analyses indicate the role of TaFKBPs, TaPars and TaPTPA in different cellular functions, understanding the mechanisms underlying these processes requires detailed analysis of *in vivo* interactions by BiFC/FRET approaches.

Conclusion

To conclude, the genes encoding different FKBP, parvulin and PTPA proteins were identified for the first time in wheat genome. Further, characterization of gene structure, domain architecture, motif analysis, cis-regulatory elements and evolutionary analysis was also carried out. The results of this study suggest functional divergence of these proteins, that needs to be elucidated by experimental approaches. The expression profiling of these genes at the seedling stage revealed that the *TaFKBP* and *TaPar* genes in wheat are differentially regulated in response to heat stress, implying their role in stress response. The findings of this study will provide the basis for further functional characterization of these gene families in wheat, that may lead to their application in improvement of crop plants through breeding and biotechnological approaches.

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

PS: conceived idea, designed the experiments, reviewing and editing of the manuscript; HS: designed the bioinformatic study, participated in writing of manuscript, data analysis; AS: conducted bioinformatic analysis and performed qRT-PCR analysis of genes, writing of manuscript, data curation and analysis; KK: helped in qRT-PCR analysis of genes, data analysis; AK: contributed in writing of manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

We gratefully acknowledge Department of Biotechnology, Government of India, New Delhi for supporting the MSc. Biotechnology Teaching Program. AS is thankful to Department of Science and Technology (DST), Govt. of India for providing the DST-INSPIRE research fellowship (Registration number. IF190202; Sanction order: DST/INSPIRE Fellowship/2019/IF190202, Dated: 30-11-2020). KK acknowledges University Grant Commission, New Delhi, Government of India for the Rajiv Gandhi National Fellowship (Sanction number: F1-17.1/2017-18/RGNF-2017-18-SC-PUN-44307/)Sa-III/website, Dated: 28-07-2017). We also acknowledge the support provided by the Department of Science and Technology, Government of India in the form of the "Fund for Improvement of Science & Technology Infrastructure (FIST)" grant (D. O. No. SR/FST/PG College/2009) to develop computational

resources at the Department of Bioinformatics, Hans Raj Mahila MahaVidyalaya, Jalandhar. The authors acknowledge Dr. Sanjay Kumar, Director CSIR- Institute of Himalayan Bioresource Technology, Palampur for the institutional support. Thanks are also due to Prof. Wen Yao, National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan, China, College of Life Sciences, Henan Agricultural University, Zhengzhou, Chin for his help in fine tuning Circos plots developed through the shinyCircos application.

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.2022.1053524/full#supplementary-material

References

Aviezer-Hagai, K., Skovorodnikova, J., Galigniana, M., Farchi-Pisanty, O., Maayan, E., Bocovza, S., et al. (2007). *Arabidopsis* immunophilins ROF1 (AtFKBP62) and ROF2 (AtFKBP65) exhibit tissue specificity, are heat-stress induced, and bind HSP90. *Plant Mol. Biol.* 63, 237–255. doi: 10.1007/s11103-006-9085-z

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucl. Acids Res.* 43, W39–W49. doi: 10.1093/nar/gkv416

Benton, A. L., Sivan, A. B., Hamsher, K. D., Varney, N. R., and Spreen, O. (1994). *Contributions to neuropsychological assessment: A clinical manual* (USA: Oxford University Press).

Boudko, S. P., Ishikawa, Y., Nix, J., Chapman, M. S., and Bächinger, H. P. (2014). Structure of human peptidyl-prolyl *cis–trans* isomerase FKBP22 containing two EF-hand motifs. *Protein Sci.* 23, 67–75. doi: 10.1002/pro.2391

Breiman, A., Fawcett, T. W., Ghirardi, M. L., and Mattoo, A. K. (1992). Plant organelles contain distinct peptidylprolyl *cis-trans* isomerases. *J. Biol. Chem.* 267, 21293–21296. doi: 10.1002/pro.2391

Carol, R. J., Breiman, A., Erel, N., Vittorioso, P., and Bellini, C. (2001). PASTICCINO1 (AtFKBP70) is a nuclear-localised immunophilin required during *Arabidopsis thaliana* embryogenesis. *Plant Sci.* 161, 527–535. doi: 10.1016/S0168-9452(01)00437-X

Carvalho, M. H. C., d'Arcy-Lameta, A., Roy-Macauley, H., Gareil, M., El Maarouf, H., Pham-Thi, A. T., et al. (2001). Aspartic protease in leaves of common bean (*Phaseolus vulgaris* l.) and cowpea (*Vigna unguiculata* l. *Walp*): enzymatic activity, gene expression and relation to drought susceptibility. *FEBS Lett.* 492, 242–246. doi: 10.1016/S0014-5793(01)02259-1

Chen, Q., Chen, Q. J., Sun, G. Q., Zheng, K., Yao, Z. P., Han, Y. H., et al. (2019). Genome-wide identification of cyclophilin gene family in cotton and expression analysis of the fibre development in *Gossypium barbadense*. *Int. J. Mol. Sci.* 20, 349. doi: 10.3390/ijms20020349

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

Dong, Q., Mao, K., Duan, D., Zhao, S., Wang, Y., Wang, Q., et al. (2018). Genome-wide analyses of genes encoding FK506-binding proteins reveal their involvement in abiotic stress responses in apple. *BMC Genomics* 19, 1–16. doi: 10.1186/s12864-018-5097-8

Eddy, S. R. (2011). Accelerated profile HMM searches. *PloS Comp. Biol.* 7, e1002195. doi: 10.1371/journal.pcbi.1002195

Feldman, M., and Levy, A. A. (2005). Allopolyploidy–a shaping force in the evolution of wheat genomes. *Cytogenet. Genome Res.* 109, 250–258. doi: 10.1159/000082407

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The pfam protein families database: towards a more sustainable future. *Nucl. Acids Res.* 44, 279–285. doi: 10.1093/nar/gkv1344

Fischer, G., Wittmann-Liebold, B., Lang, K., Kiefhaber, T., and Schmid, F. X. (1989). Cyclophilin and peptidyl-prolyl *cis-trans* isomerase are probably identical proteins. *Nature* 337, 476–478. doi: 10.1038/337476a0

Galat, A., and Rivière, S. (1998). *Peptidyl-prolyl cis/trans isomerases*: Protein profile (Oxford University Press Oxford).

Gasser, C. S., Gunning, D. A., Budelier, K. A., and Brown, S. M. (1990). Structure and expression of cytosolic cyclophilin/peptidyl-prolyl *cis-trans* isomerase of higher plants and production of active tomato cyclophilin in. *Escherichia coli. Proc. Natl. Acad. Sci.* 87, 9519–9523. doi: 10.1073/pnas.87.24.9519

Geisler, M., and Bailly, A. (2007). Tete-a-tete: the function of FKBPs in plant development. *Trends Plant Sci.* 12, 465–473. doi: 10.1016/j.tplants.2007.08.015

Geisler, M., Girin, M., Brandt, S., Vincenzetti, V., Plaza, S., Paris, N., et al. (2004). Arabidopsis immunophilin-like TWD1 functionally interacts with vacuolar ABC transporters. *Mol. Bio. Cell.* 15, 3393–3405. doi: 10.1091/mbc.E03–11–0831

Geisler, M., Kolukisaoglu, H. U., Bouchard, R., Billion, K., Berger, J., Saal, B., et al. (2003). TWISTED DWARF1, a unique plasma membrane-anchored immunophilin-like protein, interacts with arabidopsis multidrug resistance-like transporters AtPGP1 and AtPGP19. *Mol. Bio. Cell.* 14, 4238–4249. doi: 10.1091/mbc.e02-10-0698

Ge, L., Zhang, K., Cao, X., Weng, Y., Liu, B., Mao, P., et al. (2020). Sequence characteristics of *Medicago truncatula* cyclophilin family members and function analysis of MsCYP20-3B involved in axillary shoot development. *Mol. Biol. Rep.* 47, 907–919. doi: 10.1007/s11033-019-05183-x

Goldberg, T., Hecht, M., Hamp, T., Karl, T., Yachdav, G., Ahmed, N., et al. (2014). LocTree3 prediction of localization. *Nucl. Acids Res.* 42, 350–355. doi: 10.1093/nar/gku396

Gollan, P. J., and Bhave, M. (2010). Genome-wide analysis of genes encoding FK506-binding proteins in rice. *Plant Mol. Biol.* 72, 1–16. doi: 10.1007/s11103-009-9547-1

Gopalan, G., He, Z., Balmer, Y., Romano, P., Gupta, R., Héroux, A., et al. (2004). Structural analysis uncovers a role for redox in regulating FKBP13, an immunophilin of the chloroplast thylakoid lumen. *Proc. Natl. Acad. Sci. U.S.A.* 101, 13945–13950. doi: 10.1073/pnas.0405240101

Göthel, S., and Marahiel, M. (1999). Peptidyl-prolyl *cis-trans* isomerases, a superfamily of ubiquitous folding catalysts. *CMLS Cell. Mol. Life Sci.* 55, 423–436. doi: 10.1007/s000180050299

Hanes, S. D., Shank, P. R., and Bostian, K. A. (1989). Sequence and mutational analysis of ESS1, a gene essential for growth in *Saccharomyces cerevisiae*. *Yeast* 5, 55–72. doi: 10.1002/yea.320050108

Hanhart, P., Thieß, M., Amari, K., Bajdzienko, K., Giavalisco, P., Heinlein, M., et al. (2017). Bioinformatic and expression analysis of the *Brassica napus l.* cyclophilins. *Sci. Rep.* 7, 1–17. doi: 10.1038/s41598-017-01596-5

Harrar, Y., Bellec, Y., Bellini, C., and Faure, J. D. (2003). Hormonal control of cell proliferation requires PASTICCINO genes. *Plant Physiol.* 132, 1217–1227. doi: 10.1002/yea.320050108

He, Z., Li, L., and Luan, S. (2004). Immunophilins and parvulins. superfamily of peptidyl prolyl isomerases in *Arabidopsis*. *Plant Physiol.* 134, 1248–1267. doi: 10.1104/pp.103.031005

Hemmings, B. A., Adams-Pearson, C., Maurer, F., Mueller, P., Goris, J., Merlevede, W., et al. (1990). Alpha.-and. beta.-forms of the 65-kDa subunit of protein phosphatase 2A have a similar 39 amino acid repeating structure. *Biochemistry* 29, 3166–3173. doi: 10.1021/bi00465a002

Holub, E. B. (2001). The arms race is ancient history in *Arabidopsis*, the wildflower. *Nat. Rev. Genet.* 2, 516–527. doi: 10.1038/35080508

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

Janssens, V., and Goris, J. (2001). Protein phosphatase 2A: a highly regulated family of serine/threonine phosphatases implicated in cell growth and signalling. *Biochem.* 353, 417–439. doi: 10.1042/0264-6021:3530417

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/koac131

Jordens, J., Janssens, V., Longin, S., Stevens, I., Martens, E., Bultynck, G., et al. (2006). The protein phosphatase 2A phosphatase activator is a novel peptidyl-prolyl *cis-trans*-isomerase. *J. Biol. Chem.* 281, 6349–6357. doi: 10.1074/jbc.M507760200

Joseph, J. D., Heitman, J., and Means, A. R. (1999). Molecular cloning and characterization of *Aspergillus nidulans* cyclophilin b. *Fungal Genet. Biol.* 27, 55–66. doi: 10.1006/fgbi.1999.1131

Juturu, V. N., Ramesh, P., Gopalakrishna, M., Mallikarjuna, G., Talakayala, A., Puli, C. O. R., et al. (2021). Pigeonpea cyclophilin (CcCYP) enhances abiotic stress tolerance in rice. *Rice Sci.* 28, 3.

Kang, C. B., Hong, Y., Dhe-Paganon, S., and Yoon, H. S. (2008). FKBP family proteins: immunophilins with versatile biological functions. *Neurosignals* 16, 318–325. doi: 10.1159/000123041

Kaur, G., Singh, S., Dutta, T., Kaur, H., Singh, B., Pareek, A., et al. (2016). The peptidyl-prolyl *cis-trans* isomerase activity of the wheat cyclophilin, TaCypA-1, is essential for inducing thermotolerance in *Escherichia coli*. *Biochimie Open* 2, 9–15. doi: 10.1016/j.biopen.2015.11.003

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

Kurek, I., Pirkl, F., Fischer, E., Buchner, J., and Breiman, A. (2002). Wheat FKBP73 functions *in vitro* as a molecular chaperone independently of its peptidyl prolyl *cis-trans* isomerase activity. *Planta* 215, 119–126. doi: 10.1007/s00425-001-0722-0

Kuzuhara, T., and Horikoshi, M. (2004). A nuclear FK506-binding protein is a histone chaperone regulating rDNA silencing. *Nat. Struct. Mol. Biol.* 11, 275–283. doi: 10.1038/nsmb733

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucl. Acids Res.* 44, W242–W245. doi: 10.1093/nar/gkw290

Li, H., and Luan, S. (2010). AtFKBP53 is a histone chaperone required for repression of ribosomal RNA gene expression in *Arabidopsis*. *Cell Res.* 20, 357–366. doi: 10.1038/cr.2010.22

Lima, A., Lima, S., Wong, J. H., Phillips, R. S., Buchanan, B. B., and Luan, S. (2006). A redox-active FKBP-type immunophilin functions in accumulation of the photosystem II supercomplex in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12631–12636. doi: 10.1073/pnas.0605452103

Liu, J., Farmer, Jr. J.D., Lane, W. S., Friedman, J., Weissman, I., and Schreiber, S. L. (1991). Calcineurin is a common target of cyclophilin-cyclosporin a and FKBP-FK506 complexes. *Cell* 66, 807–815. doi: 10.1016/0092-8674(91)90124-h

Liu, H., Shen, J., Yuan, C., Lu, D., Acharya, B. R., Wang, M., et al. (2021). The cyclophilin ROC3 regulates ABA-induced stomatal closure and the drought stress response of *Arabidopsis thaliana*. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.668792

Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z., et al. (2015). IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics* 31, 3359–3361. doi: 10.1093/bioinformatics/btv362

Liu, Z., Xin, M., Qin, J., Peng, H., Ni, Z., Yao, Y., et al. (2015). Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and sdrought acclimation in wheat (*Triticum aestivum l.*). *BMC Plant Biol.* 5, 152. doi: 10.1186/s12870-015-0511-8

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

Luan, S., Albers, M. W., and Schreiber, S. L. (1994). Light-regulated, tissue-specific immunophilins in a higher plant. *Proc. Natl. Acad. Sci. U.S.A.* 91, 984–988. doi: 10.1073/pnas.91.3.984

Luan, S., Kudla, J., Gruissem, W., and Schreiber, S. L. (1996). Molecular characterization of a FKBP-type immunophilin from higher plants. *Proc. Natl. Acad. Sci. U.S.A.* 93, 6964–6969. doi: 10.1073/pnas.93.14.6964

Lu, Y. F., Tomizawa, K., Moriwaki, A., Hayashi, Y., Tokuda, M., Itano, T., et al. (1996). Calcineurin inhibitors, FK506 and cyclosporin a, suppress the NMDA receptor-mediated potentials and LTP, but not depotentiation in the rat hippocampus. *Brain Res.* 729, 142–146. doi: 10.1016/0006-8993(96)00568-9

Magnusdottir, A., Stenmark, P., Flodin, S., Nyman, T., Hammarstrom, M., Ehn, M., et al. (2006). The crystal structure of a human PP2A phosphatase activator reveals a novel fold and highly conserved cleft implicated in protein-protein interactions. *J. Biol. Chem.* 281, 22434–22438. doi: 10.1074/jbc.C600100200

Mainali, H. R., Chapman, P., and Dhaubhadel, S. (2014). Genome-wide analysis of cyclophilin gene family in soybean (*Glycine max*). *BMC Plant Biol.* 14, 282. doi: 10.1186/s12870-014-0282-7

Meiri, D., and Breiman, A. (2009). *Arabidopsis* ROF1 (FKBP62) modulates thermotolerance by interacting with HSP90.1 and affecting the accumulation of HsfA2-regulated sHSPs. *Plant J.* 59, 387–399. doi: 10.1111/j.1365-313X.2009.03878.x

Meiri, D., Tazat, K., Cohen-Peer, R., Farchi-Pisanty, O., Aviezer-Hagai, K., Avni, A., et al. (2010). Involvement of *Arabidopsis* ROF2 (FKBP65) in thermotolerance. *Plant Mol. Biol.* 72, 191–203. doi: 10.1007/s11103-009-9561-3

Mueller, J. W., Kessler, D., Neumann, D., Stratmann, T., Papatheodorou, P., Hartmann-Fatu, C., et al. (2006). Characterization of novel elongated parvulin isoforms that are ubiquitously expressed in human tissues and originate from alternative transcription initiation. *BMC Mol. Biol.* 7, 1–9. doi: 10.1186/1471-2199-7-9

Myers, T. A., Chanock, S. J., and Machiela, M. J. (2020). LDlinkR: an R package for rapidly calculating linkage disequilibrium statistics in diverse populations. *Front. Genet* 11, 157. doi: 10.3389/fgene.2020.00157

Nigam, N., Singh, A., Sahi, C., Chandramouli, A., and Grover, A. (2008). SUMO-conjugating enzyme (Sce) and FK506-binding protein (FKBP) encoding rice (*Oryza sativa* l.) genes: genome-wide analysis, expression studies and evidence for their involvement in abiotic stress response. *Mol. Genet. Genom.* 279, 371–383. doi: 10.1007/s00438-008-0318-5

O'Keefe, S. J., Tamura, J. I., Kincaid, R. L., Tocci, M. J., and O'Neill, E. A. (1992). FK-506-and CsA-sensitive activation of the interleukin-2 promoter by calcineurin. *Nature* 357, 692–694. doi: 10.1038/357692a0

Olson, S. A. (2002). Emboss opens up sequence analysis. *Brief. Bioinform.* 3, 87–91. doi: 10.1093/bib/3.1.87

Park, H. J., You, Y. N., Lee, A., Jung, H., Jo, S. H., Oh, N., et al. (2020). OsFKBP20-1b interacts with the splicing factor OsSR45 and participates in the environmental stress response at the post-transcriptional level in rice. *Plant J.* 102, 992–1007. doi: 10.1111/tpj.14682

Pearce, S., Vazquez-Gross, H., Herin, S. Y., Hane, D., Wang, Y., Gu, Y. Q., et al. (2015). WheatExp: an RNA-seq expression database for polyploid wheat. *BMC Plant Biol.* 15, 299. doi: 10.1186/s12870-015-0692-1

Pirkl, F., and Buchner, J. (2001). Functional analysis of the Hsp90-associated human peptidyl prolyl *cis-trans* isomerases FKBP51, FKBP52 and Cyp40. *J. Mol. Biol.* 308, 795–806. doi: 10.1006/jmbi.2001.4595

Qazi, S. R., Ahmad, S., and Shakeel, S. N. (2020). HSEAT: a tool for plant heat shock element analysis, motif identification and analysis. *Curr. Bioinform.* 15, 196–203. doi: 10.2174/1574893614666190102151956

Rahfeld, J. U., Rücknagel, K. P., Schelbert, B., Ludwig, B., Hacker, J., Mann, K., et al. (1994). Confirmation of the existence of a third family among peptidyl-prolyl *cis-trans* isomerases amino acid sequence and recombinant production of parvulin. *FEBS Lett.* 352, 180–184. doi: 10.1016/0014-5793(94)00932-5

Reddy, R. K., Kurek, I., Silverstein, A. M., Chinkers, M., Breiman, A., and Krishna, P. (1998). High-molecular-weight FK506-binding proteins are components of heat-shock protein 90 heterocomplexes in wheat germ lysate. *Plant Physiol.* 118, 1395–1401. doi: 10.1104/pp.118.4.1395

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucl. Acids Res.* 42, W320–W324. doi: 10.1093/nar/gku316

Rodrigo, I., Vera, P., and Conejero, V. (1989). Degradation of tomato pathogenesis-related proteins by an endogenous aspartyl endoproteinase. *Eur. J. Biochem.* 184, 663–669. doi: 10.1111/j.1432-1033.1989.tb15064.x

Rodrigo, I., Vera, P., Van Loon, L. C., and Conejero, V. (1991). Degradation of tobacco pathogenesis-related proteins: evidence for conserved mechanisms of degradation of pathogenesis-related proteins in plants. *Plant Physiol.* 95, 616–622. doi: 10.1104/pp.95.2.616

Romano, P. G. N., Horton, P., and Gray, J. E. (2004). The arabidopsis cyclophilin gene family. *Plant Physiol.* 134, 1268–1282. doi: 10.1104/pp.103.022160

Schöffl, F., Prandl, R., and Reindl, A. (1998). Regulation of the heat-shock response. *Plant Physiol.* 117, 1135–1141. doi: 10.1104/pp.117.4.1135

Shakeel, S., Haq, N. U., Heckathorn, S. A., Hamilton, E. W., and Luthe, D. S. (2011). Ecotypic variation in chloroplast small heat-shock proteins and related thermotolerance in *Chenopodium album*. *Plant Physiol. Biochem.* 49, 898–908. doi: 10.1016/j.plaphy.2011.05.002

Shao, W., Chen, W., Zhu, X., Zhou, X., Jin, Y., Zhan, C., et al. (2021). Genome-wide identification and characterization of wheat 14-3-3 genes unravels the role of TaGRF6-a in salt stress tolerance by binding MYB transcription factor. *Int. J. Mol. Sci.* 22, 1904. doi: 10.3390/ijms22041904

Sharma, A. D., and Singh, P. (2003). Effect of water stress on expression of a 20 kD cyclophilin-like protein in drought susceptible and tolerant cultivars of sorghum. *J. Plant Biochem. Biotechnol.* 12, 77–80. doi: 10.1007/BF03263165

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol. Syst. Biol.* 7, 539. doi: 10.1038/msb.2011.75

Sigrist, C. J., De Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., et al. (2012). New and continuing developments at PROSITE. *Nucl. Acids Res.* 41, 344–347. doi: 10.1093/nar/gks1067

Singh, A. K., Datta, A., Jobichen, C., Luan, S., and Vasudevan, D. (2020). AtFKBP53: a chimeric histone chaperone with functional nucleoplasmin and PPIase domains. *Nucl. Acids Res.* 48, 1531–1550. doi: 10.1093/nar/gkz1153

Singh, M., Kaur, K., Sharma, A., Kaur, R., Joshi, D., Chatterjee, M., et al. (2021). Genome-wide characterization of peptidyl-prolyl *cis-trans* isomerases in *Penicillium* and their regulation by salt stress in a halotolerant *P. oxalicum*. *Sci. Rep.* 11, 1–19. doi: 10.1038/s41598-021-91602-8

Singh, H., Kaur, K., Singh, S., Kaur, P., and Singh, P. (2019). Genome-wide analysis of cyclophilin gene family in wheat and identification of heat stress responsive members. *Plant Gene* 19, 100197. doi: 10.1016/J.PLGENE.2019.100197

Singh, H., Kaur, K., Singh, M., Kaur, G., and Singh, P. (2020). Plant cyclophilins: Multifaceted proteins with versatile roles. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.585212

Singh, K., Zouhar, M., Mazakova, J., and Rysanek, P. (2014). Genome wide identification of the immunophilin gene family in *Leptosphaeria maculans*: a causal agent of blackleg disease in oilseed rape (*Brassica napus*). *Omics J. Integr. Biol.* 18, 645–657. doi: 10.1089/omi.2014.0081

Stevens, R., Grelon, M., Vezon, D., Oh, J., Meyer, P., Perennes, C., et al. (2004). A CDC45 homolog in *Arabidopsis* is essential for meiosis, as shown by RNA interference-induced gene silencing. *Plant Cell* 16, 99–113. doi: 10.1105/tpc.016865

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucl. Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131

Thirumalaikumar, V. P., Gorka, M., Schulz, K., Masclaux-Daubresse, C., Sampathkumar, A., Skirycz, A., et al. (2021). Selective autophagy regulates heat stress memory in arabidopsis by NBR1-mediated targeting of HSP90. 1 and ROF1. *Autophagy* 17, 2184–2199. doi: 10.1080/15548627.2020.1820778

Trivedi, D. K., Yadav, S., Vaid, N., and Tuteja, N. (2012). Genome wide analysis of cyclophilin gene family from rice and arabidopsis and its comparison with yeast. *Plant Signal. Behav.* 7, 1653–1666. doi: 10.4161/psb.22306

Unger, T., Dym, O., Albeck, S., Jacobovitch, Y., Bernehim, R., Marom, D., et al. (2010). Crystal structure of the three FK506 binding protein domains of wheat FKBP73: evidence for a unique wFK73_2 domain. *J. Struct. Funct. Genomics* 11, 113–123. doi: 10.1007/s10969-010-9085-8

Van Hoof, C., and Goris, J. (2003). Phosphatases in apoptosis: to be or not to be, PP2A is in the heart of the question. *Biochim. Biophys. Acta* 1640, 97–104. doi: 10.1016/S0167-4889(03)00029-6

Voorrips, R. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *Heredity* 93, 77–78. doi: 10.1093/jhered/93.1.77

Vucich, V. A., and Gasser, C. S. (1996). Novel structure of a high molecular weight FK506 binding protein from *Arabidopsis thaliana*. *Mol. Gen. Genet.* 252, 510–517. doi: 10.1007/BF02172397

Walter, G., Ferre, F., Espiritu, O., and Carbone-Wiley, A. (1989). Molecular cloning and sequence of cDNA encoding polyoma medium tumor antigen-associated 61-kDa protein. *Proc. Natl. Acad. Sci.* 86, 8669–8672. doi: 10.1073/pnas.86.22.8669

Wang, W. W., Ma, Q., Xiang, Y., Zhu, S. W., and Cheng, B. J. (2012). Genome-wide analysis of immunophilin FKBP genes and expression patterns in *Zea mays*. *Genet. Mol. Res.* 11, 1690–1700. doi: 10.4238/2012.June.25.2

Wang, X. F., Qiao, Z. W., Wang, D. R., Wang, X., and You, C. X. (2022). Genome-wide identification and stress response analysis of cyclophilin gene family in apple (*Malus domestica*).

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. *Nucl. Acids Res.* 40, 49–49. doi: 10.1093/nar/gkr1293

Wang, W., Vinocur, B., and Altman, A. (2003). Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* 218, 1–14. doi: 10.1007/s00425-003-1105-5

Wang, C., Wang, G., Qu, X., Zhang, X., Ji, W., and Zhang, H. (2021). Genome-wide identification and expression analysis of the invertase gene family in common wheat. *bioRxiv*. doi: 10.1101/2021.12.29.474404

Waseem, M., Ahmad, F., Habib, S., Gao, Y., and Li, Z. (2018). Genome-wide identification of FK506-binding domain protein gene family, its characterization, and expression analysis in tomato (*Solanum lycopersicum* l.). *Gene* 678, 143–154. doi: 10.1016/j.gene.2018.08.021

Weber, H., Chételat, A., Reymond, P., and Farmer, E. E. (2004). Selective and powerful stress gene expression in *Arabidopsis* in response to malondialdehyde. *Plant J.* 37, 877–888. doi: 10.1111/j.1365-313x.2003.02013.x

Wojtkiewicz, P., Biernacka, D., Gorzelak, P., Stupak, A., Klein, G., and Raina, S. (2020). Multicopy suppressor analysis of strains lacking cytoplasmic peptidyl-prolyl *cis-trans* isomerases identifies three new PPIase activities in *Escherichia coli* that includes the DksA transcription factor. *Int. J. Mol. Sci.* 21, 5843. doi: 10.3390/ijms21165843

Xu, Q., Liang, S., Kudla, J., and Luan, S. (1998). Molecular characterization of a plant FKBP12 that does not mediate action of FK506 and rapamycin. *Plant J.* 15, 511–519. doi: 10.1046/j.1365-313x.1998.00232.x

Yadav, S., Centola, M., Glaesmann, M., Pogoryelov, D., Ladig, R., Heilemann, M., et al. (2022). Cyclophilin anaCyp40 regulates photosystem assembly and phycobilisome association in a cyanobacterium. *Nat. Commun.* 13, 1–17. doi: 10.1038/s41467-022-29211-w

Yeh, E., and Means, A. (2007). PIN1, the cell cycle and cancer. *Nat. Rev. Cancer* 7, 381–388. doi: 10.1038/nrc2107

Yu, Y., Li, Y., Jia, F., Zhao, M., Li, W., Sun, Q., et al. (2017). ZmFKBP20-1 improves the drought and salt tolerance of transformed. *Arabidopsis. J. Plant Biol.* 60, 558–570. doi: 10.1007/s12374-017-0262-1

Yu, Y., Zhang, H., Li, W., Mu, C., Zhang, F., Wang, L., et al. (2012). Genome-wide analysis and environmental response profiling of the FK506-binding protein gene family in maize (*Zea mays* l.). *Gene* 498, 212–222. doi: 10.1016/j.gene.2012.01.094

Zhang, X., Deng, Y., Chang, H., Ji, C., Zhang, M., Peng, J., et al. (2014). The specific and rapid labeling of cell surface proteins with recombinant FKBP-fused fluorescent proteins. *Protein Cell.* 5, 800–803. doi: 10.1007/s13238-014-0090-8

Zhang, Y., Han, J., Liu, D., Wen, X., Li, Y., Tao, R., et al. (2014). Genome-wide identification and analysis of FK506-binding protein gene family in peach (*Prunus persica*). *Gene* 536, 416–424. doi: 10.1016/j.gene.2013.10.059

Zhou, Q. Q., Wang, Y., Hu, J. J., Zhang, L., Li, J. B., Xu, Y. J., et al. (2022). Identification and characterization of a cyclophilin a gene from Chinese shrimp fenneropenaeus chinensis: sequence features and expression profiles. *Invertebr. Surviv. J.*, 19, 105–114. doi: 10.25431/1824-307X/isj.v19i1.105-114

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 22 December 2022
DOI 10.3389/fpls.2022.1087839

OPEN ACCESS

EDITED BY
Mehanathan Muthamilarasan,
University of Hyderabad, India

REVIEWED BY
Roshan Singh,
National Institute of Plant Genome Research (NIPGR), India
Pooja Choudhary,
Jaypee Institute of Information Technology, India
Sumi Rana,
University of Hyderabad, India

*CORRESPONDENCE
Wenxing Liu
liuwx@qau.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 02 November 2022
ACCEPTED 05 December 2022
PUBLISHED 22 December 2022

CITATION
Feng X, Meng Q, Zeng J, Yu Q, Xu D, Dai X, Ge L, Ma W and Liu W (2022) Genome-wide identification of sucrose non-fermenting-1-related protein kinase genes in maize and their responses to abiotic stresses. *Front. Plant Sci.* 13:1087839. doi: 10.3389/fpls.2022.1087839

Genome-wide identification of sucrose non-fermenting-1-related protein kinase genes in maize and their responses to abiotic stresses

Xue Feng[1†], Quan Meng[1†], Jianbin Zeng[1], Qian Yu[1], Dengan Xu[1], Xuehuan Dai[1], Lei Ge[1], Wujun Ma[1,2] and Wenxing Liu[1,3*]

[1]College of Agronomy, Qingdao Agricultural University, Qingdao, China, [2]State Agricultural Biotechnology Centre, College of Science, Health, Engineering and Education, Murdoch University, Perth, WA, Australia, [3]The Key Laboratory of the Plant Development and Environmental Adaptation Biology, Ministry of Education, School of Life Sciences, Shandong University, Qingdao, Shandong, China

Introduction: Protein kinases play an important role in plants in response to environmental changes through signal transduction. As a large family of protein kinases, sucrose non-fermenting-1 (SNF1)-related kinases (SnRKs) were found and functionally verified in many plants. Nevertheless, little is known about the *SnRK* family of *Zea mays*.

Methods: Evolutionary relationships, chromosome locations, gene structures, conserved motifs, and cis-elements in promoter regions were systematically analyzed. Besides, tissue-specific and stress-induced expression patterns of *ZmSnRKs* were determined. Finally, functional regulatory networks between *ZmSnRKs* and other proteins or miRNAs were constructed.

Results and Discussion: In total, 60 *SnRK* genes located on 10 chromosomes were discovered in maize. *ZmSnRKs* were classified into three subfamilies (*ZmSnRK1*, *ZmSnRK2*, and *ZmSnRK3*), consisting of 4, 14, and 42 genes, respectively. Gene structure analysis showed that 33 of the 42 *ZmSnRK3* genes contained only one exon. Most *ZmSnRK* genes contained at least one ABRE, MBS, and LTR cis-element and a few *ZmSnRK* genes had AuxRR-core, P-box, MBSI, and SARE ciselements in their promoter regions. The Ka:Ks ratio of 22 paralogous *ZmSnRK* gene pairs revealed that the *ZmSnRK* gene family had experienced a purifying selection. Meanwhile, we analyzed the expression profiles of *ZmSnRKs*, and they exhibited significant differences in various tissues and abiotic stresses. In addition, A total of eight ZmPP2Cs, which can interact with *ZmSnRK* proteins, and 46 miRNAs, which can target 24 *ZmSnRKs*, were identified. Generally, these results provide valuable information for further function verification of *ZmSnRKs*, and improve our understanding of the role of *ZmSnRKs* in the climate resilience of maize.

KEYWORDS

maize (*Zea mays*), SnRK, genome-wide, expression patterns, climate-resilience

Introduction

Terrestrial plants must cope with environmental changes including drought; excess salt, heat, cold, and toxic metals; and nutrient deficiency (Zhu, 2016). Protein phosphatase and kinase-regulated signal transduction play a vital role in climate resilience in plants (Hunter, 1995; Cohen, 1988). In recent years, several protein kinases have been identified, including receptor protein kinase (RLK) (Tör et al., 2009), mitogen-activated protein kinase (MAPK) (Zhang and Zhang, 2022), calcium-dependent protein kinase (CDPK) (Bender et al., 2018), and sucrose non-fermenting-1 (SNF1)-related kinase (SnRK) (Maszkowska et al., 2021).

SnRKs in plants have a high sequence identity with SNF1 of fungi and AMP-activated protein kinase (AMPK) of mammals (Halford and Hardie, 1998). The first plant SnRK genes were identified in *Secale cereale*, and showed 48% amino acid homology with SNF1 and AMPK (Alderson et al., 1991). In plants, the SnRK family has expanded and diverged into three subfamilies (SnRK1, SnRK2, and SnRK3) (Hrabak et al., 2003). The N-terminal of three SnRK subfamilies have a conserved domain with serine/threonine protein kinase, while the C-terminal displayed variation. In detail, SnRK1s were most similar to SNF1 and AMPK in terms of structure and function. SnRK2s contain acidic amino acids in the C-terminal domain, either Glu or Asp (Kulik et al., 2011). SnRK3s were widely recognized as CIPKs (CBL-interacting protein kinases) (Tominaga et al., 2010), which have two conserved domains in the C-terminal: NAF and PPI (Masaru et al., 2003; Albrecht et al., 2014).

SnRK1s are evolutionarily conserved energy-sensing protein kinases in plants. When energy supplies become limited, SnRK1s can activate transcriptional regulatory networks and maintain cellular energy homeostasis (Wurzinger et al., 2018; Wang et al., 2021). Therefore, SnRK1s were known as conserved energy sensors. For example, Wang et al. (2021) reported that the α-subunit of OsSnRK1 could phosphorylate JMJ705 to stimulate its demethylase activity, thus targeting and activating a series of low-energy stress-responsive transcription factors. SnRK1 phosphorylates the transcription factor basic leucine zipper63 (bZIP63), which could directly bind the promoter of auxin response factor19 (ARF19) and activate gene expression to mediate lateral root development (Muralidhara et al., 2021). In peach fruit, SnRK1 could phosphorylate sorbitol dehydrogenase and lead to sugar accumulation (Yu et al., 2021). In maize, SnRK1 could coordinate carbon and nitrogen absorption in seeds by phosphorylating E3 ubiquitin ligase ZmRFWD3 (Li et al., 2020).

SnRK2s are much smaller than SnRK1 and are associated with the ABA signal transduction pathway (Cutler et al., 2010; Maszkowska et al., 2021). In rice, a mutation of *OsSAPK8*, a member of SnRK2, exhibited sensitivity to diverse abiotic stresses at the vegetative stages, such as drought stresses, excess of salt, and low temperature (Zhong et al., 2020). In Arabidopsis, overexpression of *AtSnRK2.8* showed higher resistance to drought. *AtSnRK2.10* was induced by salt stress and was involved in salt tolerance through increasing photosynthesis and protecting against oxidative damage (Mazur et al., 2021). The activity of SnRK2.4 was significantly increased in roots after 30s of salt exposure, but it disappeared 1h later (Mcloughlin et al., 2012). In rice, SnRK2.6/OST1 was immediately activated in an ABA non-responsive pathway to elevate low temperature adaptation (Ding et al., 2015).

SnRK3s could combine with Ca^{2+}-dependent CBL to regulate ion transport and enhance environmental adaption in plants, especially in abiotic stress. For instance, AtCIPK24 protein could interact with AtCBL4 to phosphorylate and activate plasma membrane H^+/Na^+ antiporter (SOS1) in a Ca^{2+}-responsive manner in Arabidopsis (Qiu et al., 2002). In rice, the expression of *OsCIPK31* was up-regulated under several abiotic stresses during germination and seedling growth (Piao et al., 2010). Ma et al. (2021) indicated that CaCIPK3 positively regulated drought tolerance by the CBL-CIPK network, thus mediating MeJA signaling and the antioxidant defense system. Moreover, the OsCBL1-OsCIPK23 complex could enhance OsAKT1-mediated K^+ uptake (Li et al., 2014).

Maize is the most commonly cultivated food crop globally. However, research concerning the role of *SnRK* genes is still rare in maize. With the rapid progress of molecular biology and the whole-genome sequencing of maize (Myburg et al., 2014), it becomes possible for us to have a comprehensive understanding of *ZmSnRK* gene families. Here, 60 candidate *ZmSnRK* genes were discovered by bioinformatic analysis. Then, evolutionary relationships, chromosome locations, gene structures, conserved motifs, and cis-elements in promoter regions were systematically analyzed. Besides, tissue-specific and stress-induced expression patterns of *ZmSnRKs* were determined. Finally, functional regulatory networks between ZmSnRKs and other proteins or miRNAs were constructed. All these results offer valuable information for further function verification of *ZmSnRKs* and help to better comprehend the role of *ZmSnRKs* in the climate resilience of maize.

Materials and methods

Identification and general characterization analysis of ZmSnRK proteins

ZmSnRK proteins were searched using query probes of AtSnRKs and OsSnRKs by BLASTP. We used Zm-B73 V5.0 database from the MaizeGDB (https://maizegdb.org/) to conduct BLASP. SnRKs in Arabidopsis and rice were extracted

from ensemble plants (https://plants.ensembl.org/). Meanwhile, the HMMER 3.0 was utilized to search the ZmSnRKs protein sequence according to the Hidden Markov Model (HMM) files of SnRKs acquired from the Pfam database (http://pfam.xfam.org/). The SMART database was used to reconfirm sequences (Letunic et al., 2012). The molecular weights and isoelectric points of the proteins were analyzed through the ExPASy tool (www.expasy.org/tools/). Subcellular localization was also predicted through WoLF PSORT (https://wolfpsort.hgc.jp/).

Phylogenetic analysis and classification of ZmSnRKs

Multiple homology alignment of 60 ZmSnRK non-redundant amino acid sequences and SnRK protein sequences from *Arabidopsis thaliana*, *Oryza sativa*, and *Zea mays* were conducted by ClustalW with default parameters (Larkin et al., 2007). Then, we used MEGA 7.0 to construct a phylogenetic tree *via* neighbor-joining (NJ) method (Kumar et al., 2016), with the following parameters: Poisson model, pairwise deletion, and 1,000 bootstrap replications.

Chromosome localization and collinearity analysis of *ZmSnRKs*

Chromosome localization of *ZmSnRKs* was conducted using MapChart v. 2.3 Software (Voorrips, 2002). The collinearity of the orthologous *SnRK* genes between *Zea mays* and other species, such as *Arabidopsis thaliana*, *Oryza sativa*, and *Hordeum vulgare*, was determined through MCScanX software (Wang et al., 2012). The non-synonymous substitution rate (Ka) and synonymous substitution rate (Ks) of ZmSnRKs in the gene collinearity were calculated by using the ParaAT tool (Zhang et al., 2012). The Ka/Ks values were calculated by Calculator 2.0 software (Wang et al., 2010). The gene duplicated time was assessed based on $Ks/2\lambda$ (Wang et al., 2010), where $\lambda = 1.5 \times 10^{-8}$.

Analysis of gene structure and motifs of *ZmSnRKs*

The exon and intron arrangement of *ZmSnRK* genes was determined *via* the Gene Structure Display Server (GSDS) tool (http://gsds.cbi.pku.edu.cn/) (Hu et al., 2015). The conserved motif sequences of ZmSnRK proteins were identified using Multiple Expectation Maximization for Motif Elicitation (MEME) online program (http://meme.sdsc.edu/meme/itro.html) (Bailey et al., 2009). InterProScan (www.ebi.ac.uk/Tools/pfa/iprscan/) was utilized to annotate the motifs (Jones et al., 2014). The gene structures and conserved motifs of *ZmSnRKs* were visualized through TBtools (Chen et al., 2020).

Predication of cis-acting elements in the promoter regions of *ZmSnRKs*

Promoter sequences (-1500 bp) of *ZmSnRK* genes were obtained from the maize B73 genome. The cis-elements were determined by PlantCARE software (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Magali et al., 2002). We then use TBtools to visualize the cis-elements (Chen et al., 2020) associated with abiotic stress, including abscisic acid response components ABRE, drought response components MBS, auxin response components AuxRR-core, flavonoid biosynthetic genes regulation components MBSI, gibberellin-response components P-box, salicylic acid response components SARE, and low-temperature response compositions LTR.

Modeling of 3D structures of ZmSnRKs

The 3D structures of ZmSnRK proteins were predicted using Swiss-Model (https://swissmodel.expasy.org/interactive/) (Arnold et al., 2006). Then, the quality of 3D protein structure was assessed *via* SAVES server (http://nihserver.mbi.ucla.edu/SAVES/).

Expression profiles and interaction networks analysis of ZmSnRKs

The specific expression patterns of *SnRKs* from various tissues and under abiotic stress in the maize B73 were obtained from qTeller in MaizeGDB (https://qteller.maizegdb.org/) (Opitz et al., 2014; Stelpflug et al., 2016). We used transcript per million (TPM) reads mapped to represent the gene expression values. The genes with $|\log2 \text{ ratio}| \geq 1$ were recognized as differentially expressed genes (DEGs). A heatmap was conducted by TBtool to exhibit expression levels in various tissues ($\log_2{}^{(TPM+1)}$)and different abiotic stresses ($\log_2{}^{\text{fold change}}$). Moreover, the protein interactions were predicted by the Search Tool for the Retrieval of Interacting Genes (STRING) (http://string-db.org/cgi) (Damian et al., 2021). We used Cytoscape software to create interaction networks (Shannon et al., 2003). A low confidence (0.15) was used as the "minimum required interaction score" when analyzing protein interaction within ZmSnRKs, and high confidence was used as "maximum required interaction score" when analyzing protein interaction between ZmSnRKs and ZmPP2Cs.

Prediction of microRNAs-*ZmSnRKs* regulatory networks

The microRNA (miRNA)-*ZmSnRK* regulatory networks were predicted on the basis of the previous methods (Su et al., 2021). In brief, the *ZmSnRK*-targeting miRNAs were predicted

with coding sequences of *ZmSnRKs* through the psRNATarget server (http://plantgrn.noble.org/psRNATarget/home) under default parameters except for that penalty for G:U pair was 1 and mismatches allowed in seed region was 0. The miRNA-targeted sites were those highly complementary to coding sequences of *ZmSnRKs*. The interaction networks were created using the Cytoscape V3.9.1 software (https://cytoscape.org/download.html).

Plant growth and treatment

The hydroponic experiment was conducted in a greenhouse at Qingdao Agricultural University, Qingdao, China. Maize inbred line B73 was used in this study. Seeds of maize were treated with 3% H_2O_2 for 10 min, and were rinsed seven times with distilled water. The seeds were sown in a controlled environment with a day/night temperature of 28/22°C on moist filter papers. After germination, uniform seedlings were transferred to 2 L pots containing 1.5 L of basic nutrient solution (BNS). On the seventh day after transplanting, PEG was added to the containers to form three treatments: BNS (control), BNS plus 10% PEG, and BNS plus 100 mM NaCl. The composition of BNS was (mg L^{-1}) $MgSO_4{\cdot}7H_2O$, 160.21; K_2SO_4, 130.7; KCl, 7.455; KH_2PO_4, 34.02; $Ca(NO_3)_2{\cdot}4H_2O$, 472.3; Fe(III)-EDTA-Na, 36.7; $MnSO_4{\cdot}H_2O$, 0.17; $ZnSO_4{\cdot}7H_2O$, 0.29; $CuSO_4{\cdot}5H_2O$, 0.025; $(NH_4)_6Mo_7O_{24}{\cdot}4H_2O$, 0.006; and HBO_3, 0.062. The solution pH was adjusted to 5.8–6.0 with NaOH or HCl, as required. Plant leaves and stems were sampled on the seventh day after transplanting. Plant root samples for RNA isolation were collected 24h after treatment. All samples were stored at −80°C for downstream analysis.

Quantitative RT–PCR validation

Quantitative real-time PCR (qRT-PCR) was performed by a QuantStudio3 PCR system (Thermo, USA). First-strand cDNA synthesis was performed using the TOROBlue® qRT Premix with gDNA Eraser 2.0 (TOROIVD, China), followed by qRT-PCR using a TOROGreen® qPCR Master Mix (TOROIVD, China) with *ZmActin* as a reference. The total PCR volume was 20 µl. The reaction process was 94°C denaturation for 1 min, followed by 40 cycles of 95°C for 10 s, 60°C for 30 s. Experiments were replicated three times with $2^{-\Delta\Delta Ct}$ relative quantification method. Primers were listed in Table S14.

Statistical analysis

Statistical analyses were performed with a Processing System statistical software package using ANOVA followed by Duncan's multiple range test.

Results

Identification and evolution of SnRKs in maize

In total, we identified 60 ZmSnRK proteins in the maize inbred line B73 genome. The detailed information in terms of genes and proteins can be seen in Tables S1 and S2. For instance, the amino acid length of 60 ZmSnRKs protein ranges from 215 to 653, and the molecular weight ranges from 24.7 to 72.3 kDa. All ZmSnRKs were predicted to be located in the nucleus, cytoplasm, or mitochondria.

To verify the evolutionary relations of SnRK proteins in angiosperms, we constructed a neighbor-joining tree using 47 rice SnRKs, 39 Arabidopsis SnRKs, and 60 maize SnRKs (Figure 1). The protein sequences in rice and Arabidopsis are listed in Table S3. Hrabak et al., (2003) reported that 39 AtSnRKs were classified into three subfamilies. Here, 60 ZmSnRKs were also clustered into three groups in accordance with phylogenetic analysis. In detail, four proteins containing Pkinase (PF00069), UBA (PF00627), and KA1 (PF02149) domains belong to the ZmSnRK1 subfamily; 14 proteins belong to the ZmSnRK2 subfamily, and the rest of 42 proteins containing Pkinase and NAF (PF03822) domains were clustered into the ZmSnRK3 subfamily (Figure 2A).

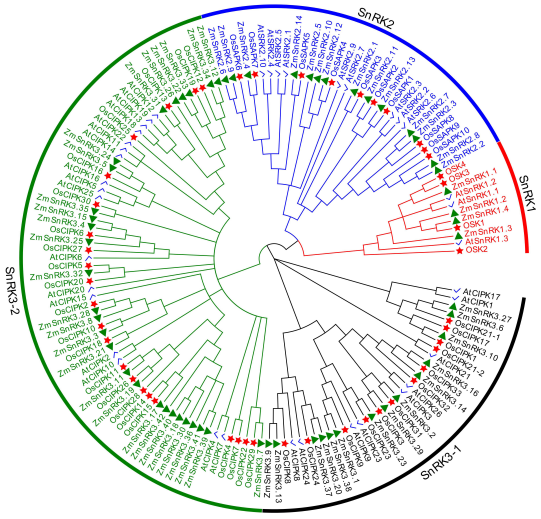


FIGURE 1
Phylogenetic tree of full-length ZmSnRK, AtSnRK and OsSnRK proteins. The different colored arcs indicate subfamilies of the SnRK proteins. Different colour shapes represent SnRKs from maize (▲), rice (☆), and Arabidopsis (√). The unrooted neighbor-joining phylogenetic tree was constructed using MEGA7 with full-length amino acid sequences and the bootstrap test replicate was set as 1,000 times.

Gene and protein structures in ZmSnRKs

The web server GSDS was used to understand the gene structure of *ZmSnRKs*, which had 1 to 14 exons unevenly. In detail, *SnRK1* subfamily genes contain 8–12 exons. *SnRK2* subfamily genes contain 7–9 exons, except *ZmSnRK2.2* and *ZmSnRK2.14* containing one exon. Interestingly, *ZmSnRK3* subfamily genes exhibit a large difference in the number of exons, ranging from 1 to 14 (Figure 2C). For example, half of the genes have only one exon, five genes have two exons, and another gene has 12–14 exons. Hence, *ZmSnRK3-1* and *ZmSnRK3-2* were classified in terms of the number of exons.

We used MEME to determine protein motifs. Generally, 10 conserved motifs were discovered in ZmSnRKs and named motifs 1–10 (Figure 2B; Table S4). Among them, the motifs ½/3/4/5 belong to phosphokinase domains. Almost all of ZmSnRKs have motifs 1, 2, 3, and 4, except ZmSnRK1.1 with motif 2/3/4, ZmSnRK3.27 and ZmSnRK2.7 with motif 1/3/4, ZmSnRK2.11 with motif 1/2/4, and ZmSnRK3.26 with motif 1/2/3. In addition, motif 6/7/9/10 exists exclusively in ZmSnRK3, except ZmSnRK1.1 and ZmSnRK1.4 containing motif 6. Moreover, motif 8 exists only in ZmSnRK1 and ZmSnRK3, except ZmSnRK3.26. In addition, 3D structures of ZmSnRK proteins were predicted and it showed that different ZmSnRKs own similar 3D structures in the N-terminal but disparate in the C-terminal (Figure 3).

Chromosomal localization and duplication of *ZmSnRK*s

The chromosome location of *ZmSnRK* genes was analyzed and 60 *ZmSnRKs* were mapped to the maize genome (Figure 4). In detail, the gene members of the *ZmSnRK1* subfamily were localized in chromosomes 1, 2, 6, and 8. Both *ZmSnRK2* and *ZmSnRK3* subfamily genes were located in all chromosomes. We used the approach of BLAST and MCScanX to search for gene duplication. Finally, 22 segmental duplication events were detected in *ZmSnRKs*, and each gene pair was located on a distinct chromosome (Figure 5; Table S5). It demonstrated that

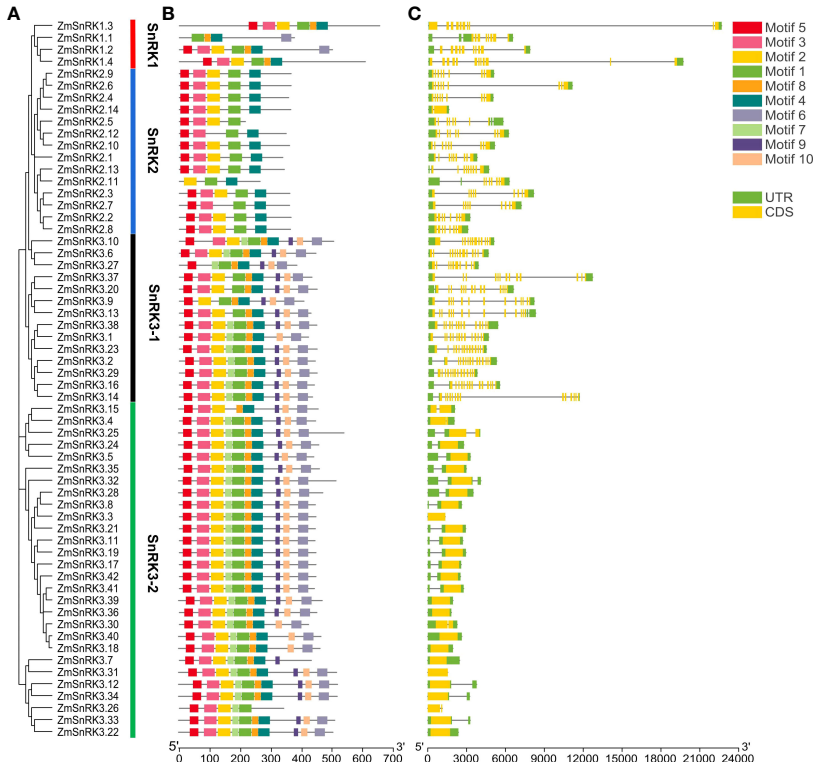


FIGURE 2
Phylogenetic relationships, architecture of conserved protein motifs and gene structure in *SnRK* genes from maize. **(A)** The phylogenetic tree was constructed based on the full-length sequences of maize SnRK proteins using MEGA 7 software. **(B)** The motif compositions of 60 ZmSnRK proteins. The motifs were identified using the MEME program. Boxes of different colors represent motifs 1 to 10. The length of the amino acid sequences can be estimated by the scale at the bottom. **(C)** Gene structures of 60 *ZmSnRK* genes. Yellow boxes represent exons, green boxes represent 5′ or 3′ untranslated regions (UTR), and black lines represent introns. The length of nucleotide sequences of exons/introns/UTRs can be estimated by the scale at the bottom.

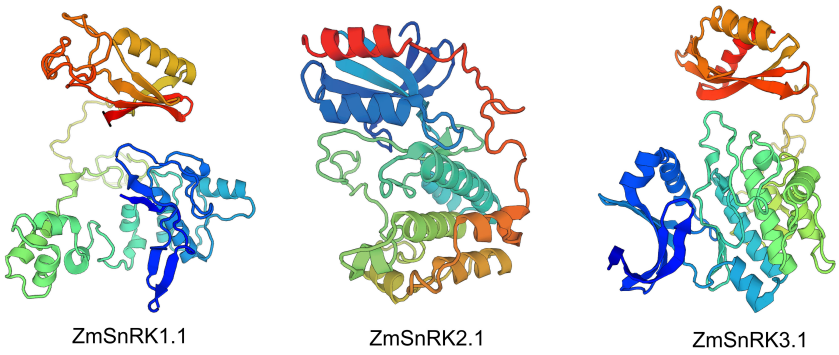


FIGURE 3
The 3D structure modeling of ZmSnRK proteins. The pymol software was used to create the structural image.

most of *ZmSnRKs* were probably derived from gene duplication, and segmental duplication events could take a dominant role in the expansion of *ZmSnRK* genes in maize. Moreover, the syntenic relations of the *SnRK* gene families between maize genome, barley genome, Arabidopsis genome, and rice genome were also determined. The results showed that two *SnRK* gene pairs were found in maize and Arabidopsis, and four *SnRK* gene pairs were discovered in maize and rice (Figure 6; Table S6, S7). For instance, *AtCIPK15* showed colinearity with *ZmSnRK3.11* and *ZmSnRK3.17*. Furthermore, 24 *ZmSnRKs* exhibited syntenic relations with *SnRKs* in barley (Figure 6; Table S8).

To understand the evolutionary constraints acting on paralogous *SnRK* family genes, Ks value, Ka value, Ka : Ks ratio, and divergence time were calculated (Table S5). Most of the segmental duplication of *ZmSnRK* gene pairs displayed Ka/Ks ratios less than 1, except *ZmSnRK3.3/ZmSnRK3.8* and *ZmSnRK3.6/ZmSnRK3.27* with 2.69 and 1.02, respectively. The average value of *ZmSnRK3* (Ka/Ks = 0.56) was higher than

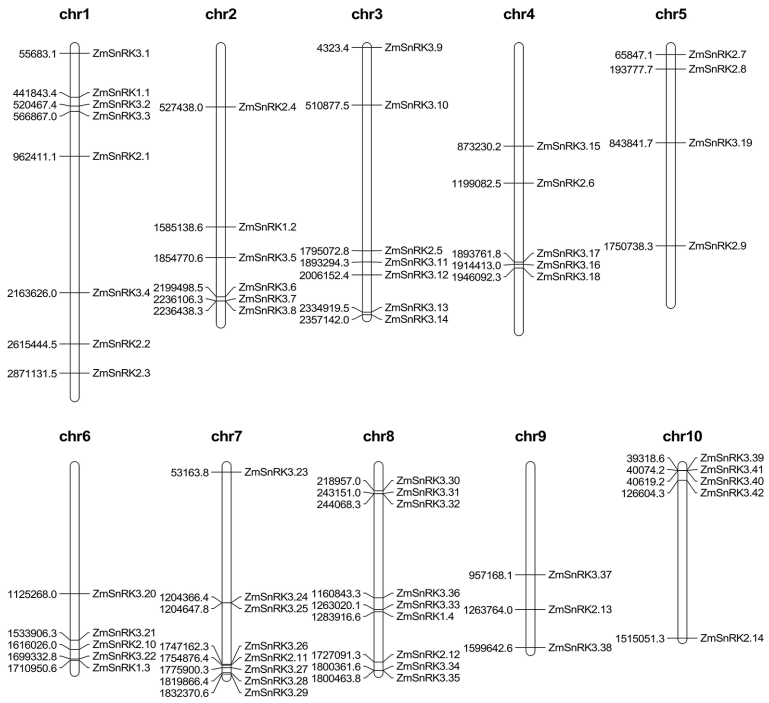


FIGURE 4
Distribution of *ZmSnRK* genes in maize chromosomes. The chromosome numbers are indicated at the top of each chromosome image.

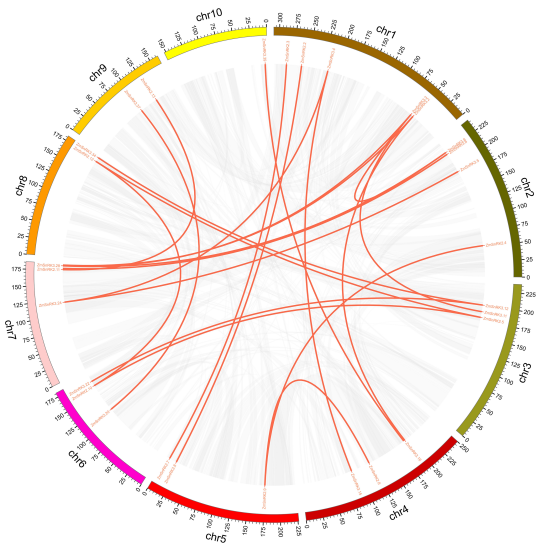


FIGURE 5
The synteny analysis of *ZmSnRK* family in maize. Gray lines indicate all synteny blocks in the maize genome. The genes linked by red lines represent homologues.

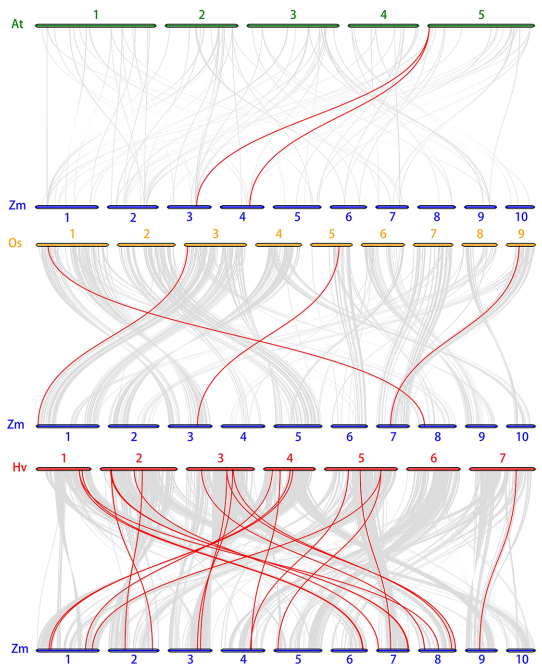


FIGURE 6
Synteny analysis of *SnRK* genes between maize, Arabidopsis, rice and barley. Gray lines: all collinear blocks within maize and other plant genomes. Red lines: the synteny of SnRK gene pairs. The specie names with the prefixes, Hv, Zm, At, and Os indicate barley, maize, Arabidopsis, and rice, respectively.

ZmSnRK2 gene pairs (Ka/Ks = 0.13). Divergence time was therefore estimated to occur between 4.304 Mya and 68.812 Mya ago. The above results implied that *ZmSnRK* gene families were likely to have suffered strong purifying selective pressure in the course of evolution.

Cis-elements analysis in the promoters of *ZmSnRKs*

Promoter sequences of all *ZmSnRK* genes were obtained from the maize B73 genome database. A total of 60 *ZmSnRKs* were assessed to analyze cis-elements, including ABRE, AuxRR-core, MBS, MBSI, P-box, SARE, LTR, involving in ABA, auxin, drought-inducibility, flavonoid biosynthetic genes regulation, gibberellin, salicylic acid and low temperature response (Figure 7; Table S9). In general, 50 *ZmSnRKs* genes (90%) owned ABRE cis-elements, 22 *ZmSnRK* genes (36.7%) carried MBS cis-elements, and 28 *ZmSnRK* genes (46.7%) had LTR cis-elements. Six *ZmSnRK* genes had AuxRR-core elements, and 12 *ZmSnRK* genes had P-box. Meanwhile, only *ZmSnRK2.4* and *ZmSnRK2.12* had MBSI and SARE elements, respectively.

Tissue-specific expression patterns of *ZmSnRKs*

We compared tissue-specific expression patterns of 60 *ZmSnRK* genes in maize B73. In terms of different expression patterns, 60 *ZmSnRK* genes were classified into three groups (Figure 8; Table S10). Group 1 contains three genes (ZmSnRK3.26/3.34/3.40), and they are not expressed in all analyzed tissues. Group 2 includes 12 genes, which were expressed only in specific tissues. For instance, *ZmSnRK3.1* was expressed only in stem and leaves, and did not display expression in other tissues. Group 3 contains 45 genes expressed in all analyzed tissues. Group 3 could be classified into three subgroups. There were 14 *ZmSnRKs* belonging to subgroup 1 with high expression patterns ($log2^{TPM+1}>2$) in all analyzed tissues, including three *ZmSnRK1s*, five *ZmSnRK2s*, and six *ZmSnRK3s*. Subgroup 2 contained only *ZmSnRK3.23* with $log2^{TPM+1}<0.5$ in all analyzed tissues. The remaining 30 genes were assigned to subgroup 3.

Expression patterns of *ZmSnRKs* under diverse abiotic stresses

We collected and analyzed the transcriptome data of maize *SnRK* genes under drought, salt, high temperature, low temperature, and ultraviolet light. In general, the expression of several *ZmSnRK* genes showed significant alterations under abiotic stresses (Figure 9; Table S11). For example,

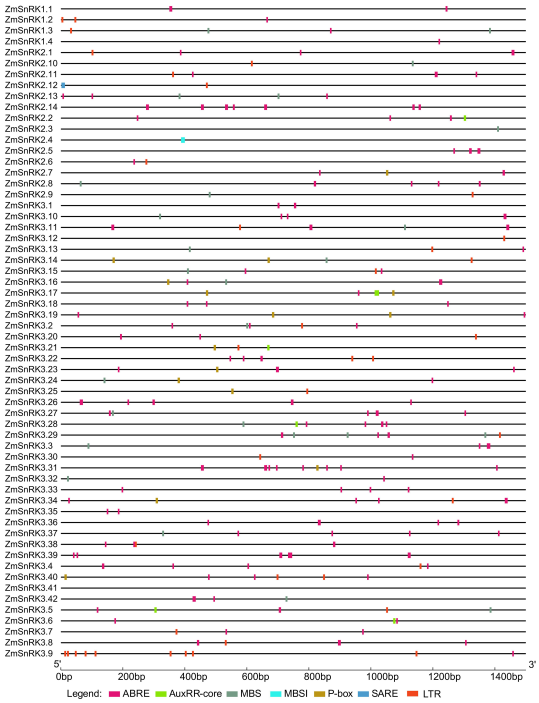


FIGURE 7
Predicted cis-regulatory elements in *ZmSnRK* promoters. Promoter sequences (about 1.5 kb) of 60 *ZmSnRK* genes were analyzed by PlantCARE. The upstream length to the translation starting site can be inferred according to the scale at the bottom.

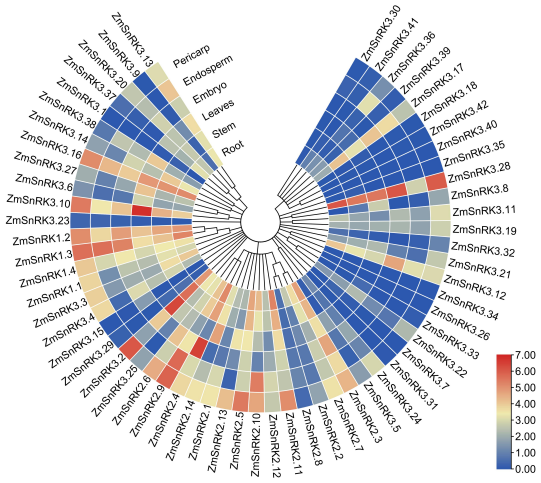


FIGURE 8
Expression profiles of the *ZmSnRK* genes in different tissues. The color scale represents expression data with row scale. Blue: low expression; red: high expression.

ZmSnRK2.13 exhibited up-regulation after drought stress. The expression pattern of *ZmSnRK2.2* increased under cold treatment. *ZmSnRK3.1* was down-regulated under cold and salt stress. On the contrary, some *ZmSnRKs* were not induced by any abiotic stresses listed in this study. For instance, 11 *ZmSnRK* genes showed almost no expression changes under all analyzed stresses, such as *ZmSnRK1.4, ZmSnRK2.3*, and *ZmSnRK3.13*. In addition, many genes showed opposite expression profiles in response to different abiotic stress. For example, *ZmSnRK2.2* displayed expression inhibition in UV treatment, but expression enhancement in cold stress. In addition, several *ZmSnRK* genes were chosen for qRT-PCR to verify the reliability of transcriptome data, and the results were uniform to the sequencing data (Figure S2, S3).

Function and regulatory network of ZmSnRKs

Protein phosphatases PP2C could dephosphorylate and inactivate plant SnRKs, the process of which prevents SnRKs from targeting ABA-dependent genes and ion channels. Hence, we used the web server of STRING to predict the protein–protein interaction (PPI) between ZmSnRKs and ZmPP2Cs to comprehend the regulatory networks of ZmSnRKs (Figure 10; Table S12). Here, eight ZmPP2Cs that could interact with ZmSnRK proteins were found. In detail, ZmPP2C65 and ZmPP2C67 could bind six ZmSnRK proteins; ZmPP2C51, ZmPP2C68, and ZmPP2C69 could bind four ZmSnRK proteins; ZmPP2C4 and ZmPP2C71 could bind three ZmSnRK proteins; PP2C17 could dephosphorylate ZmSnRK2.11 specifically. Moreover, there are also interactions within ZmSnRK proteins. For instance, ZmSnRK1.1 could bind with ZmSnRK2.14 and ZmSnRK3.12 (Figure S1).

We constructed the miRNA-mRNA regulatory network between 321 published miRNA in maize and 60 *ZmSnRKs*. A total of 46 putative miRNAs were identified to have a potential capacity to target and regulate 24 *ZmSnRKs* (Figure 11; Table S13). The miRNAs of *ZmSnRKs* could be grouped into 11 networks: group 1 involving two *ZmSnRKs* (*ZmSnRK3.1* and *ZmSnRK3.38*), group 2 involving seven *ZmSnRKs* (*ZmSnRK* 2.2/2.4/2.6/2.8 and *ZmSnRK* 3.26/3.33), group 3 involving four *ZmSnRKs* (*ZmSnRK* 2.11/2.13 and *ZmSnRK* 3.6/3.20), group 5 involving *ZmSnRK3.2* and *ZmSnRK3.28*, and group 7 involving *ZmSnRK* 3.30/3.40/3.42, group 4/6/8/9/10/11 involving *ZmSnRK* 3.27/1.2/3.19/3.31/3.17, respectively.

Discussion

Protein kinases and phosphatases can recognize and transduce stress signals to diverse cellular compartments. SNF1-related kinases (SnRKs) universally exist in all eukaryotes, especially in plants. To date, the SnRK family has

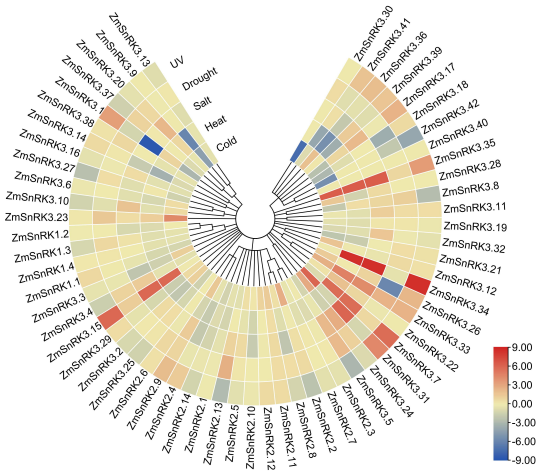


FIGURE 9
Expression profiles of the *ZmSnRK* genes under different abiotic stresses. Expression data were the ratio to control values. The color scale represents expression levels from upregulation (red) to downregulation (blue).

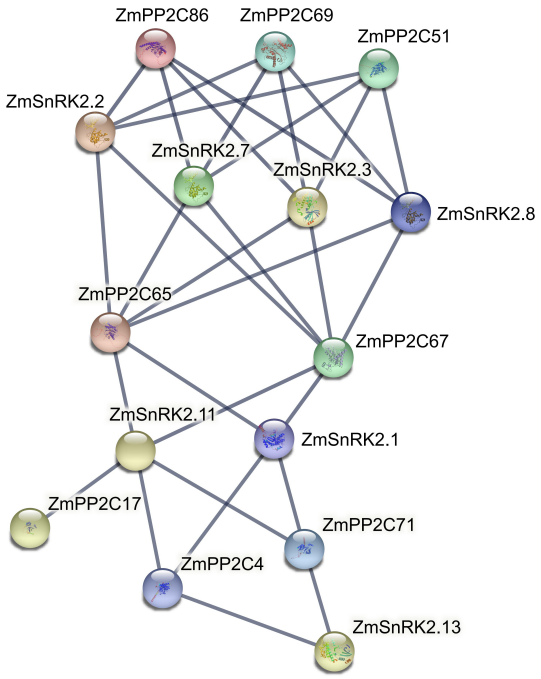


FIGURE 10
Interaction networks between ZmSnRK and ZmPP2C proteins. Information on ZmPP2Cs was shown in Table S12.

been identified in a number of plant species, including *Arabidopsis thaliana* (Hrabak et al., 2003), *Oryza sativa* (Kobayashi et al., 2004), *Brassica napus* (Zhu et al., 2020), and *Brachypodium distachyon* (Wang et al., 2015). However, the report concerning whole-genome analysis of the *ZmSnRK* genes has not yet been published in maize. In the current study, 60 *ZmSnRKs* have been found in the *Zea mays* genome, which were considered as *ZmSnRK1* to *ZmSnRK60* based on their chromosome distribution (Figures 1, 4). Previously, *SnRK* genes were classified into three subfamilies based on different domains in C-terminus in Arabidopsis and rice (Kolukisaoglu et al., 2004). In the maize genome, 60 *ZmSnRK* genes were also classified into three subfamilies, including 4 *ZmSnRK1s*, 14 *ZmSnRK2s*, and 42 *ZmSnRK3s* (Figures 1, 2; Table S2). Then, *ZmSnRK* genes family was systematically investigated, including evolutionary relationships, conserved motifs, subcellular localization, chromosome location, gene divergence time, promoter elements analysis, and expression patterns.

The plant SnRK1 subfamily is the key regulator of energy homeostasis (Crepin and Rolland, 2019). SnRK1 controls the activity of many enzymes and the transcription activity of several

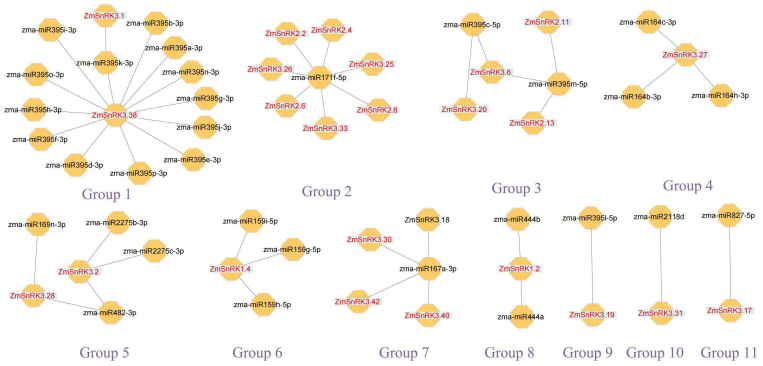


FIGURE 11
Interaction networks between miRNAs and miRNAs-acted *ZmSnRKs*. Information on miRNAs and miRNAs-acted *ZmSnRKs* was shown in Table S13.

genes (Baena-González et al., 2007; Nukarinen et al., 2016), such as bZIP, ethylene insensitive3 (EIN) and indeterminate domain (IDD) (Jeong et al., 2015; Zhai et al., 2017; Muralidhara et al., 2021). For example, SnRK1 delayed senescence through phosphorylating the EIN3 transcription factor and led to its destabilization (Zhai et al., 2017). SnRK1 strongly phosphorylated bZIP63 and enhanced its dimer formation, either homodimerization or heterodimerization with other bZIPs, thus contributing to DNA binding (Mair et al., 2015). Meanwhile, Jeong et al. (2015) reported that SnRK1 was responsible for regulating various developmental processes in plants, such as hypocotyl elongation or blooming (Simon et al., 2018). In the current study, *ZmSnRK1* has four members and exhibited a high expression pattern in all tissues, but showed no alteration upon abiotic stress (Figures 1, 8, 9; Figure S2, S3; Table S2, S10, S11). Moreover, ZmSnRK1.3 and ZmSnRK1.4 were predicted to be located in mitochondria (Table S1). ZmSnRK1.1 and ZmSnRK1.2 were likely to be located in the nucleus and cytoplasm but with 0.657 and 0.568 reliable-index if located in mitochondria (data not shown). It suggests ZmSnRK1 may also be crucial in plant development as an energy sensor.

Plant SnRK2s, a key component of ABA signaling pathways, are regulated by ABA receptors (PYR/PYL/RCAR) and protein phosphatase 2Cs (PP2CAs). In this study, eight ZmPP2Cs were predicted to interact with seven ZmSnRK2s (Figure 10). It suggests that ZmPP2Cs could also dephosphorylate ZmSnRKs in maize. The SnRK2 subfamily in plants is extremely responsible for responding to abiotic stress, such as drought (Umezawa et al., 2004; Barajas-Lopez et al., 2018; Hasan et al., 2021) and salt (Zhong et al., 2020). When plants are subjected to drought stresses, they generate more ABA, thus leading to defensive stress responses to activate many SnRK2s *via* ABA-dependent or ABA-independent pathways (Hasan et al., 2022). Here, the expression level of *ZmSnRK2.13* was significantly increased after drought stress (Figure 9; Figure S2; Table S11), which was likely to be dephodphrylated by ZmPP2C4 and ZmPP2C71 (Figure 10). Moreover, the promoter of *ZmSnRK2.13* had three ABRE cis-elements and one MBS cis-element, which was associated with ABA responsiveness and drought stress (Figure 7; Table S9). One segmental duplication event was detected between *ZmSnRK2.11* and *ZmSnRK2.13* (Figure 5; Table S5), but drought exposure does not change the gene expression in *ZmSnRK2.11* (Figure 9; Table S11), indicating that *ZmSnRK2.13* was positively selected when subjected to drought stress even though it can be cleaved by the same miRNA (Figure 11; Table S13). Belda-Palazón et al. (2022) reported that SnRK2 could interact with SnRK1 and consequently cause subcellular localization changes of the SnRK1 α-subunit, the process of which are vital for repressing the target of rapamycin (TOR) and inhibiting root growth in an ABA-dependent manner. In this study, ZmSnRK2.13 could also bind to ZmSnRK1.1 and ZmSnRK2.8 (Figure S1). It suggested that SnRK1 could interact with SnRK2 and jointly participate in plant growth in maize. In rice, *OsSAPK8* (orthologous to *ZmSnRK2.3* and *ZmSnRK2.7*) acts as a positive regulator when plants are subjected to cold, drought, and salt stress (Zhong et al., 2020). Meanwhile, Frazier et al. (2011) reported that miR395 was up-regulated under drought stress. The current study showed that miR395 could target and cut the mRNA of *ZmSnRK2.3* (Figure 11; Table S13). *ZmSnRK2.3* and *ZmSnRK2.7*, as a duplicated gene pair, do not respond to abiotic stress (Figures 5, 9; Table S5, S11), but showed high expression in all tested tissues (Figure 8; Table S10), indicating different mechanism in diverse species. Interestingly, *ZmSnRK2.4* have only MBSI cis-element, which is involved in flavonoid biosynthesis (Figure 7; Figure S2; Table S9), and its expression level was repressed under salt stress (Figure 9; Table S11). Watkins et al. (2017) found that flavonoids are involved in plant development through affecting auxin, ethylene, and ABA signaling. HSFB2b and GmMYB173 activated a number of flavonoid biosynthesis-related genes and promoted flavonoid accumulation, thus positively regulating salt tolerance (Pi et al., 2018; Bian et al., 2020). This indicates that *ZmSnRK2.4* may also be involved in flavonoid biosynthesis and regulate plant growth under salt stress.

The SnRK3 subfamily was widely recognized as CBL-interacting protein kinases (CIPKs). In plants, CBL-CIPK signaling network regulates multiple stimuli or signals (Ma et al., 2020). For example, Qiu et al. (2002) reported that AtCIPK24 could interact with AtCBL4 and activate the H^+/Na^+ antiporter. In the current study, *ZmSnRK3.33* showed a high expression level under five abiotic stresses (Figure 9; Table S11), which can be regulated by zma-miR171. Besides, *ZmSnRK2.2, ZmSnRK2.4, ZmSnRK2.6*, and *ZmSnRK2.8* were also cut by zma-miR171 (Figure 11; Table S13). These indicate that *ZmSnRK3* and *ZmSnRK2* may have a similar function, at least in one aspect. In soybean, GmCIPK2 acts as a positive regulator in response to drought stress in an ABA-dependent manner (Xu et al., 2021). Here, *ZmSnRK3.26* showed down-regulation with 5.9-fold changes after drought stress (Figure 9; Table S11), the promoter of which has five ABRE elements (Figure 7; Table S9). Hence, ZmSnRK3 could also respond to abiotic stresses through an ABA-dependent pathway.

In short, this research carried out a comprehensive analysis of the *SnRK* gene family in *Zea mays*. With the above results, we deduced that *ZmSnRKs* might play a pivotal role in the long-term climate resilience of maize. Numerous functional verification work is still necessary for comprehending the biological functions of *ZmSnRKs* in future research.

Conclusion

SnRK genes play important roles in signaling pathways, including responses to abiotic stresses in plants. In the present study, genome-wide identification and characterization of *SnRK* genes were conducted in *Zea mays*. A total of 60 *ZmSnRK* genes were characterized and divided into three subfamilies. Phylogenetic and synteny analysis of *SnRK* genes among Arabidopsis, rice, and maize provide valuable clues for the evolutionary characteristics of the *ZmSnRK* genes. Moreover, the cis-acting elements, gene expression, and regulatory network of ZmSnRK families were also determined. These results provide insights into the functional differences, evolutionary relationships, and expression profiles of *SnRKs* in *Zea mays*.

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

XF and WL conceived and designed the research. XF, QM, QY, JZ, DX, and XD performed the experiments and data analyses. WL, XF, and QM wrote the article. LG and WM revised the manuscript. All authors read and approved the final article.

Funding

This work was supported by the National Natural Science Foundation of China (grant No. 32001449; 32101660) and the Natural Science Foundation of Shandong Province (grant No. ZR2021QC052).

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.2022.1087839/full#supplementary-material

References

Albrecht, V., Ritz, O., Linder, S., Harter, K., and Kudla, J. (2014). The NAF domain defines a novel protein-protein interaction module conserved in Ca^{2+}-regulated kinases. *EMBO J.* 20, 1051–1063. doi: 10.1093/emboj/20.5.1051

Alderson, A., Sabelli, P. A., Dickinson, J. R., Cole, D., Richardson, M., Kreis, M., et al. (1991). Complementation of snf1, a mutation affecting global regulation of carbon metabolism in yeast, by a plant protein kinase cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 88, 8602–8605. doi: 10.1073/pnas.88.19.8602

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* 22, 195–201. doi: 10.1093/bioinformatics/bti770

Baena-González, E., Rolland, F., Thevelein, J. M., and Sheen, J. (2007). A central integrator of transcription networks in plant stress and energy signalling. *Nature* 23, 938–942. doi: 10.1038/nature06069

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, 202–208. doi: 10.1093/nar/gkp335

Barajas-Lopez, J. D., Moreno, J. R., Gamez-Arjona, F. M., Pardo, J. M., Punkkinen, M., Zhu, J. K., et al. (2018). Upstream kinases of plant SnRKs are involved in salt stress tolerance. *Plant J.* 93, 107–118. doi: 10.1111/tpj.13761

Belda-Palazón, B., Costa, M., Beeckman., T., Rolland, F., and Baena-González, E. (2022). ABA represses TOR and root meristem activity through nuclear exit of the SnRK1 kinase. *Proc. Natl. Acad. Sci. U.S.A.* 119, e2204862119. doi: 10.1073/pnas.2204862119

Bender, K. W., Zielinski, R. E., and Huber, S. C. (2018). Revisiting paradigms of Ca^{2+} signaling protein kinase regulation in plants. *Biochem. J.* 475, 207–223. doi: 10.1042/BCJ20170022

Bian, X. H., Li, W., Niu, C. F., Wei, W., Hu, Y., Han, J. Q., et al. (2020). A class b heat shock factor selected for during soybean domestication contributes to salt tolerance by promoting flavonoid biosynthesis. *New Phytol.* 225 (1), 268–283. doi: 10.1111/nph.16104

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

Cohen, P. (1988). Review lecture: Protein phosphorylation and hormone action. *Proc. R. Soc Lond.* 234, 115–44. doi: 10.1098/rspb.1988.0040

Crepin, N., and Rolland, F. (2019). SnRK1 activation, signaling, and networking for energy homeostasis. *Curr. Opin. Plant Biol.* 51, 29–36. doi: 10.1016/j.pbi.2019.03.006

Cutler, S. R., Rodriguez, P. L., Finklestein, R. R., and Abrams, S. R. (2010). Abscisic acid: emergence of a core signaling network. *Ann. Rev. Plant Biol.* 61, 651–679. doi: 10.1146/annurev-arplant-042809-112122

Damian, S., Gable, A. L., Nastou, K. C., David, L., Rebecca, K., Sampo, P., 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, 605–612. doi: 10.1093/nar/gkab835

Ding, Y., Li, H., Zhang, X., Xie, Q., Gong, Z., and Yang, S. (2015). OST1 kinase modulates freezing tolerance by enhancing ICE1 stability in arabidopsis. *Dev. Cell.* 32, 278–289. doi: 10.1016/j.devcel.2014.12.023

Frazier, T. P., Sun, G., Burklew, C. E., and Zhang, B. (2011). Salt and drought stresses induce the aberrant expression of microRNA genes in tobacco. *Mol. Biotechnol.* 49, 159–165. doi: 10.1007/s12033-011-9387-5

Halford, N. G., and Hardie, D. G. (1998). SNF1-related protein kinases: global regulators of carbon metabolism in plants? *Plant Molec Biol.* 37, 735–748. doi: 1023/A:1006024231305

Hasan, M. M., Gong, L., Nie, Z. F., Feng, X., Ahammed, G. J., and Fang, X. W. (2021). ABA-induced stomatal movements in vascular plants during dehydration versus rehydration. *Environ. Exp. Bot.* 186, 104436. doi: 10.1016/j.envexpbot.2021.104436

Hasan, M. M., Liu, X. D., Waseem, M., Guang-Qian, Y, Alabdallah, N. M., Jahan, M. S., and Fang, X. W. (2022). ABA activated SnRK2 kinases: an emerging role in plant growth and physiology. *Plant Signal Behav.* 17, 2071024. doi: 10.1080/15592324.2022.2071024

Hrabak, E. M., Chan, C. W., Gribskov, M., Harper, J. F., Choi, J. H., Halford, N., et al. (2003). The arabidopsis CDPK-SnRK superfamily of protein kinases. *Plant Physiol.* 132, 666–680. doi: 10.1104/pp.102.011999

Hu, B., Jin, J., Guo, A., 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

Hunter, T. (1995). Protein kinases and phosphatases: The yin and yang of protein phosphorylation and signaling. *Cell* 80, 225–236. doi: 10.1016/0092-8674(95)90405-0

Jeong, E. Y., Seo, P. J., Woo, J. C., and Park, C. M. (2015). AKIN10 delays flowering by inactivating IDD8 transcription factor through protein phosphorylation in arabidopsis. *BMC Plant Biol.* 15, 110. doi: 10.1186/s12870-015-0503-8

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Kobayashi, Y., Yamamoto, S., Minami, H., Kagaya, Y., and Hattori, T. (2004). Differential activation of the rice sucrose nonfermenting1-related protein kinase2 family by hyperosmotic stress and abscisic acid. *Plant Cell.* 16, 1163–1177. doi: 10.1105/tpc.019943

Kolukisaoglu, U., Weinl, S., Blazevic, D., Batistic, O., and Kudla, J. (2004). Calcium sensors and their interacting protein kinases: Genomics of the arabidopsis and rice CBL-CIPK signaling networks. *Plant Physiol.* 134, 43–58. doi: 10.1104/pp.103.033068

Kulik, A., Wawer, I., Krzywi ´nska, E., Bucholc, M., and Dobrowolska, G. (2011). SnRK2 protein kinases–key regulators of plant response to abiotic stresses. *Omics* 15, 859. doi: 10.1089/omi.2011.0091

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

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., Mcwilliam, H., et al. (2007). Clustal W and clustal X version 2.0. *Bioinformatics* 23, 2947–29348. doi: 10.1093/bioinformatics/btm404

Letunic, I., Doerks, T., and Bork, P. (2012). SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* 40, 302–305. doi: 10.1093/nar/gkr931

Li, J., Long, Y., Qi, G. N., Li, J., Xu, Z. J., Wu, W. H., et al. (2014). The os-AKT1 channel is critical for k^+ uptake in rice roots and is modulated by the rice CBL1-CIPK23 complex. *Plant Cell.* 26, 3387–3402. doi: 10.1105/tpc.114.123455

Li, C., Qi, W., Liang, Z., Yang, X., Ma, Z., and Song, R. (2020). A SnRK1-ZmRFWD3-Opaque2 signaling axis regulates diurnal nitrogen accumulation in maize seeds. *Plant Cell.* 32, 2823–2841. doi: 10.1105/tpc.20.00352

Magali, L., Patrice, D., Gert, T., Kathleen, M., Yves, M., Yves, V. D. P., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. doi: 10.1093/nar/30.1.325

Mair, A., Pedrotti, L., Wurzinger, B., Anrather, D., Simeunovic, A., Weiste, C., et al. (2015). SnRK1-triggered switch of bZIP63 dimerization mediates the lowenergy response in plants. *eLife* 4, e05828. doi: 10.7554/eLife.05828.043

Ma, X., Li, Y., Gai, W. X., Li, C., and Gong, Z. H. (2021). The CaCIPK3 gene positively regulates drought tolerance in pepper. *Hortic. Res.* 8, 216. doi: 10.1038/s41438-021-00651-7

Ma, X., Li, Q. H., Yu, Y. N., Qiao, Y. M., Haq., S. U., and Gong, Z. H. (2020). The CBL-CIPK pathway in plant response to stress signals. *Int. J. Mol. Sci.* 215668. doi: 10.3390/ijms21165668

Masaru, O., Yan, G., Ursula, H., and Jian-Kang, Z. (2003). A novel domain in the protein kinase SOS2 mediates interaction with the protein phosphatase 2C ABI2. *Proc. Natl. Acad. Sci. U.S.A.* 100, 11771–11776. doi: 10.1073/pnas.2034853100

Maszkowska, J., Szymańska, K. P., Kasztelan, A., Krzywińska, E., Sztatelman, O., and Dobrowolska, G. (2021). The multifaceted regulation of SnRK2 kinases. *Cell* 102180. doi: 10.3390/cells10092180

Mazur, R., Maszkowska, J., Anielska-Mazur, A., Garstka, M., Polkowska-Kowalczyk, L., Czajkowska, A., et al. (2021). The SnRK2.10 kinase mitigates the adverse effects of salinity by protecting photosynthetic machinery. *Plant Physiol.* 187, 2785–2802. doi: 10.1093/plphys/kiab438

Mcloughlin, F., Galvan-Ampudia, C. S., Julkowska, M. M., Caarls, L., van der Does, D., Laurière, C., et al. (2012). The Snf1-related protein kinases SnRK2.4 and SnRK2.10 are involved in maintenance of root system architectureduring salt stress. *Plant J.* 72, 436–449. doi: 10.1111/j.1365-313X.2012.05089.x

Muralidhara, P., Weiste, C., Collani, S., Krischke, M., Kreisz, P., Draken, J., et al. (2021). Perturbations in plant energy homeostasis prime lateral root initiation *via* SnRK1-bZIP63-ARF19 signaling. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2106961118. doi: 10.1073/pnas.2106961118

Myburg, A. A., Dario, G., Tuskan, G. A., Uffe, H., Hayes, R. D., Jane, G., et al. (2014). The genome of eucalyptus grandis. *Nature* 510, 356–362. doi: 10.1038/nature13308

Nukarinen, E., Nägele, T., Pedrotti, L., Wurzinger, B., Mair, A., Landgraf, R., et al. (2016). Quantitative phosphoproteomics reveals the role of the AMPK plant ortholog SnRK1 as a metabolic master regulator under energy deprivation. *Sci. Rep.* 6, 31697. doi: 10.1038/srep31697

Opitz, N., Paschold, A., Marcon, C., Malik, W. A., Lanz, C., Piepho, H. P., et al. (2014). Transcriptomic complexity in young maize primary roots in response to low water potentials. *BMC Genomics* 15, 741. doi: 10.1186/1471-2164-15-741

Piao, H. L., Xuan, Y. H., Park, S. H., Je, B. I., Park, S. J., Park, S. H., et al. (2010). OsCIPK31, a CBL-interacting protein kinase is involved in germination and seedling growth under abiotic stress conditions in rice plants. *Mol. Cell.* 30, 19–27. doi: 10.1007/s10059-010-0084-1

Pi, E., Zhu, C., Fan, W., Huang, Y., Qu, L., Li, Y., et al. (2018). Quantitative phosphoproteomic and metabolomic analyses reveal GmMYB173 optimizes flavonoid metabolism in soybean under salt stress. *Mol. Cell Proteomics.* 17, 1209–1224. doi: 10.1074/mcp.RA117.000417

Qiu, Q. S., Guo, Y., Dietrich, M. A., Schumaker, K. S., and Zhu, J. K. (2002). Regulation of SOS1, a plasma membrane Na^+/H^+ exchanger in arabidopsis thaliana, by SOS2 and SOS3. *Proc. Natl. Acad. Sci. U.S.A.* 99, 8436–8441. doi: 10.1073/pnas.122224699

Shannon, P., Markeil, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Simon, N. M. L., Kusakina, J., Fernández-López, Á., Chembath, A., Belbin, F. E., and Dodd, A. N. (2018). The energy-signaling hub SnRK1 is important for sucrose-induced hypocotyl elongation. *Plant Physiol.* 176, 1299–1310. doi: 10.1104/pp.17.01395

Stelpflug, S. C., Sekhon, R. S., Vaillancourt, B., Hirsch, C. N., Buell, C. R., de Leon, N., et al. (2016). An expanded maize gene expression atlas based on RNA sequencing and its use to explore root development. *Plant Genome.* 9. doi: 10.3835/plantgenome2015.04.0025

Su, W., Raza, A., Zeng, L., Gao, A., Lv, Y., Ding, X., et al. (2021). Genome-wide analysis and expression patterns of lipid phospholipid phospholipase gene family in *Brassica napus* l. *BMC Genomics* 22, 548. doi: 10.1186/s12864-021-07862-1

Tominaga, M., Harada, A., Kinoshita, T., and Shimazaki, K. (2010). Biochemical characterization of calcineurin b-like-interacting protein kinase in vicia guard cells. *Plant Cell Physiol.* 51, 408–421. doi: 10.1093/pcp/pcq006

Tör, M., Lotze, M. T., and Holton, N. (2009). Receptor-mediated signalling in plants: Molecular patterns and programmes. *J. Exp. Bot.* 60, 3645–3654. doi: 10.1093/jxb/erp233

Umezawa, T., Yoshida, R., Maruyama, K., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2004). SRK2C, a SNF1-relatedprotein kinase 2, improves drought tolerance by controlling stress-responsive gene expression in arabidopsis thaliana. *Proc. Natl. Acad. Sci. U.S.A.* 101, 17306–17311. doi: 10.1073/pnas.0407758101

Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77

Wang, 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. *Nucleic Acids Res.* 40, 49. doi: 10.1093/nar/gkr1293

Wang, L., Hu, W., Sun, J., Liang, X., Yang, X., Wei, S., et al. (2015). Genome-wide analysis of *SnRK* gene family in brachypodium distachyon and functional characterization of *BdSnRK2.9*. *Plant Sci.* 237, 33–45. doi: 10.1016/j.plantsci.2015.05.008

Wang, W., Lu, Y., Li, J., Zhang, X., Hu, F., Zhao, Y., et al. (2021). SnRK1 stimulates the histone H3K27me3 demethylase JMJ705 to regulate a transcriptional switch to control energy homeostasis. *Plant Cell.* 33, 3721–3742. doi: 10.1093/plcell/koab224

Wang, D., Zhang, Z., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genom Proteom Bioinf.* 8, 77–80. doi: 10.1016/S1672-0229(10)60008-3

Watkins, J. M., Chapman, J. M., and Muday, G. K. (2017). Abscisic acid-induced reactive oxygen species are modulated by flavonols to control stomata aperture. *Plant Physiol.* 175, 1807–1825. doi: 10.1104/pp.17.01010

Wurzinger, B., Nukarinen, E., Nagele, T., Weckwerth, W., and Teige, M. (2018). The SnRK1 kinase as central mediator of energy signaling between different organelles. *Plant Physiol.* 176, 1085–1094. doi: 10.1104/pp.17.01404

Xu, M., Li, H., Liu, Z., Wang, X. H., Xu, P., Dai, S. J., et al. (2021). The soybean CBL-interacting protein kinase, GmCIPK2, positively regulates drought tolerance and ABA signaling. *Plant Physiol. Biochem.* 167, 980–989. doi: 10.1016/j.plaphy.2021.09.026

Yu, W., Peng, F., Wang, W., Liang, J., Xiao, Y., and Yuan, X. (2021). SnRK1 phosphorylation of SDH positively regulates sorbitol metabolism and promotes sugar accumulation in peach fruit. *Tree Physiol.* 41, 1077–1086. doi: 10.1093/treephys/tpaa163

Zhai, Z., Liu, H., and Shanklin, J. (2017). Phosphorylation of WRINKLED1 by KIN10 results in its proteasomal degradation, providing a link between energy homeostasis and lipid biosynthesis. *Plant Cell.* 29, 871–889. doi: 10.1105/tpc.17.00019

Zhang, Z., Xiao, J., Wu, J., Zhang, H., Liu, G., Wang, X., et al. (2012). ParaAT: a parallel tool 665 for constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.* 419, 779–781. doi: 10.1016/j.bbrc.2012.02.101

Zhang, M., and Zhang, S. (2022). Mitogen-activated protein kinase cascades in plant signaling. *J. Integr. Plant Biol.* 64, 301–341. doi: 10.1111/jipb.13215

Zhong, R., Wang, Y., Gai, R., Xi, D., Mao, C., and Ming, F. (2020). Rice SnRK protein kinase OsSAPK8 acts as a positive regulator in abiotic stress responses. *Plant Sci.* 292, 110373. doi: 10.1016/j.plantsci.2019.110373

Zhu, J. K. (2016). Abiotic stress signaling and responses in plants. *Cell* 167 (2), 313–324. doi: 10.1016/j.cell.2016.08.029

Zhu, W., Wu, D., Jiang, L., and Ye, L. (2020). Genome-wide identification and characterization of SnRK family genes in brassica napus. *BMC Plant Biol.* 20, 287. doi: 10.1186/s12870-020-02484-3

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 13 January 2023
DOI 10.3389/fpls.2022.1083971

OPEN ACCESS

EDITED BY
Mehanathan Muthamilarasan,
University of Hyderabad, India

REVIEWED BY
Vinod Kumar,
Academy of Sciences of the Czech Republic, Czechia
Ravi Gupta,
Kookmin University, South Korea

*CORRESPONDENCE
Anil Grover
agrover@south.du.ac.in
Palak Chaturvedi
palak.chaturvedi@univie.ac.at

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

RECEIVED 29 October 2022
ACCEPTED 02 December 2022
PUBLISHED 13 January 2023

CITATION
Kumar R, Ghatak A, Goyal I, Sarkar NK, Weckwerth W, Grover A and Chaturvedi P (2023) Heat-induced proteomic changes in anthers of contrasting rice genotypes under variable stress regimes.
Front. Plant Sci. 13:1083971.
doi: 10.3389/fpls.2022.1083971

Heat-induced proteomic changes in anthers of contrasting rice genotypes under variable stress regimes

Ritesh Kumar [1], Arindam Ghatak [2], Isha Goyal [1], Neelam K. Sarkar [1], Wolfram Weckwerth [2,3], Anil Grover [1]* and Palak Chaturvedi [2]*

[1]Department of Plant Molecular Biology, University of Delhi, New Delhi, India, [2]Molecular Systems Biology Lab (MOSYS), Department of Functional and Evolutionary Ecology, University of Vienna, Vienna, Austria, [3]Vienna Metabolomics Center (VIME), University of Vienna, Vienna, Austria

Heat stress drastically affects anther tissues resulting in poor plant fertility, necessitating an urgent need to determine the key proteome regulation associated with mature anther in response to heat stress. We identified several genotype - specific protein alterations in rice anthers of Moroberekan (Japonica, heat sensitive), IR64 (Indica, moderately heat tolerant), and Nagina22 (Aus, heat tolerant) in the short-term (ST_HS; one cycle of 42°C, 4 hours before anthesis) and long-term (LT_HS; 6 cycles of 38°C, 6 hours before anthesis) heat stress. The proteins upregulated in long-term heat stress in Nagina22 were enriched in biological processes related to unfolded protein binding and carboxylic acid metabolism, including amino acid metabolism. In short-term heat stress, Nagina22 anthers were enriched in proteins associated with vitamin E biosynthesis and GTPase activator activity. In contrast, downregulated proteins were related to ribosomal proteins. The expression of different Hsp20 and DnaJ was genotype specific. Overall, the heat response in Nagina22 was associated with its capacity for adequate metabolic control and cellular homeostasis, which may be critical for its higher reproductive thermotolerance. This study improves our understanding of thermotolerance mechanisms in rice anthers during anthesis and lays a foundation for breeding thermotolerant varieties *via* molecular breeding.

KEYWORDS

anthers, heat stress, LC-MS/MS, proteomics, reproductive thermotolerance, rice

1 Introduction

Heat stress (HS) is one of the major environmental factors influencing the geographical distribution of plants. According to climate change projections, extreme temperatures may negatively affect plant cultivation in the future (http://climate.nasa.gov/evidence/; Chaturvedi et al., 2021). Rice is the most consumed grain in the world. Every degree Celsius increase in global mean temperature reduces rice yields by 3.2% (Zhao et al., 2017). Several physiological processes in rice plants are affected by HS, including stomatal opening, photosynthetic activity, and growth (Jagadish et al., 2014; Kumar et al., 2022). The reproductive stage of the plants is overly sensitive to HS (Zinn et al., 2010; Bokszczanin, 2013; Chaturvedi et al., 2016; Pazhamala et al., 2020; Chaturvedi et al., 2021). The effect of HS on the physiology and genetics of rice plants at the reproductive stage has been studied extensively (Jagadish and Pal, 2009; Tenorio et al., 2013; Jagadish et al., 2014; Das et al., 2015). Rice male reproductive organs like anther and pollen are more sensitive to HS than the female reproductive organs (Imin et al., 2001; Fujita et al., 2006; Prasad et al., 2006; Jagadish et al., 2010; Jagadish et al., 2014; Shi et al., 2022). Recent work has suggested that the female gametophyte is also affected irreversibly by heat stress (Shi et al., 2022). HS reduces fertility and the number of florets per plant in rice (Kobata et al., 2013). The duration of the HS susceptible period during anthesis differs between indica and japonica rice varieties (Satake and Yoshida, 1978; Matsui et al., 2000). Generally, tolerant varieties show early morning anthesis to avoid HS. HS produces poor anther dehiscence, reducing rice pollination (Satake and Yoshida, 1978; Matsui et al., 1997) by inhibiting pollen swelling and the normal function of the thecae (Matsui et al., 2000).

Nagina22 (N22) is a highly heat-tolerant rice variety; thus, it is a good source for mining genes for tolerance against heat and drought stress (Bahuguna et al., 2014; Ye et al., 2012). Early anther dehiscence, while anthers are still inside the glumes, leads to HS avoidance and greater pollination in N22 (Satake and Yoshida, 1978; Mackill et al., 1982). The metabolome and transcriptome analyses have revealed sugar starvation as a major process involved in reproductive failure: higher expression of sugar transporter (MST8) and a cell wall invertase (INV4) were identified in N22 in response to combined heat and drought stress (Li et al., 2015). Jagadish et al. (2010) showed that Moroberekan (Japonica), Indica Rice 64 (IR64, Indica) and N22 (Aus) demonstrated 18%, 48% and 71% spikelet fertility when exposed to 6 h of heat stress (38°C), at anthesis stage, respectively. Two-dimensional gel electrophoresis-based proteome analysis of anthers revealed the presence of 46 proteins, including significant upregulation of the cold-shock and heat-shock proteins in N22 (Jagadish et al., 2010). However, the latter study did not divulge the global perspective of the anther proteomes in these varieties in response to HS. In another study, anther proteome of thermotolerant Dianxi 4 rice genotype (Japonica) stressed at 38°C demonstrated the proteins responsible for maintaining proteostasis, like sHsps, Hsp70 and trehalose synthase, were upregulated (Kim et al., 2015). Interestingly, the comparative proteome between heat-sensitive and heat-tolerant rice varieties are not yet explored.

In the present study, we have used a shotgun proteomics approach to analyze the rice anther proteome in contrasting genotypes such as heat-sensitive Moroberekan, moderately heat-tolerant IR64 and heat-tolerant N22. Our analysis focused on anther dehiscence stage, as this stage is crucial for the fertilization process in rice. The anther proteome changes were identified in response to two stress regimes: short-term heat stress (ST_HS) and long-term heat stress (LT_HS). This study aimed to determine the key regulators involved in heat stress responses to develop rice varieties with better thermotolerance through modern molecular breeding strategies.

2 Material and methods

2.1 Plant material and growth conditions

Three rice genotypes Moroberekan, IR64 and N22 were used in this study. The details of the growth conditions followed in the analysis of Moroberekan, IR64 and N22 rice genotypes, as well as the relationship of the pollen grain stages on the progression to anthesis in rice to the stress regimes followed in this study, are shown in Figure 1. Several independent sets of plants were transferred to growth chambers for the HS treatment to obtain plants with a variable time of flowering. For the ST_HS, anthers were harvested at 'just before anthesis' stage from plants subjected to one cycle of 42°C (4 h) HS (Figure 1A). For the LT_HS, anthers were harvested at a similar growth stage from plants subjected to 6 days of consecutive 38°C (6 h) HS (Figure 1A). In LT_HS treatment, the temperature regime followed was 28°C at 6 AM, gradually increased to 38°C by 9 AM, maintained at 38°C till 3 PM and gradually lowered to 28°C and maintained overnight. In the unstressed control (CON) condition, the temperature followed was: 28°C at 6 AM, gradually increased to 30°C and maintained till 6 PM and gradually brought down to 28°C overnight. Plants (15 to 35) were harvested (for each treatment), and flowering spikelets were collected on ice. The spikelets were dissected to separate the pollinated pistils and anthers. The anthers were harvested from only the middle portion of the whole spikelet to maintain uniformity in the development stage. Dissected anthers were collected in tubes containing liquid nitrogen and stored at –80°C until further use (Figures 1B, C).

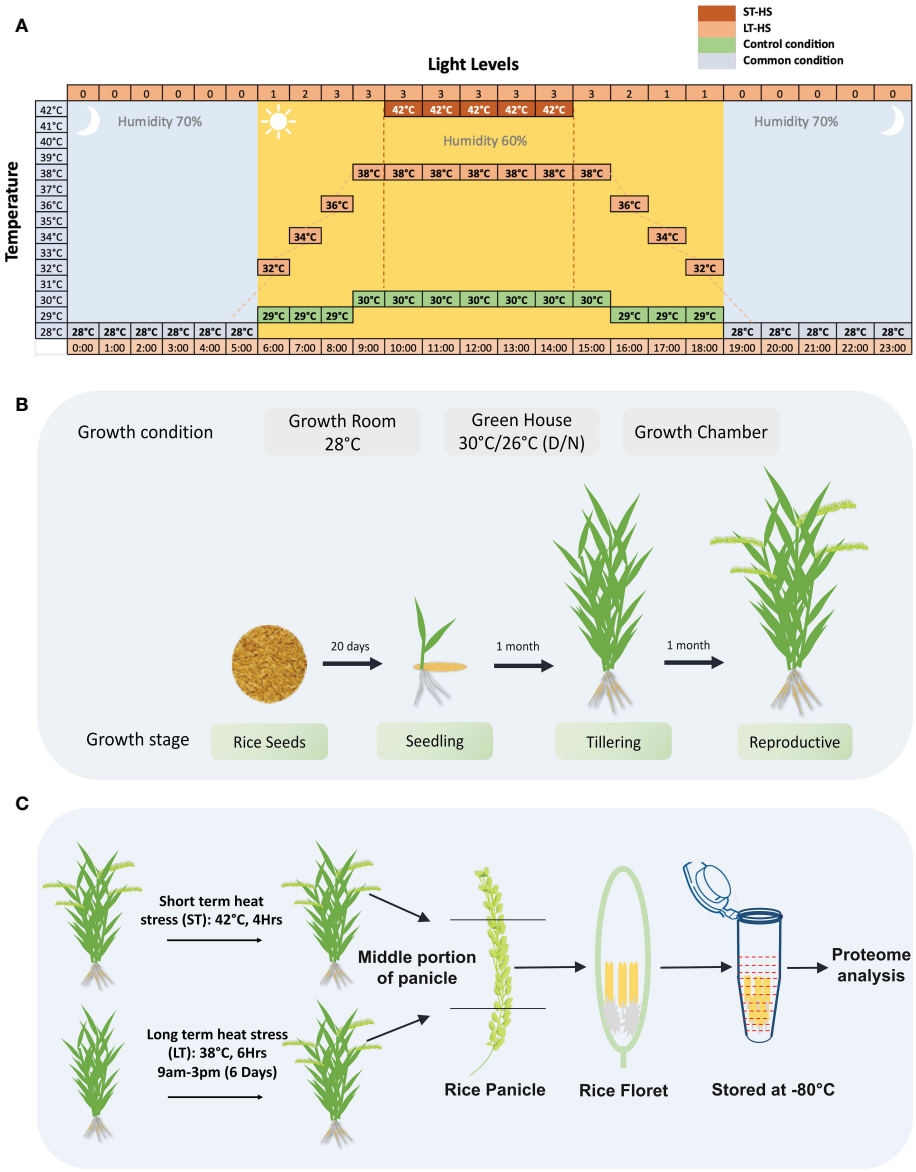


FIGURE 1
Details of the growth and stress conditions followed in the experiment. **(A)** Illustration showing the stress regimes used in the current study. In the short-term heat stress, plants were exposed to 42°C for 4 hours, and in long-term heat stress, plants were exposed to 38°C temperature for 6 hours. **(B)** Growth conditions for growing rice plants for the study. **(C)** The stress regimes used in ST-HS (short-term heat stress) and LT-HS (long-term heat stress) treatments followed for the proteome analysis are shown. The plants at the reproductive stage were stressed, and the anthers were dissected just before the anthesis.

2.2 Protein isolation and western blotting analysis

Harvested anthers (~20 mg) were lysed and homogenized in TRIzol™ Reagent (Invitrogen™) to extract proteins for Western blotting. Extracted proteins were dissolved in 0.25% SDS and protein was quantified using Bradford reagent. For Western blotting, 10 µg of total soluble protein was loaded on the SDS-gel. Proteins were transferred to the nitrocellulose membrane (Amersham, UK). Hsp101 protein was probed using polyclonal anti-AtHsp101/ClpB primary antibodies (#AS08283, Agrisera, Sweden) and horseradish peroxidase (HRP) -conjugated anti-rabbit secondary antibodies (Sigma-Aldrich, USA), both at 1:10000 dilutions. The same membrane was probed with anti-tubulin alpha chain antibodies (# AS10680, Agrisera) for loading control. The blots were developed by enhanced chemiluminescence (ECL).

2.3 Protein extraction and pre-fractionation

Proteins were extracted and quantified using a shotgun proteomics approach (Chaturvedi et al., 2013; Valledor and Weckwerth, 2014; Chaturvedi et al., 2015). Anther samples were freeze-dried and ground for 2 min in a shaking mill using steel balls (2 mm diameter). The homogenized samples were resuspended in 200 µL of protein extraction buffer (100 mM Tris- HCl, pH 8.0; 5% SDS, 10% glycerol; 10 mM DTT; 1% plant protease inhibitor cocktail (Sigma P9599) and incubated at room temperature for 5 min followed by incubation for 2.5 min at 95°C and centrifugation at 21000 × g for 5 min at room temperature. The supernatant was carefully transferred to a fresh tube. Two hundred microliter of 1.4 M sucrose was added to the supernatant. Proteins were extracted twice with 200 µL TE buffer-equilibrated phenol followed by counter extraction with 400 µL of 0.7 M sucrose. Phenol phases were combined and mixed with 2.5 volumes of 0.1 M ammonium acetate in methanol to precipitate proteins. After 16 h of incubation at -20°C, samples were centrifuged for 5 min at 5000 × g. The pellet was washed twice with 0.1 M ammonium acetate, once with acetone and air-dried at room temperature. The pellet was re-dissolved in 6 M urea and 5% SDS and protein concentration were determined using the bicinchoninic acid assay. Proteins were pre-fractionated by SDS-polyacrylamide gel electrophoresis. Forty micrograms of total protein were loaded onto a gel and run for 1.5 cm. Gels were fixed and stained with methanol: acetic acid: water: Coomassie Brilliant Blue R-250 (40:10:50:0.001). Gels were destained in methanol: water (40:60), and then each lane was divided into two fractions (Ghatak et al., 2021).

2.4 Protein digestion and LC-MS/MS

Gel pieces were destained, equilibrated and digested with trypsin (using Trypsin Sequencing Grade from Roche [11418475001]), desalted employing Bond-Elute C-18 SPEC plate (Agilent Technologies, Santa Clara, CA, USA) and concentrated in a Speed Vac concentrator (SCANVAC Cool Safe 110-4, Speed Vacuum concentrator, Labogene). Prior to mass spectrometric measurement, the tryptic peptide pellets were dissolved in 4% (v/v) acetonitrile, 0.1% (v/v) formic acid. One µg of the digested peptide (2 biological replicates from each condition) was loaded on a C18 reverse-phase column (Thermo Scientific, EASY-Spray 500 mm, 2 µm particle size). The separation was achieved with a 90 min gradient from 98% solution A (0.1% formic acid) and 2% solution B (90% ACN and 0.1% formic acid) at 0 min to 40% solution B (90% ACN and 0.1% formic acid) at 90 min with a flow rate of 300 nl min^{-1}. nESI-MS/MS measurements were performed on Orbitrap Elite (Thermo Fisher Scientific, Bremen, Germany) with the following settings: Full scan range 350–1800 m/z resolution 120000 max. 20 MS2 scans (activation type CID), repeat count 1, repeat duration 30 s, exclusion list size 500, exclusion duration 30 s, charge state screening enabled with the rejection of unassigned and +1 charge states, minimum signal threshold 500.

2.5 Peptide and protein identification

For protein identification, raw data were searched with the SEQUEST algorithm present in Proteome Discoverer version 1.3 (Thermo, Germany) as described in (Valledor and Weckwerth, 2014) using protein FASTA. We have used the following settings in Proteome Discoverer for data analysis which include: Peptide confidence: High, which is equivalent to 1% FDR, and Xcorr of 2, 3, 4, 5, 6 for peptides of charge 2, 3, 4, 5, 6. The variable modifications were set to acetylation of the N-terminus and oxidation of methionine, with a mass tolerance of 10 ppm for the parent ion and 0.8 Da for the fragment ion. The number of missed and/or non-specific cleavages permitted was 2. There were no fixed modifications, as dynamic modifications were used. The identified proteins were quantitated based on total ion count, followed by an NSAF normalization strategy (Paoletti et al., 2006).

$$(NASF)\kappa = (PSM/L)\kappa/\sum i = 1N(PSM/L)i$$

The total number of spectra counts for the matching peptides from protein k (PSM) was divided by the protein length (L), then divided by the sum of PSM/L for all N proteins.

All the MS/MS spectra of the identified proteins and their meta-information were further uploaded to the PRIDE repository, Project accession: PXD035952.

2.6 Statistics for proteome data analysis and bioinformatics

A t-test was applied to find significantly altered proteins between control and stress condition in each genotype, and proteins with p-value < 0.05 were further considered. Differentially expressed proteins (DEPs), fold change (FC) were calculated for all identified proteins in HS compared to CON. Proteins with FC ≥ 2 (upregulation) or ≤ 0.5 (downregulation) were considered. Gene ontology was determined using RiceNetDB (http://bis.zju.edu.cn/ricenetdb/documentation.php). Only gene ontology (GO) terms with a p-value ≤ 0.01 were used for the analysis. Gene ontology network analysis for the identified DEPs was performed using Cytoscape software (Version 3.7.1) with ClueGO (Version 2.5.4) plug-in. ClueGo was used to visualize non-redundant biological terms for large clusters of genes in a functionally grouped network. The *Oryza sativa* (Japonica) marker list was used for searching

biological processes, molecular function and cellular components related to proteins identified. *P*-value was calculated using a two-sided hypergeometric test and Benjamini-Hochberg adjustment for multiple test corrections. GO terms with P-value<0.05 were considered significant. The protein-protein interactions network in each set of up and down-regulated proteins was generated using the STRING database (Version 11.0). We restricted the parameter with the highest confidence to 0.9 and neglected the single molecules found in the interaction map to best predict the interactions.

3 Results

3.1 Proteome alteration in anther tissue of the three rice genotypes in response to heat stress

Heat stress demonstrated phenotypic changes in N22 plants where mature spikelets were deformed, and pollen viability was reduced, especially after long-term heat stress compared to short-term heat stress (Figure S1).

The detailed lists of the identified proteins in Moroberekan, IR64 and N22, are provided in supporting information (Tables S1–S3). In total, proteins identified in Moroberekan were 2689 in control (referred to as MCT for Moroberekan control), 2529 in ST_HS (MST for Moroberekan ST_HS) and 2687 in LT_HS (MLT for Moroberekan LT_HS). Similarly, proteins identified in IR64 were 2739 in ICT (IR64 control), 2690 in IST (IR64 ST_HS) and 2973 in ILT (IR64 LT_HS) and in N22, 3072 proteins were identified in NCT (N22 control), 3009 in NST (N22 ST_HS) and 3045 in NLT (N22 LT_HS) as shown in Figure S2A. Thus, the number of identified proteins was in the order of N22>IR64>Moroberekan. The number of unique and common proteins in each rice genotypes is represented using the Venn diagram in Figure S2B. Between the control and ST_HS regimes, the number of overlapping proteins was higher in N22 compared to Moroberekan and IR64. More proteins were common in the control and LT_HS in Moroberekan than in the other two rice genotypes. Between ST_HS and LT_HS, overlapping protein numbers were higher in Moroberekan than in the other two rice genotypes. The identified proteome from three rice genotypes was compared with transcripts expressed during different developmental stages of rice plants. For this, co-expression analysis tools were used from the rice genome annotation project database. As an output, a z-score trend plot of the locus IDs of proteins was used. In all the rice genotypes, there was a clear dense peak just above the inflorescence P6 samples (mature pollen stage) in the plot generated using the inflorescence and seed developmental series (GSE6893) dataset and panicle 5 (heading panicle stage) and stamen (stamen 1 day before flowering) in the plot generated using tissue atlas from Minghui 63 rice (GSE19024) (Figure S3). These results suggested that the proteins identified in the present work corresponded greatly with the transcripts expressed in anthers of the mature pollen stage.

3.2 Identification of differentially expressed proteins (DEPs)

The lists of the differentially expressed proteins (DEPs) in Moroberekan, IR64 and N22 rice types are provided in supporting information Table S4. The total identified proteins were sorted into DEPs between the control and stressed samples by applying *p*-value<0.05 (Figure 2A). In Moroberekan, 734 DEPs were identified in MST and 254 in MLT. Of the 734 MST DEPs, 130 proteins were upregulated >2 fold, and of these, 103 proteins were exclusively identified in ST_HS. In the same dataset, 144 proteins were downregulated, and of these, 104 proteins were present only in the control but not in the HS condition. Of the 254 MLT DEPs, 74 proteins were upregulated, of which 64 were uniquely identified in LT_HS. In this analysis, 94 proteins were downregulated and of which 69 proteins were exclusively present in the control condition. In IR64, 243 DEPs were identified in IST, out of which 72 were upregulated, and 60 were noted exclusively in IST condition. In the same dataset, 85 DEPs were downregulated, of which 52 were exclusively present only in the control condition. Similarly, in ILT, 316 DEPs were identified, which include 131 upregulated, of which 105 were exclusively present in the LT_HS condition. In the same dataset, 79 DEPs were downregulated, and 42 were exclusively present in the control condition. In N22, 243 DEPs were identified in NST. At FC >2, 56 DEPs were upregulated and 33 were exclusively present in the NST. Under the NST regime, 82 DEPs were downregulated, and 71 were exclusively present in the control condition. Out of 312 DEPs identified in NLT, 78 were upregulated, of which 50 were unique under LT_HS. In the same dataset, 64 DEPs were downregulated, of which 29 were present in control but not in HS condition.

The analysis of the overlapping DEPs among ST_HS and LT_HS datasets showed that in Moroberekan, 30 upregulated proteins were common between MST and MLT treatments. In comparison, 55 downregulated proteins were common between MST and MLT datasets (Figure 2B). In IST and ILT, 22 DEPs were common. In NST and NLT datasets, 9 DEPs were commonly upregulated and 31 were downregulated (Figure 2B). The number of common DEPs between all three rice genotypes was smaller in ST_HS than in LT_HS treatments. Among the three rice genotypes, there were no common upregulated DEPs, while two common DEPs down-regulated were β-amylase (Os07g35880) and cytochrome P450 (Os02g38290) in the ST_HS treatment. In ST_HS, the numbers of uniquely upregulated DEPs were 127, 66 and 31 in Moroberekan, IR64 and N22, respectively (Figure S4). In ST_HS, unique downregulated proteins were 129, 68 and 74 in

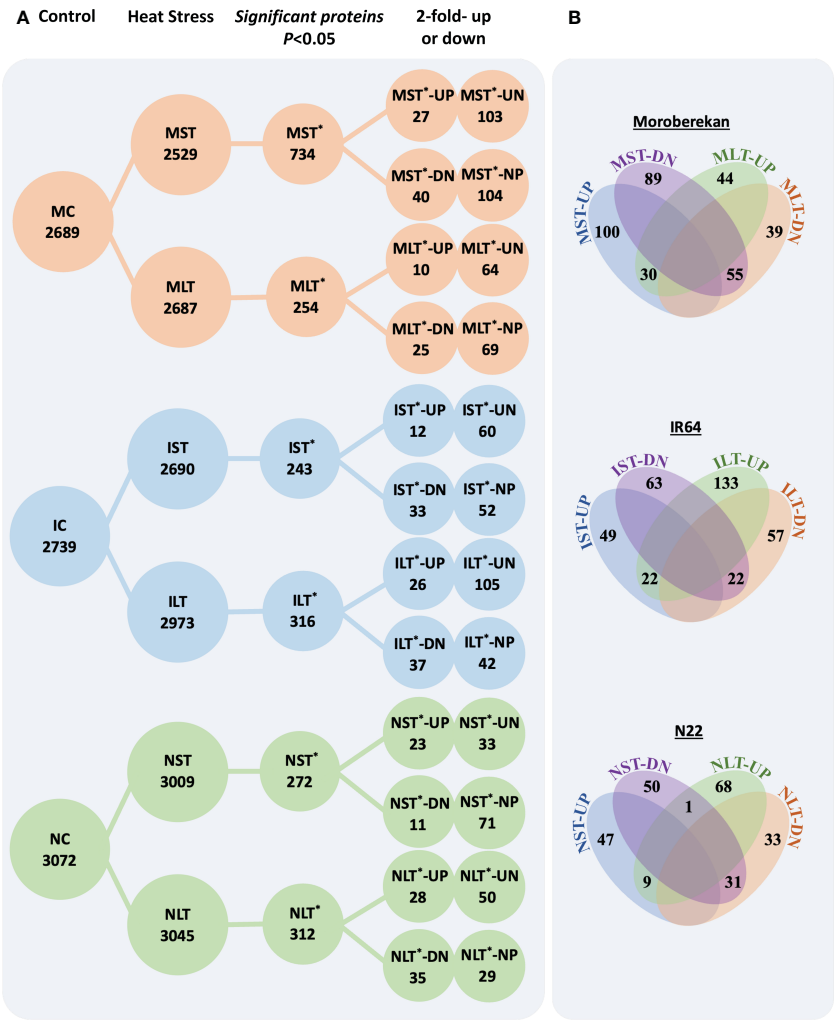


FIGURE 2
(A) The differentially expressed proteins in different rice genotypes under different heat stress regimes. **(B)** Venn diagrams show the overlapping significant and differentially expressed proteins between different datasets in each genotype. ST, Short term heat stress; LT, Long-term heat stress; MC, Moroberekan control; MST, Moroberekan ST_HS; MLT, Moroberekan LT_HS; IC, IR64 control; IST, IR64 ST_HS; ILT, IR64 LT_HS, NC, N22 control; NST, N22 ST_HS; NLT, N22 LT_HS; UP, upregulated; DN, downregulated, *-differentially expressed proteins numbers; UN-uniquely present in HS, proteins present in the heat stress sample but not in control; NP-not present in HS, proteins present in control but not detected in the heat stress condition.

Moroberekan, IR64 and N22, respectively. Overall, in LT_HS, unique upregulated proteins were 68, 126 and 70 in Moroberekan, IR64 and N22, respectively, and downregulated proteins were 88, 71 and 61 in Moroberekan, IR64 and N22, respectively (Figure S4).

3.3 Expression pattern of Hsp101 in rice anthers

Cytosolic ClpB/HSP100 family proteins are molecular chaperones that promote the renaturation of protein aggregates and are required to develop acquired thermotolerance (Hong and Vierling, 2000). To validate the protein abundance of Hsp 101, a western blot analysis of control and heat-stressed anther tissues was performed. Moroberekan and IR64 showed higher levels of Hsp101 than N22 under ST_HS (Figure 3, upper panel). Similarly, the Hsp101 protein showed higher levels in Moroberekan and IR64 than N22 (Figure 3, lower panel).

3.4 Functional regulation of the differentially expressed proteins (DEPs)

The relative abundance of the selected proteins differentially expressed in at least one of the rice genotypes and at least one

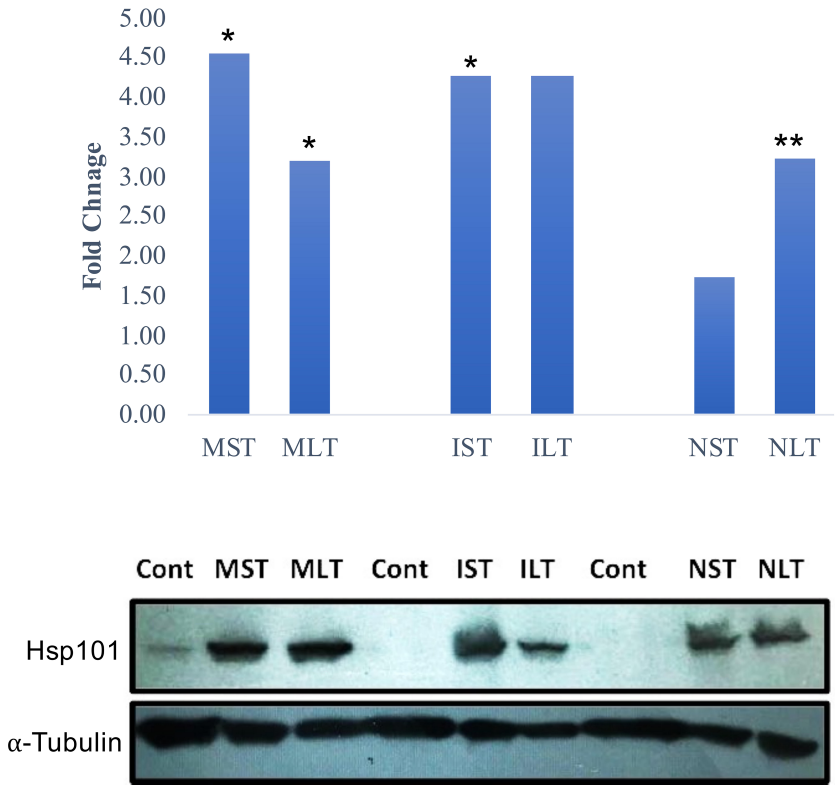


FIGURE 3
Hsp101 expression profile in Moroberekan, IR64 and N22 rice types post short-term (ST) and long-term (LT) heat stress conditions. The upper panel shows Hsp101 protein expression found in proteome study with significance (*- p value<0.05; **- p value< 0.01). The lower panel shows western blotting for Hsp101 in the anther samples of Moroberekan, IR64 and N22 rice types. Alpha-Tubulin was used as a loading control for the blot to normalize the protein levels detected across the gel. ST, Short term heat stress; LT, Long-term heat stress; MC, Moroberekan control; MST, Moroberekan ST_HS; MLT, Moroberekan LT_HS, IC, IR64 control; IST, IR64 ST_HS; ILT, IR64 LT_HS, NC, N22 control; NST, N22 ST_HS; NLT, N22 LT_HS.

condition related to diverse cellular pathways was used for constructing the heat maps for their relative expression (Figure 4). In amino acid metabolism, most of the proteins were downregulated. S-adenosylmethionine synthetase was upregulated in all the rice genotypes except in MST. Amino methyltransferase, a glycine cleavage T family protein, was upregulated in both NST and NLT but downregulated in MST, IST and ILT. Phospho-2-dehydro-3-deoxyheptonate aldolase was downregulated in all the rice genotypes. GHMP kinase and ATP binding protein were specifically upregulated in NST and NLT. Several stress-associated genes were upregulated. These included Hsp20/alpha crystallin family protein, DnaK family protein, DnaJ homolog subfamily B member 11 and Hsp101. Genotype-specific Hsp20/alpha-crystallin family protein levels were identified in ST_HS and LT_HS of Moroberekan, IR64 and N22 (Figure 4). On the other hand, the proteins in the biotic and drought/salt categories were down-regulated.

HVA22 (ABA and stress-inducible gene) was downregulated in MST, and no change was noted in NST and NLT. Endothelial differentiation-related factor 1 was upregulated in Moroberekan and downregulated in IR64 and N22. Most of the DEPs involved in fatty acid biosynthesis were unchanged or downregulated. Interestingly, most ribosomal proteins were downregulated in the three rice genotypes in all conditions (Figure 4). However, a few ribosomal proteins, including 60S ribosomal protein subunits, 60S ribosomal protein L36-2 and ribosomal protein L7Ae were heat induced in NLT. Proteins related to post-translational modification were also heat inducible, especially in the case of NLT. Most photosynthesis-related proteins were downregulated except for a few chlorophyll-binding proteins, specifically in NLT.

3.5 Gene ontology (GO) enrichment analysis of the DEPs

GO enrichment analysis determined that three major biological processes (BP): cellular process, metabolism, and physiological process were enriched (Figure 5). These terms were enriched in both up-and down-regulated DEPs of all three rice genotypes, except the cellular process category which was not represented in

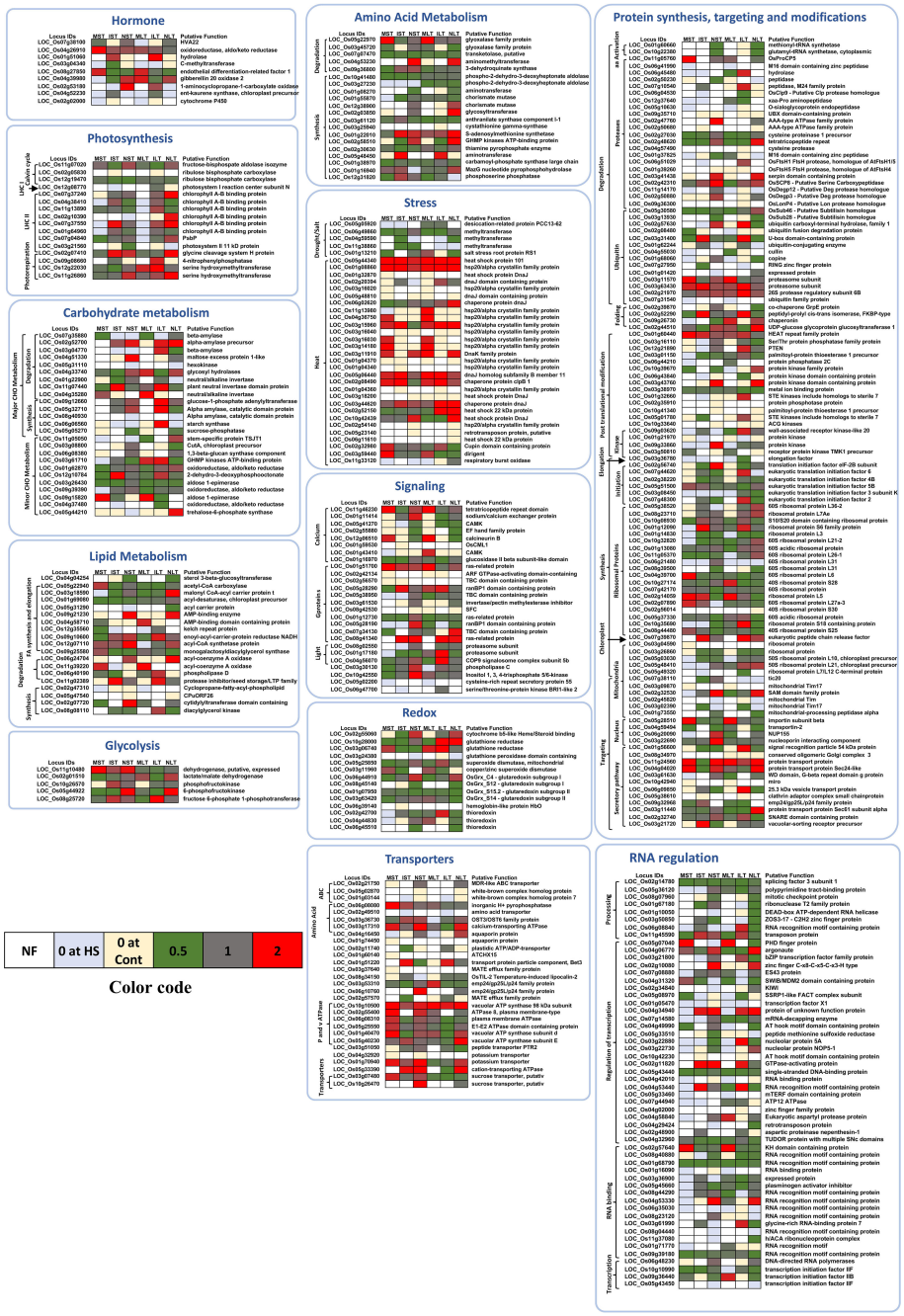


FIGURE 4
Heat maps depicting the relative abundance of proteins related to different pathways. The proteins significantly changed and differentially abundant in at least one genotype and at least in one condition are shown here. The color bar indicates the color codes used in making the heat map. White color- protein not found (NF); light blue- not found in HS; yellow- exclusively found in HS; green- downregulated (below 0.5 folds); grey- no change; red- upregulated (above 2 folds). ST, Short term heat stress; LT, Long-term heat stress; MC, Moroberekan control; MST, Moroberekan ST_HS; MLT, Moroberekan LT_HS; IC, IR64 control; IST, IR64 ST_HS; ILT, IR64 LT_HS; NC, N22 control; NST, N22 ST_HS; NLT, N22 LT_HS; UP, upregulated; DN, downregulated.

IST-UP (upregulated in IST of IR64). In response to the abiotic stimulus category, entries were identified in MST-UP and IST-UP, while no entries were found for it in NST-UP. Interestingly, the biological process of transport was enriched in MST-UP (with 26 proteins) and NST-UP (with 14 proteins) and ILT-DN. In the molecular function (MF), the transport category was enriched in MST-UP. In the cellular component, GO terms related to plastids, protein complexes, ribonucleoprotein, ribosomes, and thylakoids

were enriched in MST-DN, NST-DN and MLT-DN. GO ontology of BP: protein biosynthesis and macromolecular biosynthesis were enriched in MST-DN only. This trend was reflected in enriched MF: structural molecule activity and RNA binding of MST-DN. The vacuole was enriched in MST-UP and MST-DN and, to a small extent, in NST-UP.

The catalytic activity was noted in all treatments except in protein regulation of LT-HS. To further interpret the biological relevance of DEPs, GO terms and pathways were functionally grouped in ClueGO (Figure 6). ClueGO integrates GO terms and KEGG pathways, generating a functionally organized GO/pathway network. In MST-UP proteins, a network of proton export across the plasma membrane was identified. A major network in the N22 proteins NST-UP set was GTPase activator activity and vitamin E biosynthetic process. GTPases play an essential role in vesicular trafficking and act as molecular switches regulating plant developmental processes, including embryonic development and reproduction. Proteins in NLT-UP were enriched in unfolded protein binding process and cellular components cytosol, endoplasmic reticulum (ER) and chloroplast. Accumulation of misfolded proteins is at the core of HS in all cellular compartments, and in the ER, it activates a signaling unfolded protein response (UPR) pathway. The upregulation of unfolded protein binding process is in line with the reports showing components of the endoplasmic reticulum-folding machinery and UPR is upregulated at specific stages of pollen development, pollen germination, polar tube growth and fertilization (Fragkostefanakis et al., 2016; Chaturvedi et al., 2021).

3.6 Interaction network analysis of the identified proteome

The STRING database was used to identify the, interactions among proteins in three rice genotypes. Different functional

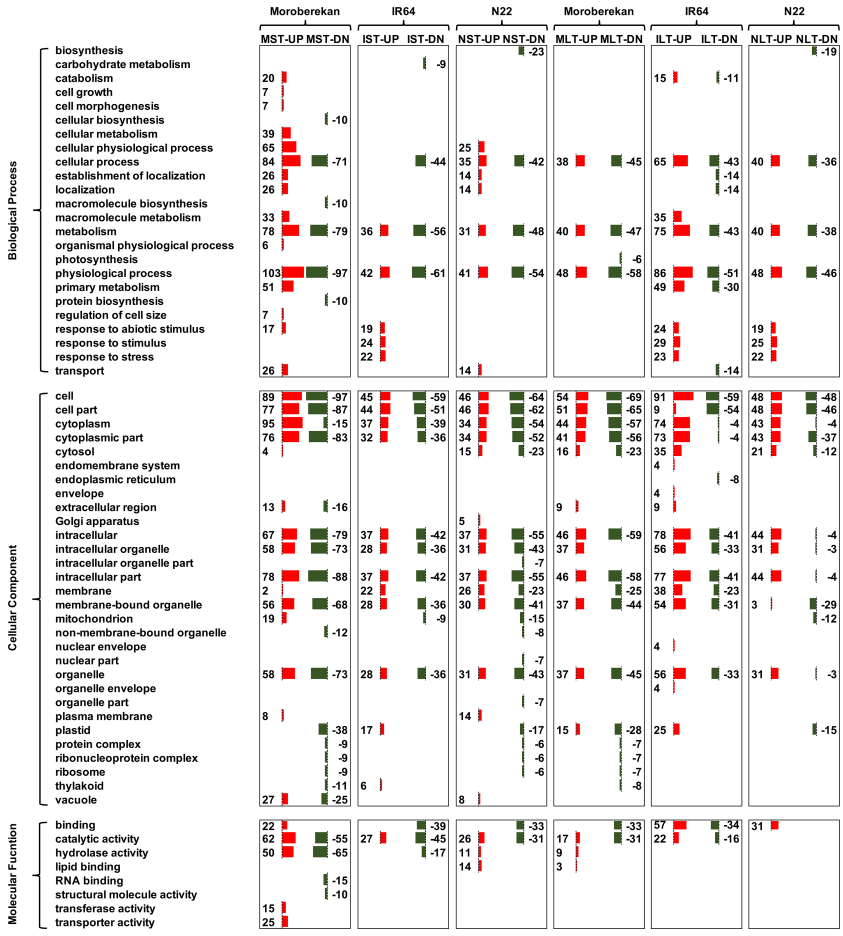


FIGURE 5
GO term enrichment analysis using RiceNetDB for the proteins up- and down regulated in short-term (ST) and long-term (LT) heat stress conditions in Moroberekan, IR64 and N22 rice genotypes. The numbers are indicative of the proteins enriched for each GO term.

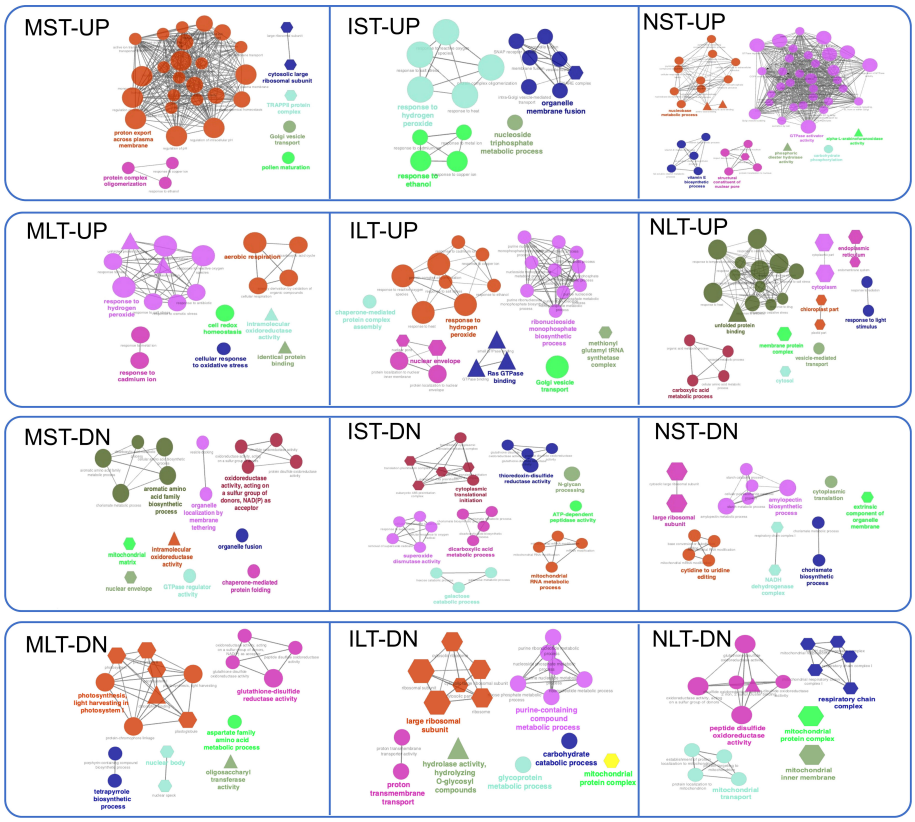


FIGURE 6
Gene Ontology (GO) network analysis of differentially abundant proteins. GO term enrichment network for the up-and down-regulated proteins in Moroberekan, IR64 and N22 short-term and long-term heat stress conditions was created using Cytoscape software (Version 3.7.1) with ClueGO (Version 2.5.4) plug-in. The colored circles represent the biological function, triangles represent the molecular function, and hexagons represent the cellular components hierarchically. ST, Short term HS; LT, Long, term heat stress; MC, Moroberekan control; MST, Moroberekan ST_HS; MLT, Moroberekan LT_HS; IC, IR64 control; IST, IR64 ST_HS; ILT, IR64 LT_HS; NC, N22 control; NST, N22 ST_HS; NLT, N22 LT_HS; UP, upregulated; DN, downregulated.

clusters emerged among the interacting proteins. Only selected interactions with more than 0.9 confidence are shown (Figure 7). In the MST-UP, nine distinct network clusters were noted (CI-CIX). These clusters included the respiratory metabolism cluster (CI), ribosomal proteins cluster (CII), and Hsps cluster (CIV). In the MST-DN, four distinct clusters were observed, including ribosomal protein (CI) and RNA modification (CII). The IST-UP revealed three clusters of interactions, including an Hsp cluster (CI). IST-DN category showed two clusters, one of which was the ribosomal proteins cluster. In both NST-UP and NST-DN, three clusters were identified. The NST-DN category contained a large cluster of ribosomal proteins, RNA splicing (CI) and starch metabolism (CIII). MLT-UP and MLT-DN two and four clusters were identified, respectively. The MLT-DN contained a cluster of ribosomal proteins (CI). ILT-UP and ILT-DN contained seven and four clusters of protein-protein interactions, respectively. ILT-UP had a proteasome proteins cluster (CI), whereas ILT-DN had a ribosomal protein interaction cluster (CI). In NLT-UP, two clusters were observed, both of which were of Hsps (CI and CII). In NLT-DN, four protein interaction clusters were noted, including one big ribosomal protein cluster (CI), one RNA splicing cluster (CIII) and one starch metabolism cluster (CIV). In the NLT-UP, cluster CI contained chaperone protein ClpB1 (Os05g44340) with 26.7 kDa sHsp (Os03g14180) and 18.0 kDa class II sHsp (Os01g08860), which further interacted with 24.1 kDa mitochondrial sHsp (Os02g52150). Similarly, 24.1 kDa sHsp was found in anther proteome of N22 in an earlier study, and its higher accumulation was linked to better thermotolerance of N22 (Jagadish et al., 2010).

4 Discussion

Heat stress response at the reproductive stage of three rice genotypes Moroberekan, IR64, and N22 demonstrate differential effects: Moroberekan is heat-sensitive, IR64 is moderately heat-tolerant, and N22 is heat-tolerant (Jagadish et al., 2010). In the

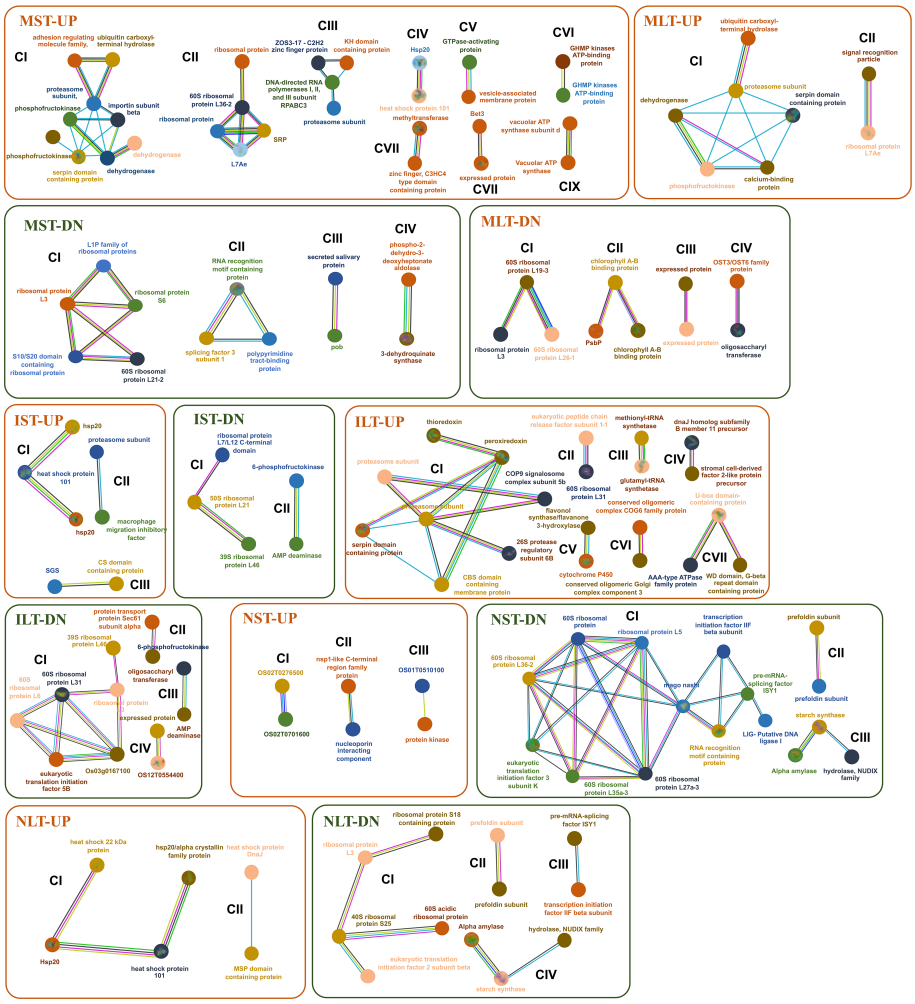


FIGURE 7
Protein interaction network profiles represent up- and down-regulated protein interactions in Moroberekan, IR64 and N22 in short-term heat stress conditions. The parameter was restricted with the highest confidence to 0.900 neglecting the single molecules found in the interaction map to best predict the interactions. Each circle represents a protein, and lines represent the type of evidence available for finding interactions between the two proteins. Redline- the presence of fusion evidence; Green line- neighborhood evidence; Blue line- cooccurrence evidence; Purple line- experimental evidence; Yellow line- text mining evidence; Light blue line- database evidence; Black line- co-expression evidence. ST, Short term heat stress; LT, Long-term heat stress; MC, Moroberekan control; MST, Moroberekan ST_HS; MLT, Moroberekan LT_HS, IC, IR64 control; IST, IR64 ST_HS; ILT, IR64 LT_HS; NC, N22 control; NST, N22 ST_HS; NLT, N22 LT_HS; UP, upregulated; DN, downregulated.

previous study, it has been reported that Nipponbare and N22 rice seedlings demonstrate differential phenotype and protein regulation in response to heat stress under basal (unprimed heat stress regime) as well as short-term and long-term acquired (primed heat stress regimes) thermotolerance assays (Lin et al., 2014). The present study attempts to understand proteome regulations associated with short-term and long-term HS responses in three rice genotypes. Here in N22 higher levels of 6-phosphofructokinase and fructose 6-phospahate 1-phosphotransferase were identified compared to Moroberekan indicating that glycolytic activity was enhanced in N22 under heat-stress conditions. This data were consistent with the previous study were N22 and Moroberekan were compared at metabolome and transcriptome levels (Li et al., 2015). The latter study found the higher expression of Hsp100 in Moroberekan, which agrees with our findings. Jagadish and coworkers also reported that IR64 and N22 were highly induced for putative low molecular weight HSPs (Jagadish et al., 2010). In the current study, we observed a similar response in the case of N22 and IR64 where many of the small sHSPs were induced under heat stress conditions.

Despite having a higher number of proteins in N22, MST showed a higher number of significantly altered proteins (734 proteins) compared to IST (243 proteins) and NST (272

proteins). Several changes in protein levels in different rice genotypes are linked to reproductive processes (Figure S2). For example, MADS2 (Os01g66030) regulation in NST-DN, NLT-DN and MST-UP. MADS2 interacts with OsMADS16 and functions in lodicule and stamen development (Kong et al., 2019). The presence of expansins (4 proteins) in the MST-UP suggests their significance in reproduction as the knockdown of Os10g44710 (beta-expansin 2; EXPB2) was reported to cause a sterile phenotype in rice (Huang et al., 2018). Expansins are involved in cell wall loosening and its extension that facilitates the pollen tube penetration through the stigma and style (Tabuchi et al., 2011). MLT-DN was enriched by cellular component-plastid (34 proteins), cytosol (23 proteins), ribosome (8 proteins) and molecular function _binding (24 proteins) and translation (8 proteins) processes. Accordingly, proteins related to these processes, like ribosomal proteins, RNA binding proteins, photosystem and chlorophyll a-b binding proteins, were down-regulated. OsPsbS1 (Os01g64960) in MST-DN and MLT-DN exert control over the CO_2 assimilation rate in fluctuating light in rice (Hubbart et al., 2012). Aspartic protease (LOC_Os04g58840) and cysteine protease CP1 (Os04g57490) were present in ILT_UP. These proteases are essential for male fertility (Huang et al., 2013). In the mutant plants of aspartic protease pollen maturation was normal, but germination was hampered (Huang et al., 2013). CP1 suppressed mutant showed a significant defect in pollen development. Rice apoptosis inhibitor 5 (LOC_Os02g20930, API5) was downregulated in MST, resulting in delayed degeneration of the tapetum due to inhibition of the tapetal programmed cell death (PCD) process leading to defects in the formation of male gametophyte. Interestingly, OsAPI5 interacts with two DEAD-box ATP-dependent RNA helicases, API5-INTERACTING PROTEIN1 (AIP1) and AIP2 form dimers that interact directly with the promoter region of CP1 gene (Li et al., 2011). In NST-DN, Os03g07140 [male sterility 2; MS2/ defective pollen wall (DPW), a fatty acyl-CoA reductase that produces 1-hexadecanol] was identified. This protein is expressed in both tapetal cells and microspores during anther development. DPW participates in a conserved step of primary fatty alcohol synthesis for anther cuticle and pollen sporopollenin biosynthesis in monocots and dicots. Male sterile mutant, *dpw*, displays defective anther development and degenerated pollen grains with an irregular exine (Shi et al., 2011).

4.1 Regulation of HSPs in three rice genotypes

In all three rice genotypes, different Hsps were highly induced. Hsps showed genotype - specific protein abundance pattern which can also be used as genotype-specific markers for the heat stress response. The abundance of Hsp20, chaperone protein DnaJ homolog subfamily B member Os05g06440.1 and Hsp101 (Os05g44340) were upregulated across all the three genotypes under stress condition. Hsp101 is an important protein in imparting thermotolerance in bacteria, yeast and Arabidopsis (Kumar et al., 2016; Kumar et al., 2020; Tiwari et al., 2021). The introduction of *Arabidopsis thaliana* Hsp101 (AtHsp101) into rice made the host variety more thermotolerant (Katiyar-Agarwal et al., 2003). Hsp20 provide thermotolerance to various levels of life forms such as bacteria and plants (Sarkar et al., 2009). Hsp20 co-aggregate with misfolded proteins (Sarkar et al., 2009; Kumar et al., 2015; Ungelenk et al., 2016; Sarkar et al., 2020). Furthermore, ATP-dependent Hsp70 and Hsp100 chaperones facilitate the solubilization and refolding of protein aggregates (Cherkasov et al., 2013; McLoughlin et al., 2016; Merret et al., 2017).

In NLT, upregulated proteins demonstrated a significant interaction with Hsps (Figure 6). This interaction was not enriched in ILT and MLT. A major interaction with the ribosomal proteins suggests reduction in translation during heat stress. Other major interactions with proteasome and dehydrogenase also indicate that MST is involved in more protein degradation. Heat-induced ribosome inhibits pre-rRNA processing which triggers the imbalance in ribosomal profiles in *Arabidopsis thaliana* (Merret et al., 2015; Darriere et al., 2022).

Another category of stress proteins is related to carbohydrate metabolism. N22 was enriched in proteins such as trehalose-6-phosphate synthase, trehalose synthase, and haloacid dehalogenase-like hydrolase. The trehalose accumulation in rice protects plants from various stresses (Benaroudj et al., 2001). HS affects rice sugar partitioning. It appears that heat-tolerant plant types are better at preserving starch and sugar levels in pollen, which helps them germinate at higher temperatures (Pressman et al., 2002; Firon et al., 2006; Chaturvedi et al., 2015; Jegadeesan et al., 2018). Protein import into the nucleus is another process enriched in N22. Hsfs and other transcription factors are dynamically transported across the nucleus to induce the expression of Hsps and other stress-responsive genes (Scharf et al., 2012). This may indicate more efficient trafficking in N22 under heat stress conditions.

4.2 Differential regulation of hormone metabolism and signaling

Thermotolerance involves different signaling pathways where phytohormones play important roles. HS leads to the abscission of reproductive organs due to increased abscisic acid and ethylene levels and reduced levels of auxins (Binder and Patterson, 2009). Also, the alterations in auxin biosynthesis and

cytokinin content reduces pollen sterility and kernel filling, respectively, in cereals (Banowetz et al., 1999; Sakata et al., 2010). In the present study, DEPs involved in hormone metabolism were mostly downregulated under HS conditions. HVA22 was specifically induced in IST and ILT but downregulated in MST and MLT, and no change was noted in NST and NLT. This protein acts downstream of GAMyb and activates programmed cell death and other GA-mediated processes (Guo and Ho, 2008). Another protein, CYTOCHROME P450 51G1 (LOC_Os11g32240.1), was highly upregulated in IST, ILT and NLT. This protein participates in brassinosteroid biosynthesis. The application of brassinolide promotes panicle ripening in rice (Saka et al., 2015). Also, brassinolide is known for its induced resistance to abiotic stress in plants.

4.3 Regulation of amino acid metabolism in three rice genotypes

Our analysis revealed significant changes in major cellular processes and pathways associated with stress regimes and rice genotypes like amino acid metabolism, maintenance of cellular homeostasis and photosynthesis (Figure 4). The amino acid metabolism is a major process affected by HS (Wang et al., 2018). In N22, higher number of proteins involved in amino acid synthesis were upregulated. S-adenosyl methionine (SAM) synthetase protein participates in various biological processes and methionine metabolic pathways. This protein was specifically downregulated in MST. Increased levels of this protein were noted in NST and NLT. This protein also plays a key role in pollen tube growth in the mutant of methionine adenosyltranferase3 (MAT3, one of the four kinds of SAM synthetase genes in *Arabidopsis*) (Chen et al., 2016). Some amino acid metabolism proteins like aminomethyltransferase, chorismate mutase, and cystathionine gamma synthase showed genotype- specific responses but not found in Moroberekan. This change in amino acid metabolism could be linked to the genetic variations in heat tolerance in rice.

4.4 Differentially regulated proteins enhancing stress tolerance in the N22 genotype

Proteins upregulated in NLT were majorly enriched in biological processes like unfolded protein binding (Figure 5). This process is critical for repairing unfolded or damaged proteins formed during HS (McLoughlin et al., 2019). Furthermore, proteins involved in carboxylic acid metabolism, including those involved in amino acid metabolism, were enriched under NLT-UP. This indicates that N22 anthers can efficiently maintain their energy metabolism under heat stress conditions. The pollen tissues at the mature stage are metabolically active and need energy and new proteins for the successful fertilization (Chaturvedi et al., 2013; Selinski and Scheibe, 2014). Proteins associated with aerobic respiration enrichment were found in MLT-UP, implicating an efficient TCA cycle for energy production. A major process noted in NST-UP was the GTPase activator. GTP metabolism is involved in protein trafficking (Hutagalung and Novick, 2011). This process may be important for maintaining cellular homeostasis during HS. It was also found that NST-UP was enriched in proteins involved in vitamin E biosynthesis (Figure 5). Vitamin E protects photosystem II against heat and drought stresses (Havaux et al., 2005).

The amount of information available on photosynthesis-related proteins in rice anthers is limited. Several proteins related to light-harvesting complex II and photorespiration were upregulated in N22, especially in NLT. The chlorophyll a-b binding protein is key to balancing the excitation energy between the two photosystems and was found upregulated in NST. Considering the importance of ABC transporter in pollen exine formation and pollen-pistil interactions (Chang et al., 2016), ABC-type transporters found in both NST and NLT demonstrate relevance in the developmental process.

5 Conclusion

This study highlights the proteins affected by HS in N22, Moroberekan and IR64 rice genotypes (Figure 8). It also demonstrates the biological activities enriched in N22 over Moroberekan and IR64 under short-term and long-term HS. We identified that the major processes critical for N22 tolerance include the repair of unfolded proteins, Vitamin E biosynthesis, and trehalose accumulation processes which help N22 maintain reproduction better than Moroberekan and IR64 under heat stress conditions. We further showed that Hsp20, DnaJ, and Hsp101 are differentially expressed in the three rice genotypes, and these can be critical for N22 thermotolerance at the reproductive stage. Future studies on these proteins may shed light on the genetic mechanisms of heat tolerance in rice. A few quantitative trait loci (QTLs) have been identified in different growth and reproductive stages of rice. A major QTL named Thermotolerance 1 (TT1) encoding α2 subunit of the 26S proteasome and involved in the degradation of ubiquitinated proteins has been found in African rice *Oryza glaberrima* (Yan et al., 2020). Overexpression of a receptor-like kinase (ERECTA) in Arabidopsis, tomato, and rice plants was developed as the potential gene for breeding thermotolerant crops with no growth penalty (Shen et al., 2015). Interestingly, some of the DEPs identified in this analysis may have links to successful reproduction under HS regimes. This study provides a

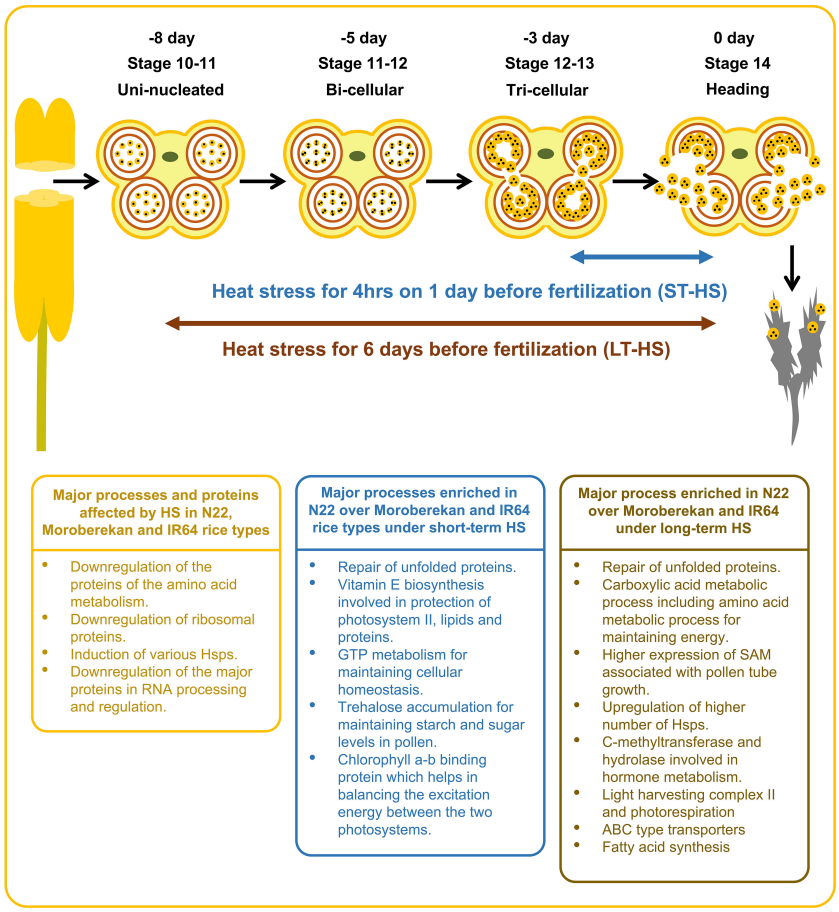


FIGURE 8
Illustration showing the effects of heat stress on major biological processes and the enriched biological processes in N22 over Moroberekan and IR64 in response to short-term and long-term HS conditions. ST, Short term heat stress and LT, Long-term heat stress conditions.

substantial amount of novel information on the proteome regulation in the anthers under heat stress conditions, facilitating a multiomics strategy that could effectively improve crop resilience (Weckwerth et al., 2020).

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

RK and IG performed heat stress treatment, sample generation experiments, and protein extraction to perform western blotting. PC and AGh performed protein extraction, quantification, digestion, LC-MS/MS measurement, peptide and protein identification. RK, NS, PC, and AGr did the data analysis. RK, IG, NS, AGr, AGh, WW, and PC drafted the manuscript. All authors have read and approved the final version of the manuscript.

Funding

RK acknowledge the University Grant Commission, the Government of India, New Delhi, for the Fellowship award. AGr acknowledges the J.C. Bose Fellowship grant from the Department of Science and Technology (DST), the Government of India, and the Center for Advanced Research and Innovation on Plant Stress and Developmental Biology, Department of Biotechnology (DBT), Government of India for the financial support. AGh is supported by the European Union Horizon 2020 research and innovation program (ADAPT) under grant agreement number GA 2020 862-858.

Acknowledgments

Special thanks to Dr. A. K. Singh of the Indian Agricultural Research Institute and Dr. H. E. Shashidhar of Gandhi Krishi Vigyan Kendra, Bengaluru, for providing rice seeds. PC is thankful to the Austrian Science Fund (FWF, DerWissenschaftsfonds), grant agreement number I 5234, for their financial 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.

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.2022.1083971/full#supplementary-material

References

Bahuguna, R. N., Jha, J., Pal, M., Shah, D., Mf, L., and Khetarpal, S. (2014). Physiological and biochemical characterization of NERICA-L-44: A novel source of heat tolerance at the vegetative and reproductive stages in rice 54, 543-559. doi: 10.1111/ppl.12299

Banowetz, G. M., Ammar, K., and Chen, D. D. (1999). Temperature effects on cytokinin accumulation and kernel mass in a dwarf wheat. *Ann. Bot.* 83, 303–307. doi: 10.1006/anbo.1998.0823

Benaroudj, N., Lee, D. H., and Goldberg, A. L. (2001). Trehalose accumulation during cellular stress protects cells and cellular proteins from damage by oxygen radicals. *J. Biol. Chem* 276, 24261–24267. doi: 10.1074/jbc.M101487200

Binder, B. M., and Patterson, S. E. (2009). Ethylene-dependent and -independent regulation of abscission. *Stewart Postharvest Rev.* 5, 1–10. doi: 10.2212/spr.2009.1.1

Bokszczanin, K. (2013). Perspectives on deciphering mechanisms underlying plant heat stress response and thermotolerance. *Front. Plant Sci.* 4. doi: 10.3389/fpls.2013.00315

Chang, Z., Chen, Z., Yan, W., Xie, G., Lu, J., Wang, N., et al. (2016). An ABC transporter, OsABCG26, is required for anther cuticle and pollen exine formation and pollen-pistil interactions in rice. *Plant Sci.* 253, 21–30. doi: 10.1016/j.plantsci.2016.09.006

Chaturvedi, P., Doerfler, H., Jegadeesan, S., Ghatak, A., Pressman, E., Castillejo, M. A., et al. (2015). Heat-Treatment-Responsive proteins in different developmental stages of tomato pollen detected by targeted mass accuracy precursor alignment (tMAPA). *J. Proteome Res.* 14, 4463–4471. doi: 10.1021/pr501240n

Chaturvedi, P., Ghatak, A., and Weckwerth, W. (2016). Pollen proteomics: from stress physiology to developmental priming. *Plant Reprod.* 29, 119–132. doi: 10.1007/s00497-016-0283-9

Chaturvedi, P., Ischebeck, T., Egelhofer, V., Lichtscheidl, I., and Weckwerth, W. (2013). Cell-specific analysis of the tomato pollen proteome from pollen mother cell to mature pollen provides evidence for developmental priming. *J. Proteome Res.* 12, 4892–4903. doi: 10.1021/pr400197p

Chaturvedi, P., Wiese, A. J., Ghatak, A., Záveská Drábková, L., Weckwerth, W., and Honys, D. (2021). Heat stress response mechanisms in pollen development. *New Phytol.* 231, 571–585. doi: 10.1111/nph.17380

Chen, Y., Zou, T., and McCormick, S. (2016). S-adenosylmethionine synthetase 3 is important for pollen tube growth. *Plant Physiol.* 172, 244–253. doi: 10.1104/PP.16.00774

Cherkasov, V., Hofmann, S., Druffel-Augustin, S., Mogk, A., Tyedmers, J., Stoecklin, G., et al. (2013). Coordination of translational control and protein homeostasis during severe heat stress. *Curr. Biol.* 23, 2452–2462. doi: 10.1016/j.cub.2013.09.058

Darriere, T., Jobet, E., Zavala, D., Escande, M. L., Durut, N., de Bures, A., et al. (2022). Upon heat stress processing of ribosomal RNA precursors into mature rRNAs is compromised after cleavage at primary p site in arabidopsis thaliana. *RNA Biol.* 19, 719. doi: 10.1080/15476286.2022.2071517

Das, S., Krishnan, P., Mishra, V., Kumar, R., Ramakrishnan, B., and Singh, N. K. (2015). Proteomic changes in rice leaves grown under open field high temperature stress conditions. *Mol. Biol. Rep.* 42, 1545–1558. doi: 10.1007/s11033-015-3923-5

Firon, N., Shaked, R., Peet, M. M., Pharr, D. M., Zamski, E., Rosenfeld, K., et al. (2006). Pollen grains of heat tolerant tomato cultivars retain higher carbohydrate concentration under heat stress conditions. *Sci. Hortic.* 109, 212–217. doi: 10.1016/j.scienta.2006.03.007

Fragkostefanakis, S., Mesihovic, A., Hu, Y., and Schleiff, E. (2016). Unfolded protein response in pollen development and heat stress tolerance. *Plant Reprod.* 29, 81–91. doi: 10.1007/s00497-016-0276-8

Fujita, D., Doi, K., Yoshimura, A., and Yasui, H. (2006). Molecular mapping of a novel gene, Grh5, conferring resistance to green rice leafhopper (Nephotettix cincticeps uhler) in rice, oryza sativa l. *Theor. Appl. Genet.* 113, 567–573. doi: 10.1007/s00122-006-0270-x

Ghatak, A., Chaturvedi, P., Bachmann, G., Valledor, L., Ramšak, Ž., Bazargani, M. M., et al. (2021). Physiological and proteomic signatures reveal mechanisms of superior drought resilience in pearl millet compared to wheat. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.600278

Guo, W. J., and Ho, T. H. D. (2008). An abscisic acid-induced protein, HVA22, inhibits gibberellin-mediated programmed cell death in cereal aleurone cells. *Plant Physiol.* 147, 1710–1722. doi: 10.1104/PP.108.120238

Havaux, M., Eymery, F., Porfirova, S., Rey, P., and Dörmann, P. (2005). Vitamin e protects against photoinhibition and photooxidative stress in arabidopsis thaliana. *Plant Cell*. doi: 10.1105/tpc.105.037036

Hong, S. W., and Vierling, E. (2000). Mutants of arabidopsis thaliana defective in the acquisition of tolerance to high temperature stress. *Proc. Natl. Acad. Sci. U.S.A.* 97, 4392–4397. doi: 10.1073/pnas.97.8.4392

Huang, J., Li, J., Zhou, J., Wang, L., Yang, S., Hurst, L. D., et al. (2018). Identifying a large number of high-yield genes in rice by pedigree analysis, whole-genome sequencing, and CRISPR-Cas9 gene knockout. *Proc. Natl. Acad. Sci. U.S.A.* 115, E7559–E7567. doi: 10.1073/pnas.1806110115

Huang, J., Zhao, X., Cheng, K., Jiang, Y., Ouyang, Y., Xu, C., et al. (2013). OsAP65, a rice aspartic protease, is essential for male fertility and plays a role in pollen germination and pollen tube growth. *J. Exp. Bot.* 64, 3351–3360. doi: 10.1093/jxb/ert173

Hubbart, S., Ajigboye, O. O., Horton, P., and Murchie, E. H. (2012). The photoprotective protein PsbS exerts control over CO2 assimilation rate in fluctuating light in rice. *Plant J* 71, 402–412. doi: 10.1111/j.1365-313X.2012.04995.x

Hutagalung, A. H., and Novick, P. J. (2011). Role of rab GTPases in membrane traffic and cell physiology. *Physiol. Rev.* 91, 119–149. doi: 10.1152/physrev.00059.2009

Imin, N., Kerim, T., Weinman, J. J., and Rolfe, B. G. (2001). Characterization of rice anther proteins expressed at the young microspore stage. *Proteomics* 1, 1149–1161. doi: 10.1002/1615-9861(200109)1:9<1149::aid-prot1149>3.0.co;2-r

Jagadish, K. S. V., Craufurd, P., Shi, W., and Oane, R. (2014). A phenotypic marker for quantifying heat stress impact during microsporogenesis in rice (Oryza sativa l.). *Funct. Plant Biol* 41, 48–55. doi: 10.1071/FP13086

Jagadish, S. V. K., Muthurajan, R., Oane, R., Wheeler, T. R., Heuer, S., Bennett, J., et al. (2010). Physiological and proteomic approaches to address heat tolerance during anthesis in rice (Oryza sativa l.). *J. Exp. Bot.* 61, 143–156. doi: 10.1093/jxb/erp289

Jagadish, S. V. K., and Pal, M. (2009). Response of rice (L.) to increasing temperature and atmospheric CO. *Climate Change and Crops. Environmental Science and Engineering.* (Berlin, Heidelberg: Springer) 63–80. doi: 10.1007/978-3-540-88246-6_4

Jegadeesan, S., Chaturvedi, P., Ghatak, A., Pressman, E., Meir, S., Faigenboim, A., et al. (2018). Proteomics of heat-stress and ethylene-mediated thermotolerance mechanisms in tomato pollen grains. *Front. Plant Sci.* 871. doi: 10.3389/FPLS.2018.01558

Katiyar-Agarwal, S., Agarwal, M., and Grover, A. (2003). Heat-tolerant basmati rice engineered by over-expression of hsp101. *Plant Mol. Biol.* 51, 677–686. doi: 10.1023/a:1022561926676

Kim, M., Kim, H., Lee, W., Lee, Y., Kwon, S. W., and Lee, J. (2015). Quantitative shotgun proteomics analysis of rice anther proteins after exposure to high temperature. *Int. J. Genomics* 2015. doi: 10.1155/2015/238704

Kobata, T., Yoshida, H., Masiko, U., and Honda, T. (2013). Spikelet sterility is associated with a lack of assimilate in high-spikelet-number rice. *Agron. J.* 105, 1821–1831. doi: 10.2134/agronj2013.0115

Kong, L., Duan, Y., Ye, Y., Cai, Z., Wang, F., Qu, X., et al. (2019). Screening and analysis of proteins interacting with OsMADS16 in rice (Oryza sativa l.). *PloS One* 14. doi: 10.1371/JOURNAL.PONE.0221473

Kumar, R., Bahuguna, R. N., Tiwari, M., Pal, M., Chinnusamy, V., Sreeman, S., et al. (2022). Walking through crossroads–rice responses to heat and biotic stress interactions. *Theor. Appl. Genet*, 135: 4065–4081. doi: 10.1007/s00122-022-04131-x

Kumar, R., Khungar, L., Shimphrui, R., Tiwari, L. D., Tripathi, G., Sarkar, N. K., et al. (2020). AtHsp101 research sets course of action for the genetic improvement of crops against heat stress. *J. Plant Biochem. Biotechnol.* 29, 715–732. doi: 10.1007/s13562-020-00624-2

Kumar, R., Lavania, D., Singh, A. K., Negi, M., Siddiqui, M. H., Al-Whaibi, M. H., et al. (2015). Identification and characterization of a small heat shock protein 17.9-CII gene from faba bean (Vicia faba l.). *Acta Physiol. Plant* 37, 1–13. doi: 10.1007/s11738-015-1943-3

Kumar, R., Singh, A. K., Lavania, D., Siddiqui, M. H., Al-Whaibi, M. H., and Grover, A. (2016). Expression analysis of ClpB/Hsp100 gene in faba bean (Vicia faba l.) plants in response to heat stress. *Saudi J. Biol. Sci.* 23, 243–247. doi: 10.1016/j.sjbs.2015.03.006

Li, X., Lawas, L. M. F., Malo, R., Glaubitz, U., Erban, A., Mauleon, R., et al. (2015). Metabolic and transcriptomic signatures of rice floral organs reveal sugar starvation as a factor in reproductive failure under heat and drought stress. *Plant Cell Environ.* 38, 2171–2192. doi: 10.1111/pce.12545

Li, X., Gao, X., Wei, Y., Deng, L., Ouyang, Y., Chen, G., et al. (2011). Rice APOPTOSIS INHIBITOR5 coupled with two DEAD-box adenosine 5′-triphosphate-dependent RNA helicases regulates tapetum degeneration. *Plant Cell* 23, 1416–1434. doi: 10.1105/tpc.110.082636

Lin, M., Chai, K., Ko, S., Kuang, L., Lur, H., and Charng, Y. (2014). A positive feedback loop between HEAT SHOCK PROTEIN101 and HEAT STRESS-ASSOCIATED 32-KD PROTEIN modulates long-term acquired thermotolerance illustrating diverse heat stress responses in rice varieties. *Plant Physiol.* 164, 2045–2053. doi: 10.1104/pp.113.229609

Mackill, D. J., Coffman, W. R., and Rutger, J. N. (1982). Pollen shedding and combining ability for high temperature tolerance in rice 1. *Crop Sci.* 22, 730–733. doi: 10.2135/cropsci1982.0011183x002200040008x

Matsui, T., Omasa, K., and Horie, T. (1997). High temperature-induced spikelet sterility of japonica rice at flowering in relation to air temperature, humidity and wind velocity conditions. *Japanese J. Crop Science* 66, 449-455. doi: 10.1626/jcs.66.449

Matsui, T., Omasa, K., and Horie, T. (2000). High temperature at flowering inhibits swelling of pollen grains, a driving force for thecae dehiscence in rice (Oryza sativa l.). *Plant Prod Sci.* 3, 430–434. doi: 10.1626/pps.3.430

McLoughlin, F., Basha, E., Fowler, M. E., Kim, M., Bordowitz, J., Katiyar-Agarwal, S., et al. (2016). Class I and II small heat shock proteins together with HSP101 protect protein translation factors during heat stress. *Plant Physiol.* 172, 1221–1236. doi: 10.1104/pp.16.00536

McLoughlin, F., Kim, M., Marshall, R. S., Vierstra, R. D., and Vierling, E. (2019). HSP101 interacts with the proteasome and promotes the clearance of ubiquitylated protein aggregates. *Plant Physiol.* 180, 1829–1847. doi: 10.1104/pp.19.00263

Merret, R., Carpentier, M. C., Favory, J. J., Picart, C., Descombin, J., Bousquet-Antonelli, C., et al. (2017). Heat shock protein HSP101 affects the release of ribosomal protein mRNAs for recovery after heat shock. *Plant Physiol.* 174, 1216–1225. doi: 10.1104/pp.17.00269

Merret, R., Nagarajan, V. K., Carpentier, M. C., Park, S., Favory, J. J., Descombin, J., et al. (2015). Heat-induced ribosome pausing triggers mRNA co-translational decay in arabidopsis thaliana. *Nucleic Acids Res.* 43, 4121–4132. doi: 10.1093/NAR/GKV234

Paoletti, A. C., Parmely, T. J., Tomomori-Sato, C., Sato, S., Zhu, D., Conaway, R. C., et al. (2006). Quantitative proteomic analysis of distinct mammalian mediator complexes using normalized spectral abundance factors. *Proc. Natl. Acad. Sci.* 103, 18928–18933. doi: 10.1073/pnas.0606379103

Pazhamala, L. T., Chaturvedi, P., Bajaj, P., Srikanth, S., Ghatak, A., Chitikineni, A., et al. (2020). Multiomics approach unravels fertility transition in a pigeonpea line for a two-line hybrid system. *Plant Genome* 13, e20028. doi: 10.1002/tpg2.20028

Prasad, P. V. V., Boote, K. J., Allen, L. H., Sheehy, J. E., and Thomas, J. M. G. (2006). Species, ecotype and cultivar differences in spikelet fertility and harvest index of rice in response to high temperature stress. *Field Crops Res* 95, 398-411. doi: 10.1016/j.fcr.2005.04.008

Pressman, E., Peet, M. M., and Pharr, D. M. (2002). The effect of heat stress on tomato pollen characteristics is associated with changes in carbohydrate concentration in the developing anthers. *Ann. Bot.* 90, 631–636. doi: 10.1093/aob/mcf240

Saka, H., Fujii, S., Imakawa, A. M., Kato, N., Watanabe, S. I., Nishizawa, T., et al. (2015). Effect of brassinolide applied at the meiosis and flowering stages on the levels of endogenous plant hormones during grain-filling in rice plant (Oryza salivaL.). *Plant Production Sci.* 6, 36–42. doi: 10.1626/PPS.6.36

Sakata, T., Oshino, T., Miura, S., Tomabechi, M., Tsunaga, Y., Higashitani, N., et al. (2010). Auxins reverse plant male sterility caused by high temperatures. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8569–8574. doi: 10.1073/pnas.1000869107

Sarkar, N. K., Kim, Y.-K., and Grover, A. (2009). Rice sHsp genes: genomic organization and expression profiling under stress and development. *BMC Genomics* 10, 393. doi: 10.1186/1471-2164-10-393

Sarkar, N. K., Kotak, S., Agarwal, M., Kim, Y.-K., and Grover, A. (2020). Silencing of class I small heat shock proteins affects seed-related attributes and thermotolerance in rice seedlings. *Planta* 251, 26. doi: 10.1007/s00425-019-03318-9

Satake, T., and Yoshida, S. (1978). High temperature-induced sterility in indica rices at flowering. *Japanese J. Crop Science.* doi: 10.1626/jcs.47.6

Scharf, K. D., Berberich, T., Ebersberger, I., and Nover, L. (2012). The plant heat stress transcription factor (Hsf) family: Structure, function and evolution. *Biochim. Biophys. Acta Gene Regul. Mech.* 1819, 104–119. doi: 10.1016/j.bbagrm.2011.10.002

Selinski, J., and Scheibe, R. (2014). Pollen tube growth: Where does the energy come from? *Plant Signal Behav.* 9, e977200–1-e977200-9. doi: 10.4161/15592324.2014.977200

Shen, H., Zhong, X., Zhao, F., Wang, Y., Yan, B., Li, Q., et al. (2015). Overexpression of receptor-like kinase ERECTA improves thermotolerance in rice and tomato. *Nat. Biotechnol.* 33, 996–1003. doi: 10.1038/nbt.3321

Shi, J., Tan, H., Yu, X. H., Liu, Y., Liang, W., Ranathunge, K., et al. (2011). Defective pollen wall is required for anther and microspore development in rice and encodes a fatty acyl carrier protein reductase. *Plant Cell* 23, 2225–2246. doi: 10.1105/tpc.111.087528

Shi, W., Yang, J., Kumar, R., Zhang, X., Impa, S. M., Xiao, G., et al. (2022). Heat stress during gametogenesis irreversibly damages female reproductive organ in rice. *Rice* 15, 32. doi: 10.1186/s12284-022-00578-0

Tabuchi, A., Li, L. C., and Cosgrove, D. J. (2011). Matrix solubilization and cell wall weakening byβ-expansin (group-1 allergen) from maize pollen. *Plant J.* 68, 546–559. doi: 10.1111/j.1365-313X.2011.04705.x

Tenorio, F. A., Ye, C., Redoña, E., Sierra, S., Laza, M., and Argayoso, M. A. (2013). Screening rice genetic resources for heat tolerance. *SABRAO J. Breed Genet.* 45, 371–381. doi: 10.1186/ s12863-015-0199-7

Tiwari, L. D., Kumar, R., Sharma, V., Sahu, A. K., Sahu, B., Naithani, S. C., et al. (2021). Stress and development phenotyping of Hsp101 and diverse other hsp mutants of arabidopsis thaliana. *J. Plant Biochem. Biotechnol.* 30, 889–905. doi: 10.1007/s13562-021-00706-9

Ungelenk, S., Moayed, F., Ho, C. T., Grousl, T., Scharf, A., Mashaghi, A., et al. (2016). Small heat shock proteins sequester misfolding proteins in near-native conformation for cellular protection and efficient refolding. *Nat. Commun.* 7, 1–14. doi: 10.1038/ncomms13673

Valledor, L., and Weckwerth, W. (2014). An improved detergent-compatible gel-fractionation LC-LTQ-Orbitrap-MS workflow for plant and microbial proteomics. *Methods Mol. Biol.* 1072, 347–358. doi: 10.1007/978-1-62703-631-3_25

Wang, J., Yuan, B., Xu, Y., and Huang, B. (2018). Differential responses of amino acids and soluble proteins to heat stress associated with genetic variations in heat tolerance for hard fescue. *J. Am. Soc. Hortic. Sci.* 143, 45–55. doi: 10.21273/JASHS04246-17

Weckwerth, W., Ghatak, A., Bellaire, A., Chaturvedi, P., and Varshney, R. K. (2020). PANOMICS meets germplasm. *Plant Biotechnol. J.* 18, 1507–1525. doi: 10.1111/pbi.13372

Yan, C., Zhan, G., Hong, X., and Yang, D. (2020). Identification and fine mapping of a major QTL, TT1-2, that plays significant roles in regulating heat tolerance in rice. *Plant Mol. Biol. Rep* 39, 376–385. doi: 10.1007/s11105-020-01256-5

Ye, C., Tenorio, F. A., Argayoso, M. A., Laza, M. A., Koh, H.-J., Redoña, E. D., et al. (2012). Identifying and confirming quantitative trait loci associated with heat tolerance at flowering stage in different rice populations 16, 1-10. doi: 10.1186/s12863-015-0199-7

Zhao, P., Zhang, J., Qian, C., Zhou, Q., Zhao, X., Chen, G., et al. (2017). SNP discovery and genetic variation of candidate genes relevant to heat tolerance and agronomic traits in natural populations of sand rice (Agriophyllum squarrosum). *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00536

Zinn, K. E., Tunc-Ozdemir, M., and Harper, J. F. (2010). Temperature stress and plant sexual reproduction: Uncovering the weakest links. *J. Exp. Bot.* 61, 1959–1968. doi: 10.1093/jxb/erq053

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 17 January 2023
DOI 10.3389/fpls.2022.1063813

OPEN ACCESS

EDITED BY
Bhanu Prakash Petla,
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

REVIEWED BY
Quanzi Li,
Chinese Academy of Forestry, China
Nannan Li,
Southwest University, China

*CORRESPONDENCE
Yuanyuan Zhao
yyzhao@bjfu.edu.cn

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

RECEIVED 07 October 2022
ACCEPTED 21 December 2022
PUBLISHED 17 January 2023

CITATION
Liao B, Wang C, Li X, Man Y, Ruan H and Zhao Y (2023) Genome-wide analysis of the *Populus trichocarpa* laccase gene family and functional identification of *PtrLAC23*. *Front. Plant Sci.* 13:1063813. doi: 10.3389/fpls.2022.1063813

Genome-wide analysis of the *Populus trichocarpa* laccase gene family and functional identification of *PtrLAC23*

Boyang Liao[1,2,3], Chencan Wang[1,2], Xiaoxu Li[1,2], Yi Man[1,2], Hang Ruan[4] and Yuanyuan Zhao[1,2]*

[1]College of Biological Science and Technology, Beijing Forestry University, Beijing, China, [2]National Engineering Research Center of Tree Breeding and Ecological Restoration, Beijing Forestry University, Beijing, China, [3]Institute of Environmental Biology and Life Support Technology, School of Biological Science and Medical Engineering, Beihang University, Beijing, China, [4]School of Cyber Science and Technology, Beihang University, Beijing, China

Introduction: Biofuel is a kind of sustainable, renewable and environment friendly energy. Lignocellulose from the stems of woody plants is the main raw material for "second generation biofuels". Lignin content limits fermentation yield and is therefore a major obstacle in biofuel production. Plant laccase plays an important role in the final step of lignin formation, which provides a new strategy for us to obtain ideal biofuels by regulating the expression of laccase genes to directly gain the desired lignin content or change the composition of lignin.

Methods: Multiple sequence alignment and phylogenetic analysis were used to classify *PtrLAC* genes; sequence features of *PtrLACs* were revealed by gene structure and motif composition analysis; gene duplication, interspecific collinearity and Ka/Ks analysis were conducted to identify ancient *PtrLACs*; expression levels of *PtrLAC* genes were measured by RNA-Seq data and qRT-PCR; domain analysis combine with cis-acting elements prediction together showed the potential function of *PtrLACs*. Furthermore, Alphafold2 was used to simulate laccase 3D structures, *proLAC23::LAC23-eGFP* transgenic Populus stem transects were applied to fluorescence observation.

Results: A comprehensive analysis of the *P. trichocarpa* laccase gene (*PtLAC*) family was performed. Some ancient *PtrLAC* genes such as *PtrLAC25*, *PtrLAC19* and *PtrLAC41* were identified. Gene structure and distribution of conserved motifs clearly showed sequence characteristics of each *PtrLAC*. Combining published RNA-Seq data and qRT-PCR analysis, we revealed the expression pattern of *PtrLAC* gene family. Prediction results of cis-acting elements show that *PtrLAC* gene regulation was closely related to light. Through above analyses, we selected 5 laccases and used Alphafold2 to simulate protein 3D structures, results showed that *PtrLAC23* may be closely related to the

lignification. Fluorescence observation of *proLAC23::LAC23-eGFP* transgenic *Populus* stem transects and qRT-PCR results confirmed our hypothesis again.

Discussion: In this study, we fully analyzed the Populus trichocarpa laccase gene family and identified key laccase genes related to lignification. These findings not only provide new insights into the characteristics and functions of Populus laccase, but also give a new understanding of the broad prospects of plant laccase in lignocellulosic biofuel production.

KEYWORDS

biofuel, *populus trichocarpa*, lignin, laccase, gene family analysis

1 Introduction

Biofuels, as a kind of sustainable renewable energy instead of fossil energy, have been widely concerned in recent years. The "second generation" biofuel is made from inedible lignocellulose, through enzymatic hydrolysis of polysaccharides, release of monoses, fermentation, and finally converted into cellulosic ethanol for application (Chang, 2007). Compared with the "first generation" biofuels which produce bioethanol by fermentation of edible parts such as sugar and starches, there is no competition with food supply, and pay more attention to ecological benefits in environmental protection (Prem et al., 2021). According to United States Department of Energy, second-generation biofuels are expected to reduce greenhouse gas emissions by up to 96% (Clark, 2008). The main component of lignocellulosic biomass is plant cell wall, which is composed of cellulose, hemicelluloses and lignin (Cosgrove, 2015). In the process of biofuel conversion, lignin can easily adsorb cellulolytic enzymes and restrict the release of cellulose, thus directly limiting fermentation yield (Zeng et al., 2014). Therefore, changing the content and composition of lignin in plants is an effective measure to boost biofuel production.

Laccase (E.C.1.10.3.2), also known as polyphenol oxidase or urushiol oxidase, was first discovered in the SAP of *Rhus Venicifera* in 1883 (Zoppellaro et al., 2001). As the largest subfamily of multicopper oxidases (LMCOs), laccase has a wide range of substrates and is widely distributed in bacteria, plants, insects and fungi (Mate et al., 2016). Commonly, the molecular size of laccase concentrated in 60-130 KD, composed of 500-600 Aa (Su et al., 2017). Although the amino acid sequence of laccase from different sources is quite different, their catalytic sites are relatively conserved (Janusz et al., 2020). Crystal structure analysis of laccase protein has been found that laccase has three copper ion binding sites (T1, T2, T3) and four copper ions, which can be divided into a type-1 Cu, a type-2 Cu, and two type-3 Cu atoms (Zoppellaro et al., 2001). It is speculated that the catalytic mechanism of laccase is copper ions at the active site of T1 absorb electrons from the reduced substrate, causing the substrate to be oxidized and form free radicals, leading to various non-enzymatic secondary reactions (Sato et al., 2020). The current research on laccase mainly focused on its physicochemical properties and biological activities. Fungal laccase has been widely used *in vitro* lignin degradation (Reyes et al., 2021), wastewater treatment (Unuofin and Okoh, 2019) and dye decolorization (Ahmed et al., 2020). Laccase CotA in *Bacillus subtilis* has been reported to play a role in pigment synthesis and protection against ultraviolet damage (Hullo et al., 2001). Insect laccase has been shown to be associated with cuticle formation in insects (Asano et al., 2019). However, the function of plant laccase need to be more determined. Several lines of evidence suggest that laccase can play a key role in plant lignification, which refers to the coupling of a lignin monomer with the terminal of a growing polymer (Bao et al., 2016). Lignin monomers, also called monolignols, are the non-methoxylated *p*-coumaryl alcohol, the monomethoxylated coniferyl alcohol and the dimethoxylated sinapyl alcohol, which forms p-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) units in lignin polymer, respectively. Laccase can catalyze the oxidative dehydrogenation of lignin monomers to form free radicals, once the radicals are generated, polymerization of lignin monomers will take place in a pure chemical reaction, no longer catalyzed by enzymes or proteins (Ralph et al., 2019).

There has been several evidence that laccase is closely associated with lignification. Berthet et al. have found that *LAC4* and *LAC17* played important roles in the lignification process of *Arabidopsis* stems, silencing of these two genes significantly reduced the biosynthesis of lignin (Berthet et al., 2011). In addition, simultaneous knockout *LAC11*, *LAC4*, and *LAC17* in *Arabidopsis* severely affected plant growth, resulting in narrowing of root diameter, closure of anthers, stagnation of vascular tissue development, and a significant reduction in lignification (Zhao et al., 2016). Richard Dixon et al. have reported that GhLAC8, a laccase specifically expressed in the

seed coats of *Cleome hassleriana*, which can facilitate the polymerization of lignin in plants by using caffeyl alcohol as the sole substrate, and determined the lignin component content in plant cell walls (Wang et al., 2020). Except for the role in lignification, laccase has many other functions also worth exploring (Wang et al., 2015a). Results show that expression of *GhLac1* in transgenic cotton could enhance the defense response of cotton to against pathogens and pests (Acid et al., 2018). Almost all of the *GmLAC* genes in soybeans are involved in responding to *P.sojae* infection (Wang et al., 2019). Some researchers inserted *Arabidopsis* laccase gene *AtLAC15* promoter into *Brassica napus*, and found that *AtLAC15* promoter could be used as a seed coat-specific promoter for canola. In addition, this coat-specific promoter can help to overexert or inhibit the expression of some genes, which can alter the metabolism and producing seed with reduced fiber content (Wu and Saleh, 2011).

In recent years, Batch anaerobic digestion experiments on 41 energy crops showed that 80% of the sample biofuel yield variation could be explained by lignin content (Dandikas et al., 2014). In addition, inhibition of lignin biosynthesis pathway gene *Pt4CL1* in transgenic poplar (*Populus tremuloides Michx.*) showed that there is a compensatory mechanism between lignin and cellulose, that is, inhibition of lignin synthesis can increase the accumulation of cellulose (Hu et al., 1999). From the perspective of energy utilization, treatment of laccase in plants can inhibit lignin deposition and increase cellulose content, which provides a new idea for the improvement of energy plants (Mai and Kharazipour, 2001).

As an important biomass energy source (Dandikas et al., 2014), woody plant is of great significance to analyze the function and mechanism of its laccase gene. To date, the *Populus trichocarpa* laccase gene family members have been identified (Lu et al., 2013) (Bryan et al., 2016), but their sequence characteristics, structures and functions have not been elucidated in detail. Here, we choose *Populus trichocarpa* as the research object, through collinearity analysis, gene structure analysis, motif and domain analysis, expression pattern analysis, cis-acting elements and protein structure prediction and other detailed analyses of *Populus trichocarpa* laccase gene family, to find out some *PtrLACs* which involved in plant lignification, in order to provide more directions for improvement of energy trees.

2 Methods

2.1 Multiple sequence alignment, phylogenetic analysis, and classification of laccase family genes

The genome of *Populus trichocarpa v3.0* and *Arabidopsis* were downloaded from Phytozome (https://phytozome.jgi.doe.gov) (Goodstein et al., 2012). According to published research, *Arabidopsis thaliana* has a total of 17 laccase genes (Mccaig et al., 2005), this number is 5 in *Zea mays* (Caparrós-Ruiz et al., 2006), and 53 in *Populus trichocarpa* (Bryan et al., 2016). Each of the 75 sequences has three domains, which named Cu-oxidase, Cu-oxidase_2, and Cu-oxidase_3 (PF00394, PF07731, and PF07732). An alignment of 17 AtLAC peptides, 5 ZmLAC peptides,53 PtrLAC peptides, 46 P. *deltoides* laccase peptides and 12 P. *tomentosa* laccase peptides were performed using ClustalW in MEGA 10.0 with the default parameters. Full-length sequences of all proteins were used to construct phylogenetic tree with the neighbor-joining(NJ) method (1000 replicates) (Kumar et al., 2018). Finally, the tree was annotated and decorated by ITOL online tools (http://itol.embl.de/) (Letunic and Bork, 2016).

2.2 Gene structure, motif composition and domain analysis

The evolutionary tree construction method of 53 genes was the same as shown in 2.1. The exon-intron organization of laccase genes was identified by the online program Gene Structure Display Server 2.0 (http://gsds.cbi.pku.edu.cn) (Hu et al., 2015). The MEME online program (http://meme.nbcr.net/meme/intro.html) was used to identify conserved motifs in *Populus trichocarpa* laccases (Bailey et al., 2009), and the maximum number of motif was set to 15. The domain information of each *Populus* laccase protein was obtained from Pfam database (https://pfam.xfam.org/search). Visualization of the motif and domain position was accomplished by TBtools software (https://github.com/CJ-Chen/TBtools) (Chen et al., 2020).

2.3 Chromosomal localization and gene duplication

Chromosome length and the distribution of 53 *LAC* genes of *P. trichocarpa* were obtained by using the Phytozome database (https://phytozome.jgi.doe.gov). Multiple Collinearity Scan toolkit (MCScanX) (Wang et al., 2012) was adopted to analyze the gene duplication events with the default parameters. The Gene Location Visualize tool of TBtools software (https://github.com/CJ-Chen/TBtools) was used to Visualize the results of tandem repeat analysis, the Circle Gene View function in TBtools was used to exhibit the collinearity of laccase gene family in *P. trichocarpa*, the *Dual Systeny Plot* in TBtools was used to show the synteny relationships between *P. trichocarpa* and other 5 species. Ka/Ks Calculator 2.0 was used to calculate the Ka/Ks ratio of *PtrLACs* to reveal the natural selection pressure experienced by laccase genes during the evolution (Wang et al., 2010).

2.4 Expression pattern of *PtrLACs*

The RNA-Seq data and FPKM (Fragments per kilobase of transcript per million fragments mapped) values of 40 samples including different tissues and developmental stages were obtained from phytozome database. The expression level of *PtrLAC* genes were presented in the form of heatmap by TBtools software.

2.5 Analysis of cis-acting elements in *PtrLAC* genes' promoters

The upstream sequences (2000 bp) of 53 *PtrLAC* genes were extracted by PGSC. Then, PlantCare (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was used for predictive analysis (Lescot et al., 2002), and 9 cis-acting elements were obtained after screening. They are abscisic acid responsiveness, auxin responsiveness, defense and stress responsiveness, drought-inducibility, gibberellin-responsiveness, light responsiveness, low-temperature responsiveness, MeJA-responsiveness, MYB binding site involved in light responsiveness.

2.6 Properties of *Populus trichocarpa* laccase proteins

The number of amino acids per protein was obtained using the phytozome database, the molecular weight (Da) and isoelectric point (pl) of the protein were predicted by using the online ExPASy tool (https://web.expasy.org/protparam/) (Gasteiger et al., 2003). Subcellular locations of PtrLAC members were calculated by the online software CELLO (http://cello.life.nctu.edu.tw/).The signal peptides of PtrLACs were predicted by using the SignalP online tool (http://www.cbs.dtu.dk/services/SignalP/) (Almagro Armenteros, 2019).

2.7 Structural prediction of laccase proteins

Protein structure was predicted by AlphaFold2. The structure prediction process was roughly as described in AlphaFold Paper 2, consisting of five steps:MSA construction, template search, five model reasoning, average PLDD-based model ordering, and constraint relaxation of the predicted structure. Structure factors of ZmLAC3, which coded as 6KLG (ZmLAC3 native), 6KLI (ZmLAC3–SinA complex) and 6KLJ (ZmLAC3–ConA complex), were downloaded from Protein Data Bank (http://www.rcsb.org/) (Andreaeopsida, 2013). AutoDock Vina software were used for molecular docking of PtrLAC with SinA and ConA (Trott and Olson, 2009). Structural comparison between ZmLAC3 and PtrLACs as well as Polar interaction between monolignols and amino acid were analyzed by PyMol 2.0 (Seeliger and De Groot, 2010).

2.8 Plant materials

45-day-old 84K poplars (*P.alba×P.glandulosa*) were used as materials for qRT-PCR. Primers gaattcgagctcGGTACC attcaaacctgcgttgatcc and aagcttgcatgcCTGCAGacactt gggaagatcagatggtgg were used to clone a 4.7 kb genomic region into *pCAMBIA1305.1-native-eGFP* for generating *proLAC23::LAC23-eGFP*. 2-mo-old *pLAC23-LAC23-EGFP* line was used for fluorescence observation.

2.9 Fluorescence microscope observation

The stem transverse section preparation was followed as described by Li et al. (Li et al., 2011). Images were observed and captured by using the LEICA DM2500 fluorescence microscope.

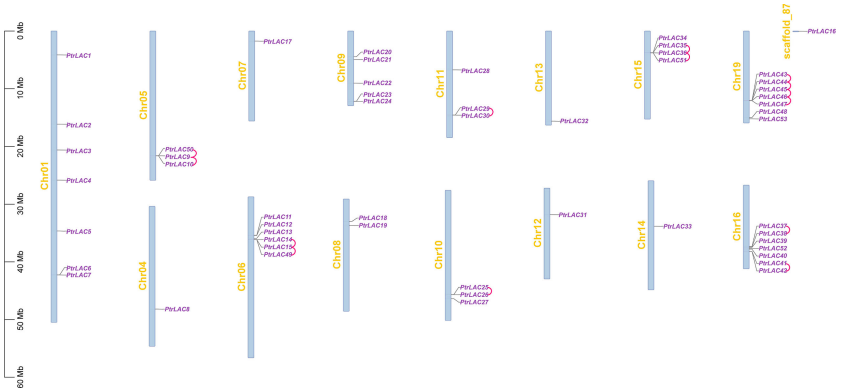


FIGURE 1
Chromosomal location and gene duplication of *P. trichocarpa* laccase. The tandem duplicated genes are connected by red curves.

The autofluorescence of lignin was observed using the third gear (UV excitation light) of the microscope, and EGFP fluorescence was observed using the fifth gear (blue excitation light) of the microscope.

2.10 Quantitative Real-Time PCR of laccase transcript abundance

Total RNA was isolated from young leaves, young stems and young roots. cDNA of these three samples were obtained by reverse transcription. Gene-specific forward and reverse primers (Table S1) were used for quantitative RT-PCR to estimate the abundance of gene-specific laccase transcripts in different tissues and stages. The *SuperReal PreMix Plus (SYBR Green)* kit used was purchased from Tiangen Biotech (Beijing) Co., LTD. The instrument used in the qRT-PCR experiment was Bio-Red.

3 Results

3.1 Chromosomal location, duplication and expansion events of *Populus* laccase gene family

It has been reported that there are 53 laccase genes in *Populus trichocarpa*, which are located on 16 chromosomes (Figure 1). Most of the laccase genes are concentrated on five chromosomes (chr 01, chr 06, chr 09, chr 16 and chr 19), and some chromosomes have only one gene (chr 04, chr 07, chr 12, chr 13, chr 14 and scaffold_87).

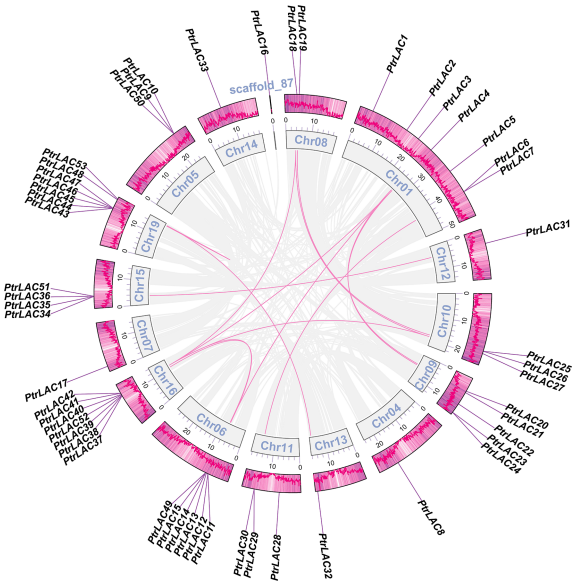


FIGURE 2
Collinearity of the *PtrLAC* gene pairs. The outermost scale shows gene density on each chromosome, genes are listed in the outer circle according to their chromosomal location. The inner circle represents the chromosomes. Gray lines indicate all synteny blocks in the *Populus trichocarpa* genome, synteny relationships between gene pairs are marked with a pink line.

Segmental and tandem duplications are usually considered to be the main causes of large gene family expansion in plants (Cannon et al., 2004), thus we analyzed the tandem and segmental duplication events of *PtrLACs* family. The results showed that 22 genes (41.5%) were considered as tandem duplicates, among which two pairs of independent tandem duplication genes were located in chr 10, chr 11 and chr 16. Three groups of three tandem duplicated genes located on chr 05, chr 06 and chr 15. Five tandem duplicated genes located on chr 19. Segmental duplications are regions of 1000 or more base pairs with a DNA sequence identity of 90% or greater that are present in more than one copy (Giannuzzi et al., 2011)˙ (Van de Peer et al., 2009). Therefore, the collinearity gene pair can be considered as the fragment duplication genes in *Populus* laccase gene family. In our study, a total of 22 genes (41.5%) were involved in fragment replication (Figure 2). Based on above results, it could be inferred that tandem duplication and fragment replication play an equal important role in the expansion of *Populus* laccase gene family.

To better understand the evolutionary constraints acting on *Populus* laccase gene family, we calculated the non-synonymous nucleotide substitution rate (Ka) and synonymous nucleotide substitution rate (Ks) between both segmental and tandem duplicated *PtrLAC* gene pairs (So, 2002). The Ka/Ks value of almost all of the gene pairs was less than one (except the tandem repeat pair *PtrLAC45*/*PtrLAC46*), which indicated that *Populus* laccase gene family underwent a strong purifying selective pressure during the evolution.

Existing molecular evolution analysis combined with fossil evidence shows that the molecular evolution rate (E value) of *Populus* is about 1/6 of that of *Arabidopsis* (Tuskan and Torr, 2012), while the E value of *Arabidopsis* is 1.5×10^{-8} replacement/synonymous replacement site/year (Koch et al., 1998). Therefore, the E value of *Populus* in this paper is 1/6 of that of *Arabidopsis*, that is, 2.5×10^{-9} replacement/synonymous replacement site/year. Salicoid duplication event occurs between about 60 and 65 MYA (Million Years Ago) (Tuskan and Torr, 2012). Estimation of the repetition time of collinearity gene pairs in *PtrLAC* gene family show that about 13.3% (4/30) gene pairs were separated during the WGD event, 20.0% (6/30) of the repeats were replicated earlier than 150 million years ago, and the rest collinear gene pairs were replicated after the WGD events. In addition, tandem repeats were often replicated later than segmental repeats, suggesting that fragment duplication and tandem duplication were the main driving forces of gene family expansion in the early and late stages, respectively (Table S2, S3). These analyses helped us screen out some ancient laccase genes, such as *PtrLAC4, PtrLAC41, PtrLAC19, PtrLAC25* and *PtrLAC18*.

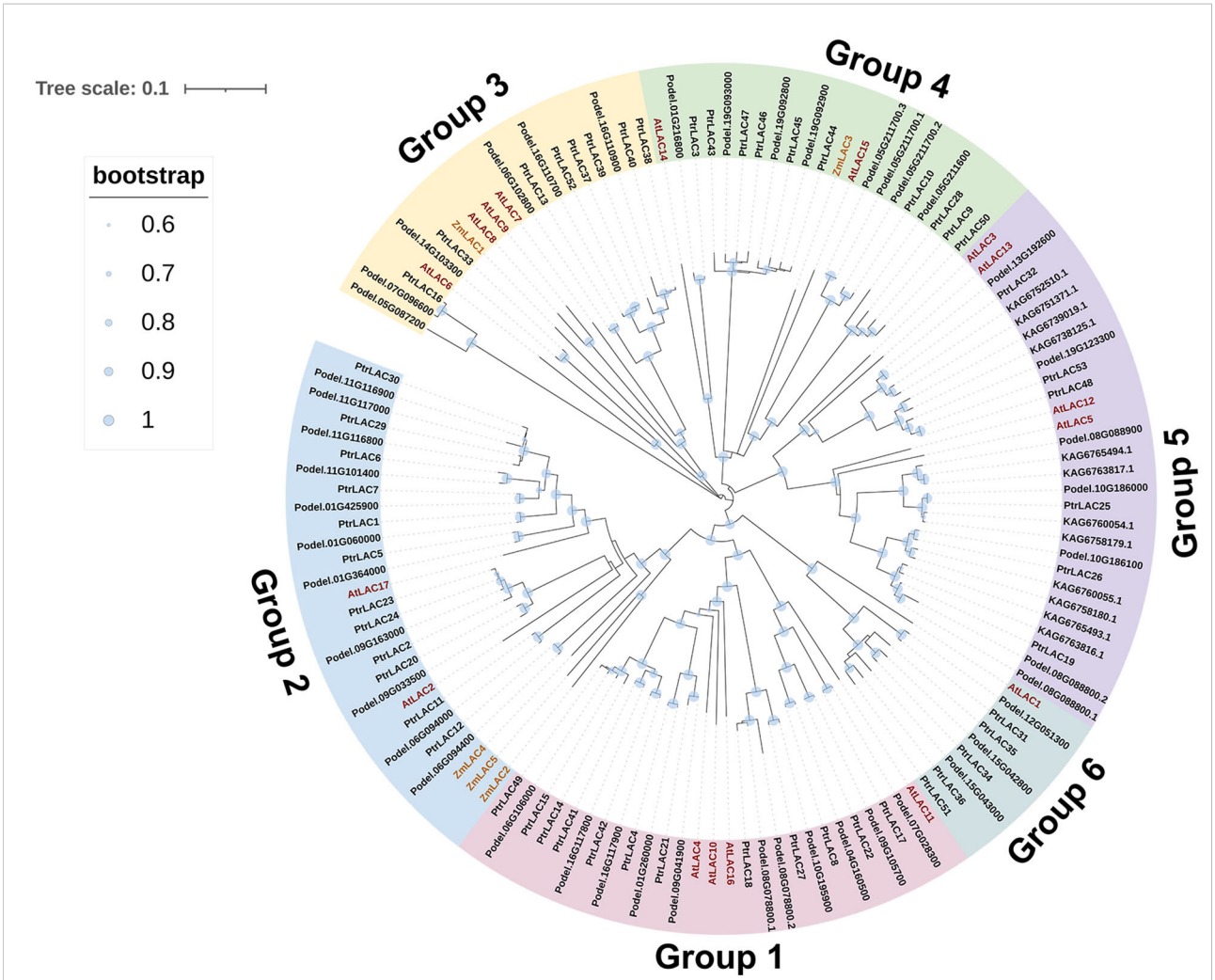


FIGURE 3
Phylogenetic tree of laccase proteins from *Arabidopsis thaliana* TAIR10 (AtLAC), *Zea mays* (ZmLAC), *Populus trichocarpa* v3.0 (PtrLAC), *Populus deltoides* WV94 v2.1 (Podel) and *Populus tomentosa* (KAG). Full length amino acid sequences of these 133 proteins were aligned using the ClustalW program. Nneighbor-joining tree was constructed using MEGA10.0. 6 groups were distinguished in different colors.

3.2 General information, phylogenetic analysis and classification of *PtrLAC* genes

To study the evolutionary relationships among the 75 laccase genes in *Populus Arabidopsis*, and maize, we built a phylogenetic tree using full-length amino acid sequences. In total, 53 sequences from*P. trichocarpa*, 17 sequences from *Arabidopsis*, 5 sequences from maize, 46 sequences from *P. deltoides*, and 12 sequences from *P. tomentosa* were assessed in the phylogenetic tree (Figure 3). The phylogenetic analysis indicated that *PtrLACs* could be divided into six large groups corresponding to group 1 to 6 in *Arabidopsis* as defined by Bonnie C (Mccaig et al., 2005). The group 1, 2 and 4 each contains 12 *PtrLAC* members, group 3 and group 5 each possesses 6 members, and group 6 contains 5 members. In addition, we found that members of group 4 involved in tandem duplication accounted for 66.7% (8/12), all the genes belong to group 5 and group 6 were involved in gene duplication events. These phenomena suggest that gene duplication is responsible for expansion of these 3 subfamilies.

The subcellular prediction shows that most of the laccase proteins are likely to be extracellular, but some have been shown to be localized on the cell membrane or in lysosomes and peroxisomes. A signal peptide analysis shows that most laccases are secretory proteins and have a signal peptide of about 27 amino acids, but the signal peptide of PtrLAC51 was not predicted, indicating that it may be intracellularly localized (Table S4). We found that most laccases in poplar are alkaline, which is similar to the results in most plants, such as *A. thaliana* and maize.

3.3 Exon-intron structure, motif distribution and domain composition of *Populus* laccase gene family

To better understand the structural features of *PtrLACs* and gain more insight into the evolution of the laccase family in *P. trichocarpa*, we analyzed the intron-exons composition of all sequences as shown in Figure 4. All the sequences in the group 1, group 3 and group 4 contains 5 introns and 6 exons, most sequences in group 2 also have such structure, except *PtrLAC8*, which contains 4 introns and 5 exons. The sequences in group 5 and most of the sequences in group 6 have four introns and five exons. Notably, in group 6, *PtrLAC51* contained only 3 exons and 2 introns, while *PtrLAC36* contained 7 exons and 6 introns. Overall, the family members with closer genetic relationships have more similar gene structures (Figure 4B).

To further clarify the characteristics of the *Populus* laccase family, we detected 15 conserved motifs in 53 sequences and analyzed their distribution in combination with the evolutionary tree. As expected, most of the laccase protein sequences in *Populus trichocarpa* had all 15 motifs. However, in group 6, none of the sequences had motif 15. In addition, *PtrLAC51* also lost motif 2 and motif 6, *PtrLAC36* doesn't include motif 6, 9 and 11. In group 4, *PtrLAC3* doesn't have motif 15, and *PtrLAC16* just contains six motifs (motif 6, 2, 3, 8, 5 and 1). These differences suggest that family members of the two branches may have lost part of their motifs during evolutionary process, resulting in a new function. (Figure 4C).

In order to define the function of each member of the *Populus* laccase gene family, we use the Pfam database (https://pfam.xfam.org/search) to analyze the domain of each sequence. Results showed that all the members have Cu-oxidase, Cu-oxidase_2, and Cu-oxidase_3 (PF00394, PF07731, and PF07732) these three conserved domains, which associated with the redox reaction of substrates. Beyond that, we also detected 9 domains and some of them have specific functions: YAF2_RYBP was a C-terminal binding motif, which is first found in YAF2 and RYBP proteins and usually forms the RYBP -/YAF2-PRC1 complex (Wang et al., 2015b); ResB domain was related to the synthesis of cytochrome C and is essential for plant growth (Le Brun et al., 2000); NAR2 is a plant protein with a C-terminal transmembrane region, which often works with NRT2 and can transport nitrate at low concentration, this NAR2-NRT2 system plays an important role in the regulation of

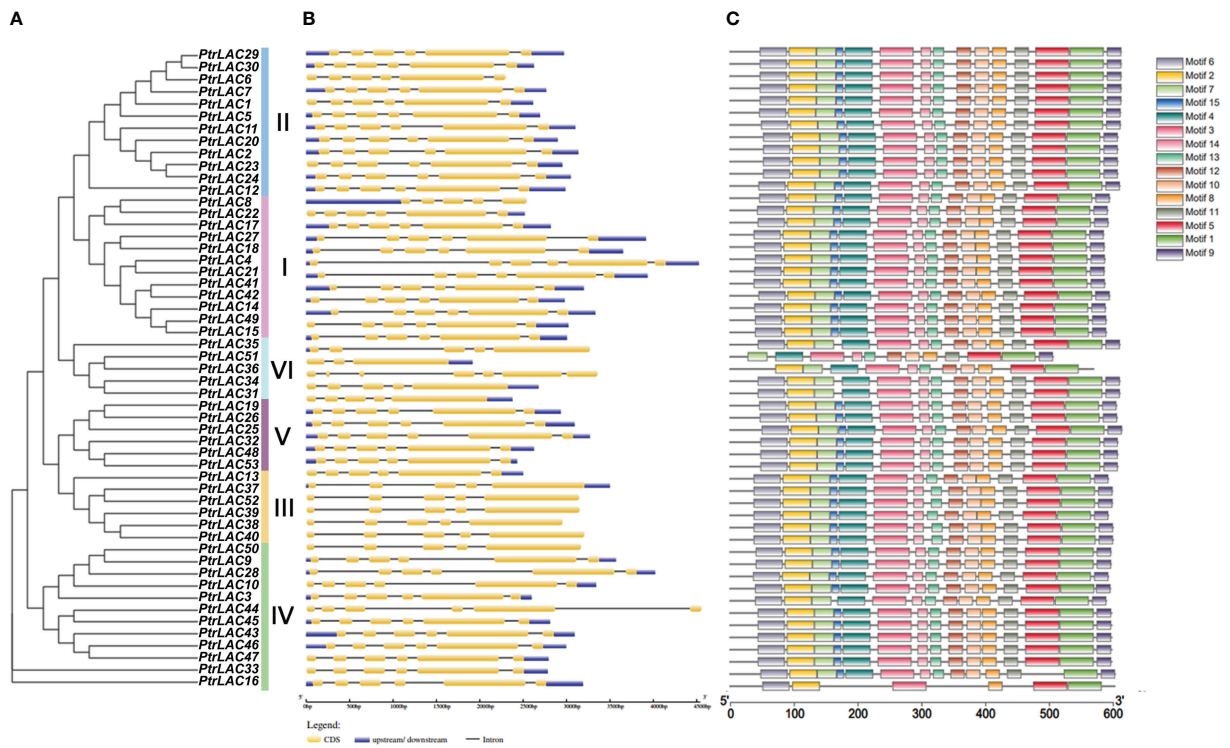


FIGURE 4
Phylogenetic relationships, gene structure and architecture of conserved protein motifs of *P. trichocarpa* laccase gene family. **(A)** Nneighbor-joining tree was constructed based on the full-length sequences of *P. trichocarpa* laccase proteins using MEGA 10.0 software. **(B)** Exon-intron structure of *P. trichocarpa* laccase genes. Blue boxes indicate untranslated 5′- and 3′-regions; yellow boxes indicate exons while black lines indicate introns. **(C)** The motif composition of *Populus trichocarpa* laccase proteins. Colored boxes represent the different motifs, indicated in the top right-hand corner. The scales at the bottom of the image indicate the estimated exon/intron length in kb and motif length in numbers of amino acids (aa), respectively.

lateral root growth (Yong et al., 2010); CBiM is a substrate specific component of the cobalt transport complex CBiMNQO, which is related to the synthesis of coenzyme B12 (Santos et al., 2008) (Figure 5). Laccases with these special domains may enable to perform other specific functions in plants, and should be the focus of future research.

3.4 Synteny analysis of *PtrLAC* genes

To further reveal the phylogenetic mechanisms of *Populus* laccase gene family, we constructed five comparative syntenic maps of *Populus* associated with five representative species, including three monocots (*Setaria viridis*, *Oryza sativa*, *Sorghum bicolor*) and two dicots (*Arabidopsis thaliana* and *Glycine max*). Results show that, 16 *PtrLACs* had a collinear relationship with *Arabidopsis thaliana*, and the number was 5, 7, 6 and 24 for sorghum, rice, *Setaria* and soybean. As shown in Figure 6, the number of orthologous gene pairs between *Populus* and other species was 22 gene pairs (with *Arabidopsis*), 10 gene pairs (with *Setaria*), 13 gene pairs (with rice), 7 gene pairs (with sorghum) and 59 gene pairs (with soybean), respectively. Some *PtrLAC* genes were found to be associated with at least three syntenic gene pairs, such as *PtrLAC19* (between *Populus* and *Arabidopsis*), *PtrLAC25* (between *Populus* and *Setaria*), *PtrLAC25* and *PtrLAC32* (between *Populus* and Rice). This phenomenon is particularly prominent between *Populus* and soybean, there are 9 collinearity gene pairs of this type in between (*PtrLAC4, PtrLAC50, PtrLAC3, PtrLAC19, PtrLAC21, PtrLAC25, PtrLAC32, PtrLAC37, PtrLAC48*).

We also found that, there were 59 collinear gene pairs of laccase genes between *Populus* and soybean, but only 7

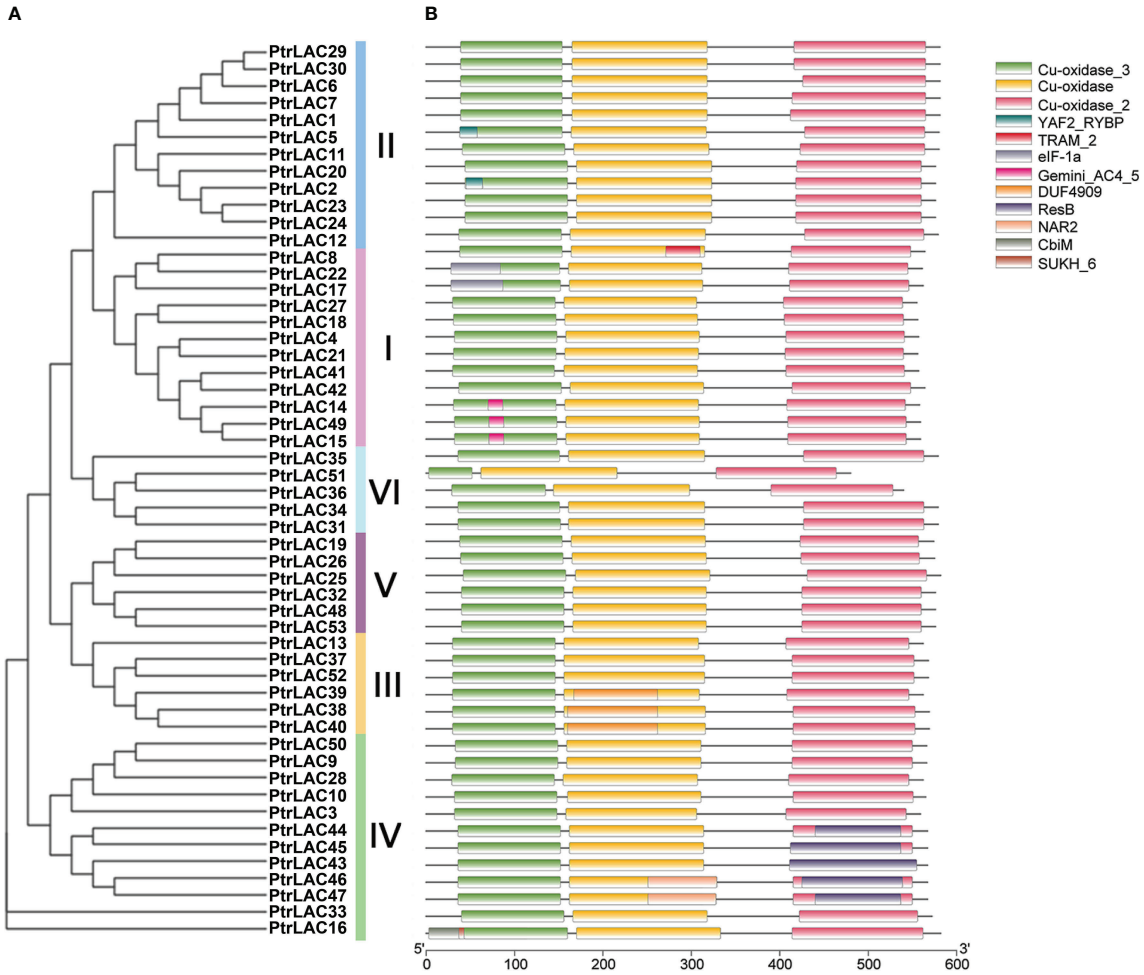


FIGURE 5
Phylogenetic relationships and conservative domain distribution of the *PtrLACs* family. **(A)** Nneighbor-joining tree was constructed based on the full-length sequences of *P. trichocarpa* laccase proteins using MEGA 10.0 software. **(B)** Domain distribution of *P. trichocarpa* laccase proteins. Heterochromatic boxes represent different domains, indicated in the top right-hand corner. The length of protein can be estimated using the scale at the bottom.

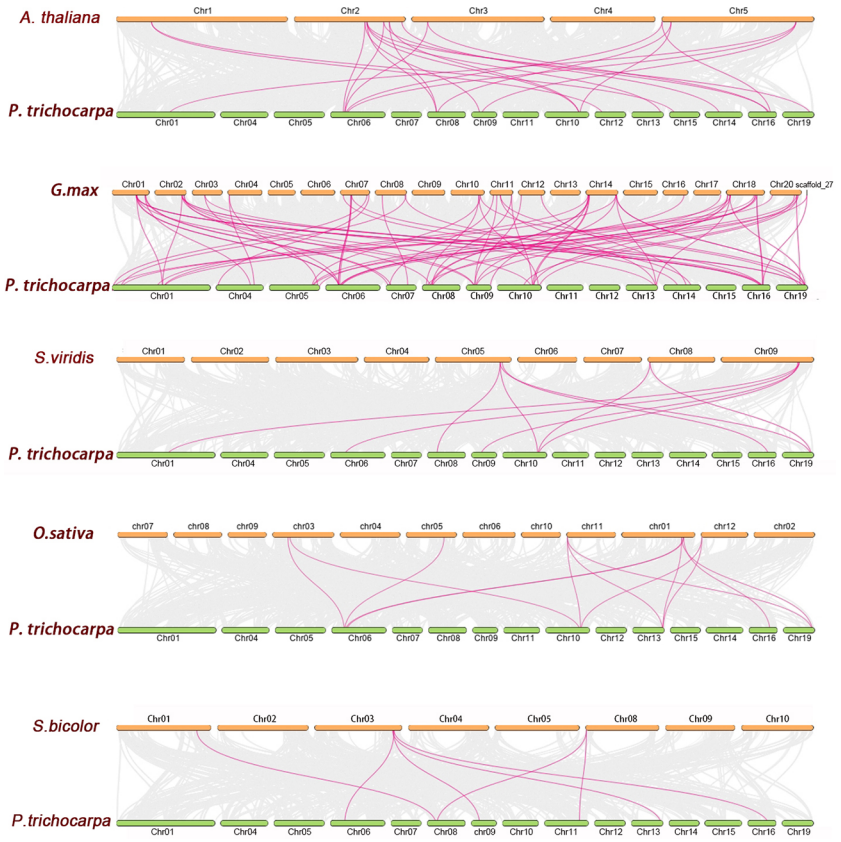


FIGURE 6
Synteny analysis of *LAC* genes between *P. trichocarpa* and five representative plant species. Gray lines indicate the collinear blocks within *P. trichocarpa* and other plant genomes, while the purple lines highlight the syntenic *LAC* gene pairs.

between *Populus* and *Sorghum*, which may relate to the phylogenetic relationships between *Populus* and other five plant species. What is noteworthy is that, many *PtrLACs* (*PtrLAC4, PtrLAC21, PtrLAC33, PtrLAC48*) that are collinearly related to *Arabidopsis* and soybean have not been found in other three monocotyledons, which may indicate that these orthologous pairs formed after the divergence of dicotyledonous and monocotyledonous plants. Some other *PtrLACs*, however, were identified have collinearity with at least four species (*PtrLAC11, PtrLAC25, PtrLAC20, PtrLAC19*), suggesting that these genes maybe present before the ancestors differentiated.

3.5 Expression patterns of *PtrLAC* genes

In order to understand the expression of *PtrLACs* in different tissues and stages, we use the RNA-seq data published in Phytozome to draw a heat map and analyzed its FPKM (Fragments Per Kilobase Million) values. we compared the expression patterns of *PtrLACs* in different developmental processes such as bud set and bud flush, male/female catkin development, leaf expansion and root/stem response to different nitrogen nutrition as shown in Figure 7.

At different stages of apical bud growth, the expression of laccase genes was more active in mid-spring. During the

FIGURE 7
Expression profiles of *PtrLAC* genes in 40 samples including different tissues and developmental stages.

formation of dormant buds, the expression level of *PtrLAC43* is higher in the pre-dormant buds and early dormant buds, which may be related to the protection of meristem in winter. In general, most *PtrLAC* genes were expressed at low levels in the apical meristem of *Populus*, which was consistent with the results in *Arabidopsis*. High expression level of laccase genes was not detected at different stages of stem development, but it was observed in some members of the group 4 (*PtrLAC44, PtrLAC45, PtrLAC43, PtrLAC47*) that the expression level was higher in late fall f2 and early winter f1, indicating a dynamic response to seasonal variation. It is worth noting that the expression levels of most *PtrLACs* are high in both nodes and internodes, which may be related to the involvement of laccase in the lignification and thickening of the plant secondary wall.

In male catkin development, *PtrLAC5* is consistently highly expressed, most laccase genes had high expression in the early stage of male catkin development and 3 genes (*PtrLAC13, 19, 25*) showed high expression level in the middle development of male catkin, showed dynamic changes of transcription abundance. In female catkin development, compared with other laccase genes, *PtrLAC30, 11, 21, 19* showed consistently higher expression, most of the *PtrLACs* showed certain dynamic changes with the developmental stage.

In addition, we also analyzed the expression of *PtrLAC* genes under different nitrogen sources. The results showed that *PtrLAC* gene was not significantly differentially expressed in stems due to different nitrogen sources, while in roots, it was found that the expression levels of some laccase genes were significantly increased under nitrate nitrogen treatment compared with other nitrogen sources, such as *PtrLAC11, PtrLAC13, PtrLAC14, PtrLAC18, PtrLAC19, PtrLAC20, PtrLAC21, PtrLAC26*, and *PtrLAC41*. Accordingly, we speculated that most *PtrLAC* genes may be involved in the assimilation of nitrate nitrogen in roots (Figure 7).

In order to understand the relative expression level of *PtrLACs* in different tissues at the early stage of secondary growth, we selected 84K poplar (*P.alba×P.glandulosa*) trees which were 45-day-old, and analyzed the relative expression levels of *PtrLACs* in root, stem and leaf by Quantitative Real-time PCR. It turns out that only part of *PtrLACs* were detected in the early stage of secondary growth, including 20 genes in roots, 17 genes in stems and 12 genes in leaves. We also found some genes with spatial expression specificity in roots and stems, such as *PtrLAC5, PtrLAC7, PtrLAC23, PtrLAC12, PtrLAC18* and *PtrLAC22*, which are currently only expressed in stems. During this period, *PtrLAC13*, *PtrLAC19, PtrLAC21, PtrLAC26, PtrLAC28, PtrLAC31, PtrLAC36, PtrLAC37, PtrLAC42* and *PtrLAC52* were only expressed in roots. In addition, the expression levels of *PtrLAC* genes were relatively close in the same sample, no significant difference was detected. By comparing the relative expression of one *PtrLAC* in different samples, we found that the expression level of most *PtrLAC* genes was higher in roots than in leaves and stems. Expression levels of *PtrLAC9, PtrLAC10, PtrLAC16, PtrLAC25, PtrLAC33, PtrLAC40* and *PtrLAC43* in roots were significantly higher than those in other tissues. (Figure 8).

3.6 Cis-elements in PtrLACs promoters

To further investigate the potential regulatory mechanisms of *PtrLACs*, we used PlantCARE to predict cis-acting elements within 2.0-kb upstream of each gene sequences, nine cis-acting elements were analyzed and displayed in Figure 8. We found that the number of cis-acting elements involved in photoresponse was significantly higher than others and some of them have *MYB* protein binding sites (Baldoni et al., 2015). In view of this interesting finding, we conjectured that the regulation of laccase expression is closely related to photosynthesis. Plastocyanin (Pc) is an important part of the electron transport chain of photosynthesis, which transfer electrons through the redox changes of copper ions in proteins (Höhner et al., 2020). The activity of laccase is also dependent on copper ion (Sun et al., 2019), which leads to the competition effect of copper ions between laccase and Pc. It has been reported that under copper ion stress, the transcriptional regulator SPL7 binds to the copper responsive cis-element (GTAC) in *miR397* promoter to promote *miR397* synthesis, thereby inhibiting the expression of laccase (Yamasaki et al., 2009). Therefore, we speculated that under copper ion stress, light may inhibit the expression of laccase genes, so that copper ion in plants can meet pc requirements and ensure the progress of photosynthesis. In addition, many *PtrLAC* photoresponsive cis-elements contain *MYB* protein binding sites, suggesting that *MYB* may be participate in regulation of this physiological process as a trans-factor. Besides, *PtrLACs* also possess cis-acting elements involved in plant hormone response, especially to jasmonic acid and abscisic acid, indicating that laccase may be related to inducing plant chemical defense and bud dormancy. We also detected the cis-elements related to drought induction, low temperature response, defense and stress response in *PtrLACs*, suggesting a potential stress response under certain conditions. Anyhow, the cis-element analysis illustrated that *PtrLACs* can respond to different abiotic factors, among which light and copper ion concentration may play a crucial role in the regulation of laccase expression. (Figure 9)

3.7 Structure modelling of representative PtrLACs

In order to explore the relationship between the structure and function of laccase proteins, we selected some representative laccases and simulated its 3D structure by

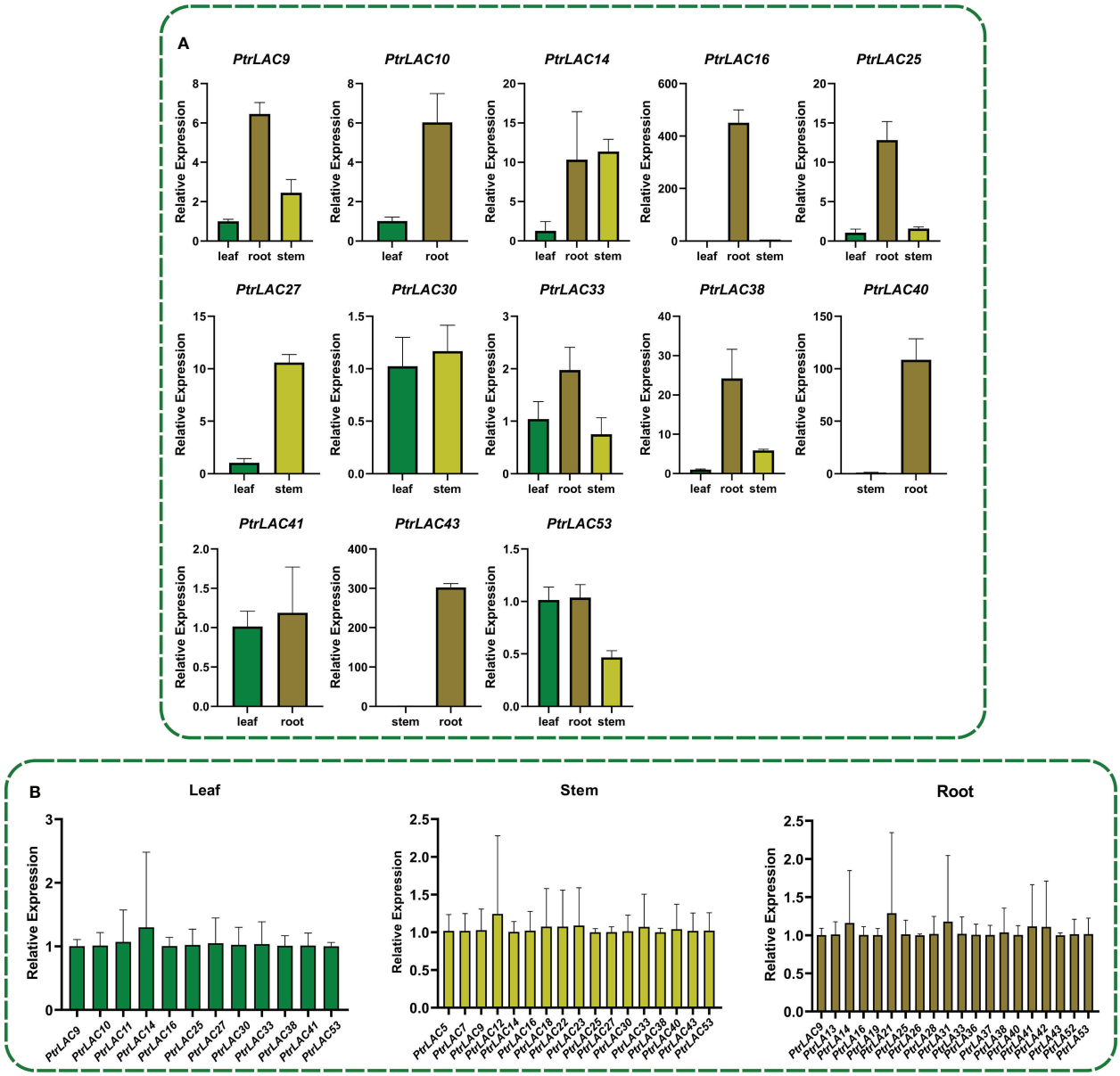


FIGURE 8
Expression analysis of *PtrLACs* by qRT-PCR. **(A)** Relative expression of *PtrLACs* in different tissues, Transcript levels in "leaf" were set to "1", except *PtrLAC40*, *PtrLAC43*, which levels in stem were set to "1". **(B)** Relative expression levels of *PtrLACs* in the same tissue. *PtrLAC9* was set to "1" in "Leaf" and "Root", while *PtrLAC5* was set to "1" in "Stem". All the data were normalized to *18S rRNA* genes, bars represent standard deviation.

using AlphaFold2 (Cramer, 2021). Older genes may be more closely related to the basic physiological functions in plants. Therefore, we selected *PtrLAC25*, a gene that had a collinear relationship with all five species in Figure 6 and participated in fragment repetition 200 million years ago, and *PtrLAC41*, a gene that had a collinear relationship with multiple *PtrLACs* both in WGD events and 200 million years ago. *AtLAC2* has been suggested an involvement in dehydration response (Cai et al., 2006). *PtrLAC12* as a member of group 2 is likely to be involved in *Populus* developmental lignification. In phylogenetic analysis, *PtrLAC12* and *AtLAC2* had a strong genetic relationship and close genetic distance. Combined with the results of cis-acting elements prediction, we speculated that *PtrLAC12* may also played a role in drought stress. Therefore, as a potentially functionally complex laccase, *PtrLAC12* was used as a member of the structural simulation. The Cu-

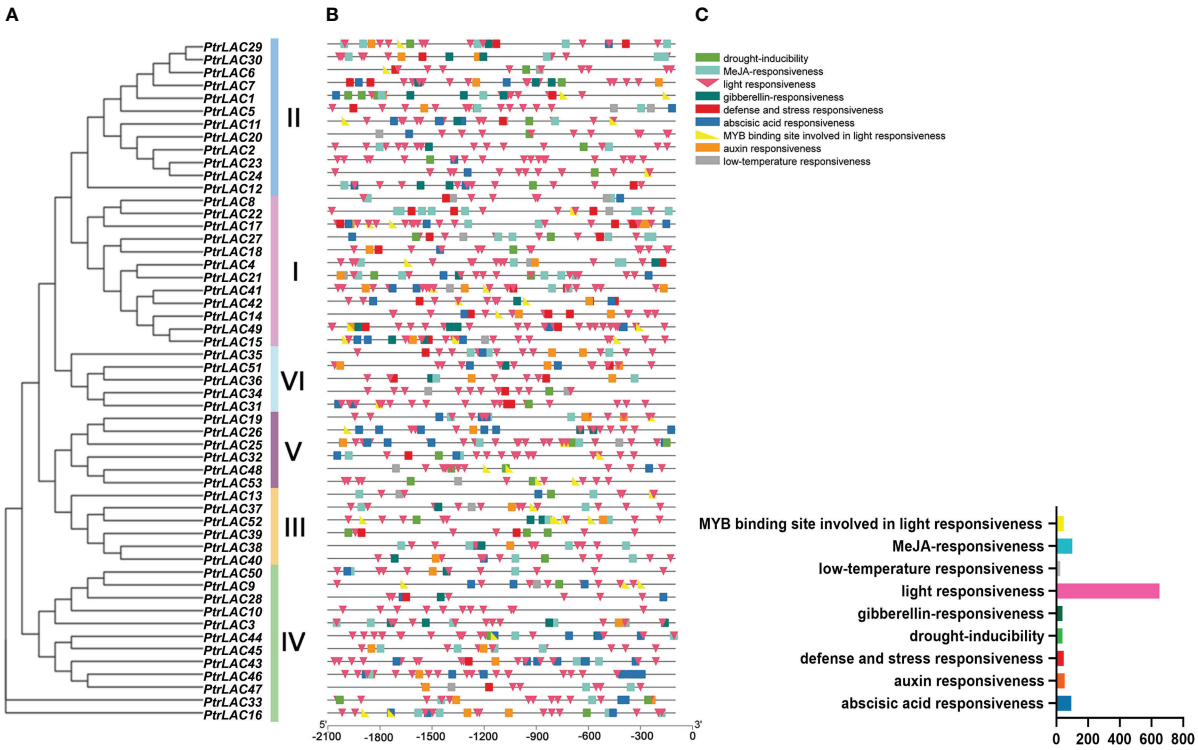


FIGURE 9
Predicted cis-elements in *PtrLAC* promoters. **(A)** Neighbor-joining tree of *PtrLAC* gene family **(B)** Promoter sequences (–2000 bp) of 53 *PtrLACs* were analyzed by PlantCARE. Diverse cis-acting elements were displayed in different colors and shapes. **(C)** Histogram of the number of different cis-acting elements. Cis-acting elements represented by different colors and shapes were shown in the top right-hand corner.

oxidase_3 domain of PtrLAC51 was severely lost, which was the most different from other laccase genes in sequence structure, so it was also within our selection range. *PtrLAC23* is also a member of group 2 and it was important to note that it did not participate in any repetition events of *PtrLAC* gene family and its sequence was complete, without motif loss. Meanwhile, it has been proved that *miRNA novel-m1190-3p* targets *PtrLAC23* in the secondary stem of *P. trichocarpa* (Wang et al., 2021), so we speculated that this gene may be related to wood formation, thus we also predict its structure.

ZmLAC3 is the only plant laccase with known structure and has been confirmed to be implicated in the polymerisation of lignin in maize. Structural comparison with ZmLAC3 protein show that five laccases (PtrLAC12, PtrLAC23, PtrLAC25, PtrLAC41, PtrLAC51) all have low RMSD values (< 3 Å) with ZmLAC3, which means their protein structures are almost the same with ZmLAC3 (Reva et al., 1998) (Pitera, 2014). *Populus* lignin is mainly composed of G and S type monomers, while H type monomer content is very low (Humphreys and Chapple, 2002). Therefore, coniferyl alcohol (ConA) and sinapyl alcohol (SinA) were selected for molecular docking with selected laccases. Results show that the selected five laccases generally prefer ConA to SinA, and PtrLAC23 has the highest affinity with ConA among them.

3.8 PtrLAC23 may be a lignin-related laccase in *Populus trichocarpa*

The results of Figure 10 indicated that *PtrLAC23* (Potri.009G156600) may be involved in the lignification process of *Populus*. To further prove this prediction, we constructed the fusion protein expression vector of *PtrLAC23* and enhanced Green Fluorescent protein (EGFP) initiated by the endogenous promoter: *pLAC23-LAC23-EGFP*, which was introduced into 84K poplar by *Agrobacterium tumefaciens* to obtain stable genetic *proLAC23::LAC23-EGFP* transgenic material. Stem sections of the transgenic plants (growing for 2 months) were observed by LEICA DM2500 fluorescence microscope. Basic principles and flow of the above operations are shown in Figure S1.

We use the green fluorescence of EGFP and spontaneous fluorescence of lignin to locate *PtrLAC23* and lignin, respectively. It turns out that *PtrLAC23* is mainly located in xylem, and a small amount is distributed in phloem region. Its distribution position is almost the same as that of lignin, which can be inferred that *PtrLAC23* has a certain correlation with the synthesis of lignin. (Figure 11). To further verify the function of

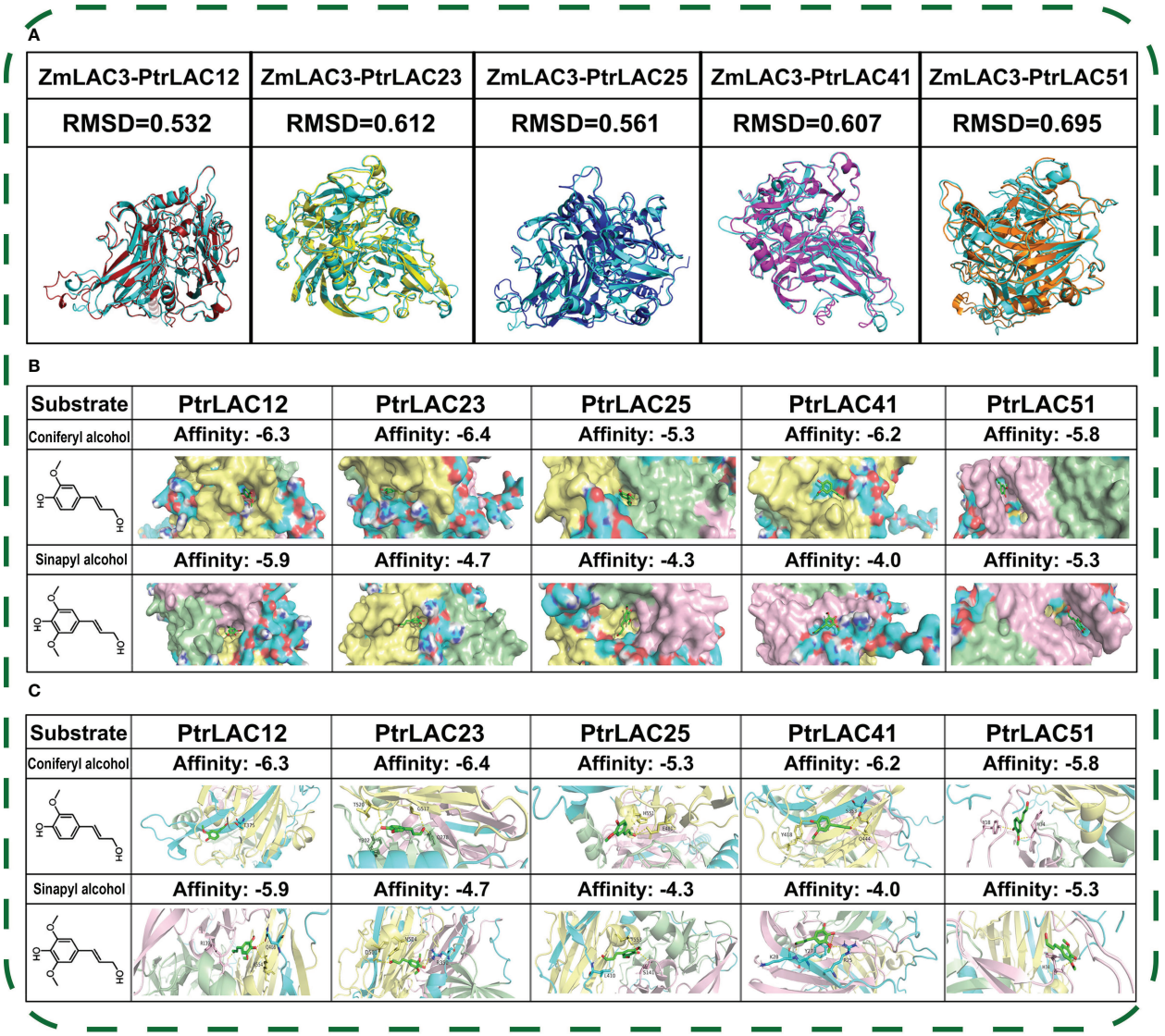


FIGURE 10
Structure modelling of PtrLAC12, PtrLAC23, PtrLAC25, PtrLAC41, PtrLAC51 by using AlphaFold2. **(A)** Structure comparison between PtrLAC and ZmLAC3, ZmLAC3 (cyan), PtrLAC12 (red), PtrLAC23 (yellow), PtrLAC25 (blue), PtrLAC41(magenta), PtrLAC51(orange); RMSD, Root Mean Squared Deviation. **(B)** Pocket shape of PtrLAC substrate binding site, proteins are shown in "surface" form. Pink region represent Cu-oxidase_3 domain, green region represent Cu-oxidase domain, yellow region represent Cu-oxidase_2 domain. Affinity, free-energy of enzyme-substrate binding (more negative value depicts better binding). **(C)** Polar interactions between substrate molecules and amino acids near the binding sites.

PtrLAC23, we determined the relative expression level of *PtrLAC23* in poplar stems at different growth stages. Results showed that compared with the 33 days, 56 days and 69 days samples, the expression level of *PtrLAC23* was significantly increased in stems of 80 day-old *Populus* (Figure 12).

4 Discussion

Laccases, as the largest subfamily of the polycopper oxidase (LMCOs) protein family, have a wide variety of substrates and are widely distributed in plants, insects and microorganisms. Previous studies have shown that plant laccase plays an important role in the oxidation and polymerization of lignin monomers. Under aerobic conditions, laccase can catalyze the oxidative dehydrogenation of lignin monomers to form free radicals, thus initiating the spontaneous polymerization process between monomers. Therefore, the regulation of plant laccase gene expression can indirectly regulate the composition and content of lignocellulose, which provides an effective basis for the improvement of energy trees. *Populus trichocarpa* is a model plant in forest research. Although its laccase gene family has

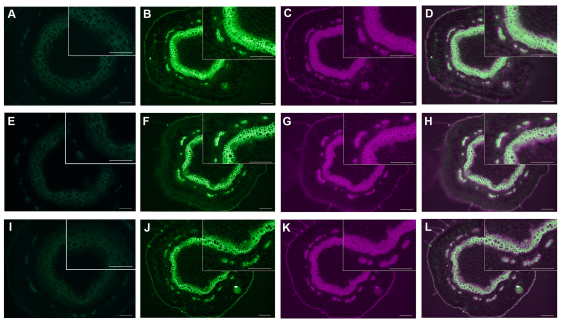


FIGURE 11
Expression and localization of *PtrLAC23* and lignin in different sections of 84k poplar stem. **(A, E, I)** Spontaneous fluorescence of wild-type poplar under blue excitation light.**(B, F, J)** Fluorescence of *pLAC23-LAC23-EGFP* fusion protein (green). **(C, G, K)** Fluorescence of lignin (magenta). **(D, H, L)** Fluorescence of *pLAC23-LAC23-EGFP* fusion protein and lignin. *Upper panel* shows the localization of *PtrLAC23* and lignin at 2.0 cm away from the bottom of the stem. Localization of *PtrLAC23* and lignin at 3.5 cm and 5.0 cm from the bottom of the stem are displayed in *middle and lower panels* respectively. Bar in **(A–L)** = 100 μm.

been identified, detailed analysis of family members was still relatively scarce.

Gene duplication is one of the main causes of gene family formation. There were 22 genes involved in tandem duplication and 22 genes involved in fragment duplication in *Populus* laccase gene family, indicating that the two gene duplication events played an equally important role in the expansion of *PtrLAC* gene family. In addition, the Ka/Ks values of almost all *PtrLAC*

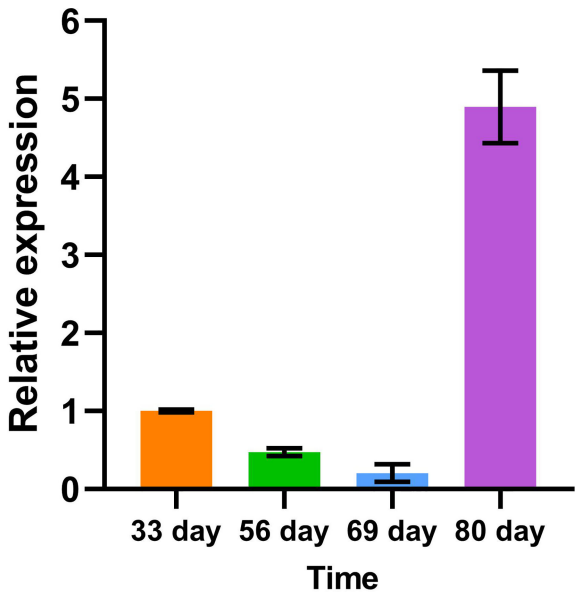


FIGURE 12
Relative expression of *PtrLAC23* in *Populus* stems at different growth stages.

duplicate gene pairs were less than 1, illustrating that *PtrLACs* were underwent a strong purifying selective pressure during the evolution. This conclusion is consistent with the results obtained in *Pyrus bretschneideri* (Cheng X et al., 2020), *Glycine max* (Wang et al., 2019) and *Setaria viridis* (Simões MS et al., 2020), suggesting that this may be a general rule in the plant kingdom. Combined with the molecular evolution rate of *Populus*, we estimated the replication time of collinear gene pairs and found that fragment duplication and tandem duplication were the main driving force of gene family expansion in the early and late stages, respectively. In addition, most *PtrLACs* were replicate after the WGD event, this is different from the results in soybean, which laccase genes replication typically occured during WGD events (Wang et al., 2019). This result also shows that the laccase gene family has different expansion modes in different species. Moreover, 20% of the duplicate gene pairs were separated before 150 MYA, indicating that laccase genes have a very long history in *Populus*, which also allows us to Preliminary lock a series of ancient *PtrLACs* represented by *PtrLAC25* and *PtrLAC41*.

In the early study, *Arabidopsis* laccase gene family was divided into 6 subfamilies, and most of the subsequent studies were based on this criterion. In our study, PtrLACs were distributed in all six subfamilies, in which gene duplication was the main cause of the expansion in group 4, group 5, and group 6. The function of acidic or basic isoforms of laccase in lignification has always been controversial, it is generally believed that acidic peroxidases have poor ability to catalyze sinapyl alcohol (Barceló AR et al., 2007). The evolutionary analysis also revealed that acidic PtrLACs mainly clustered within group 3 and group 4, which happened to be the group of ZmLAC1 and ZmLAC3, respectively. Moreover, ZmLAC3 as an acidic laccase has been confirmed to be involved in injury-induced lignification in maize. Therefore, we hypothesized that the acidic laccases in *Populus* may be functionally specific, and the acidic laccases belonging to the group 3 may also play a role in the process of defense-induced lignification. Although, ZmLAC2, ZmLAC4, and ZmLAC5, which belong to the group 2 have not been confirmed to be involved in the process of injury-induced lignification, their expression patterns were similar to the one expected for genes involved in lignification (Caparrós-Ruiz et al., 2006). Therefore, we speculated that they might play a role in the developmental lignification, and PtrLACs in group 2 may also potentially involved in this process. In conclusion, we believe that PtrLACs in group 3 and 4 may be involved in the defense-induced lignification in *Populus*, and PtrLACs in Group 2 may be involved in the developmental lignification in *Populus*.

Motif is generally considered to be conserved sequence with biological functions, which may contain specific binding sites or common sequence segments involved in a specific biological process. Adjacent *PtrLACs* on the evolutionary tree generally have more similar gene structure and motif composition. However, a few of members in group 4 and group 6 have lost

part of motifs, indicating that these laccases may have acquired new functions in the evolution. In fact, loss events of genes occur frequently during the evolution. Due to the whole-genome triplication (WGT), genes in *Arabidopsis* should have three homologs in *B. rapa* and *B. oleracea*, but in fact, some *AtLACs* can not even have homologs in either species (Lu et al., 2019). The genome of *Populus* is about four times as large as that of *Arabidopsis*, while the number of laccase genes in *Populus* is only three times as large as in *Arabidopsis*. All of these phenomena indicate that gene loss events existed in *LAC* gene family universally.

A large number of studies have shown that laccase exists widely in plants and is of great significance for maintaining plant growth and development. Synteny analysis between *Populus* and five other species showed that some *PtrLACs* (*PtrLAC11, PtrLAC25, PtrLAC20, PtrLAC19*) may have existed before the divergence of monocots and dicots ancestors. Combined with the estimation of the replication time of fragment duplication genes, we found that *PtrLAC25* is not only a relatively old gene in *Populus*, but also has homology with laccase genes in five other species, which deserves more attention.

Gene expression pattern can reflect the function of genes to a certain extent. RNA-Seq data showed that most *PtrLAC* genes were highly expressed in nodes, internodes and catkins, which again indicated the important role of laccase in the polymerization of lignin monomers. Lu et al. have found that 49 laccase genes were all expressed in the 6-month-old *Populus*. However, we could only detect the expression of a few of laccase genes in the early stage of *Populus* secondary growth. *PtrLACs* (*PtrLAC5, PtrLAC7, PtrLAC23, PtrLAC12, PtrLAC18* and *PtrLAC22*) specifically expressed in roots and stems are both members of group 1 and group 2, which is consistent with the findings in *Setaria* (Simões MS et al., 2020). Among them, *PtrLAC5*, *PtrLAC7*, *PtrLAC23* and *PtrLAC12* are both members of group 2. The specific expression of these genes in stem, a lignin-accumulating tissue, again proves the conclusion that members of group 2 may be related to the developmental lignification in *Populus*. In addition, expression level of some *PtrLAC* genes were significantly higher in roots. These laccase genes, with the exception of *PtrLAC25*, all belong to group 3 and group 4. Cotton laccase GhLAC1 is also preferentially expressed in roots. Previous studies have shown that it was involved in the lignification process induced by defense and thus improves the resistance of cotton to biotic stress (Acid et al., 2018).Therefore, we believe that these *PtrLACs*, which are mainly expressed in roots at the initial stage of secondary growth, may also play a potential role in plant defense besides participating in root lignification. This also suggests that members of group 3 and group 4 are likely to be involved in defense-induced lignification in *Populus*.

It has been shown that there is a dynamic equilibrium between photosynthesis and cell wall synthesis, and copper ion concentration in plants is the key to regulate this equilibrium. When copper ions are deficient in plants, *miRNA* (*AtmiR397*, *AtmiR398*, *AtmiR408*, *AtmiR857* in *Arabidopsis*; *PtmiR397*, *PtmiR408*, *PtmiR1444* in *Populus*) will inhibit the expression of other copper-containing proteins, and preferentially supply copper to the essential copper-containing proteins in the photosynthetic electron transport chain represented by PC. Cis-acting elements prediction showed that *PtrLAC* genes had a large number of photoresponsiveness cis-acting elements. In addition, the activity of laccase also depends on the copper ion center, which suggests that there is a relationship between photosynthesis, copper ion concentration and the regulation of laccase gene expression. We also found that MYB protein may be involved in this process as a transcription factor. The network mechanism of MYB family members regulating plant secondary cell wall formation has been fully revealed. However, there are still few reports on the interaction between *MYB* protein and laccase, which needs to be verified by subsequent studies (Xiao et al., 2021).

Understanding the structure of a protein is crucial for defining its function. With the development of artificial intelligence (AI), AlphaFold2 made a breakthrough in the protein folding problem (Marcu et al., 2022). In CASP14, AlphaFold2 achieved a remarkable result. Some of its predicted protein structures were almost indistinguishable from the experimental results (Flower and Hurley, 2021). ZmLAC3 in maize is the only plant laccase which structure has been resolved. Moreover, ZmLAC3 has been proved to catalyze the oxidative polymerization of lignin monomers. The structure prediction of five *Populus* laccase proteins showed that the overall structure of the PtrLAC12, PtrLAC23, PtrLAC25, PtrLAC41 and PtrLAC51 is very close to that of ZmLAC3. These findings indicate that laccase is structurally conserved, and suggest that these proteins may play a potential role in the polymerization of lignin monomers. In addition, PtrLAC23 has the strongest affinity to lignin monomer ConA, which may be a key laccase related to lignification in *Populus*.

At present, there is a lot of evidence that laccase is involved in cell wall lignification in *Populus*. Bryan et al. found that *PdLAC2* plays a role in overcoming plant cell wall recalcitrance in *Populus deltoides* (Bryan et al., 2016). Liu et al. showed that PtrLAC16 can polymerize lignin monomers *in vitro*, inhibition the expression of *PtrLAC16* led to a significant decrease in lignin content and altered cell wall structure (Liu et al., 2021). Niu et al. proved that overexpression of *PeuLAC2* increased the secondary cell wall thickening, fiber cell length and stem tensile strength (Niu et al., 2021). Tissue specificity analysis of these laccases showed that *PtrLAC16* is specifically localized in the xylem and phloem of the stem. Expression of *PeuLAC2* in *P. euphratica* tissues mainly occurred in the xylem. In *P. deltoides*, *LAC2* had the highest expression in xylem compared to other tissues. In our study, fluorescence observation of cross sections of *proLAC23::LAC23-EGFP* line showed that the expression localization of *PtrLAC23* was mainly in xylem and a few areas of the phloem, and was

highly consistent with the deposition site of lignin. These results are similar to the tissue localization analysis of laccases above, which again proved the correlation between *PtrLAC23* and *Populus* wood formation. In addition, relative expression of *PtrLAC23* was significantly increased in the stem of 80-day-old *Populus* than that of young *Populus*, suggesting that the role of *PtrLAC23* in *Populus* stem's secondary growth should not be ignored. The predicted results showed that *PtrLAC23* was likely an extracellular protein. Subcellular localization of PtrLAC16 and PeuLAC2 both showed that laccase protein was probably secreted extracellular and transported to the cell wall to play a role (Niu et al., 2021) (Liu et al., 2021). Here, we also observed the fluorescence of proLAC23::LAC23-EGFP transgenic *Populus* protoplasts (Figure S2). In our results, only chloroplast spontaneous fluorescence was observed and no EGFP signal was found, which also confirmed that PtrLAC23 should be an extracellular protein that probably be transported to play a role on the cell wall.

5 Conclusions

In summary, sequence characteristics, structures and functions of the 53 members of the *PtrLAC* gene family have been elucidated in detail. They can be divided into 6 subfamilies according to their phylogenetic relationships with *Arabidopsis* and maize. Gene duplication analysis indicated that some colinear gene pairs separated 200 million years ago. Moreover, in the process of gene family expansion, fragment repetition is the main driving force in the early stage while tandem repetition plays this role in the late stage. Most laccase proteins are likely to be extracellular, but some have been shown to localize to cell membranes or to lysosomes and peroxisomes. Gene structure and motif analysis demonstrated that adjacent members in the phylogenetic tree had similar intro-exon structure and motif composition. Domain analysis, expression pattern analysis and cis-acting elements prediction shed light on the complex function of *Populus* laccase. Protein structure simulation and molecular docking analysis of PtrLAC12, PtrLAC23, PtrLAC25, PtrLAC41 and PtrLAC51 indicated that PtrLAC23likely to involved in the lignification process. Fluorescence observation of *proLAC23:: LAC23-EGFP* transgenic *Populus* stem transects showed that the location of PtrLAC23 highly overlapped with the lignin deposition. This indicates that PtrLAC23 may play an important role in *Populus* lignin synthesis. Plant laccase is of great significance to improve the utilization efficiency of lignocellulose. Regulating the expression of plant laccase can indirectly influence the content of lignin and cellulose, which can provide the theoretical support for us to obtain ideal biomass materials. These results also provides a broad prospect for us to obtain more ideal biofuels through genetic engineering in the future.

Data availability statement

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

Author contributions

BL conceived the idea, performed the experimental work and wrote the original draft; CW and XL contributed to data curation, review and editing original draft, YM contributed to the design, supervision and direction of experiment, HR provided equipment and technical support for the operation of AlphaFold2, YZ Contributed to project management, material provision and financial support. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (31870571) and National College Students Innovation and Entrepreneurship Training Program (S202010022058).

Acknowledgments

We thanks Haonan Zheng and Hong Li for their help in making the chart, thanks Hailin Hu and Li Xu for their contributions to plant culture, thanks Pan Zhao for providing reference primers.

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.2022.1063813/full#supplementary-material

References

Acid, J., Open, S., Hu, Q., Min, L., Yang, X., Jin, S., et al. (2018). Laccase GhLac1 modulates broad-spectrum biotic stress tolerance *via* manipulating phenylpropanoid pathway. *Plant Physiology* 176, 1808–1823. doi: 10.1104/pp.17.01628

Ahmed, S., Zakaria, G., Tajudin, A. A., Shukuri, M., Ali, M., and Shariff, F. M. (2020). Current development in decolorization of synthetic dyes by immobilized laccases. *Front. Microbiol.* 11, 1–8. doi: 10.3389/fmicb.2020.572309

Allouche, A. (2012). Software news and updates gabedit — a graphical user interface for computational chemistry softwares. *J. Comput. Chem.* 32, 174–182. doi: 10.1002/jcc

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z

Andreaeopsida, B. (2013). 1, 1–14. doi: 10.1007/978-1-4939-7000-1

Asano, T., Seto, Y., Hashimoto, K., and Kurushima, H. (2019). Mini-review an insect-specific system for terrestrialization: Laccase-mediated cuticle formation. *Insect Biochem. Mol. Biol.* 108, 61–70. doi: 10.1016/j.ibmb.2019.03.007

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME s UITE: Tools for motif discovery and searching. *Nucleic Acids Research* 37, 202–208. doi: 10.1093/nar/gkp335

Baldoni, E., Genga, A., and Cominelli, E. (2015). Plant MYB transcription factors: Their role in drought response mechanisms. *Int. J. Mol. Sci.* 16, 15811–15851. doi: 10.3390/ijms160715811

Bao, W., Malley, D. M. O., Whetten, R., and Sederoff, R. R. (2016). *A laccase associated with lignification in loblolly pine xylem* Vol. 260 (American Association for the Advancement of Science), 672–674. Available at: http://www.jstor.org/stable.

Barceló, A. R., Ros, L. V. G., and Carrasco, A. E. (2007). Looking for syringyl peroxidases. *Trends Plant Science.* 12, 486–491. doi: 10.1016j.tplants.2007.09.002

Berthet, S., Demont-caulet, N., Pollet, B., Bidzinski, P., Ce, L., Blondet, E., et al. (2011). Disruption of LACCASE4 and 17 results in tissue-specific alterations to lignification of arabidopsis thaliana stems. *The Plant Cell* 23, 1124–1137. doi: 10.1105/tpc.110.082792

Bryan, A. C., Jawdy, S., Gunter, L., Gjersing, E., Sykes, R., Hinchee, M. A. W., et al. (2016). Knockdown of a laccase in populus deltoides confers altered cell wall chemistry and increased sugar release. *Plant Biotechnol. J.* 14, 2010–2020. doi: 10.1111/pbi.12560

Cai, X., Davis, E. J., Ballif, J., Liang, M., Bushman, E., and Haroldsen, V. (2006). Mutant identification and characterization of the laccase gene family in arabidopsis. *J. Exp. Bot.* 57, 2563–2569. doi: 10.1093/jxb/erl022

Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., and May, G. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in arabidopsis thaliana. *BMC Plant Biol.* 21, 1–21. doi: 10.1186/1471-2229-4-10

Caparrós-Ruiz, D., Fornalé, S., Civardi, L., Puigdomènech, P., and Rigau, J. (2006). Isolation and characterisation of a family of laccases in maize. *Plant Sci.* 171, 217–225. doi: 10.1016/j.plantsci.2006.03.007

Chang, M. C. Y. (2007). Harnessing energy from plant biomass. *Curr. Opin. Chem. Biol.* 11, 677–684. doi: 10.1016/j.cbpa.2007.08.039

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

Cheng, X., Li, G., Ma, C., Abdullah, M., Zhang, J., Zhao, H., et al. (2020). Comprehensive genome-wide analysis of the pear (Pyrus bretschneideri) laccase gene (PbLAC) family and functional identification of PbLAC1 involved in lignin biosynthesis. *PloS One* 15, 1–24. doi: 10.1371/journal.pone.0228183

Clark, W. C. (2008). Biofuels and sustainable development citation. *SSRN Electronic Journal* 174, 1–15. doi: 10.2139/ssrn.1282111

Cosgrove, D. J. (2015). Plant cell wall extensibility: Connecting plant cell growth with cell wall structure, mechanics, and the action of wall- modifying enzymes. *J. Exp. Bot.* doi: 10.1093/jxb/erv511

Cramer, P. (2021). AlphaFold2 and the future of structural biology. *Nat. Struct. Mol. Biol.* 28, 704–705. doi: 10.1038/s41594-021-00650-1

Dandikas, V., Heuwinkel, H., Lichti, F., Drewes, J. E., and Koch, K. (2014). Bioresource technology correlation between biogas yield and chemical composition of energy crops. *Biores. Technol.* 174, 316–320. doi: 10.1016/j.biortech.2014.10.019

Flower, T. G., and Hurley, J. H. (2021). Crystallographic molecular replacement using an in silico-generated search model of SARS-CoV-2 ORF8. *Protein Sci.* 30, 728–734. doi: 10.1002/pro.4050

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., Bairoch, A., et al. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research* 31, 3784–3788. doi: 10.1093/nar/gkg563

Giannuzzi, G., Addabbo, P. D., Gasparro, M., Martinelli, M., Carelli, F. N., Antonacci, D., et al. (2011). Analysis of high-identity segmental duplications in the grapevine genome. *BMC Genomics.* 12, 436–450. doi: 10.1186/1471-2164-12-436

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: S comparative platform for green plant genomics. *Nucleic Acids Research.* 40, 1178–1186. doi: 10.1093/nar/gkr944

Höhner, R., Pribil, M., Herbstová, M., Susanna, L., and Kunz, H. (2020). Plastocyanin is the long-range electron carrier between photosystem II and photosystem I in plants. *PNAS.* 117, 1–9. doi: 10.1073/pnas.2005832117

Hu, W., Harding, S. A., Lung, J., Popko, J. L., Ralph, J., Stokke, D. D., et al. (1999). Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nature Biotechnology.* 17, 808-812. doi: 10.1038/11758

Hu, B., Jin, J., Guo, A., Zhang, H., and Luo, J. (2015). Genome analysis GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

Hullo, M. F., Moszer, I., Danchin, A., and Martin-Verstraete, I. (2001). CotA of bacillus subtilis is a copper-dependent laccase. *J. Bacteriol.* 183, 5426–5430. doi: 10.1128/JB.183.18.5426-5430.2001

Humphreys, J. M., and Chapple, C. (2002). Rewriting the lignin roadmap. *Curr. Opin. Plant Biol.* 5, 224–229. doi: 10.1016/S1369-5266(02)00257-1

Janusz, G., Pawlik, A., Świderska, U., Polak, J., and Sulej, J. (2020). Laccase properties, physiological functions, and evolution. *Int. J. Mol. Sci.* 21, 1–25. doi: 10.3390/ijms21030966

Koch, M. A., Haubold, B., and Mitchell-olds, T. (1998). Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in arabidopsis, arabis, and related genera (Brassicaceae). *Mol. Biol. Evol.* 17, 1483–1498. doi: 10.1093/oxfordjournals.molbev.a026248

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

Le Brun, N. E., Bengtsson, J., and Hederstedt, L. (2000). Genes required for cytochrome c synthesis in bacillus subtilis. *Mol. Microbiol.* 36, 638–650. doi: 10.1046/j.1365-2958.2000.01883.x

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

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research.* 44, 242–245. doi: 10.1093/nar/gkw290

Li, Q., Min, D., Wang, J. P. Y., Peszlen, I., Horvath, L., Horvath, B., et al. (2011). Down-regulation of glycosyltransferase 8D genes in populus trichocarpa caused reduced mechanical strength and xylan content in wood. *Tree Physiol.* 31, 226–236. doi: 10.1093/treephys/tpr008

Liu, Y., Cao, S., Liu, X., Li, Y., Wang, B., Sun, Y., et al. (2021). PtrLAC16 plays a key role in catalyzing lignin polymerization in the xylem cell wall of populus. *Int. J. Biol. Macromol.* 188, 983–992. doi: 10.1016/j.ijbiomac.2021.08.077

Lu, S., Li, Q., Wei, H., Chang, M., Tunlaya-anukit, S., Kim, H., et al. (2013). Ptr-miR397a is a negative regulator of laccase genes affecting lignin content in populus trichocarpa. *PNAS.* 110, 10848–10853. doi: 10.1073/pnas.1308936110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1308936110

Mai, C., and Kharazipour, A. (2001). Modification of lignin for the production of new compounded materials. *Appl. Microbiol. Biotechnol.* 55, 387–394. doi: 10.1007/s002530000590

Marcu, Ş.B., Tăbîrcă, S., and Tangney, M. (2022). An overview of alphafold's breakthrough. *Front. Artif. Intell.* 5. doi: 10.3389/frai.2022.875587

Mate, D. M., and Alcalde, M. (2016). Minireview laccase: A multi-purpose biocatalyst at the forefront of biotechnology. *Microb. Biotechnol.* doi: 10.1111/1751-7915.12422

Mccaig, B. C., Meagher, Æ.R.B., and Dean, J. F. D. (2005). Gene structure and molecular analysis of the laccase-like multicopper oxidase (LMCO) gene family in arabidopsis thaliana. *Planta.* 221, 619–636. doi: 10.1007/s00425-004-1472-6

Niu, Z., Li, G., Hu, H., Lv, J., Zheng, Q., Liu, J., et al. (2021). A gene that underwent adaptive evolution, LAC2 (LACCASE), in populus euphratica improves drought tolerance by improving water transport capacity. *Horticult. Res.* 8, 1–17. doi: 10.1038/s41438-021-00518-x

Pitera, J. W. (2014). Expected distributions of root-mean-square positional deviations in proteins. *J. Phys. Chem. B* 118, 6526–6530. doi: 10.1021/jp412776d

Prem, N., Pabbathi, P., Velidandi, A., Tavarna, T., Gupta, S., and Raj, R. S. (2021). Role of metagenomics in prospecting novel endoglucanases, accentuating functional metagenomics approach in second-generation biofuel production: A review. *Biomass Convers. Biorefin.* doi: 10.1007/s13399-020-01186-y

Ralph, J., Lapierre, C., and Boerjan, W. (2019). Lignin structure and its engineering OH gymnosperm. *Curr. Opin. Biotechnol.* 56, 240–249. doi: 10.1016/j.copbio.2019.02.019

Reva, B. A., Finkelstein, A. V., and Skolnick, J. (1998). What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å? *Folding Design* 3, 141–147. doi: 10.1016/S1359-0278(98)00019-4

Reyes, C., Poulin, A., Nyström, G., Schwarze, F. W. M. R., and Ribera, J. (2021). Enzyme activities of five white-rot fungi in the presence of nanocellulose. *Journal of Fungi.* 7, 9–13. doi: 10.3390/jof7030222

Santos, F., Vera, J. L., van der Heijden, R., Valdez, G., de Vos, W. M., Sesma, F., et al. (2008). The complete coenzyme b 12 biosynthesis gene cluster of lactobacillus reuteri CRL1098. *Microbiology* 154, 81–93. doi: 10.1099/mic.0.2007/011569-0

Sato, S., Nakane, K., and Nakamura, H. (2020). A laccase-catalysed tyrosine click reaction. *Org. Biomol. Chem.* 18, 3664-3668. doi: 10.1039/d0ob00650e

Simões, M. S., Carvalho, G. G., Ferreira, S. S., Hernandes-Lopes, J., de Setta, N., and Cesarino, I. (2020). Genome-wide characterization of the laccase gene family in setaria viridis reveals members potentially involved in lignification. *Planta.* 251, 1–18. doi: 10.1007/s00425-020-03337-x

Seeliger, D., and De Groot, B. L. (2010). Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J. Computer-Aided Mol. Design* 24, 417–422. doi: 10.1007/s10822-010-9352-6

So, C. (2002). The ka / ks ratio: Diagnosing the form of sequence evolution. *TRENDS in Genetics.* 18, 486–487. doi: 10.1016/s0168-9525(02)02722-1

Su, J., Fu, J., Wang, Q., Silva, C., Cavaco-paulo, A., Su, J., et al. (2017). Critical reviews in biotechnology laccase: A green catalyst for the biosynthesis of polyphenols. *Crit. Rev. Biotechnol.* 38, 294–307. doi: 10.1080/07388551.2017.1354353

Sun, K., Li, S., Yu, J., Gong, R., Si, Y., Liu, X., et al. (2019). Chemosphere Cu 2 þ -assisted laccase from trametes versicolor enhanced self-polyreaction of triclosan. *Chemosphere* 225, 745–754. doi: 10.1016/j.chemosphere.2019.03.079

Tuskan, G. A., and Torr, P. (2012). The genome of black cottonwood. *Science* 313, 1596–1604. doi: 10.1126/science.1128691

Trott, O., and Olson, A. J. (2009). AutoDock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi: 10.1002/jcc.21334

Unuofin, J. O., and Okoh, A. I. (2019). Aptitude of oxidative enzymes for treatment of wastewater Pollutants : A laccase perspective. *Molecules* 24, 1–36. doi: 10.3390/molecules24112064

Van de Peer, Y., Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* 10, 725–732. doi: 10.1038/nrg2600

Wang, J., Feng, J., Jia, W., Chang, S., Li, S., and Li, Y. (2015a). Lignin engineering through laccase modification: A promising field for energy plant improvement. *Biotechnol. Biofuels* 8, 1–11. doi: 10.1186/s13068-015-0331-y

Wang, Q., Li, G., Zheng, K., Zhu, X., Ma, J., Wang, D., et al. (2019). The soybean laccase gene family: Evolution and possible roles in plant defense and stem strength selection. *Genes.* 10, 1–19. doi: 10.3390/genes10090701

Wang, R., Reng, M., Tian, S., Liu, C., Cheng, H., Liu, Y., et al. (2021). Transcriptome-wide identification and characterization of microRNAs in diverse phases of wood formation in populus trichocarpa. *G3: Genes Genomes Genet.* 11, 1-8. doi: 10.1093/g3journal/jkab195

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. *Nucleic Acids Research.* 40, 1–14. doi: 10.1093/nar/gkr1293

Wang, R., Taylor, A. B., Leal, B. Z., Chadwell, L. V., Robinson, A. K., Schirf, V., et al. (2015b). Polycomb Group Targeting through Different Binding Partners of RING1B C-Terminal Domain. *Structurel* 18, 966–975. doi: 10.1016/j.str.2010.04.013

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs _ calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies 0–3. doi: 10.1016/S1672-0229(10)60008-3

Wang, X., Zhuo, C., Xiao, X., and Wang, X. (2020). Substrate specificity of LACCASE8 facilitates polymerization of caffeyl alcohol for c-lignin biosynthesis in the seed coat of cleome hassleriana. *Plant Cell* 32, 3825–3845. doi: 10.1105/tpc.20.00598

Wu, L., and Saleh, A. E. (2011). A seed coat outer integument-specific promoter for brassica napus. *Plant Cell Rep.* 30, 75–80. doi: 10.1007/s00299-010-0945-2

Xiao, R., Zhang, C., Guo, X., and Li, H. (2021). MYB transcription factors and its regulation in secondary cell wall formation and lignin biosynthesis during xylem development. *Int. J. Mol. Sci.* 22, 1–19. doi: 10.3390/ijms22073560

Yamasaki, H., Hayashi, M., Fukazawa, M., Kobayashi, Y., and Shikanai, T. (2009). SQUAMOSA promoter binding protein – Like7 is a central regulator for copper homeostasis in arabidopsis. *The Plant Cell.* 21, 347–361. doi: 10.1105/tpc.108.060137

Yong, Z., Kotur, Z., and Glass, A. D. M. (2010). Characterization of an intact two-component high-affinity nitrate transporter from arabidopsis roots. *The Plant J.* 1, 739–748. doi: 10.1111/j.1365-313X.2010.04278.x

Zeng, Y., Zhao, S., Yang, S., and Ding, S. (2014). ScienceDirect lignin plays a negative role in the biochemical process for producing lignocellulosic biofuels. *Curr. Opin. Biotechnol.* 27, 38–45. doi: 10.1016/j.copbio.2013.09.008

Zhao, Q., Nakashima, J., Chen, F., Yin, Y., Fu, C., Yun, J., et al. (2016). LACCASE is necessary and nonredundant with PEROXIDASE for lignin polymerization during vascular development in arabidopsis and nonredundant with LACCASE is necessary PEROXIDASE for lignin polymerization during in ara bid ops is vascular development 25, 3976–3987. doi: 10.1105/tpc.113.117770

Zoppellaro, G., Sakurai, T., and Huang, H. (2001). A novel mixed valence form of rhus vernicifera laccase and its reaction with dioxygen to give a peroxide intermediate bound to the trinuclear center. *Journal of Biochemistry.* 953, 949–953. doi: 10.1093/oxfordjournals.jbchem.a002942

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 07 February 2023
DOI 10.3389/fpls.2023.1068202

OPEN ACCESS

EDITED BY
Venkateswara Rao,
University of Hyderabad, India

REVIEWED BY
Prafull Salvi,
National Agri-Food Biotechnology Institute, India
Ranay Mohan Yadav,
University of Hyderabad, India

*CORRESPONDENCE
Anete Pereira de Souza
anete@unicamp.br

†These authors contributed equally to this work and 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 12 October 2022
ACCEPTED 18 January 2023
PUBLISHED 07 February 2023

CITATION
Santos LB, Aono AH, Francisco FR, da Silva CC, Souza LM and Souza AP (2023) The rubber tree kinome: Genome-wide characterization and insights into coexpression patterns associated with abiotic stress responses.
Front. Plant Sci. 14:1068202.
doi: 10.3389/fpls.2023.1068202

The rubber tree kinome: Genome-wide characterization and insights into coexpression patterns associated with abiotic stress responses

Lucas Borges dos Santos[1†], Alexandre Hild Aono[1†], Felipe Roberto Francisco[1†], Carla Cristina da Silva[1], Livia Moura Souza[1,2] and Anete Pereira de Souza[1,3*]

[1]Center for Molecular Biology and Genetic Engineering, State University of Campinas, Campinas, Brazil, [2]São Francisco University (USF), Itatiba, Brazil, [3]Department of Plant Biology, Biology Institute, University of Campinas (UNICAMP), Campinas, Brazil

The protein kinase (PK) superfamily constitutes one of the largest and most conserved protein families in eukaryotic genomes, comprising core components of signaling pathways in cell regulation. Despite its remarkable relevance, only a few kinase families have been studied in *Hevea brasiliensis*. A comprehensive characterization and global expression analysis of the PK superfamily, however, is currently lacking. In this study, with the aim of providing novel inferences about the mechanisms associated with the stress response developed by PKs and retained throughout evolution, we identified and characterized the entire set of PKs, also known as the kinome, present in the Hevea genome. Different RNA-sequencing datasets were employed to identify tissue-specific expression patterns and potential correspondences between different rubber tree genotypes. In addition, coexpression networks under several abiotic stress conditions, such as cold, drought and latex overexploitation, were employed to elucidate associations between families and tissues/stresses. A total of 1,809 PK genes were identified using the current reference genome assembly at the scaffold level, and 1,379 PK genes were identified using the latest chromosome-level assembly and combined into a single set of 2,842 PKs. These proteins were further classified into 20 different groups and 122 families, exhibiting high compositional similarities among family members and with two phylogenetically close species *Manihot esculenta* and *Ricinus communis*. Through the joint investigation of tandemly duplicated kinases, transposable elements, gene expression patterns, and coexpression events, we provided insights into the understanding of the cell regulation mechanisms in response to several conditions, which can often lead to a significant reduction in rubber yield.

KEYWORDS

coexpression networks, *Hevea brasiliensis*, kinase family, RNA-sequencing, tandem duplications, transposanable elements

1 Introduction

Rubber is one of the world's major commodities and is extensively used in various industrial and domestic applications, yielding more than 40 billion dollars annually (Board, 2018) The major source of latex for rubber production is *Hevea brasiliensis* (Hbr), commonly referred to as rubber tree, a perennial native plant from the Amazon rainforest belonging to the *Euphorbiaceae* family (Priyadarshan and Goncalves, 2003). Although the warm and humid weather in the Amazon region offers a favourable climate for Hbr growth and propagation, large-scale cultivation of Hbr is unviable due to the incidence of a highly pathogenic fungus, *Pseudocercospora ulei* (da Hora Júnior et al., 2014). Thus, rubber tree plantations were transferred to other countries and regions, which could not offer optimal conditions for developing tropical crops due to the low temperatures during winter, dry periods, and elevated wind incidence (Hoa et al., 1998). Exposure to these abiotic stresses often leads to a significant reduction in latex production in most Hbr wild varieties, which has stimulated the development of breeding programs with a focus on stress-tolerant cultivars (Pushparajah, 1983; Priyadarshan and Goncalves, 2003).

Different types of abiotic stresses may trigger several physiological responses in susceptible rubber tree genotypes and often impact their survival, growth and productivity, depending on the age and vigour of the affected plant (Kuruvilla et al., 2017). In general, cold and drought stresses result in the inhibition of photosynthesis and chlorophyll degradation (Devakumar et al., 2002). Water deficit may affect the plant growth and canopy architecture of trees, and its impact during tapping seasons tends to be more severe due to the deviation of resources (carbon and water) caused by wounding stress (Devakumar et al., 1999; Kunjet et al., 2013). Cold damage leads to a decrease in membrane permeability (Meti et al., 2003; Sevillano et al., 2009), together with photosynthesis inhibition, causing more critical injuries to the plant, such as the wilting and yellowing of leaves, interveinal chlorosis, darkening of the green bark, reduction of latex flow and dieback of shoots (Meti et al., 2003).

Projections indicate that climate changes caused by global warming will enable the expansion of areas suitable for rubber tree plantation, especially in regions with greater production of natural rubber (Zomer et al., 2014; Yang et al., 2019). However, the real impact of climate change on the origin of rubber tree diversity is still unknown, but some studies suggest that these changes will have severe consequences for rubber tree biodiversity, mainly due to changes in the water regime (Marengo et al., 2018; do Prado Tanure et al., 2020). These alterations are even worse considering that the rubber tree is still being domesticated and has little genetic variability explored (De Souza et al., 2018).

The ability to sense and adapt to adverse conditions relies on the activation of complex signaling networks that protect plants from potential damage caused by these environmental changes (Kovtun et al., 2000). Protein kinases (PKs) comprise one of the most diverse protein superfamilies in eukaryotic organisms (Liu et al., 2015) and act as key components in stimulus perception and signal transduction through a chain of phosphorylation events, resulting in the activation of genes and several cellular responses (Colcombet and Hirt, 2008). The expansion of this family underlies several mechanisms of gene duplication throughout the evolutionary history of eukaryotes, including chromosomal and whole-genome duplication, tandem duplication in linked regions and retroposition events, leading to more specialized or novel gene functions (Zhang, 2003).

In the rubber tree, several kinase families have been characterized, including the mitogen-activated protein kinase (MAPK) (Jin et al., 2017), calcium-dependent protein kinase (CDPK) (Xiao et al., 2017; Zhu et al., 2018b), CDPK-related protein kinase (CPK) (Xiao et al., 2017), and sucrose non-fermenting 1-related protein kinase 2 (SnRK2) (Guo et al., 2017). These studies revealed contrasting expression patterns of the kinase families among tissues, in addition to the elevated expression of the SnRK2 and CPK families in laticifers in response to ethylene, ABA, and jasmonate stimulation (Guo et al., 2017; Zhu et al., 2018b), suggesting their potential participation during several developmental and stress-responsive processes. However, the comprehensive identification and characterization of rubber tree PKs have not yet been performed and would greatly benefit plant science research to promote a better understanding of the molecular mechanisms underlying the stress response. Furthermore, the complete characterization of the rubber tree kinome has the potential to highlight important PKs associated with stress resilience across evolution, which is of great interest for plant breeding efforts, especially considering the current genetic engineering and genome editing methodologies available (Pandita and Wani, 2021). Due to the long breeding cycles for the development of rubber tree cultivars [ranging from 25 to 30 years (Priyadarshan, 2017)] and the need for introgression of resistant alleles to more productive varieties, the definition of key molecular mechanisms signaling stress response represents an important contribution to target gene candidates for molecular breeding approaches.

In this study, we investigated the kinase diversity present in the Hbr genome through a characterization of its PKs, including the subfamily classification and the prediction of several protein properties, such as molecular weight, subcellular localization, and biological functions. The rubber tree kinome, defined as the complete repertoire of PKs, was estimated using a combined analysis with the two major genome assemblies of the rubber plant and comparative analyses with cassava (*Manihot esculenta* (Mes)) and castor plant (*Ricinus communis* (Rco)) kinomes. Furthermore, RNA sequencing (RNA-Seq) data from different Hbr genotypes were used to identify expression patterns of the kinase subfamilies, followed by the construction of gene coexpression networks for control and abiotic stress conditions. In addition to fully characterizing the Hbr kinome, our study aims to provide novel inferences about the mechanisms associated with the stress response developed by PKs and retained throughout evolution, assessed by a joint analysis considering tandemly duplicated PK subfamilies, transposable elements and expression patterns. Our study provides broad resources for future functional investigations and valuable insights into the major components associated with cell adaptation in response to environmental stresses.

2 Material and methods

2.1 Data acquisition

Sequence and annotation files of Hbr, Mes, and Rco were downloaded from the NCBI (Geer et al., 2010) and Phytozome v.13 (Goodstein et al., 2012) databases. We selected the latest genomes of

cassava v.7.1 (Bredeson et al., 2016) and castor plant v.0.1 (Chan et al., 2010), as well as two major genomes of the rubber tree: the latest chromosome-level genome (Liu et al., 2020b) (Hb chr) and the reference scaffold-level assembly (Tang et al., 2016) (Hb scaf), under accession numbers PRJNA587314 and PRJNA310386 in GenBank, respectively. The same data analysis procedures for PK identification and characterization were applied to Hbr, Mes and Rco.

2.2 Kinome definition

The hidden Markov models (HMMs) of the two typical kinase domains, Pkinase (PF00069) and Pkinase Tyr (PF07714), were retrieved from the Pfam database (El-Gebali et al., 2019). To select putative proteins having one or more kinase domains, protein sequences were aligned to each HMM profile using HMMER v.3.3 (Finn et al., 2011) (E-value of 1.0E-10). We retained only sequences covering at least 50% of the respective domain and the longest isoform.

The Hbr kinome was created as a combination of putative PKs identified from two different genomic datasets: Hb_chr and Hb_scaf. To avoid redundancy, we combined the sets using CD-HIT v.4.8.1 software (Fu et al., 2012) with the following selection criteria: (i) for proteins present in both sets as a single copy, the longest sequence was retained, and the other one was discarded; and (ii) when putative duplications were present, i.e., there were protein clusters with significant similarities in both Hb_chr and Hb_scaf, and all proteins from the largest set were retained. For pairwise comparisons, we set a minimum sequence identity threshold of 95% and a maximum length difference of 75%.

2.3 Kinase characterization and phylogenetic analyses

The PKs were classified into groups and subfamilies according to the HMMs of each family built from four plant model species (*Arabidopsis thaliana*, *Chlamydomonas reinhardtii*, *Oryza sativa*, and *Physcomitrella patens*) and supported among 21 other plant species (Lehti-Shiu and Shiu, 2012). The classification was further validated through phylogenetic analyses. The domain sequences from all PKs were aligned using Muscle v.8.31 (Edgar, 2004), and a phylogenetic tree was constructed for each kinase dataset using the maximum likelihood approach in FastTree v.2.1.10 software (Price et al., 2010) with 1,000 bootstraps and default parameters through the CIPRES gateway (Miller et al., 2011). The resulting dendrograms were visualized and plotted using the statistical software R (Ihaka and Gentleman, 1996) together with the ggtree (Yu et al., 2017) and ggplot2 (Wickham, 2009) packages.

For each PK, we obtained the following characteristics: (a) gene location and intron number, according to the GFF annotation files; (b) molecular weight and isoelectric point with ExPASy (Gasteiger et al., 2003); (c) subcellular localization prediction using CELLO v.2.5 (Yu et al., 2006) and LOCALIZER v.1.0.4 (Sperschneider et al., 2017) software; (d) the presence of transmembrane domains using TMHMM Server v.2.0 (Krogh et al., 2001); (e) the presence of N-terminal signal peptides with SignalP Server v.5.0 (Armenteros et al., 2019); and (f) gene ontology (GO) term IDs using Blast2GO software (Conesa and Götz, 2008) with the SwissProt Viridiplantae protein dataset (Consortium, 2019).

2.4 Duplication events in the rubber tree kinome

We determined duplication events of the PK superfamily in Hbr based on the physical location of PK genes and their compositional similarities assessed through comparative alignments with the Hbr genome using the BLASTn algorithm (Altschul et al., 1990). Tandem duplications were defined as PK pairs separated by a maximum distance of 25 kb on the same chromosome and with the following: (i) a minimum similarity identity of 95%; (ii) an E-value cutoff of 1.0E-10; and (iii) a 75% minimum sequence length coverage. The chromosomal location of putative tandemly duplicated PK genes was illustrated using MapChart v.2.32 (Voorrips, 2002), and synteny relationships of the PKs were visualized using Circos software v.0.69 (Krzywinski et al., 2009).

2.5 Transposable element search

We searched for transposable elements (TEs) in the Hbr genome using TE data of 40 plant species obtained from the PlanNC-TE v3.8 database (Pedro et al., 2018). For this purpose, we performed a comparative alignment between the TEs retrieved and the *H. brasiliensis* reference chromosomes using BLASTn (Altschul et al., 1990) for short sequences (blastn-short) with the following parameters: (i) minimum coverage of 75%; (ii) word size of 7; and (iii) an E-value cutoff of 1.0E-10. We selected TEs located within a 100 kb window from Hbr PK genes. The chromosomal localization of TEs was illustrated using MapChart v.2.32 (Voorrips, 2002)

2.6 RNA-seq data collection

Several publicly available Hbr RNA-Seq experiments were collected from the NCBI Sequence Read Archive (SRA) database (Leinonen et al., 2010). The samples consisted of a wide range of tissues and comprised various genotypes. In total, we obtained 129 samples from 10 studies (Lau et al., 2016; Li et al., 2016; Tang et al., 2016; Tan et al., 2017; Cheng et al., 2018; Deng et al., 2018; Montoro et al., 2018; Sathik et al., 2018; Mantello et al., 2019; Rahman et al., 2019) that evaluated control and stress conditions (cold, drought, latex overexploitation, jasmonate, and ethylene treatments).

2.7 Expression analysis

The raw sequence data were submitted to a sequence quality control assessment using the FastQC tool (Andrews, 2010), following a low-quality read filtering and adapter removal step using Trimmomatic software v.0.39 (Bolger et al., 2014). After removing adapter sequences, we retained only reads larger than 30 bp and bases with Phred scores above 20. The corresponding coding sequences

(CDSs) from Hb_chr and Hb_scaf were used as a reference for the quantification step using Salmon software v.1.1.0 (Patro et al., 2017) with the k-mer length parameter set to 31. The expression values of each PK transcript were normalized using the transcript per million (TPM) metric, and samples containing biological replicates were combined by defining the mean value among replicates. To visualize the expression of each kinase subfamily among different tissues and cultivars, we generated two heatmap figures for control and stressed samples using the R package pheatmap (Kolde and Kolde, 2015).

2.8 Coexpression network construction

Two coexpression networks of Hbr PK subfamilies, corresponding to control and abiotic stress situations, were modeled and visualized using the R package igraph (Csardi and Nepusz, 2006) with the minimum Pearson correlation coefficient set to 0.7. To assess the structure of each network and specific subfamily attributes, we estimated the hub scores of each PK subfamily from Kleinberg's hub centrality scores (Kleinberg, 1999) and edge betweenness values from the number of geodesics passing through each edge (Brandes, 2001).

3 Results

3.1 Genome-wide identification, classification and characterization of PKs

Based on the established pipeline, we identified 2,842 typical putative PK genes in Hbr (Supplementary Table S1A), 1,531 in Mes (Supplementary Table S1B), and 863 in Rco (Supplementary Table S1C). The rubber tree kinome resulted from 1,206 (42.43%) proteins from the Hb_scaf dataset and 1,636 (57.57%) from Hb_chr. Interestingly, we also identified several PKs containing multiple kinase domains in all three datasets (191, 91, and 44 in Hbr, Mes and Rco, respectively) (Supplementary Tables S2), which were probably retained during evolution due to their action with specific substrates. Typical PKs were defined as protein sequences presenting high similarity to a given kinase domain with minimum coverage of 50%. The atypical PKs of Hbr (728), Mes (230), and Rco (95) were removed from subsequent analyses, being considered as probable pseudogenes.

Typical Hbr, Mes, and Rco PKs were further classified into groups and subfamilies based on the HMM profiles of 127 kinase subfamilies defined by Lehti-Shiu and Shiu (2012). The PK domain classification was validated by phylogenetic analyses (Supplementary Figures S1, S2). Thus, PKs were grouped into 20 major groups: PKs A, G and C (AGC), Aurora (Aur), budding uninhibited by benzimidazoles (BUB), calcium- and calmodulin-regulated kinases (CAMK), casein kinase 1 (CK1), cyclin-dependent, mitogen-activated, glycogen synthase, and CDC-like kinases (CMGC), plant-specific, inositol-requiring enzyme 1 (IRE1), NF-kB-activating kinase (NAK), NIMA-related kinase (NEK), Pancreatic eIF-2α kinase (PEK), Receptor-like kinase (RLK)-Pelle, *Saccharomyces cerevisiae* Scy1 kinase (SCY1), Serine/threonine kinase (STE), Tyrosine kinase-like kinase (TKL), Tousled-like kinases (TLK), Threonine/tyrosine kinase (TTK), Unc-51-like kinase (ULK), Wee1, Wee2, and Myt1 kinases (WEE), and with no lysine-K (WNK). We also identified 72 PKs in Hbr (2.5%), 30 in Mes (2.0%), and 22 in Rco (2.5%) that did not cluster in accordance with any subfamily classification and were placed in the "Unknown" category (Supplementary Table S3). This category represents species-specific PKs, which might be related to new gene families.

The RLK-Pelle was the most highly represented group in all three species, as evidenced in Figure 1, and was divided into 59 different subfamilies, accounting for 65.5%, 68.1%, 65.2% of all rubber tree, cassava, and castor plant PKs, respectively, followed by the CMGC (6.4% in Hbr, 5.9% in Mes, 7.5% in Rco), CAMK (5.9% in Hbr, 6.5% in Mes, 6.5% in Rco), TKL (4.9% in Hbr, 4.9% in Mes, 5.6% in Rco) and others (Supplementary Table S4). Such similarity in the distribution of PK subfamilies corroborates the phylogenetic proximity between these species and the high conservation of PKs. We investigated the chromosomal positions, intron distribution and structural properties of Hbr, Mes, and Rco PKs using several approaches (Supplementary Tables S5-S7). Hbr and Mes PK genes were distributed along all Hbr and Mes chromosomes (Supplementary Figure S3) with an apparent proportion to the chromosome length. There was also a higher concentration of PKs in the subtelomeric regions, which may indicate chromosomal rearrangements and increased variability. Most PK genes contained at least 1 intron, and only 284 (10.0%), 229 (14.9%), and 140 (16.2%) intronless genes were found in Hbr, Mes, and Rco, respectively. In specific subfamilies, the distribution of introns presented a similar profile, e.g. the CK1_CK1 subfamily (21 members) had an average of 14 introns per gene with a coefficient of variation of about 10%. Interestingly, members of the SCY1_SCYL2, TKL-Pl-7, RLK-Pelle_LRR-XIIIb, and RLK-Pelle_LRR-VII-3 subfamilies presented the same quantity of introns, which might be related to conserved plant roles across evolution.

Most interestingly, the protein characteristics of all three kinomes were highly comparable. Several PKs had predicted transmembrane domains (45.6% in Hbr, 50.5% in Mes, and 48.2% in Rco) and N-terminal signal peptides (29.6%, 37.3%, and 33.5%, respectively). Similarly, the distribution of molecular weights and isoelectric points were relatively uniform (Supplementary Figure S4), equally observed for the subfamily divisions. Moreover, the subcellular localization predictions performed with the selected software were mostly on the plasma membrane, cytoplasm and nucleus (Supplementary Figure S5), in accordance with the enriched "cellular component" GO category (Supplementary Table S8; Supplementary Figure S5). Indeed, PKs have a more pronounced presence across the plasma membrane with specific members acting on other subcellular components.

Finally, we investigated the domain composition of PKs based on the complete set of conserved domains present in the Pfam database. In total, we identified 1,472 PKs containing additional conserved domains in Hbr (52.8%), 827 in Mes (54.0%) and 442 in Rco (51.2%) (Supplementary Tables S9, S10). This varied profile confirms the diverse functions of PKs, their speciation and importance for plant adaptation. Interestingly, we identified a significant portion of members from groups CAMK (58.0% in Hbr, 61.6% in Mes, and 64.3% in Rco), RLK-Pelle (63%, 65.5%, and 63.9%, respectively), and TKL (48.2%, 48%, and 52%) with additional domains. Based on these findings, it can be observed that the evolutionary history of PKs is also based on specific domain arrangements.

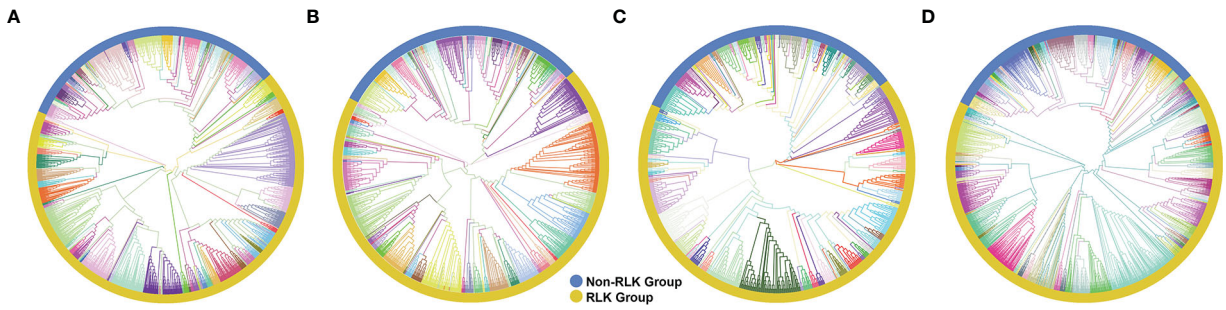


FIGURE 1
Phylogenetic analyses of putative typical protein kinases (PKs) identified in the *Hevea brasiliensis* (Hbr), *Manihot esculenta* (Mes), and *Ricinus communis* (Rco) genomes. **(A)** Phylogenetic tree constructed with 2,842 Hbr PKs organized into 123 subfamilies. **(B)** Phylogenetic tree of the 1,531 Mes PKs organized into 123 subfamilies. **(C)** Phylogenetic tree of the 863 Rco PKs organized into 125 subfamilies. **(D)** Phylogenetic tree of all Hbr, Mes, and Rco PKs. Kinase subfamilies are represented by different branch colors."1 Phylogenetic analyses of putative typical protein kinases (PKs) identified in the Hevea brasiliensis (Hbr), Manihot esculenta (Mes), and Ricinus communis (Rco) genomes.

3.2 Kinase duplication events in *H. brasiliensis*

To examine the expansion of PK subfamilies, tandemly duplicated kinase genes were identified based on their physical localization on Hbr chromosomes, and protein similarities were assessed through comparative alignments. Taken together, we found that 339 of the 2,842 Hbr PK genes (~11.9%) were arranged in clusters of highly similar gene sequences among the 18 reference chromosomes, which are likely to represent tandem duplication events (TDEs) of the kinase superfamily in rubber tree (Figure 2A). These genes were dispersed in 145 separate clusters and comprised members of 63 kinase subfamilies (Supplementary Tables S11, S12). Chromosome 14 showed the highest number of TDEs (19), containing 47 PK genes. In contrast, chromosome 1 contained the least number of TDEs (2). We found that for some kinase subfamilies, a large portion of their members originated from TDEs, suggesting that such duplication events influenced the expansion of specific groups due to putative functional implications. A total of 100%, 100%, and 75% of TTK, ULK_Fused, and CMGC_CDKL-Os members were tandemly organized, while other subfamilies, such as RLK-Pelle_DLSV, showed the largest absolute number of TDEs (45) distributed across 9 chromosomes, although it accounted for only 16.8% (45/268) of its total size.

Segmental duplication events were estimated based on sequence similarities between two or more PKs separated by a genomic window larger than 100 kb or present in different chromosomes. Genomic correspondences increased as the sequence similarity decreased (Figures 2B–D). In total, we identified 858 kinase correspondences with compositional similarity greater than 90%, 1,673 for 75% and 10,121 for 50%. Clearly, the expansion of the rubber tree kinome is mainly caused by segmental duplication events, probably related to the *Hevea* paleopolyploid genome origin (Pootakham et al., 2017). To further investigate potential biological processes associated with duplicated kinase genes, we performed a functional annotation pipeline on tandemly duplicated PKs and selected GO terms related to the "biological process" category (Supplementary Figure S6). The findings were consistent with those resulting from the analysis performed using the complete set of Hbr PKs (Supplementary Figure S7). Although there is a subset of PK subfamilies with members originated by TDEs, their functional profile represents the entire repertoire of PK GO terms, which shows that, even less pronounced, TDEs in rubber tree play an important role in the kinome expansion.

3.3 Transposable elements in *H. brasiliensis* genome

Due to the high abundance of TEs in the rubber tree genome and current evidence associating these elements with phenotypic modifications wu2020structural, francisco2021unravelling, we predicted TEs near Hbr PK genes using a comprehensive database that combined data from overlapping regions of TE features and several classes of noncoding RNAs (ncRNAs). Overall, the percentage of TEs associated with PK genes in the rubber tree was reduced (23.7%) when compared to overlapping ncRNAs (76.3%) (Supplementary Table S13), which have a broader set of regulatory roles hadjiargyrou2013intertwining. Out of the 8,457 annotated TEs in the reference genome, 88% were classified as long terminal repeat (LTR) retrotransposons. These elements appeared to be associated (within a 100 kb genomic window) with 362 (12.7%) kinase genes (Figure 2A; Supplementary Table S14), of which 56 (15.5%) were tandemly duplicated. Such findings highlight the importance of these joint mechanisms for understanding the complex dynamics behind rubber tree kinome expansion. Nearly 73.2% of these duplicated genes were members of the RLK-Pelle group, reiterating its importance.

3.4 Expression patterns of PK subfamilies

In order to supply a broad evaluation of PK expression, we analyzed the expression levels of 118 kinase subfamilies among 129 samples related to control and different abiotic stress conditions (Supplementary Table S15). The resulting dataset comprised transcriptomic data of 14 different cultivars from various tissues and organs, including leaf, petiole, bark, latex, seed, male and female flowers, in early and mature developmental stages. After filtering out low-quality reads and removing adapter sequences, we mapped the filtered reads to the complete set of CDS sequences in Hb_chr and Hb_scaf reference genomes separately and further generated a subset of the quantifications corresponding to PK genes present in the Hbr kinome. Such an approach enabled a deep

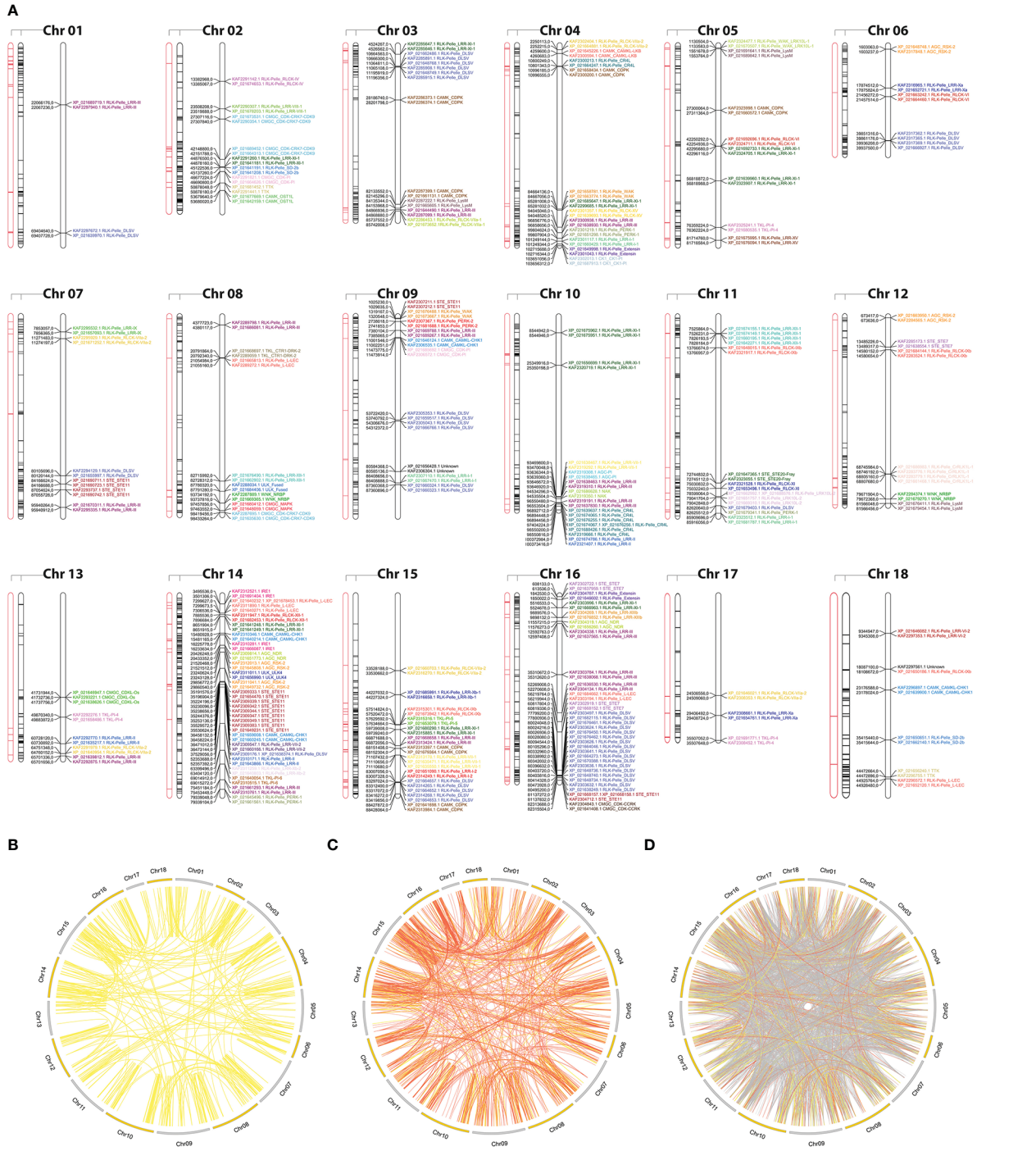


FIGURE 2
(A) Kinase distribution along *Hevea brasiliensis* chromosomes. For each chromosome, from left to right: (i) transposable elements located within a 100 kb window around kinase genes are highlighted in red; (ii) all genes with kinase domains are highlighted in black; and (iii) tandemly duplicated kinase genes are colored and labeled according to the kinase subfamily classification. **(B)** Potential segmental duplication events in the *H brasiliensis* genome considering similarities greater than 90% (yellow); **(C)** 75% (orange); and **(D)** 50% (gray).

characterization of PK gene expression, yet under-explored. The results were normalized to TPM values (Supplementary Tables S16, S17), and 3 samples presenting significantly low quantifications were excluded (Hb_Bark_3001_normal_rep1, Hb_Latex_712_normal_rep2, Hb_Latex_2025_normal_rep1). For cases where replicates were present, the expression values were averaged.

For both control and stress heatmaps (Supplementary Figures S8, S9, respectively), samples belonging to the same tissues were clustered together based on Euclidean distance measures. In general, we observed similar patterns of expression of each kinase subfamily within samples of a given tissue; however, specific experimental conditions of each RNA-Seq dataset may have influenced the expression levels, leading to inconsistent patterns in some cases. In addition to the high genetic diversity among the rubber tree genotypes used in the datasets, some PK subfamilies may have a different response to each type of stress present in our study, as observed in

the different clustering profiles (Supplementary Figure S10). From left to right in the heatmap containing all experiments (Supplementary Figure S10), there were 5 major clusters separated into the following categories: (i) latex; (ii) leaf and seed tissues; (iii) bark, root, male and female flowers; (iv) leaf; and (v) samples from latex and petiole.

Interestingly, several subfamilies were highly expressed in nearly all samples, including AGC_PKA-PKG, TKL-Pl-1, RLK-Pelle_LRR-VIII-1, RLK-Pelle_RLCK-IXa, and Aur, which probably represent PK subfamilies not affected by stress stimuli, but related to basal plant functions. Therefore, distinctions in PK expression between leaf and latex samples were clear. Latex and bark tissues presented lower expression in most subfamilies; however, we detected a few cases where the expression in latex and bark was significantly higher than in leaves, such as STE_STE-Pl and RLK-Pelle_LRR-VIII-1, providing indicative targets of the roles of specific PK subfamilies that affect the secondary metabolism. Additionally, we found a small number of subfamilies with elevated expression in bark (TKL-Pl-7 and ULK_ULK4), which depends directly on the seasonality of the climate (Budzinski et al., 2016). Overall, in the analysis of the expression levels under abiotic stress conditions, the number of subfamilies that presented moderately high (dark orange) and high (red) expression increased when compared to control samples (highlighted in blue), showing the activation of PK subfamilies under stress.

3.5 Coexpression networks in response to abiotic stresses

The quantification analysis revealed different expression profiles of PK subfamilies among different tissues, genotypes and conditions. To expand our understanding of how these proteins interact under exposure to abiotic conditions, we further investigated potential relationships between kinase subfamilies by constructing coexpression networks based on the expression data described above. Using the Hbr PK set, two independent networks were constructed: one for control and one for abiotic stress conditions. Using such an approach, we were able to investigate PK subfamily interactions activated under abiotic stress conditions. For each network, we used the following conventions: (i) kinase subfamilies were represented by separate nodes; (ii) the node size corresponded to the mean gene expression value; (iii) the edges represented coexpression events determined by pairwise expression correlations between subfamilies with a minimum Pearson correlation coefficient of 0.7; and (iv) the edge thickness corresponded to the degree of correlation, from moderate (minimum PCC of 0.7) to relatively strong (minimum PCC of 0.9) correlations.

We observed a different number of edges between networks (1,162 in control and 704 in stress). The presence of a reduced number of connections in the stress coexpression network might indicate the impact of stress conditions on PK subfamilies, changing their expression profile and consecutively their functional interaction with other subfamilies. Moreover, we found 15 elements in each network that were disconnected from the main core (i.e., kinase subfamilies with no significant correlation in the expression); however, they were related to different subfamilies in each network (Figure 3), showing that under specific conditions PK subfamilies might activate a different communication mechanism, overcoming external factors. Figs. 3B and 3D highlight the red correlation similarities between control and stress coexpression networks, while dark grey edges represent unique connections for each condition. It is possible to observe that there is a common core of connections between the two modeled networks, with specific interactions related to the samples used to model these systems, i.e. the presence or absence of stress.

To obtain an overview of the most influential subfamilies in PK processes, we first calculated hub centrality scores within each network, which are represented by the node colors in Figure 3 (Supplementary Table S18). Elevated hub scores (highlighted in dark gray in Figure 3) indicate PK subfamilies with a significant number of connections, probably developing a broad set of molecular interactions and being an important component in PK communication. Interestingly, out of the 10 subfamilies with the highest hub scores (ranging from 0.85 to 1) in the Hbr control network, eight belonged to members of the RLK-Pelle group (RLK-Pelle_L-LEC, RLK-Pelle_RLCK-V, RLK-Pelle_RLCK-VIIa-2, RLK-Pelle_LRR-I-1, RLK-Pelle_LysM, RLK-Pelle_SD-2b, RLK-Pelle_DLSV, and RLK-Pelle_RLCK-VIIa-1), while the others were CAMK-CDPK and CK1_CK1-Pl. In addition to being the most abundant and associated with TDEs, the importance of the RLK-Pelle group is evidenced by its putative interaction with several PK subfamilies. In contrast, under adverse conditions, the hub scores of the top 10 subfamilies varied from 0.75 to 1; only 6 of them were members of the RLK-Pelle group (RLK-Pelle_RLCK-X, RLK-Pelle_PERK-1, RLK-Pelle_LysM, RLK-Pelle_SD-2b, RLK-Pelle_L-LEC, RLK-Pelle_RLCK-VIIa-1), while the others were CAMK_CDPK, TKL-Pl-4, STE_STE7, and CK1_CK1-Pl.

Ultimately, we investigated network structural weaknesses by measuring edge betweenness centrality scores among kinase subfamily interactions (edges) within each network (Supplementary Table S19). The edges presenting elevated betweenness values (colored in light blue in Figure 3) indicate relationships sustained by few connections possibly related to a greater flow of interaction into the complex system modeled. Such connections represent specific interactions between PK subfamilies with a significant importance for the whole system communication, i.e. subfamilies acting in different biological processes through indirect implications. Overall, under adverse conditions, PK subfamilies tended to arrange into a less cohesive network architecture, as evidenced by a large number of scattered connections. Through betweenness measures, we observed that other influential subfamily pairs of the control network were RLK-Pelle_LRR-XII-1/RLK-Pelle_RLCK-XVI, RLK-Pelle_LRR-VII-3/RLK-Pelle_RLCK-XVI, and RLK-Pelle_CR4L/RLK-Pelle_LRR-Xb-1, in contrast to the ones found during abiotic stress situations: RLK-Pelle_RLCK-Os/TKL-Pl-6, CAMK_CAMK1-DCAMKL/RLK-Pelle_RLCK-Os, and RLK-Pelle_RLCK-V/TKL-Pl-6. These differences in the top betweenness connections demonstrate the distinct interactions of PK subfamilies during the stress response.

4 Discussion

4.1 Rubber tree kinome

In the last decades, an increasing number of initiatives have been established to produce a high-quality reference genome for the rubber

10.3389/fpls.2023.1068202

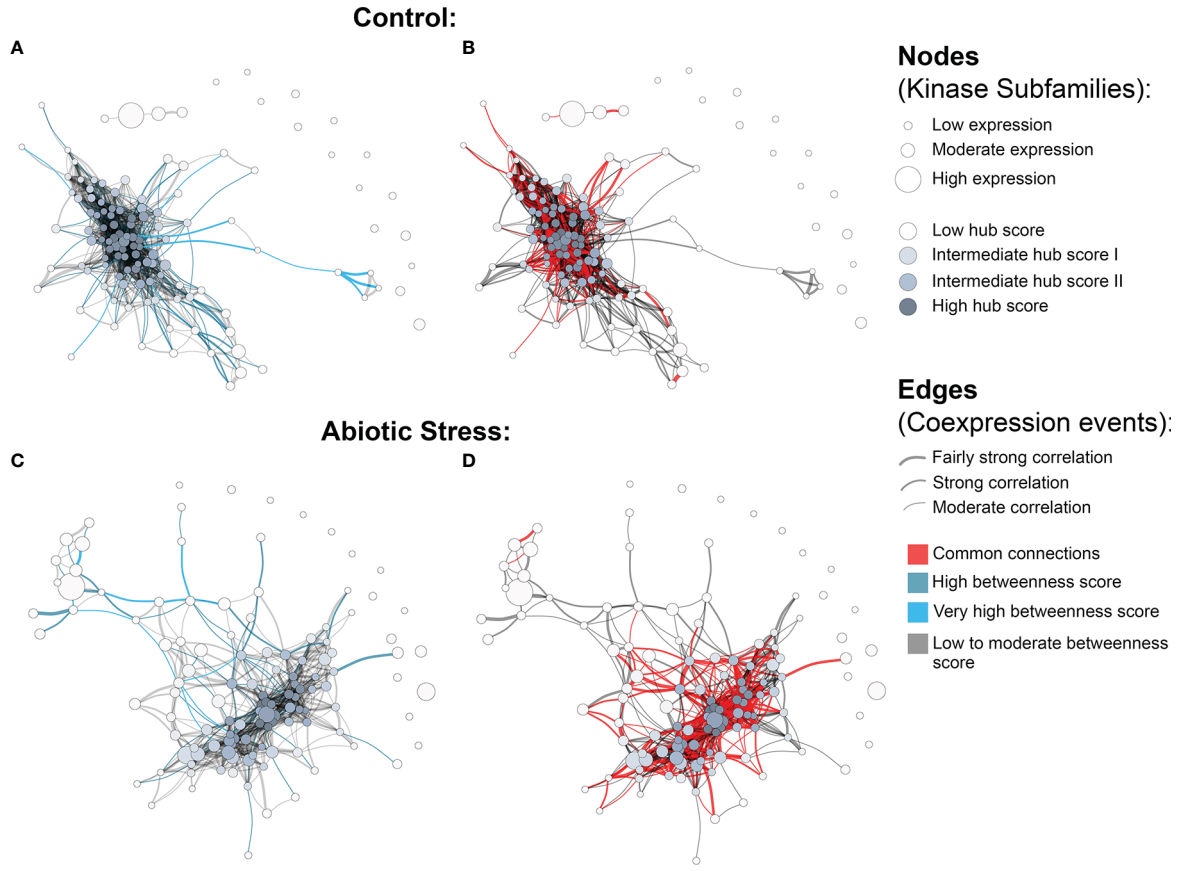


FIGURE 3
Coexpression networks for *H brasiliensis* (Hbr) kinase subfamilies. **(A)** Hbr control network with betweenness values highlighted in light blue. **(B)** Hbr control network indicating edge similarities (red) with the Hbr stress network. **(C)** Hbr abiotic-stress network with betweenness values highlighted in light blue. **(D)** Hbr abiotic stress network indicating edge similarities (red) with the Hbr control network."3 Coexpression networks for H brasiliensis (Hbr) kinase subfamilies.

tree (Rahman et al., 2013; Tang et al., 2016; Pootakham et al., 2017; Liu et al., 2020b). However, the high complexity of the Hbr genome introduces many challenges that have hampered the ability to obtain contiguous genomic sequences and complete gene annotation (English et al., 2012; Pootakham et al., 2017). Although the most recent version of the Hbr genome (Liu et al., 2020b) provided the assembly of contiguous sequences for chromosomes for the first time, there is still a lack of knowledge about its gene content and functional implications, highlighting the need for efforts to profile and fully characterize important protein families, such as PKs. Here, we established a combined approach to generate a comprehensive and diverse kinase database for Hbr. Joining two independent rubber tree genomic resources (Tang et al., 2016; Liu et al., 2020b) and comparing them with kinomes from two other members of the *Euphorbiaceae* family (Mes and Rco) enabled an in-depth investigation of the rubber tree kinome, supplying a large and reliable reservoir of data. We suggest that such a combination approach could be a valuable strategy for other species with limited genomic resources, increasing the identification and definition of important molecular elements.

Plant kinomes have been studied in several other species, including 942 members in A. *thaliana* (Zulawski et al., 2014), 2,166 in soybean (Liu et al., 2015), 1,168 in grapevine (Zhang, 2003), and 1,210 in sorghum (Aono et al., 2021). In this study, we identified 2,842 PKs in the rubber tree, a considerably larger size when compared to 1,531 in cassava and 863 in castor bean. Although the rubber tree possesses a large genome (1.47 Gb), which is nearly 3-4.5 times larger than the cassava (495 Mb) and castor bean (320 Mb) genomes (Chan et al., 2010; Bredeson et al., 2016; Tang et al., 2016), the large kinome size in Hbr resulted from the combination of two sources of PKs related to different rubber tree genotypes (Reyan7-33-97 and GT1) (Tang et al., 2016; Liu et al., 2020b). When analyzing the two Hbr PK sets separately, we found a much smaller number of PKs in each of them (1,809 and 1,379), placing the Hbr kinome in the range of other plant species. The discrepancy found between the two sources of data reinforces the differences in completeness among genome assemblies, which we believe that could potentially mislead further genomic investigations, especially considering the elevated heterozygosity levels and the high amount of repetitive elements in the rubber tree genome (Gouvêa et al., 2010; Lau et al., 2016), ˆ mainly caused by the demographic history and genetic diversity of the rubber tree. Additionally, a recent study showed that the number of PKs in sugarcane was significantly decreased on the allelic level when compared to those found for all allele copies of a given gene (Aono et al., 2021), demonstrating that the redundancy in PK datasets may contribute to the overestimation of kinome sizes.

Similar to the results of other kinome studies (Wei et al., 2014; Zulawski et al., 2014; Liu et al., 2015; Yan et al., 2018; Zhu et al., 2018a; Liu et al., 2020a; Aono et al., 2021; Ferreira-Neto et al., 2021), RLK-Pelle

was the most pronounced group in Hbr, Mes and Rco kinomes. Considering the diverse functions of PKs (Lehti-Shiu and Shiu, 2012), our study corroborates the remarkable role of this group in the Hbr response to stress. We also suggest that such importance transcends its high abundance, including its expansion through TDEs and also the way the RLK-Pelle subfamilies functionally interact with other PKs, as evidenced by the coexpression relationships found. One notable feature observed across Hbr PKs was the large diversity in domain configuration. Most PKs (56.9%) in rubber trees had two or more functional domains incorporated with them, similar to what has been observed in cassava (59.2%), castor bean (55%), soybean (56.5%) (Liu et al., 2015), and pineapple (50.7%) (Zhu et al., 2018a). By combining different protein properties, we inferred a considerable presence of extracellular domains (ECDs), mainly because of: (i) the large diversity of additional domains in PK genes; (ii) the detection of signal peptides and transmembrane regions; and (iii) the wide range of subcellular localizations predicted (Supplementary Figure S5). PKs combined with ECDs may broaden the scope of functionality within signaling networks by sensing new extracellular signals and their aggregation to existing response networks (Gish and Clark, 2011; Lehti-Shiu and Shiu, 2012), which makes our results a valuable source of data for establishing PKs sensing different environmental stimuli and signalling important metabolic mechanisms, such as the rubber biosynthesis.

Our comparative analyses of the PKs of the three *Euphorbiaceae* species revealed a high degree of similarity in their kinase subfamily compositions, protein characteristics and gene organization (Supplementary Figures S2–S4). Our integrative approach allowed us to corroborate the validity of the Hbr kinome, which was composed of different data sources and had evident resemblances with closely related phylogenetic species. However, Mes was more similar to Hbr in kinome size than Rco. This pattern of gene expansion within the *Euphorbiaceae* clade was also observed in other gene families, including the SWEET and SBP-box families (Cao et al., 2019; Li et al., 2019). Although the focus of our study was not to evaluate the PK evolutionary divergences between Hbr, Mes and Rco, the similarities that we observed between the kinomes are corroborated by the evolutionary history of such taxonomic groups. Phylogenetic studies indicated that Hevea and Manihot underwent a whole-genome duplication event before their divergence of approximately 36 million years ago (MYA), while the Ricinus lineage diverged from other Euphorbia members approximately 60 MYA (Bredeson et al., 2016; Shearman et al., 2020). In this sense, the increase in the size of the PK superfamily could be partially attributed to the expansion of several gene families through duplication events during the evolutionary history of these species, as already reported by Lehti-Shiu and Shiu (2012).

4.2 Duplication events in the rubber tree kinome

Our analysis suggested that segmental duplications mostly accounted for Hbr kinome expansion (Figure 2B), with ~58.9% of PKs displaying more than 75% compositional similarities. TDEs, on the other hand, seemed to contribute to the expansion of the PK superfamily to a lesser extent and were restricted to a few subfamilies (Supplementary Table S11), accounting for the generation of nearly 11.9% of PK genes in Hbr. As we observed the same functional profile in the subgroup of PKs tandemly duplicated and throughout the kinome, we infer that such TDEs might be important for the maintenance of PK activities throughout evolution. This TDE rate was within the range of what has been reported for other higher plants, such as 10.6% in soybean (Liu et al., 2015), 12.5% in pineapple (Zhu et al., 2018a), and 14.8% in strawberry (Liu et al., 2020a).

Tandemly duplicated kinases have been associated with stress responses (Freeling, 2009), which is of great interest for molecular breeding. Taking this association into account, we suggest an important role of RLK-Pelle_DLSV, RLK-Pelle_LRR-III, RLK-Pelle_LRR-XI-1, STE_STE11, and CAMK_CDPK subfamilies in stress signaling, which should be investigated in further studies. Additionally, in rubber tree research, different initiatives have brought to light the importance of PKs in the configuration and maintenance of economically important traits, including not only resistance to different types of stress (Duan et al., 2010; Venkatachalam et al., 2010; Jin et al., 2017; Mantello et al., 2019) but also plant performance in the field (Francisco et al., 2021; Bini et al., 2022). Together with these findings, recent contributions have pinpointed the role of TEs beyond Hbr genomic organization, suggesting a potential influence on the configuration of desirable rubber tree traits (Wu et al., 2020; Francisco et al., 2021). We also identified a considerable number of PKs (15.5%) associated with TEs, which we also suggest is related to the kinome expansion. In In Hbr, Wu et al. (2020) showed that TEs located in gene regulatory regions were involved in latex production through cis regulation, which would explain the differential gene expression among contrasting genotypes. The incidence of PKs close to TEs pinpoints the importance of such elements on PK functionality, as already demonstrated by other studies describing TE-mediated regulation in kinases (Zayed et al., 2007; Fan et al., 2019).

It has been well established that TEs are abundant in the rubber tree genome, and the proportion of TE types found in our study was similar to those found by other authors (Tang et al., 2016; Liu et al., 2020b; Wu et al., 2020). Due to the mutagenic potential of TEs caused by epigenetic mechanisms, these elements can alter regulatory networks and confer genetic adaptations, leading to important phenotypic variations (Lisch, 2013; Wei and Cao, 2016; Wu et al., 2020); this has currently received great attention in genetic improvement programs for several species (Lee et al., 2006; Domínguez et al., 2020; Wang et al., 2020). Similar to PKs, which are especially active during abiotic stress (Morris, 2001; Jaggi, 2018), TEs are related to plant adaptations throughout evolution (Naito et al., 2009; Casacuberta and González, 2013; Lisch, 2013; Negi et al., 2016; Dubin et al., 2018). We found an association between TEs and specific kinase subfamilies, such as those present in the RLK-Pelle group. As expected due to the occurrence of duplication events caused by TEs (Flagel and Wendel, 2009), we also observed an association of these elements with tandemly duplicated PKs, enabling the elucidation of diverse biological mechanisms favoring stress resistance.

Considering the growing demand for latex production, the interest in developing cold-resistant cultivars for expanding *Hevea* plantation motivates the comprehension of the molecular mechanisms associated with resistance. Such a scientific challenge is directly related to the signaling mechanisms of PKs, with the direct action of specific subfamilies. We believe that the association of TEs, TDEs and specific PK subfamilies provides deeper insights into the role and maintenance of PKs for stress resilience. In addition to the genetic adaptation of the rubber tree, TEs have increasingly been attracting attention, mainly due

to their association with (i) intron gain (Gozashti et al., 2022); (ii) increased heterozygosity in stress-responsive genes (De Kort et al., 2022); (iii) increases in gene expression variation (Uzunović et al., 2019); and (iv) important polymorphisms associated with the genetic effects of complex traits(Vourlaki et al., 2022). These facts, together with the association of TDEs and stress responses, enabled the establishment of important PK subfamilies with high potential to be investigated for climate resilience.

4.3 Gene expression evaluations

Differentially expressed gene (DEG) analyses are based on statistical tests performed on gene expression quantifications measured under certain conditions, contrasting physiological contexts and different stimuli, enabling the evaluation of increased gene expression (Casassola et al., 2013; Costa-Silva et al., 2017). Although we did not perform such an analysis of Hbr PKs due to the different experiments and datasets employed, it was possible to visualize a distinct overall expression profile for subfamily expression across the samples employed, illustrating putative molecular mechanisms adopted by PK subfamilies to overcome stress conditions (Mantello et al., 2019).

The subfamilies CAMP_AMPK, CMGC_PITthe, CK1_CK1, RLK-Pelle_LRR-XIIIb, and RLK-Pelle_URK-2 exhibited more pronounced expression in samples under stress conditions, as already reported in other studies (Hawley et al., 2005; Saito et al., 2019). Even without statistical evaluations, it is possible to infer an impact of stress conditions on the molecular mechanisms of these PK subfamilies. CAMP AMPK has been described as an important energy regulator in eukaryotes, coordinating metabolic activities in the cytosol with those in mitochondria and plastids, possibly allocating energy expenditure to overcome these adversities (Hawley et al., 2005; Suzuki et al., 2012; Roustan et al., 2016). Although members of the CK1 subfamily were highly conserved in eukaryotes and involved in various cellular, physiological and developmental processes, their functions in plant species are still poorly understood (Saito et al., 2019). Studies indicated that CK1 members in *A. thaliana* are involved in several processes related to the response to environmental stimuli, such as regulation of stomatal opening (Zhao et al., 2016), signaling in response to blue light (Tan et al., 2013), organization and dynamics of cortical microtubules (Ben-Nissan et al., 2008), and ethylene production (Tan and Xue, 2014). In this context, our study corroborates the putative role of the CK1 subfamily, highlighting its change in expression during stress response.

As a complementary approach to elucidate different patterns in the expression of PK subfamilies in control and abiotic stress-related samples, we employed gene coexpression networks. Through a graph representation of the PK subfamily interactions in these two groups of samples, we estimated coexpression patterns inherent to each network, inferring functional implications through the network topology. As the modeled coexpression networks represent the interaction between the PK subfamilies, we suggest that, by contrasting the topological differences between these networks, it is possible to infer changes in the interaction of the PK subfamilies induced by stress conditions. Even with a common set of interactions (Figures 3B, D), it is possible to note fewer associations in the network modeled with stress samples (a loss of ~40%). During stress conditions, PKs act through a signaling system to activate specific cellular responses to overcome these adversities, affecting molecular interactions. In the network modeled with samples affected by stress, it is possible to visualize such alterations, i.e. the cohesive set of interactions between the PK subfamilies becomes more sparse, evidencing that specific subfamilies start to perform different functions.

In a complex network structure, nodes with the largest number of connections (high degree) are called hubs, which are elements recognized as critical to network maintenance (Barabasi and Oltvai, 2004). Therefore, PK subfamilies with the highest hub scores are considered to be important regulators over the set of biological mechanisms affected by PKs (Barabasi and Oltvai, 2004; Azuaje, 2014; Van Dam et al., 2018), which provides additional insights into key mechanisms over PKs' action (Vandereyken et al., 2018).

In both networks modeled, we found that the CAMK CDPK subfamily had the highest hub score, suggesting the importance of calcium signals over Hbr PK activities, as already reported in soybean (Liu et al., 2015). Interestingly, members of the CAMK CDPK subfamily were found to be tandemly duplicated and also associated with TEs. Considering the previously described importance of TDEs and TEs in the Hbr kinome, we suggest that such a subfamily plays a key role in activating other PK subfamilies to overcome the stress response. Members of the RLK-Pelle group were also identified as hubs in both networks, reinforcing the primary and secondary metabolic functions of this group (Bolhassani et al., 2021). Additionally, TKL-Pl-4 was among the hubs in the stress-related network, corroborating the already described upregulation of members of this subfamily in stress conditions (Yan et al., 2017). Similar to CAMK CDPK, the TKL-Pl-4 subfamily was also related to TDEs and TEs, which shows the conservation of such a subfamily throughout evolution. Furthermore, under stress conditions, we observed that the TKL-Pl-4 subfamily starts to play a more prominent role in the network, probably indicating its signaling activity on other PK subfamilies in response to stress.

Another measure evaluated in the modeled networks was edge betweenness scores. In a complex network structure, edges with high betweenness indicate points of vulnerability in the network structure, i.e. connections that, if removed, have a larger probability of causing network separation. We suggest that in the networks modeled for PK subfamily interactions, the identification of such edges can supply indicators of subfamilies mediating a significant amount of mechanisms over a larger set of PKs. Additionally, the comparison of edges with high betweenness between networks can provide clues about the change of molecular mechanisms affected by stress. Interestingly, the two highest betweenness values for the control network (RLK-Pelle_LRR-XII-1/RLK-Pelle_RLCK-XVI and RLK-Pelle_LRR-VII-3/RLK-Pelle_RLCK-XVI edges) were disrupted in the stress-associated network. The RLK-Pelle_RLCK-XVI subfamily did not have any connection in the stress network, showing a change in the interaction of this subfamily under stress. We infer that such a change is caused by stress conditions, i.e. the putative functions of the RLK-Pelle_RLCK-XVI subfamily are related to biological processes that are directly affected by the external stimuli. Interestingly, RLCK members have already been shown to be related to plant growth and vegetative development (Gao and Xue, 2012; Yan et al., 2018), which is directly impacted by stress. We suggest that such a subfamily is an interesting target for understanding the impact of climate changes on Hbr.

Additionally, in the stress-related network, we found that the PK subfamily pairs RLK-Pelle_RLCK-Os—TKL-Pl-6, CAMK_CAMK1-

DCAMKL—RLK-Pelle_RLCK-Os, and RLK-Pelle_RLCK-V—TKL-Pl-6 had the largest betweenness scores. All these subfamilies presented a similar number of connections in the control network; however, they were not considered vulnerability points. This result indicated that under stress, established connections might become sensitive and cause network breaks in more adverse conditions, representing PK subfamilies that can have their functions altered under high stress, such as the RLK-Pelle_RLCK-XVI subfamily in the change from control to stress. Given the described downregulation of TKL-Pl-6 expression during stress (Yan et al., 2017), our results corroborate the impact of stress not only on its activity, but also on the way this subfamily interacts with other PK subfamilies. Additionally, CAMK CAMK1-DCAMKL has already been described as induced during stress (Liu et al., 2015), and such a fact can be observed in our network, with changes in the network configuration.

Given the importance of rubber trees, the rising demand for latex production, and the elevated complexity of the Hbr genome (Tang et al., 2016; Board, 2018), providing resources for understanding stress responses is of great interest for Hbr breeding programs (Priyadarshan and Goncalves, 2003). Our work provided a rich and large reservoir of data for Hbr research. In the first study to profile the complete set of PKs in Hbr, we combined different data sources to provide a wider PK characterization, taking advantage of the resources available and contrasting our results with two phylogenetically close species. From a set of 2,842 PKs classified into 20 groups and distributed along all Hbr chromosomes; our findings demonstrated the high diversity and scope of functionality of Hbr PKs. Additionally, we provided different insights across stress responses in rubber trees through the association of tandemly duplicated PKs, TEs, gene expression patterns, and coexpression events.

Data availability statement

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

Author contributions

LBS, AA and FF performed all analyses and wrote the manuscript. AS, CS and LMS conceived of the project. All authors reviewed, read and approved the manuscript.

Funding

This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Computational Biology Programme). LBS received an undergraduate fellowship from FAPESP (2019/19340-2); AA received a PhD fellowship from FAPESP (2019/03232-6); FF received a PhD fellowship from FAPESP (2018/18985-7); and AS received research fellowships from CNPq (312777/2018).

Acknowledgments

We would like to acknowledge the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

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.2023.1068202/full#supplementary-material

References

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

Andrews, S. (2010) *Fastqc: A quality control tool for high throughput sequence data. version 0.11. 2.* Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Aono, A. H., Pimenta, R. J. G., Garcia, A. L. B., Correr, F. H., Hosaka, G. K., Carrasco, M. M., et al. (2021). The wild sugarcane and sorghum kinomes: Insights into expansion, diversification, and expression patterns. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.668623

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). Signalp 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z

Azuaje, F. J. (2014). Selecting biologically informative genes in co-expression networks with a centrality score. *Biol. direct* 9, 1–23. doi: 10.1186/1745-6150-9-12

Barabasi, A.-L., and Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113. doi: 10.1038/nrg1272

Ben-Nissan, G., Cui, W., Kim, D.-J., Yang, Y., Yoo, B.-C., and Lee, J.-Y. (2008). Arabidopsis casein kinase 1-like 6 contains a microtubule-binding domain and affects the organization of cortical microtubules. *Plant Physiol.* 148, 1897–1907. doi: 10.1104/pp.108.129346

Bini, K., Saha, T., Radhakrishnan, S., Ravindran, M., and Uthup, T. K. (2022). Development of novel markers for yield in hevea brasiliensis muell. arg. based on candidate genes from biosynthetic pathways associated with latex production. *Biochem. Genet.* 1–29, 2171-2199. doi: 10.1007/s10528-022-10211-w

Board, M. R. (2018). *Natural rubber statistics 2018* (Kuala Lumpur: Malaysian Rubber Board).

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

Bolhassani, M., Niazi, A., Tahmasebi, A., and Moghadam, A. (2021). Identification of key genes associated with secondary metabolites biosynthesis by system network analysis in valeriana officinalis. *J. Plant Res.* 134, 625–639. doi: 10.1007/s10265-021-01277-5

Brandes, U. (2001). A faster algorithm for betweenness centrality. *J. Math. sociology* 25, 163–177. doi: 10.1080/0022250X.2001.9990249

Bredeson, J. V., Lyons, J. B., Prochnik, S. E., Wu, G. A., Ha, C. M., Edsinger-Gonzales, E., et al. (2016). Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat. Biotechnol.* 34, 562–570. doi: 10.1038/nbt.3535

Budzinski, I. G. F., Moon, D. H., Morosini, J. S., Lindén, P., Bragatto, J., Moritz, T., et al. (2016). Integrated analysis of gene expression from carbon metabolism, proteome and metabolome, reveals altered primary metabolism in eucalyptus grandis bark, in response to seasonal variation. *BMC Plant Biol.* 16, 1–15. doi: 10.1186/s12870-016-0839-8

Cao, Y., Liu, W., Zhao, Q., Long, H., Li, Z., Liu, M., et al. (2019). Integrative analysis reveals evolutionary patterns and potential functions of sweet transporters in euphorbiaceae. *Int. J. Biol. macromolecules* 139, 1–11. doi: 10.1016/j.ijbiomac.2019.07.102

Casacuberta, E., and González, J. (2013). The impact of transposable elements in environmental adaptation. *Mol. Ecol.* 22, 1503–1517. doi: 10.1111/mec.12170

Casassola, A., Brammer, S. P., Chaves, M. S., Martinelli, J. A., Grando, M. F., and Denardin, N. (2013). Gene expression: A review on methods for the study of defense-related gene differential expression in plants. *American Journal of Plant Sciences* 4 (12C), 64-73. doi: 10.4236/ajps.2013.412A3008

Chan, A. P., Crabtree, J., Zhao, Q., Lorenzi, H., Orvis, J., Puiu, D., et al. (2010). Draft genome sequence of the oilseed species ricinus communis. *Nat. Biotechnol.* 28, 951–956. doi: 10.1038/nbt.1674

Cheng, H., Chen, X., Fang, J., An, Z., Hu, Y., and Huang, H. (2018). Comparative transcriptome analysis reveals an early gene expression profile that contributes to cold resistance in hevea brasiliensis (the para rubber tree). *Tree Physiol.* 38, 1409–1423. doi: 10.1093/treephys/tpy014

Colcombet, J., and Hirt, H. (2008). Arabidopsis mapks: A complex signalling network involved in multiple biological processes. *Biochem. J.* 413, 217–226. doi: 10.1042/BJ20080625

Conesa, A., and Götz, S. (2008). Blast2go: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008, 619832-619844. doi: 10.1155/2008/619832

Consortium, U. (2019). Uniprot: A worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. doi: 10.1093/nar/gky1049

Costa-Silva, J., Domingues, D., and Lopes, F. M. (2017). Rna-seq differential expression analysis: An extended review and a software tool. *PloS One* 12, e0190152. doi: 10.1371/journal.pone.0190152

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Syst.* 1695, 1–9.

da Hora Júnior, B. T., de Macedo, D. M., Barreto, R. W., Evans, H. C., Mattos, C. R. R., Maffia, L. A., et al. (2014). Erasing the past: A new identity for the damoclean pathogen causing south american leaf blight of rubber. *PloS One* 9, e104750. doi: 10.1371/journal.pone.0104750

De Kort, H., Legrand, S., Honnay, O., and Buckley, J. (2022). Transposable elements maintain genome-wide heterozygosity in inbred populations. *Nat. Commun.* 13, 1–11. doi: 10.1038/s41467-022-34795-4

Deng, X., Wang, J., Li, Y., Wu, S., Yang, S., Chao, J., et al. (2018). Comparative transcriptome analysis reveals phytohormone signalings, heat shock module and ros scavenger mediate the cold-tolerance of rubber tree. *Sci. Rep.* 8, 1–16. doi: 10.1038/s41598-018-23094-y

De Souza, L. M., Dos Santos, L. H., Rosa, J. R., Da Silva, C. C., Mantello, C. C., Conson, A. R., et al. (2018). Linkage disequilibrium and population structure in wild and cultivated populations of rubber tree (hevea brasiliensis). *Front. Plant Sci.* 9, 815. doi: 10.3389/fpls.2018.00815

Devakumar, A., Prakash, P. G., Sathik, M., and Jacob, J. (1999). Drought alters the canopy architecture and micro-climate of hevea brasiliensis trees. *Trees* 13, 161–167. doi: 10.1007/PL00009747

Devakumar, A., Sathik, M. M., Sreelatha, S., Thapliyal, A., and Jacob, J. (2002). Photosynthesis in mature trees of hevea brasiliensis experiencing drought and cold stresses concomitant with high light in the field. *Indian J. Nat. Rubb. Res.* 15, 1–13.

Domínguez, M., Dugas, E., Benchouaia, M., Leduque, B., Jiménez-Gómez, J. M., Colot, V., et al. (2020). The impact of transposable elements on tomato diversity. *Nat. Commun.* 11, 1–11. doi: 10.1038/s41467-020-17874-2

do Prado Tanure, T. M., Miyajima, D. N., Magalhães, A. S., Domingues, E. P., and Carvalho, T. S. (2020). The impacts of climate change on agricultural production, land use and economy of the legal amazon region between 2030 and 2049. *EconomiA* 21, 73–90. doi: 10.1016/j.econ.2020.04.001

Duan, C., Rio, M., Leclercq, J., Bonnot, F., Oliver, G., and Montoro, P. (2010). Gene expression pattern in response to wounding, methyl jasmonate and ethylene in the bark of hevea brasiliensis. *Tree Physiol.* 30, 1349–1359. doi: 10.1093/treephys/tpq066

Dubin, M. J., Scheid, O. M., and Becker, C. (2018). Transposons: A blessing curse. *Curr. Opin. Plant Biol.* 42, 23–29. doi: 10.1016/j.pbi.2018.01.003

Edgar, R. C. (2004). Muscle: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995

English, A. C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., et al. (2012). Mind the gap: upgrading genomes with pacific biosciences rs long-read sequencing technology. *PloS One* 7, e47768. doi: 10.1371/journal.pone.0047768

Fan, Y., Bazai, S. K., Daian, F., Arechederra, M., Richelme, S., Temiz, N. A., et al. (2019). Evaluating the landscape of gene cooperativity with receptor tyrosine kinases in liver tumorigenesis using transposon-mediated mutagenesis. *J. Hepatol.* 70, 470–482. doi: 10.1016/j.jhep.2018.11.027

Ferreira-Neto, J. R. C., Borges, A. N., d., C., da Silva, M. D., de Lima Morais, D. A., Bezerra-Neto, J. P., et al. (2021). The cowpea kinome: Genomic and transcriptomic analysis under biotic and abiotic stresses. *Front. Plant Sci.* 945. doi: 10.3389/fpls.2021.667013

Finn, R. D., Clements, J., and Eddy, S. R. (2011). Hmmer web server: Interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367

Flagel, L. E., and Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytol.* 183, 557–564. doi: 10.1111/j.1469-8137.2009.02923.x

Francisco, F. R., Aono, A. H., Da Silva, C. C., Gonçalves, P. S., Junior, E. J. S., Le Guen, V., et al. (2021). Unravelling rubber tree growth by integrating gwas and biological network-based approaches. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.768589

Freeling, M. (2009). Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* 60, 433–453. doi: 10.1146/annurev.arplant.043008.092122

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). Cd-hit: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565

Gao, L.-L., and Xue, H.-W. (2012). Global analysis of expression profiles of rice receptor-like kinase genes. *Mol. Plant* 5, 143–153. doi: 10.1093/mp/ssr062

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003). Expasy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31, 3784–3788. doi: 10.1093/nar/gkg563

Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., et al. (2010). The ncbi biosystems database. *Nucleic Acids Res.* 38, D492–D496. doi: 10.1093/nar/gkp858

Gish, L. A., and Clark, S. E. (2011). The rlk/pelle family of kinases. *Plant J.* 66, 117–127. doi: 10.1111/j.1365-313X.2011.04518.x

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944

ouvêa, L. R. L., Rubiano, L. B., Chioratto, A. F., Zucchi, M. I., and de Souza Gonçalves, P. (2010). Genetic divergence of rubber tree estimated by multivariate techniques and microsatellite markers. *Genet. Mol. Biol.* 33, 308–318. doi: 10.1590/S1415-47572010005000039

Gozashti, L., Roy, S. W., Thornlow, B., Kramer, A., Ares, M.Jr., and Corbett-Detig, R. (2022). Transposable elements drive intron gain in diverse eukaryotes. *Proc. Natl. Acad. Sci.* 119, e2209766119. doi: 10.1073/pnas.2209766119

Guo, D., Li, H.-L., Zhu, J.-H., Wang, Y., An, F., Xie, G.-S., et al. (2017). Genome-wide identification, characterization, and expression analysis of snrk2 family in hevea brasiliensis. *Tree Genet. Genomes* 13, 1–12. doi: 10.1007/s11295-017-1168-2

Hawley, S. A., Pan, D. A., Mustard, K. J., Ross, L., Bain, J., Edelman, A. M., et al. (2005). Calmodulin-dependent protein kinase kinase-β is an alternative upstream kinase for amp-activated protein kinase. *Cell Metab.* 2, 9–19. doi: 10.1016/j.cmet.2005.05.009

Hoa, T., Tuy, L., Duong, P., Phuc, L., and Truong, V. (1998). "Selection of hevea clones for the 1998–2000 planting recommendation in vietnam," in Proc. IRRDB Symposium on Natural Rubber (Hertford, UK: IRRDB , 1998), Vol. 1. 164–177.

Ihaka, R., and Gentleman, R. (1996). R: A language for data analysis and graphics. *J. Comput. graphical Stat* 5, 299–314.

Jaggi, M. (2018). "Recent advancement on map kinase cascade in biotic stress," in *Molecular aspects of plant-pathogen interaction* (Singapore: Springer), 139–158. doi: 10.1007/978-981-10-7371-7_6

Jin, X., Zhu, L., Yao, Q., Meng, X., Ding, G., Wang, D., et al. (2017). Expression profiling of mitogen-activated protein kinase genes reveals their evolutionary and functional diversity in different rubber tree (hevea brasiliensis) cultivars. *Genes* 8, 261. doi: 10.3390/genes8100261

Kleinberg, J. M. (1999). Hubs, authorities, and communities. *ACM computing surveys (CSUR)* 31, 5–es. doi: 10.1145/345966.345982

Kolde, R., and Kolde, M. R. (2015). *Package 'pheatmap'. r package*, Vol. 1. 790. Available at: https://cran.r-project.org/web/packages/pheatmap/index.html.

Kovtun, Y., Chiu, W.-L., Tena, G., and Sheen, J. (2000). Functional analysis of oxidative stress-activated mitogen-activated protein kinase cascade in plants. *Proc. Natl. Acad. Sci.* 97, 2940–2945. doi: 10.1073/pnas.97.6.2940

Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315

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

Kunjet, S., Thaler, P., Gay, F., Chuntuma, P., Sangkhasila, K., and Kasemsap, P. (2013). Effects of drought and tapping for latex production on water relations of hevea brasiliensis trees. *Agric. Natural Resour.* 47, 506–515.

Kuruvilla, L., Sathik, M. M., Thomas, M., Luke, L. P., and Sumesh, K. (2017). Identification and validation of cold responsive micrornas of hevea brasiliensis using high throughput sequencing. *J. Crop Sci. Biotechnol.* 20, 369–377. doi: 10.1007/s12892-017-0062-0

Lau, N.-S., Makita, Y., Kawashima, M., Taylor, T. D., Kondo, S., Othman, A. S., et al. (2016). The rubber tree genome shows expansion of gene family associated with rubber biosynthesis. *Sci. Rep.* 6, 1–14. doi: 10.1038/srep28594

Lee, J.-K., Park, J.-Y., Kim, J.-H., Kwon, S.-J., Shin, J.-H., Hong, S.-K., et al. (2006). Genetic mapping of the isaac-cacta transposon in maize. *Theor. Appl. Genet.* 113, 16–22. doi: 10.1007/s00122-006-0263-9

Lehti-Shiu, M. D., and Shiu, S.-H. (2012). Diversity, classification and function of the plant protein kinase superfamily. *Philos. Trans. R. Soc. B: Biol. Sci.* 367, 2619–2639. doi: 10.1098/rstb.2012.0003

Leinonen, R., Sugawara, H., Shumway, M.Collaboration, I. N. S. D (2010). The sequence read archive. *Nucleic Acids Res.* 39, D19–D21. doi: 10.1093/nar/gkq1019

Li, J., Gao, X., Sang, S., and Liu, C. (2019). Genome-wide identification, phylogeny, and expression analysis of the sbp-box gene family in euphorbiaceae. *BMC Genomics* 20, 1–15. doi: 10.1186/s12864-019-6319-4

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

Liu, J., Chen, N., Grant, J. N., Cheng, Z.-M., Stewart, C. N.Jr., and Hewezi, T. (2015). Soybean kinome: functional classification and gene expression patterns. *J. Exp. Bot.* 66, 1919–1934. doi: 10.1093/jxb/eru537

Liu, H., Qu, W., Zhu, K., and Cheng, Z.-M. M. (2020a). The wild strawberry kinome: identification, classification and transcript profiling of protein kinases during development and in response to gray mold infection. *BMC Genomics* 21, 1–14. doi: 10.1186/s12864-020-07053-4

Liu, J., Shi, C., Shi, C.-C., Li, W., Zhang, Q.-J., Zhang, Y., et al. (2020b). The chromosome-based rubber tree genome provides new insights into spurge genome evolution and rubber biosynthesis. *Mol. Plant* 13, 336–350. doi: 10.1016/j.molp.2019.10.017

Li, D., Zeng, R., Li, Y., Zhao, M., Chao, J., Li, Y., et al. (2016). Gene expression analysis and snp/indel discovery to investigate yield heterosis of two rubber tree f1 hybrids. *Sci. Rep.* 6, 1–12. doi: 10.1038/srep24984

Mantello, C. C., Boatwright, L., da Silva, C. C., Scaloppi, E. J., de Souza Goncalves, P., Barbazuk, W. B., et al. (2019). Deep expression analysis reveals distinct cold-response strategies in rubber tree (hevea brasiliensis). *BMC Genomics* 20, 1–20. doi: 10.1186/s12864-019-5852-5

Marengo, J. A., Souza, C. M.Jr., Thonicke, K., Burton, C., Halladay, K., Betts, R. A., et al. (2018). Changes in climate and land use over the amazon region: Current and future variability and trends. *Front. Earth Sci.* 6, 228. doi: 10.3389/feart.2018.00228

Meti, S., Meenattoor, J., Mondal, G., and Chaudhuri, D. (2003). Impact of cold weather condition on the growth of hevea brasiliensis clones in northern west bengal. *Indian J. Natural Rubber Res.* 16, 53–59.

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2011). "The cipres science gateway: A community resource for phylogenetic analyses," in Proceedings of the 2011 TeraGrid Conference: extreme digital discovery (Salt Lake City, Utah, ACM New York, NY, USA). 1–8.

Montoro, P., Wu, S., Favreau, B., Herlinawati, E., Labrune, C., Martin-Magniette, M.-L., et al. (2018). Transcriptome analysis in hevea brasiliensis latex revealed changes in hormone signalling pathways during ethephon stimulation and consequent tapping panel dryness. *Sci. Rep.* 8, 1–12. doi: 10.1038/s41598-018-26854-y

Morris, P. C. (2001). Map kinase signal transduction pathways in plants. *New Phytol.* 151, 67–89. doi: 10.1046/j.1469-8137.2001.00167.x

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

Negi, P., Rai, A. N., and Suprasanna, P. (2016). Moving through the stressed genome: Emerging regulatory roles for transposons in plant stress response. *Front. Plant Sci.* 7, 1448. doi: 10.3389/fpls.2016.01448

Pandita, D., and Wani, S. H. (2021). "Osmosensing and signalling in plants: Potential role in crop improvement under climate change," in *Compatible solutes engineering for crop plants facing climate change* (Cham: Springer), 11–46. doi: 10.1007/978-3-030-80674-3_2

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197

Pedro, D. L. F., Lorenzetti, A. P. R., Domingues, D. S., and Paschoal, A. R. (2018). Planc-te: a comprehensive knowledgebase of non-coding rnas and transposable elements in plants. *Database* 2018, 1-7. doi: 10.1093/database/bay078

Pootakham, W., Sonthirod, C., Naktang, C., Ruang-Areerate, P., Yoocha, T., Sangsrakru, D., et al. (2017). *De novo* hybrid assembly of the rubber tree genome reveals evidence of paleotetraploidy in hevea species. *Sci. Rep.* 7, 1–15. doi: 10.1038/srep41457

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). Fasttree 2–approximately maximum-likelihood trees for large alignments. *PloS One* 5, e9490. doi: 10.1371/journal.pone.0009490

Priyadarshan, P. (2017). Refinements to hevea rubber breeding. *Tree Genet. Genomes* 13, 1–17. doi: 10.1007/s11295-017-1101-8

Priyadarshan, P., and Goncalves, P. (2003). Hevea gene pool for breeding. *Genet. Resour. Crop Evol.* 50, 101–114. doi: 10.1023/A:1022972320696

Pushparajah, E. (1983). Problems and potentials for establishing hevea under difficult environmental conditions. *Planter* 50, 242-251.

Rahman, S. N. A., Bakar, M. F. A., Singham, G. V., and Othman, A. S. (2019). Single-nucleotide polymorphism markers within mva and mep pathways among hevea brasiliensis clones through transcriptomic analysis. *3 Biotech.* 9, 1–10. doi: 10.1007/s13205-019-1921-3

Rahman, A. Y. A., Usharraj, A. O., Misra, B. B., Thottathil, G. P., Jayasekaran, K., Feng, Y., et al. (2013). Draft genome sequence of the rubber tree hevea brasiliensis. *BMC Genomics* 14, 1–15. doi: 10.1186/1471-2164-14-75

Roustan, V., Jain, A., Teige, M., Ebersberger, I., and Weckwerth, W. (2016). An evolutionary perspective of ampk–tor signaling in the three domains of life. *J. Exp. Bot.* 67, 3897–3907. doi: 10.1093/jxb/erw211

Saito, A. N., Matsuo, H., Kuwata, K., Ono, A., Kinoshita, T., Yamaguchi, J., et al. (2019). Structure–function study of a novel inhibitor of the casein kinase 1 family in arabidopsis thaliana. *Plant direct* 3, e00172. doi: 10.1002/pld3.172

Sathik, M. M., Luke, L. P., Rajamani, A., Kuruvilla, L., Sumesh, K., and Thomas, M. (2018). *De novo* transcriptome analysis of abiotic stress-responsive transcripts of hevea brasiliensis. *Mol. Breed.* 38, 1–17. doi: 10.1007/s11032-018-0782-5

Sevillano, L., Sanchez-Ballesta, M. T., Romojaro, F., and Flores, F. B. (2009). Physiological, hormonal and molecular mechanisms regulating chilling injury in horticultural species. postharvest technologies applied to reduce its impact. *J. Sci. Food Agric.* 89, 555–573. doi: 10.1002/jsfa.3468

Shearman, J. R., Pootakham, W., and Tangphatsornruang, S. (2020). The bpm 24 rubber tree genome, organellar genomes and synteny within the family euphorbiaceae. *Rubber Tree Genome* 55, 55–66. doi: 10.1007/978-3-030-42258-5_4

Sperschneider, J., Catanzariti, A.-M., DeBoer, K., Petre, B., Gardiner, D. M., Singh, K. B., et al. (2017). Localizer: subcellular localization prediction of both plant and effector proteins in the plant cell. *Sci. Rep.* 7, 1–14. doi: 10.1038/srep44598

Suzuki, N., Koussevitzky, S., Mittler, R., and Miller, G. (2012). Ros and redox signalling in the response of plants to abiotic stress. *Plant Cell Environ.* 35, 259–270. doi: 10.1111/j.1365-3040.2011.02336.x

Tan, D., Hu, X., Fu, L., Kumpeangkeaw, A., Ding, Z., Sun, X., et al. (2017). Comparative morphology and transcriptome analysis reveals distinct functions of the primary and secondary laticifer cells in the rubber tree. *Sci. Rep.* 7, 1–17. doi: 10.1038/s41598-017-03083-3

Tan, S.-T., and Xue, H.-W. (2014). Casein kinase 1 regulates ethylene synthesis by phosphorylating and promoting the turnover of acs5. *Cell Rep.* 9, 1692–1702. doi: 10.1016/j.celrep.2014.10.047

Tan, S.-T., Dai, C., Liu, H.-T., and Xue, H.-W. (2013). Arabidopsis casein kinase1 proteins ck1. 3 and ck1. 4 phosphorylate cryptochrome2 to regulate blue light signaling. *Plant Cell* 25, 2618–2632. doi: 10.1105/tpc.113.114322

Tang, C., Yang, M., Fang, Y., Luo, Y., Gao, S., Xiao, X., et al. (2016). The rubber tree genome reveals new insights into rubber production and species adaptation. *Nat. Plants* 2, 1–10. doi: 10.1038/nplants.2016.73

Uzunović, J., Josephs, E. B., Stinchcombe, J. R., and Wright, S. I. (2019). Transposable elements are important contributors to standing variation in gene expression in capsella grandiflora. *Mol. Biol. Evol.* 36, 1734–1745. doi: 10.1093/molbev/msz098

Van Dam, S., Vosa, U., van der Graaf, A., Franke, L., and de Magalhaes, J. P. (2018). Gene co-expression analysis for functional classification and gene–disease predictions. *Briefings Bioinf.* 19, 575–592. doi: 10.1093/bib/bbw139

Vandereyken, K., Van Leene, J., De Coninck, B., and Cammue, B. P. (2018). Hub protein controversy: taking a closer look at plant stress response hubs. *Front. Plant Sci.* 9, 694. doi: 10.3389/fpls.2018.00694

Venkatachalam, P., Geetha, N., Priya, P., and Thulaseedharan, A. (2010). Identification of a differentially expressed thymidine kinase gene related to tapping panel dryness syndrome in the rubber tree (hevea brasiliensis muell. arg.) by random amplified polymorphic dna screening. *Int. J. Plant Biol.* 1, e7. doi: 10.4081/pb.2010.e7

Voorrips, R. (2002). Mapchart: Software for the graphical presentation of linkage maps and qtls. *J. heredity* 93, 77–78. doi: 10.1093/jhered/93.1.77

Vourlaki, I.-T., Castanera, R., Ramos-Onsins, S. E., Casacuberta, J. M., and Pérez-Enciso, M. (2022). Transposable element polymorphisms improve prediction of complex agronomic traits in rice. *Theor. Appl. Genet.* 135, 3211–3222. doi: 10.1007/s00122-022-04180-2

Wang, J., Lu, N., Yi, F., and Xiao, Y. (2020). Identification of transposable elements in conifer and their potential application in breeding. *Evolutionary Bioinf.* 16, 1176934320930263. doi: 10.1177/1176934320930263

Wei, L., and Cao, X. (2016). The effect of transposable elements on phenotypic variation: Insights from plants to humans. *Sci. China Life Sci.* 59, 24–37. doi: 10.1007/s11427-015-4993-2

Wei, K., Wang, Y., and Xie, D. (2014). Identification and expression profile analysis of the protein kinase gene superfamily in maize development. *Mol. Breed.* 33, 155–172. doi: 10.1007/s11032-013-9941-x

Wickham, H. (2009). *Ggplot2: Elegant Graphics for Data Analysis. 2nd Edition* (New York, NY: Springer) 2, 1–189. doi: 10.1007/978-0-387-98141-98143

Wu, S., Guyot, R., Bocs, S., Droc, G., Oktavia, F., Hu, S., et al. (2020). Structural and functional annotation of transposable elements revealed a potential regulation of genes involved in rubber biosynthesis by te-derived sirna interference in hevea brasiliensis. *Int. J. Mol. Sci.* 21, 4220. doi: 10.3390/ijms21124220

Xiao, X.-H., Yang, M., Sui, J.-L., Qi, J.-Y., Fang, Y.-J., Hu, S.-N., et al. (2017). The calcium-dependent protein kinase (cdpk) and cdpk-related kinase gene families in hevea brasiliensisâ€"comparison with five other plant species in structure, evolution, and expression. *FEBS Open Bio* 7, 4–24. doi: 10.1002/2211-5463.12163

Yan, J., Li, G., Guo, X., Li, Y., and Cao, X. (2018). Genome-wide classification, evolutionary analysis and gene expression patterns of the kinome in gossypium. *PloS One* 13, e0197392. doi: 10.1371/journal.pone.0197392

Yan, J., Su, P., Wei, Z., Nevo, E., and Kong, L. (2017). Genome-wide identification, classification, evolutionary analysis and gene expression patterns of the protein kinase gene family in wheat and aegilops tauschii. *Plant Mol. Biol.* 95, 227–242. doi: 10.1007/s11103-017-0637-1

Yang, X., Blagodatsky, S., Marohn, C., Liu, H., Golbon, R., Xu, J., et al. (2019). Climbing the mountain fast but smart: Modelling rubber tree growth and latex yield under climate change. *For. Ecol. Manage.* 439, 55–69. doi: 10.1016/j.foreco.2019.02.028

Yu, C.-S., Chen, Y.-C., Lu, C.-H., and Hwang, J.-K. (2006). Prediction of protein subcellular localization. *Proteins: Structure Function Bioinf.* 64, 643–651. doi: 10.1002/prot.21018

Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T.-Y. (2017). Ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8, 28–36. doi: 10.1111/2041-210X.12628

Zayed, H., Xia, L., Yerich, A., Yant, S. R., Kay, M. A., Puttaraju, M., et al. (2007). Correction of dna protein kinase deficiency by spliceosome-mediated rna trans-splicing and sleeping beauty transposon delivery. *Mol. Ther.* 15, 1273–1279. doi: 10.1038/sj.mt.6300178

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

Zhao, S., Jiang, Y., Zhao, Y., Huang, S., Yuan, M., Zhao, Y., et al. (2016). Casein kinase1-like protein2 regulates actin filament stability and stomatal closure *via* phosphorylation of actin depolymerizing factor. *Plant Cell* 28, 1422–1439. doi: 10.1105/tpc.16.00078

Zhu, L., Jin, X., Xie, Q., Yao, Q., Wang, X., and Li, H. (2018b). Calcium-dependent protein kinase family genes involved in ethylene-induced natural rubber production in different hevea brasiliensis cultivars. *Int. J. Mol. Sci.* 19, 947. doi: 10.3390/ijms19040947

Zhu, K., Liu, H., Chen, X., Cheng, Q., and Cheng, Z.-M. M. (2018a). The kinome of pineapple: catalog and insights into functions in crassulacean acid metabolism plants. *BMC Plant Biol.* 18, 1–16. doi: 10.1186/s12870-018-1389-z

Zomer, R. J., Trabucco, A., Wang, M., Lang, R., Chen, H., Metzger, M. J., et al. (2014). Environmental stratification to model climate change impacts on biodiversity and rubber production in xishuangbanna, yunnan, china. *Biol. Conserv.* 170, 264–273. doi: 10.1016/j.biocon.2013.11.028

Zulawski, M., Schulze, G., Braginets, R., Hartmann, S., and Schulze, W. X. (2014). The arabidopsis kinome: Phylogeny and evolutionary insights into functional diversification. *BMC Genomics* 15, 1–15. doi: 10.1186/1471-2164-15-548

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 09 February 2023
DOI 10.3389/fpls.2023.1108588

OPEN ACCESS

EDITED BY
Mehanathan Muthamilarasan,
University of Hyderabad, India

REVIEWED BY
Zhanqi Wang,
Huzhou University, China
Roshan Singh,
National Institute of Plant Genome Research (NIPGR), India
Pooja Choudhary,
Jaypee Institute of Information Technology, India

*CORRESPONDENCE
Daping Gong
gongdaping@caas.cn
Quan Sun
sunquan@cqupt.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 26 November 2022
ACCEPTED 30 January 2023
PUBLISHED 09 February 2023

CITATION
Gu L, Hou B, Chen X, Wang Y, Chang P, He X, Gong D and Sun Q (2023) The Bcl-2-associated athanogene gene family in tobacco (*Nicotiana tabacum*) and the function of *NtBAG5* in leaf senescence. *Front. Plant Sci.* 14:1108588. doi: 10.3389/fpls.2023.1108588

The Bcl-2-associated athanogene gene family in tobacco (*Nicotiana tabacum*) and the function of *NtBAG5* in leaf senescence

Linxin Gu[1†], Bing Hou[1†], Xiao Chen[1], Yu Wang[1], Pingan Chang[1], Xiaohong He[1], Daping Gong[2*] and Quan Sun[1*]

[1]Chongqing Key Laboratory of Big Data for Bio Intelligence, College of Bioinformation, Chongqing University of Posts and Telecommunications, Nan'an, Chongqing, China, [2]Tobacco Research Institute, Chinese Academy of Agricultural Sciences, Qingdao, China

Leaf senescence in tobacco is closely related to leaf maturation and secondary metabolites. Bcl-2-associated athanogene (BAG) family members are highly conserved proteins and play key roles in senescence, growth and development, and resistance to biotic and abiotic stresses. Herein, the BAG family of tobacco was identified and characterized. In total, 19 tobacco BAG protein candidate genes were identified and divided into two classes, class I comprising *NtBAG1a–e*, *NtBAG3a–b*, and *NtBAG4a–c* and class II including *NtBAG5a–e*, *NtBAG6a–b*, and *NtBAG7*. Genes in the same subfamily or branch of the phylogenetic tree exhibited similarities in gene structure and the *cis*-element on promoters. RNA-seq and real-time quantitative reverse transcription polymerase chain reaction (qRT-PCR) revealed that the expression of *NtBAG5c–f* and *NtBAG6a–b* was upregulated in senescent leaves, implying that they play a role in regulating leaf senescence. *NtBAG5c* was localized in the nucleus and cell wall as a homology of leaf senescence related gene *AtBAG5*. Further, the interaction of NtBAG5c with heat-shock protein 70 (HSP70) and sHSP20 was demonstrated using yeast two-hybrid experiment. Virus-induced gene silencing indicated that *NtBAG5c* reduced the lignin content and increased superoxide dismutase (SOD) activity and hydrogen peroxide (H_2O_2) accumulation. In *NtBAG5c*-silenced plants, the expression of multiple senescence-related genes cysteine proteinase (*NtCP1)*, *SENESCENCE 4* (*SEN4*) and *SENESCENCE-ASSOCIATED GENE 12* (*SAG12*) was downregulated. In conclusion, tobacco BAG protein candidate genes were identified and characterized for the first time.

KEYWORDS

tobacco, BAG protein, leaf senescence, *Nicotiana tabacum*, Bcl-2-associated athanogene

Introduction

Leaf senescence is a common process during plant growth and development and is regulated by variety of intrinsic factors, such as cell death, hormones, physiological and biochemical metabolism, senescence-related gene regulation, and environmental factors (Jahan et al., 2021). Leaf senescence is usually accompanied with a visible color change from green to yellow or brown (Koyama, 2018). The internal structure of the leaf undergoes significant changes during senescence, manifested by an abnormal chloroplast structure and decreased in chlorophyll content (Jahan et al., 2021). Decreased of nucleic acid and protein contents and cytoprotective enzyme activity as well as enhanced membrane lipid peroxidation are also observed during plant leaf senescence. The expression of numerous genes, such as senescence-associated genes (SAGs), is upregulated at the onset of senescence, whereas that of photosynthesis- and chlorophyll biosynthesis-related genes is downregulated (Hortensteiner, 2006; Li et al., 2020). Leaf senescence also causes the excessive accumulation of reactive oxygen species (ROS) (Wu et al., 2012; Gutle et al., 2016). In Arabidopsis, ROS are considered as signaling molecules during leaf senescence (Cui et al., 2013).

The Bcl-2-associated athanogene (BAG) protein family acts as a cochaperone that participates in various cellular processes, including stress responses, proliferation, migration, and cell death (Takayama and Reed, 2001; Doong et al., 2002; Kabbage and Dickman, 2008). The first BAG protein, BAG-1, was identified using the protein interacting cloning technique (Takayama et al., 1995). To date, six members of the BAG family have been identified in humans (Song et al., 2001; Takayama and Reed, 2001; Doukhanina et al., 2006; Anderson et al., 2010; Bruchmann et al., 2013). BAG family genes are evolutionarily conserved, with homologs found from yeast to animals. BAG proteins in plants are similar to those in animals and are mainly involved in cell development and apoptotic cell death. However, unlike their homologs in animals, there has been limited study on plant BAG proteins, although the first plant BAG protein was identified in *Arabidopsis* in 2006 (Doukhanina et al., 2006). Subsequently, seven additional members of the *Arabidopsis* BAG family were identified (Fang et al., 2013). Four of these family members (*AtBAG1–4*) encode domain structures outside of the BAG domain (BD) that are similar to those in animal BAGs (Kabbage et al., 2017). AtBAG5–7 is characterized by a tightly distributed calmodulin-binding motif (IQ motif) upstream of the BD. This IQ motif is a specific calmodulin-binding site, which is unique to plant BAG proteins (Kabbage and Dickman, 2008; Kabbage et al., 2016; Kabbage et al., 2017).

The BAG protein family is characterized by the presence of a conserved C-terminal BAG structural domain, which interacts with the ATPase structural domain of heat-shock protein 70 (HSP70/HSC70) to regulate the activity of this chaperone protein (Wang et al., 1996; Takayama et al., 1999; Kabbage and Dickman, 2008). To date, it has been confirmed that BAG proteins can interact with HSP70. In *Arabidopsis*, BAG proteins act as a cofactor in Hsp70-mediated proteasomal degradation of unimported plastid proteins or in plant growth and development (Fang et al., 2013; Lee et al., 2016). AtBAG4 interacts with SNF7.1, NBR1, and Hsp70 and is involved in abiotic stress tolerance (Doukhanina et al., 2006; Arabidopsis Interactome Mapping, Consortium 2011). The CaM/BAG5/Hsc70 complex in *Arabidopsis* regulates leaf senescence (Li et al., 2016a). In addition, AtBAG6 and AtBAG7 may play roles in disease resistance, autophagy, and heat and cold tolerance through the potential interactor C2 GRAM domain protein and an aspartyl protease (Li et al., 2016b), immunoglobulin-binding protein 2 (Williams et al., 2010), WRKY29, and small ubiquitin-like modifier (Li et al., 2017b).

In tobacco, one of the oldest model plants, leaf senescence is associated with the maturation of leaves and rapid formation of secondary metabolites. To date, studies on the senescence in plant leaves have been mainly performed in the model plant *Arabidopsis* and various cash crops, such as rice, wheat, and cotton. However, to the best of our knowledge, there have been no systematic studies on BAG proteins in tobacco, an important protein family with regard to leaf senescence. To bridge this research gap, we identified *BAG* gene family members in tobacco for the first time and analyzed their phylogenetic relationships, gene structures, and chromosomal locations. Moreover, we determined the expression profile of *NtBAG* genes and the protein interactions of candidate NtBAGs. These results obtained here may provide valuable information for exploring the molecular mechanisms of leaf senescence.

Materials and methods

Plant materials and growth conditions

The tobacco (*Nicotiana tabacum*) cultivar K326 was used to analyze gene expression patterns. Seeds were sowed into 2:1 vermiculite:turfy soil and cultured at a constant temperature of 22° C under a 16/8 h light/dark cycle in a culture room. All materials were immediately frozen in liquid nitrogen and stored at −70°C until RNA isolation.

Identification of BAG proteins in tobacco

The genome sequences of *Nicotiana tabacum* (version Nitab4.5) were downloaded from the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA376174) or Sol Genomics Networ (SGN) (web: https://solgenomics.net/organism/Nicotiana_tabacum/genome and ftp://ftp.sgn.cornell.edu/genomes/Nicotiana_tabacum/edwards_et_al_2017/) (Sierro et al., 2014; Edwards et al., 2017). In addition, the BD sequence was downloaded from Pfam database (accession no. PF02179.17) (Finn et al., 2010) and used as a query for a search in the *Nicotiana tabacum* protein data using HMMER 3.0 with an E-value of<$1e^{-5}$ (Finn et al., 2011). To confirm the results obtained using the HMMER algorithm, the BD sequence was further verified using Pfam and Smart databases (Finn et al., 2010; Letunic et al., 2015; Letunic and Bork, 2018). The BAG protein sequences of *Arabidopsis thaliana* were downloaded from the *Arabidopsis* information resource website (https://www.arabidopsis.org).

Three-dimensional structural analysis

The structures of NtBAG proteins were predicted using Iterative Threading Assembly Refinement (I-TASSER) server (http://zhanglab.

ccmb.med.umich.edu/I-TASSER) (Yang and Zhang, 2015). This server is an online resource for automated prediction of protein structures and structure-based functional annotation. In I-TASSER server, structural templates are first identified from the Protein Data Bank using the multithreaded alignment method, after which full-length structural models are constructed *via* iterative fragment assembly simulations. Considering the confidence level (C-score) of the prediction model, we obtained five simulated structures for each of the provided NtBAG protein sequences. For each sequence, the predicted result with the highest C-score value was selected for further analysis.

Sequence and phylogenetic analysis

The subcellular localization of proteins was predicted using Plant-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/#) (Chou and Shen, 2010). We used the ClustalW to perform multiple sequence alignments of BAG proteins obtained from *N. tabacum* and *A. thaliana* (Thompson et al., 1997). A phylogenetic tree was constructed using MEGA 7.0 software and the maximum likelihood method based on the Poisson correction model, with 1000 bootstrap replicates (Tamura et al., 2013). A schematic of the gene structure was constructed using the online software of the GSDS2.0 server (http://http://gsds.gao-lab.org/index.php) (Hu et al., 2015). Data regarding the chromosomal location of *NtBAG* were obtained from the genome of *N. tabacum*. We subsequently mapped these *BAG* genes using MapInspect software. Conserved protein motifs were identified using default parameters for Multiple Em for Motif Elicitation (http://meme-suite.org/) program, with a maximum of 12 motifs. The subcellular localization of NtBAGs was predicted using ProtComp9.0 (www.softberry.com), and the identified protein motifs were further annotated using Weblogo (http://weblogo.berkeley.edu/). Finally, a 1500-bp segment of the 5′ sequence was used as the promoter region of each BAG gene to analyze the *cis*-acting elements using PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al., 2002).

Expression profile of *BAG* genes based on RNA sequencing data

RNA sequencing (RNA-seq) data were downloaded from the NCBI Sequence Read Archive database (http://www.ncbi.nlm.nih.gov/sra/) with the accession numbers SRP029183 and SRP029184. Clean reads filtered from raw reads were mapped onto the *N. tabacum* genome using HISAT (version Nitab4.5) with the default parameters (Kim et al., 2015). The levels of expression of individual genes were quantified in terms of transcripts per kilobase of exon model per million mapped read (TPM) values, which were obtained using StringTie with the default parameters (Pertea et al., 2015; Pertea et al., 2016).

Real-time quantitative reverse transcription polymerase chain reaction analysis

Total RNA was extracted from different plant materials using RNA Plant Kit (Takara, Qingdao, China) and treated with DNase I (Takara, Qindao, China) to remove genomic DNA. Reverse transcription was performed using the HiScript II 1st strand complementary DNA (cDNA) synthesis kit (Vazyme, Nanjing, China). qRT-PCR was performed using a 20 μL reaction volume comprising 10 μL SYBR qPCR Master Mix (Vazyme), 6.4 μL of ddH_2O, 0.8 μL of forward primer (10 μmol/L), 0.8 μL of reverse primer, and 2 μL of template cDNA. Nt36s was used as the internal reference gene for qRT-PCR. The gene-specific primers are listed in Table S1.

Three replicates corresponding to each period were subjected to amplification using Bio-Rad IQ5 Real-Time PCR instrument (Bio-Rad Laboratories, Hercules, CA, USA). The amplification parameters were as follows: activation at 50°C for 2 min, predenaturation at 95°C for 2 min, denaturation at 95°C for 15 s, and annealing at 60°C for 1 min (40 cycles). Finally, the relative gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

Subcellular localization of *NtBAG5c* protein

NtBAG5c was cloned into the PCAMBIA1300-35S-GFP vector and transformed into *Agrobacterium tumefaciens* strain LBA4404. The primers were designed based on *NtBAG5c* sequence (Table S1). *Agrobacterium* containing the control GFP vector or NtBAG5c-GFP recombinant vector was injected into the abaxial side of the leaves (2–4 weeks old). The leaves were then incubated in the dark for 2–3 days. Subsequently, 4′,6-diamidino-2-phenylindole (DAPI) (Solarbio, Beijing, China) was used for nuclear counterstaining. Approximately 5 min after staining, the tissue was washed twice or thrice with phosphate-buffered saline. The plasmolysis experiments were identified according to methods previously reported (Zhang et al., 2022). Finally, the tissue was sealed with a cover slip and examined under a microscope (IX73, OLYMPUS, Japan).

Yeast two-hybrid experiment

Total leaf RNA was extracted from *N. benthamiana* and reverse transcribed to obtain cDNA. The coding sequences of the *NtBAG5c* and *HSP70* genes were amplified *via* PCR using the designed primers (Table S1), which were cloned into the *Eco*R I and *Bam*H I sites of the pGBKT7 and pGADT7 vectors, respectively. The yeast receptor cells transformed with plasmid combinations pGBKT7-NtBAG5c + pGADT7-HSP70, pGADT7 + pGBKT7-NtBAG5c, pGBKT7 + pGADT7-HSP70, pGADT7-T + pGBKT7-53, pGADT7-T + pGBKT7-Lam, and pGADT7 + pGBKT7 were inoculated onto two-deficiency synthetic defined (SD) medium and incubated upside-down at 30°C for 2–3 days. Then, colonies with a diameter of >2 mm were transferred to a four-deficiency SD medium and incubated upside-down at 30°C for 4–5 days.

Virus-induced gene silencing of *NtBAG5c*

To explore the role of *NtBAG5c* gene in leaf senescence, tobacco rattle virus (TRV)::NtBAG5c vector was firstly constructed and empty vector (TRV::00) was used as a control. The tobacco phytoene desaturase (PDS) fragments were cloned into TRV vectors to construct TRV::PDS, which was used as a positive control. The

constructed cloning vector was transformed into *Agrobacterium tumefaciens* GV3101. The primers used in this experiment are shown in Table S1. Virus-induced gene silencing (VIGS) was performed as described previously (Bachan and Dinesh-Kumar, 2012). Then, a buffer containing pTRV1 was mixed with TRV::00, TRV::PDS, or TRV::NtBAG5c at a ratio of 1:1 by volume. The plants in which the fourth leaf had fully expanded were used for VIGS. Small holes were punched with a needle on the underside of the leaves to facilitate infiltration. The inoculated plants were grown at 20°C for 24 h under relative humidity of 60% in the dark, and then placed in a growth room at 25°C with a 16-h light/8-h dark photoperiod. Ten days later, the leaves were obtained for subsequent experiments. The assays were performed with at least ten plants for each vector, and the experiments were repeated at least three times.

Determination of chlorophyll content

On day 10 after of VIGS, the third functional leaf (0.2 g, from the top to bottom) was sliced into pieces and 20 mL of acetone:ethanol: water (4.5:4.5:1 [v:v]) was added to the leaves; the mixture were stored in the dark. The OD_{645} and OD_{663} values were measured after the leaves turned white. The chlorophyll content was calculated using the following formulas (He et al., 2021):

$$Chl\ a(mg/g) = (12.7 \times OD_{663} - 2.69 \times OD_{645}) \times V(mL)/[1000 \times m(g)]$$

$$Chl\ b(mg/g) = (22.9 \times OD_{645} - 4.68 \times OD_{663}) \times V(mL)/[1000 \times m(g)]$$

Determination of lignin content

Lignin content was measured using ultraviolet spectrophotometry in accordance with the manufacturer's instructions for the kit (BC4200; Beijing Solarbio Science Technology Co., Ltd., Beijing, China) (Liang et al., 2020).

Superoxide dismutase activity assay

The total SOD activity was determined using a kit (Nanjing Jiancheng Bioengineering Institute, Nanjing, China). One unit of SOD activity is defined as the amount of enzyme required for 1 mg of tissue proteins in 1 ml reaction volume to achieve a SOD inhibition rate of 50%, as monitored at 550 nm. SOD activity is presented as U mg^{-1} protein.

Analysis of hydrogen peroxidase accumulation

H_2O_2 accumulation was detected using 3,3′-diaminobenzidine (DAB) staining. The fifth and sixth leaves of negative control plants (pTRV2:00) and NtBAG5c-silenced plants were immersed overnight in DAB solution (1 mg/mL; pH 3.8) overnight to detect the *in-situ* accumulation of H_2O_2. Leaves were fixed with 100% ethanol for 3 h and then boiled in 95% ethanol for 10 min to remove chlorophyll before imaging (Li et al., 2016a).

Statistical analyses

Standard error of mean was calculated using GraphPad Prism, and statistical significance was determined using paired *t*-test *via* Excel. All data are presented as the mean (± SEM) of three independent biological determinations.

Results

Identification of BAG family members in *N. tabacum*

To identify BAG proteins in *N. tabacum*, we screened out 19 genes and confirmed the domains using Pfam and Smart databases. On the basis of similarity with homologous genes in *A. thaliana*, the 19 *NtBAG* genes were named *NtBAG1a–NtBAG7* (Table S2). The coding proteins range from 196 to 1273 amino acids (aa), with a molecular weight of 22.4–141.76 kDa and an isoelectric point (pI) of 4.59–9.56. We also predicted that most NtBAG were localized in the nucleus, while NtBAG3a, NtBAG5c, and NtBAG7 were also localized in chloroplasts and NtBAG3a was localized in mitochondria (Table S2). Of the 19 BAG genes, 9 were located on 7 chromosomes, whereas the other genes were anchored in multiple scaffolds.

Phylogenetic tree of *NtBAG*s and *AtBAG*s

To assess the phylogenetic relationships among the members of the BAG family, we used the predicted BAG protein sequences from *N. tabacum* and *A. thaliana* to construct a phylogenetic tree. The BAG proteins could be divided into two groups: class I (gray) and class II (blue) (Figure 1A). The class I BAG proteins mainly comprised AtBAG1 to AtBAG4, clustered with AtBAG1/AtBAG2 homologous proteins NtBAG1a–e and NtBAG3a–b as well as AtBAG4 homologous proteins NtBAG4a–c. The class II BAG proteins mainly included AtBAG5–7 clustered with their homologous proteins, NtBAG5a–e, NtBAG6a–b, and NtBAG7 (Figure 1A).

NtBAGs typically have three helices in the conserved domain and often possess Glu and Asp residues at positions 11 and 21 of the conserved domain, respectively (Figures 1B, C).

Three-dimensional structural analysis

Structural analysis of NtBAGs was performed based on predicted structures obtained using I-TASSER server. The predicted structure with the highest C-score was selected as a representative structure for further analysis (Figure 1D). The structures of NtBAG1a, NtBAG1b, NtBAG1c, NtBAG1d, NtBAG1e, NtBAG3a, NtBAG3b, NtBAG4a,

10.3389/fpls.2023.1108588

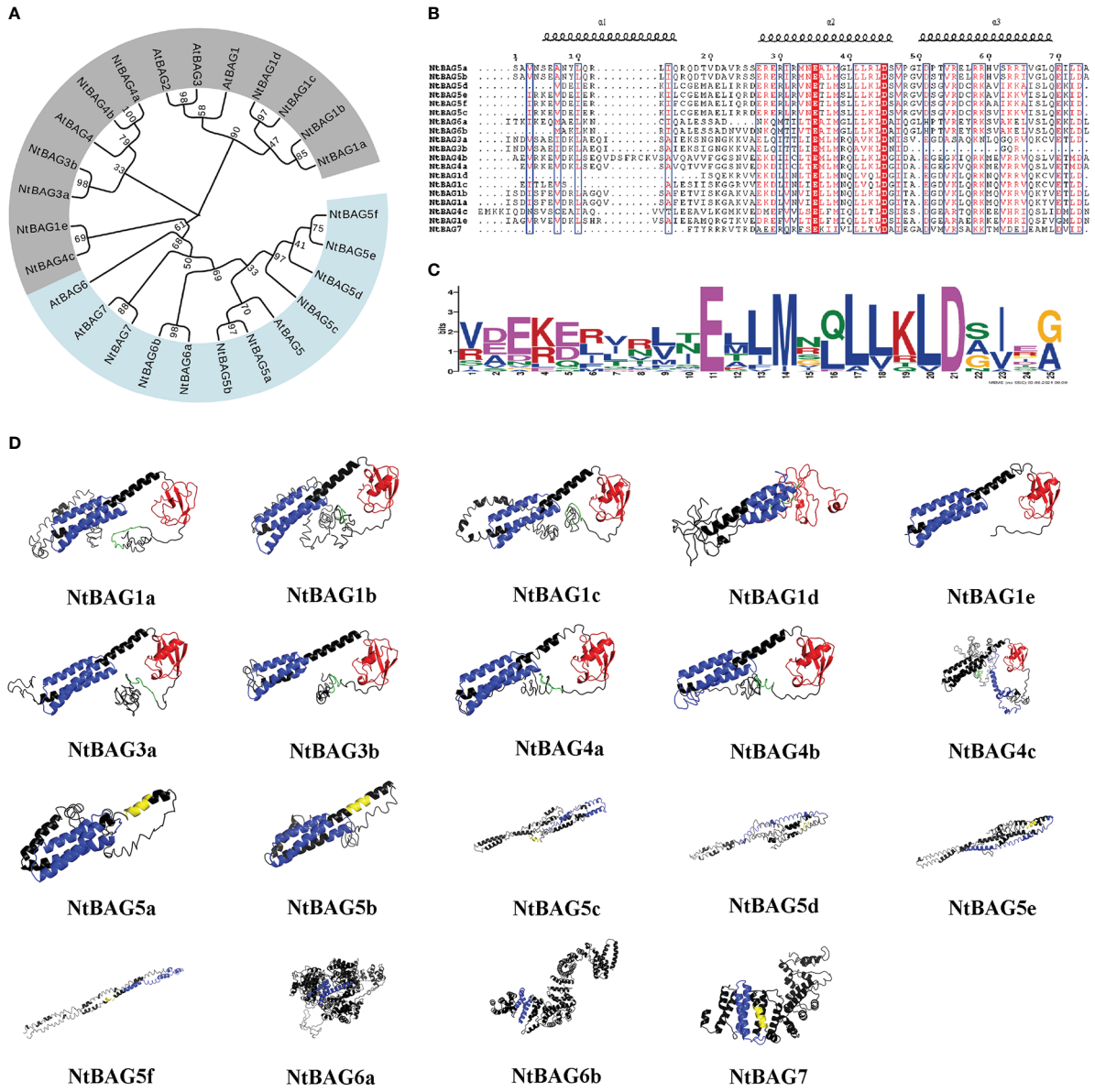


FIGURE 1
Phylogenetic and three-dimensional structure analysis of the NtBAG proteins. **(A)** The evolutionary history was inferred using the maximum likelihood method based on the Poisson correction model. The bootstrap consensus tree inferred from 1,000 replicates is taken to represent the evolutionary history of the analyzed taxa. Branches corresponding to partitions reproduced in less than 50% of bootstrap replicates are collapsed. Initial trees for the heuristic search were obtained automatically by applying neighbor-joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with the superior log likelihood value. The analysis involved 86 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 66 positions in the final dataset. Evolutionary analyses were conducted in MEGA7. **(B)** Alignment of the domains conserved within the BAG protein family. The locations of the three helices are shown above the alignment. **(C)** A motif conserved among the BAG family proteins. The consensus sequences were displayed using Weblogo. **(D)** Prediction of three-dimensional structure of NtBAG proteins. The BDs of the NtBAG proteins are depicted in blue, the ubiquitin domain (UBD) is depicted in red, and the IQ motif is depicted in yellow. A conserved sequence motif of 12 amino acids (ExRPGG[ML/VV]QxR) is depicted in green.

NtBAG4b, and NtBAG4c were comparable in terms of the presence and organization of the major structural domains. Their structures contained a highly organized BD and an ubiquitin-like domain. The conserved sequence pattern of the characteristic 12 aa in the N-terminal domain, which is found in all family members except NtBAG1e, showed a hairpin loop pattern. The highly conserved ubiquitin-like structural domain strictly exhibited four β-sheets and two α-sheets. Of the four β-sheets, two were organized as a central β-sheet with an α-helix on either side. The BAG structural domain of all NtBAGs consists of a typical three α-helix bundle structure. The IQ-calmodulin binding pattern in NtBAG5c is a small α-helix connected by two hairpin loops, whereas that in NtBAG5a, NtBAG5b and NtBAG7 is a complete single helix.

Structures and conserved motif analysis of *NtBAGs*

To further analyze the characteristics of *NtBAGs*, we explored the conserved motifs and gene structure in terms of exons and introns.

10.3389/fpls.2023.1108588

The conserved motifs in these genes also showed similarities within the same subgroup, such as six similar motifs in all NtBAG1–4 homologous proteins, except for NtBAG1e and NtBAG4c (Figure 2A, B). NtBAG6a–b had three motifs, whereas NtBAG7 only had one of the ten motifs. Among the six NtBAG5 homologous proteins, NtBAG5a–b had three motifs, whereas NtBAG5c–e had four to five motifs. Gene structure analysis indicated that most *NtBAG1–4* homologous genes had four exons, except for *NtBAG1e* and *NtBAG4c*, which contained five and nine exons, respectively. NtBAG5a–b, NtBAG5c–f, and NtBAG6–7 contained two, one, and two to three exons, respectively. Further, we revealed that the genetic structures of all BAG family proteins from *N. tabacum* were similar, with them showing close evolutionary relationships. Moreover, genes within the same subfamily often showed similar gene structures (Figures 2A, C).

Promoter *cis*-acting element analysis of *NtBAGs*

The *cis*-acting elements in the promoter usually regulate gene expression and function. Multiple *cis*-acting elements, such as plant hormone, light, and stress response elements, were identified in BAG gene promoters (Figure S1 and Table S3).

Regarding hormone-related *cis*-acting elements, we identified at least one abscisic acid response element (ABREs) in the promoters of *N. tabacum* BAG genes, except for *NtBAG1e*, *NtBAG5b–c*, and *NtBAG5e–f*. Further, only one auxin response element (TGA element) was detected in the promoters of *NtBAG3a*, *NtBAG4b*, *NtBAG5d*, and *NtBAG7* promoters. Meanwhile, at least one of the gibberellin response elements, such as TARC-box, P-box, or GARE, was found in the promoters of most genes, except for *NtBAG1a–b*,

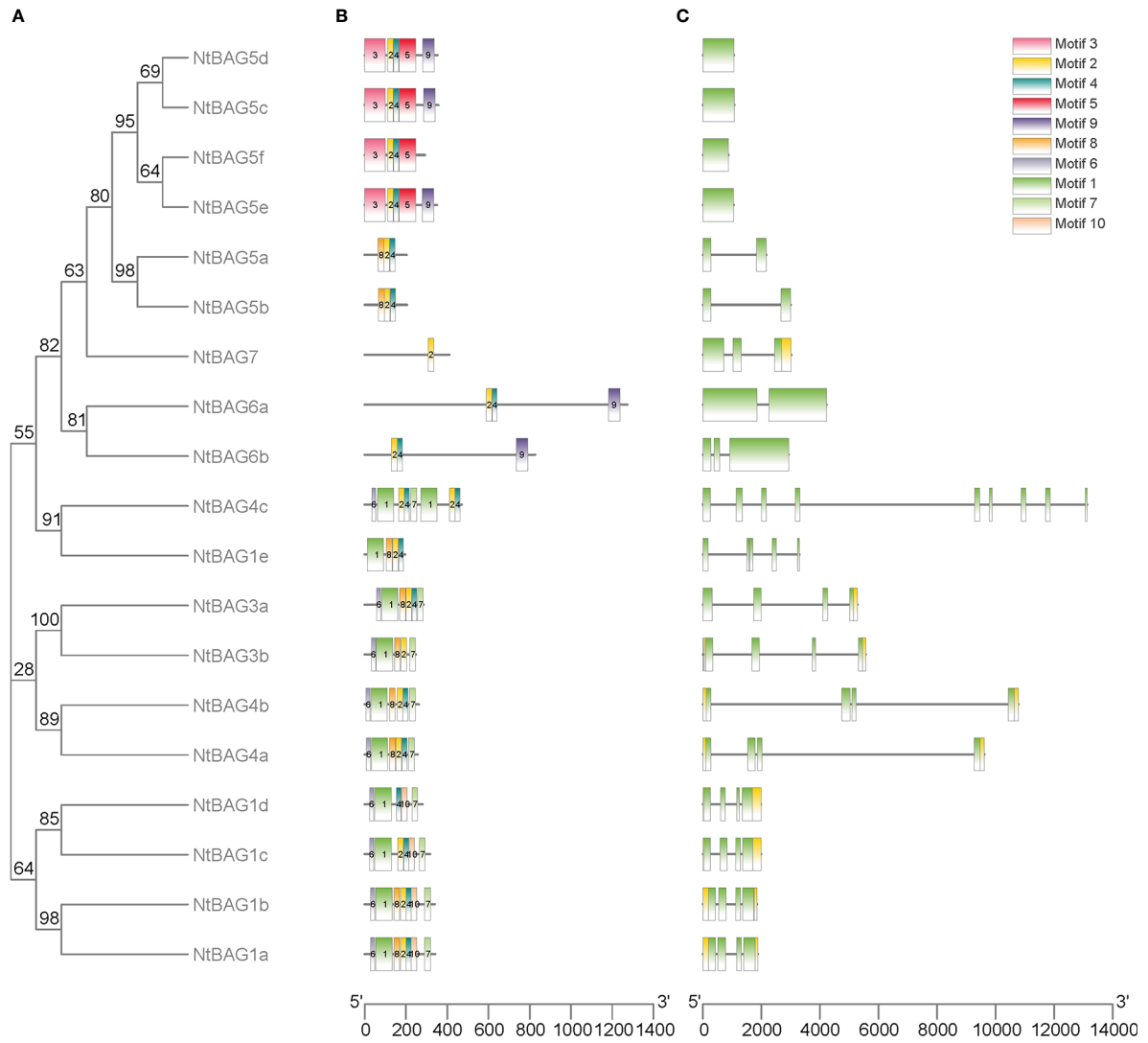


FIGURE 2
Genomic structure and motif composition of NtBAGs. **(A)** The phylogenetic tree of NtBAG proteins. **(B)** The conserved motifs in NtBAG proteins were identified using MEME. Each motif is represented by a specific color and the nucleotide sequence numbers are shown below. **(C)** Genomic structure of NtBAG family members in tobacco. Exons and introns are represented by green boxes and black lines, while yellow box shows the noncoding region.

NtBAG4a, *NtBAG4c*, *NtBAG5c–e*, and *NtBAG6a*. The Methyl Jasmonate (MeJA) response elements CGTCA and TGACG, were identified in the promoters of most genes, except for *NtBAG4a*, *NtBAG4c*, *NtBAG5a*, and *NtBAG6a*. Further, we found several other hormone-related *cis*-elements, such as the salicylic acid (SA) response element TCA element, in some *NtBAG* promoters.

In addition, we identified numerous *cis*-acting elements related to light response in these promoters, including ACE, AE box, TCT motif, ATC motif, Box 4, GATA motif, G Box, and GT1 motif. Among them, G box is distributed in the promoters of most members of the *NtBAG* family. We also detected stress response-related *cis*-acting elements in these promoters, including LTR, MBS, and TC-rich repeats, along with other elements, such as WUN motif (related to wounds); circadian element (related to circadian control); and Myb, Myc, STRE, TC-rich, W box, and ARE elements. In particular, we identified Myb and Myc-motif elements in almost all BAG promoters.

Tissue-specific expression profiles of *NtBAG*s

Based on the previously obtained RNA-seq data (Sierro et al., 2014), the expression patterns of all *BAG*s in different developmental periods and tissues were analyzed. *BAG*s were divided into four clusters according to their expression trends. The first cluster contained NtBAG5c–d and NtBAG6a–b, which were highly expressed in senescent flower (SF), young leaf (YL), mature leaf (ML), senescent leaf (SL), root (R), stem (S), dry capsule (DC), and other tissues. Moreover, the expression patterns of these genes tended to increase with the aging of flowers and leaf tissues. The second cluster comprised NtBAG1a–e, NtBAG3a, NtBAG4a–b, and NtBAG5a–b, which showed low expression levels in all tissues. NtBAG3b and NtBAG7 constituted the third cluster and showed high expression in R and S tissues. In addition, NtBAG7 was highly expressed in immature flower (IF) tissues, whereas it showed low expression in leaf and flower tissues with aging. The fourth

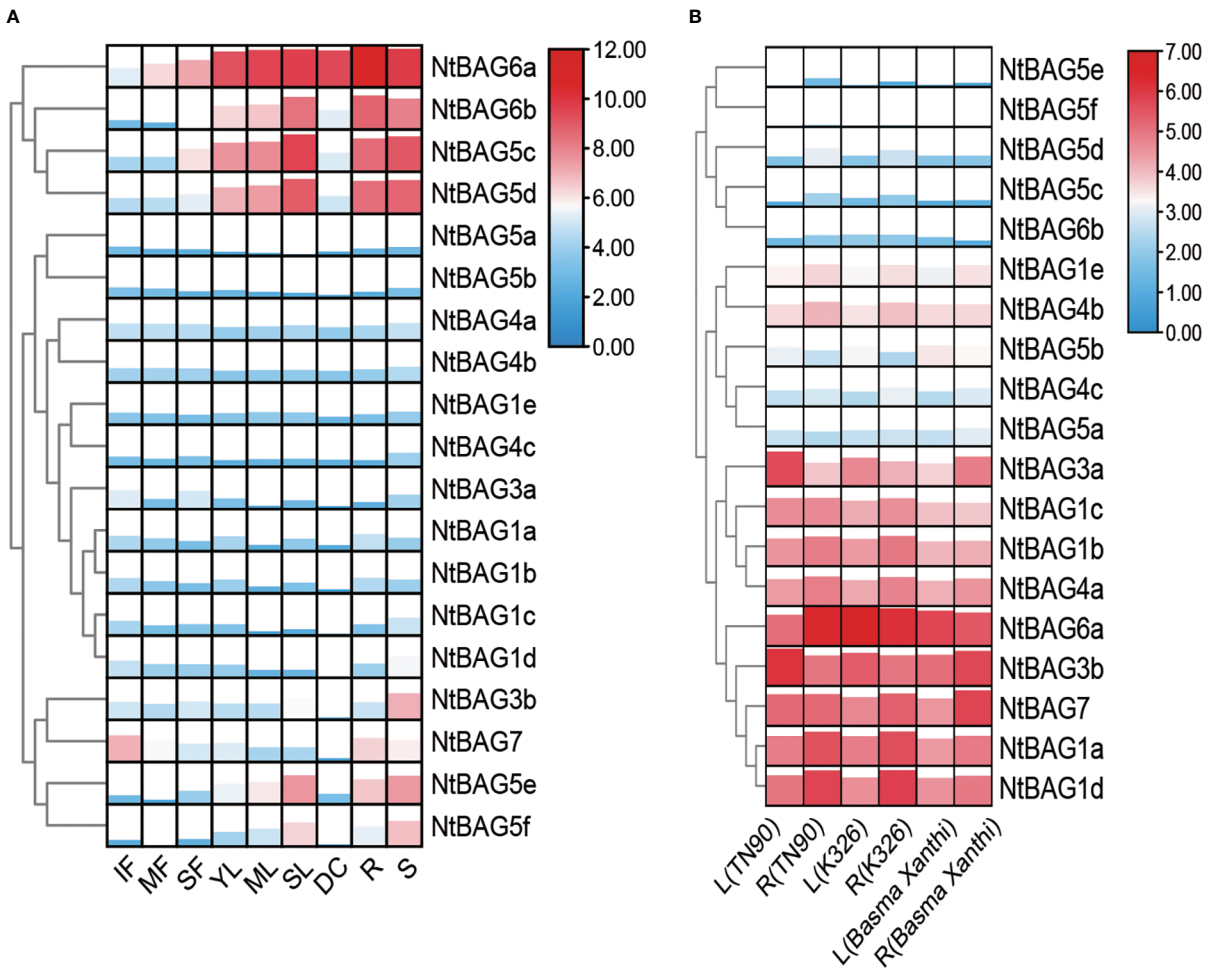


FIGURE 3
Expression patterns of *BAG* genes. **(A)** Expression patterns in different tissues and developmental stages of tobacco. **(B)** *BAG* gene expression patterns of leaf and root in different varieties. MF, mature flower; R, root; DC, dry capsule; ML, mature leaf; SF, senescent flower; SL, senescent leaf; S, stem; IF, immature flower; YL, young leaf. The expression levels are represented by the color bar (log2-transformed).

cluster included NtBAG5e–f, which showed a lower expression trend similar to that of NtBAG5c-d (Figure 3A). Further analysis revealed a similar in the expression pattern in both root and stem tissues in different varieties (Figure 3B).

Expression of *NtBAG*s during leaf senescence

Based on the abovementioned results, multiple *NtBAGs* may be associated with aging. qRT-PCR was used to further identify the genes related to leaf senescence. Compared with young leaves, the expression of *NtBAG1a* was significantly downregulated in mature leaves but was upregulated in senescent leaves (Figure 4A). Meanwhile, the expression levels of *NtBAG1c* and *NtBAG1e* were significantly decreased compared with that in young leaves (Figures 4B, C). Compared with young leaves, the expression level of *NtBAG3a* was significantly decreased in mature leaves but significantly increased in senescent leaves (Figure 4D). The expression patterns of *NtBAG4a*, *NtBAG4c*, and *NtBAG5b* were also highly similar to that of NtBAG1c (Figures 4E–G). The expression of *NtBAG5c* and *NtBAG6b* was significantly decreased in mature leaves but was significantly increased in senescent leaves, consistent with the expression of *NtBAG3a* (Figures 4H, I). Interestingly, the expression of NtBAG5c showed a sharp variation as it was upregulated in senescent leaves compared with young leaves by approximately 20 times, which was much higher than the expression of *NtBAG3a* and *NtBAG6b* (Figures 4D, H, I). The expression of *NtBAG7* was high in young leaves and significantly low in mature and senescent leaves (Figure 4J). The expression trends of these genes were consistent with the previous RNA-seq data.

NtBAG5c localizes in the cell wall and interacts with HSP proteins

Localization prediction using Plant-mPLoc revealed that NtBAG5c was localized in the nucleus and chloroplast. NtBAG5c-GFP fusion protein was observed to be accumulated in the nucleus, cell membrane, and cell wall (Figure 5A). After separation of the cell wall, NtBAG5c-GFP fusion protein was localized in the cell wall and nucleus (Figure 5B). The nucleus targeting of NtBAG5c-GFP was consistent with the previous prediction (Figures 5A, B).

Most members of the BAG family can interact with HSPs (Doong et al., 2000; Alberti et al., 2003; Chung and Dawson, 2004; Davidson et al., 2016). We analyzed the interaction of NtBAG5c proteins with the candidate proteins HSP70 and HSP20 using yeast two-hybrid experiment. The results showed that NtBAG5c can interact with HSP70 and HSP20 *in vitro* (Figures 5C, D).

NtBAG5c is involved in leaf senescence

NtBAG5c was silenced in *N. tabacum* using a TRV-based VIGS method. *NtBAG5c*-silenced plants showed no significant phenotypic differences compared with the control plants (Figures 6A–C). qRT-PCR analysis revealed that the expression of the *NtBAG5c* in *NtBAG5c*-silenced plants was significantly downregulated compared with that in the negative control (Figure 6D), thus confirming the successful silencing of *NtBAG5c*.

The expression of SAGs, such as *NtCP1*, *NtCP2*, *SAG12*, and *SEN4*, was analyzed using further tests. In senescent leaves, the expression levels of *NtCP1*, *SAG12*, and *SEN4* were significantly high and that of *NtCP2* was low (Figures 6E–H). These results were consistent with those in previous reports (Nam, 1997; Beyene et al., 2006; Jiang et al., 2014). In *NtBAG5c*-silenced plants, the expression of *NtCP1*, *SEN4*, and *SAG12* was significantly decreased (Figures 6I, K, L) but that of *NtCP2* was upregulated (Figure 6J). Thus, *NtBAG5c* may act as an upstream regulator to participate in leaf senescence. Although there were evident changes in the expression levels of senescence-related genes, no significant differences in phenotypes were observed between the *NtBAG5c*-silenced and the control plants (Figures 6A–C). This was further confirmed by the chlorophyll contents in the leaves of *NtBAG5c*-silenced plants and control

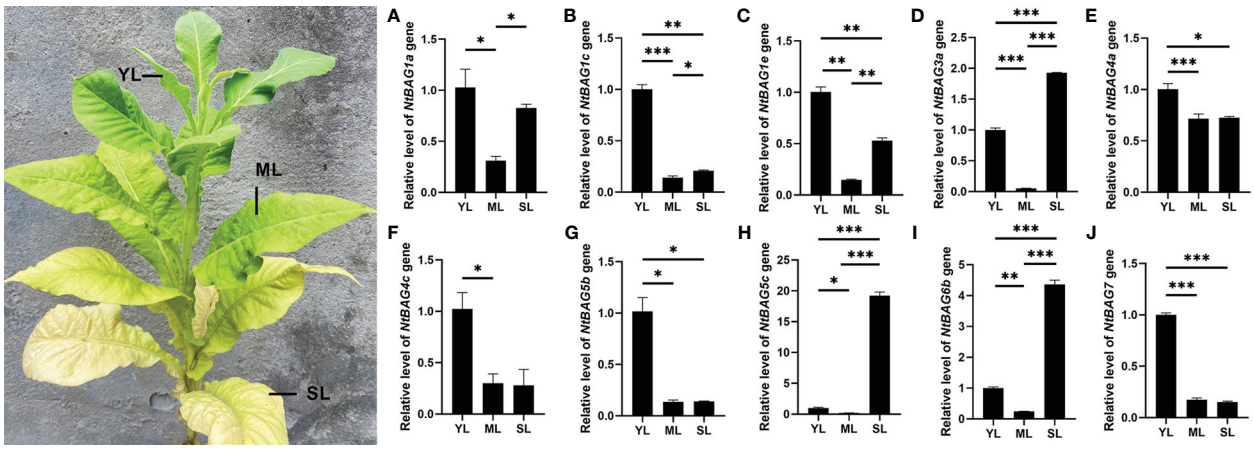


FIGURE 4
The relative expression pattern of *NtBAG* in wild type tobacco. The expression levels of NtBAG1a **(A)**, NtBAG1c **(B)**, NtBAG1e **C)**, NtBAG3a **(D)**, NtBAG4a **(E)**, NtBAG4c **(F)**, NtBAG5b **(G)**, NtBAG5c **(H)**, NtBAG6b **(I)** and NtBAG7 **(J)** in plant leaves at different developmental periods, as determined by qRT-PCR. YL, young leaf; ML, mature leaf; SL, senescent leaf. All data are presented as the mean (± SEM) of three independent biological determinations and were analyzed by Student's *t*-test (*p< 0.05, **p< 0.01, ***p< 0.001).

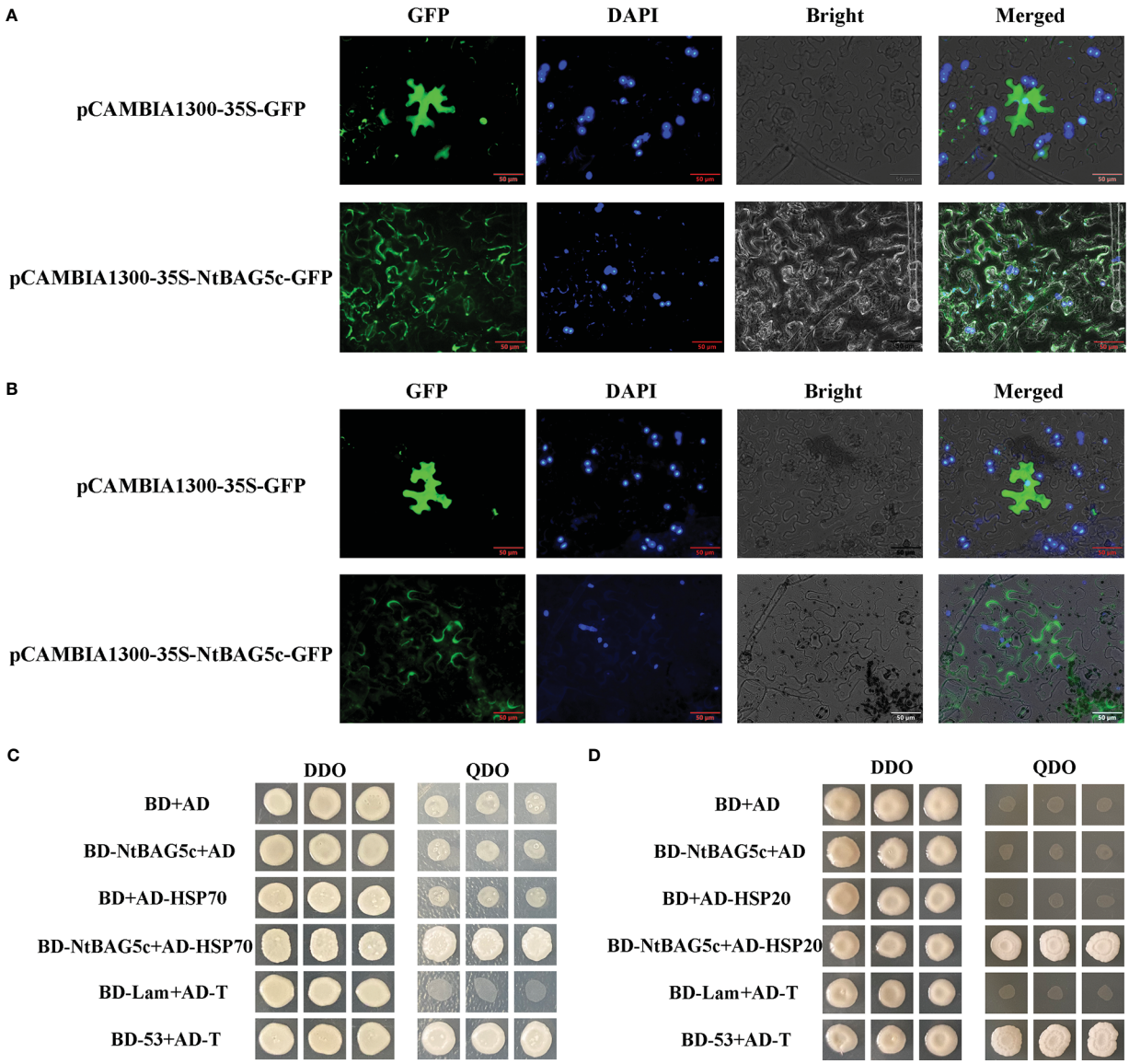


FIGURE 5
Localization of *NtBAG5c* in epidermal cells of *N. benthamiana*. **(A)** Subcellular localization analysis demonstrated that NtBAG5c is located in the cell membrane and cell wall. **(B)** After the wall separation, subcellular localization analysis indicated that NtBAG5c is located in the cell wall. GFP, green fluorescent protein; DAPI, fluorescent dye capable of binding strongly to DNA; Bright, white light; Merged, superposition of GFP, DAPI, and Bright. **(C, D)** Yeast two-hybrid assay. **(C)** The interaction of NtBAG5c and HSP70 in yeast cells. **(D)** The interaction of NtBAG5c and HSP20 in yeast cells. BD-53 + AD-T and BD + AD were used as positive and negative controls, respectively. The yeast co-transformed groups were grown on the SD Leu-Trp medium [double dropout (DDO), without leucine and tryptophan], and then grown on SD-Leu-Trp-His-Ade medium [quadruple dropout (QDO), with leucine, tryptophan, histidine, and adenine].

plants. Thus, the silencing of *NtBAG5c* did not affect the chlorophyll content in plants (Figures 6M, N).

NtBAG5c inhibits lignin formation and promotes H_2O_2 generation

The lignin content is lower in senescent leaves of maize mutant is lower than that in the wild type maize (Jiao et al., 2019). Compared with control plants, a significant increase in lignin content was observed in the leaves of *NtBAG5c*-silenced plants (Figure 6O), suggesting that *NtBAG5c* plays a role in lignin synthesis.

SOD activity is closely related to leaf senescence and regulates the level of H_2O_2 accumulation (Cui et al., 2013; Wang et al., 2016). The SOD activity in *NtBAG5c*-silenced plants was significantly lower than that in control plants (Figure 6P), implying that NtBAG5c promotes SOD enzyme activity. Further assays of H_2O_2 levels *via* DAB staining revealed that the *NtBAG5c*-silenced leaves were more weakly stained than the control leaves, indicating that NtBAG5c can promote H_2O_2 accumulation (Figure 6Q).

Discussion

The formation and composition of secondary metabolites in tobacco leaves during post-maturation can affect the quality of the leaves. Leaf senescence is associated with the maturation of leaves and

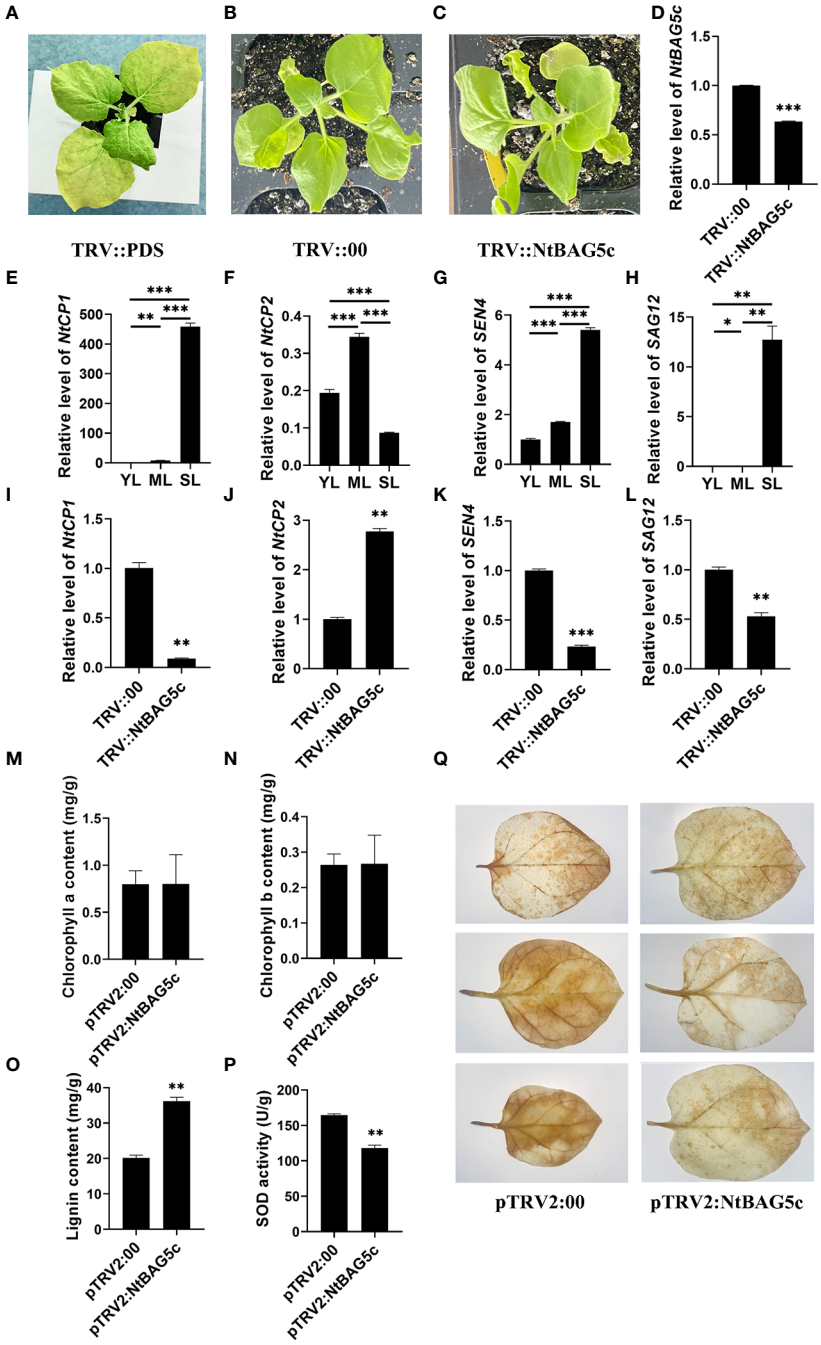


FIGURE 6
Silencing of NtBAG5c in *Nicotiana tabacum* using a tobacco rattle virus (TRV)-based VIGS system. **(A–C)** Phenotypes of tobacco plants after 10 days of VIGS. **(A)** PDS control plants (*TRV::PDS*); **(B)** negative control plants (*TRV::00*); **(C)** NtBAG5c-silenced plants (*TRV::NtBAG5c*). The expression levels of NtBAG5c **(D)**, NtCP1 **(E)**, NtCP2 **(F)**, SEN4 **(G)**, and SAG12 **(H)** in plant leaves at different developmental periods, as determined by qRT-PCR. The expression levels of NtCP1 **(I)**, NtCP2 **(J)**, SEN4 **(K)**, and SAG12 **(L)** in NtBAG5c-silenced plants by qRT-PCR. **(M)** Chlorophyll a content. **(N)** Chlorophyll b content. **(O)** Lignin content (mg/g). **(P)** SOD activity (U/g). **(Q)** DAB staining of isolated leaves of pTRV2:00 and pTRV2:NtBAG5c. Results were reproduced in three independent experiments using three plants each. All data are presented as the mean (± SEM) of three independent biological determinations and were analyzed using Student's t-test (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

the rapid formation of secondary metabolites. *BAGs* are involved in plant growth, development, and stress response (Thanthrige et al., 2020). In plants, BAG1 was the first identified member of the BAG family and was shown to interact with BCL-2 by enhancing its antiapoptotic function (Takayama and Reed, 2001). Homologs of the BAG family have been identified in rice (Rashid Mehmood, 2012), tomato (Jiang et al., 2022), wheat (Ge et al., 2016), banana (Dash and Ghag, 2022), and *Arabidopsis* (Yan et al., 2003; Doukhanina et al., 2006; Nawkar et al., 2017). Seven BAG genes were identified in *Arabidopsis*, but we identified 19 homologs in common tobacco. Among these genes, six homologs of AtBAG5 may be the main reason for the increased number of BAG genes. In *Arabidopsis*, the function

of AtBAG5 is mainly associated with leaf senescence (Cui et al., 2015; Li et al., 2016a). It remains to unclear whether the direction of selection by scientists during breeding leads to an increase in the number of BAG5 homologs.

Phylogenetic analysis confirmed that BAGs are highly conserved throughout the plant kingdom. The 19 NtBAGs were divided into two groups. The first group comprised *NtBAG1a–e*, *NtBAG3a–b*, and *NtBAG4a–c* clustered with *AtBAG1–4*, whereas the second group comprised *NtBAG5a–e*, *NtBAG6a–b*, and *NtBAG7* clustered with *AtBAG5–7*. Moreover, the structure and conserved motifs of *NtBAGs* showed similarities within the same subgroup. These results imply that *BAGs* in tobacco have highly similar functions to those of the homologous genes in *Arabidopsis*.

Cis-acting elements regulate gene transcription. We identified a series of abiotic stress- and hormone-related elements in most *NtBAG* promoters, such as the phytohormone response elements ABRE, CGTCA, and TGACG. Meanwhile, NtBAG1a–e, NtBAG3a–b, NtBAG4a–c, and NtBAG5a had two or more G Box *cis*-acting elements, and transcriptomic data revealed that the expression of these genes was relatively stable over multiple periods and tissues, showing a slight variation compared with that of other genes. The characteristics of continuous and constant expression of these genes are also highly similar to those of AtBAG1–3 in *Arabidopsis* (Vivancos et al., 2010). NtBAG5b–f and NtBAG6a–b have less than one G box *cis*-acting element, and their expression varies dramatically during maturation and senescence. However, it remains unclear whether this element is related to the regulation of senescence during maturation. In addition, stress response-related *cis*-acting elements, including LTR, MBS, and TC-rich repeat sequences, were detected in the promoters of multiple *NtBAGs*, except for NtBAG3b, NtBAG4a, NtBAG4c, and NtBAG5c, suggesting that multiple *NtBAGs* are involved in the plant stress response. This finding is consistent with that of previous studies on *Arabidopsis BAGs*, which also contain stress response elements, including ABRE, ERE, CGTCA motifs, MBS, and TC-rich repeat sequences (Doukhanina et al., 2006; Nawkar et al., 2017). This suggests that *BAGs* play an important role in the resistance of plants to stress.

Based on the expression patterns obtained using RNA-seq data and slight differences in gene sequences within the same cluster in the phylogenetic tree, 10 *NtBAGs* were screened for qRT–PCR analysis. These genes were divided into two categories based on their expression patterns. The first category included genes with high expression in young leaves but significantly low expression in mature and senescent leaves, namely, *NtBAG1a*, *NtBAG1c*, *NtBAG1e*, *NtBAG4a*, *NtBAG4c*, *NtBAG5b*, and *NtBAG7*. The second category comprised genes that showed a significantly higher expression in senescent leaves than in young and mature leaves, namely, *NtBAG3a*, *NtBAG5c*, and *NtBAG6b*. In *Arabidopsis*, BAGs can also be divided into two broad groups based on the organization of their domains. Notably, the expression of NtBAG5c increased dramatically in senescent leaves, which is similar to the expression trend of marker genes of senescence in plants, namely, *NtCP1*, *SEN4*, and *SAG12* (Park et al., 1998; Grbić, 2002; Beyene et al., 2006).

However, regarding the localization of BAGs of *Arabidopsis*, BAG5 was accumulated in mitochondria (Li et al., 2016a), whereas NtBAG5c was mainly localized in the nucleus and cell wall, which may be related to the functional differentiation of multiple BAG5 homologs in tobacco.

The BAG family proteins of *Arabidopsis*, rice (Zhou et al., 2021), and other plants can act as a backbone to link Hsp70 and small heat stress proteins, such as sHsp20, *via* protein interactions, as determined using yeast two-hybrid and immunoprecipitation experiments. The IQ motif in the AtBAG5 can bind to calcium-regulated proteins and thus participate in the regulation of the calcium signaling pathway, forming a complex with Hsp70 and regulating leaf senescence (Li et al., 2016a). Furthermore, the results of yeast two-hybrid experiment showed that NtBAG5c can interact with Hsp70 and Hsp20 *in vitro*, which may form a complex involved in the regulation of leaf senescence in tobacco.

In *Arabidopsis*, BAG5 regulated the production of ROS and expression of SAGs (Li et al., 2016a). H_2O_2 can induce ROS production and accelerate leaf senescence. A previous study revealed that NtBAG5c-silenced plants exhibit low levels of H_2O_2, which is consistent with the findings of a study on AtBAG5 (Li et al., 2016a). This indicates that NtBAG5c can promote H_2O_2 accumulation and accelerate leaf senescence. Moreover, we revealed that NtBAG5c-silenced plants exhibit lower SOD activity and higher lignin content than control plants, suggesting that NtBAG5 can inhibit lignin accumulation and promote SOD activity. Further, a study reported that CsMYB4a overexpressed in tobacco plants can promote senescence and reduce total lignin content (Li et al., 2017a). In tobacco, NtBAG5c may regulate lignin content through its homology with CsMYB4a. In particular, *CP1*, *SEN4*, and *SAG12* are often highly expressed in senescent leaves (Park et al., 1998; Grbić, 2002; Grbić, 2003; Beyene et al., 2006). The expression of *NtCP1*, *SEN4*, and *SAG12* was significantly downregulated in *NtBAG5c*-silenced plants, indicating that NtBAG5c acts as a positive regulator and is involved in the regulation of leaf senescence.

Conclusions

This study identified and screened 19 *BAG* family members from the tobacco genome. The gene structure, structural domains, physicochemical properties, and expression patterns of tobacco *BAGs* were characterized. Further analysis of the expression of these *NtBAGs* using qRT–PCR revealed that multiple *NtBAGs* may be related to tobacco leaf senescence. Among them, NtBAG5c interacts with Hsp70 and Hsp20, indicating that NtBAG5c affects tobacco leaf senescence by forming a complex with Hsp70 or Hsp20. This study provides a theoretical basis for the further investigation of the *NtBAG* gene family and provides important guidance for molecular breeding.

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.

Ethics statement

The authors declare that all methods were carried out in accordance with relevant guidelines and regulations.

Author contributions

QS and DG designed the study. QS carried out bioinformation analyses. LG, BH, YW carried out the qRT-PCR analyses. BH, YW, XH and PC collected plant materials. LG, BH and XC carried out other experiments. LG wrote the original draft. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by The Key Science and Technology Project of China National Tobacco Corporation (grant No.110202001023(JY-06) and 110201901018(JY-05)). The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication. All authors declare no other competing interests.

Acknowledgments

We appreciate the reviewers and editors for the patience to the work.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1108588/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Cis-acting elements on promoters (1500 bp) of *NtBAG* genes. The number shows the number of *cis*-acting elements.

SUPPLEMENTARY TABLE 1
Information of primer sequences

SUPPLEMENTARY TABLE 2
BAG protein family members in *Nicotiana tabacum*

SUPPLEMENTARY TABLE 3
The predicted cis-element of each *NtBAG* genes

References

Alberti, S., Esser, C., and Hohfeld, J. (2003). BAG-1–a nucleotide exchange factor of Hsc70 with multiple cellular functions. *Cell Stress Chaperones* 8, 225–231. doi: 10.1379/1466-1268(2003)008<0225:BNEFOH>2.0.CO;2

Anderson, L. R., Sutherland, R. L., and Butt, A. J. (2010). BAG-1 overexpression attenuates luminal apoptosis in MCF-10A mammary epithelial cells through enhanced RAF-1 activation. *Oncogene* 29, 527–538. doi: 10.1038/onc.2009.362

Arabidopsis Interactome Mapping, Consortium. (2011). Evidence for network evolution in an arabidopsis interactome map. *Science* 333, 601–607. doi: 10.1126/science.1203877

Bachan, S., and Dinesh-Kumar, S. P. (2012). Tobacco rattle virus (TRV)-based virus-induced gene silencing. *Methods Mol. Biol.* 894, 83–92. doi: 10.1007/978-1-61779-882-5_6

Beyene, G., Foyer, C. H., and Kunert, K. J. (2006). Two new cysteine proteinases with specific expression patterns in mature and senescent tobacco (Nicotiana tabacum l.) leaves. *J. Exp. Bot.* 57, 1431–1443. doi: 10.1093/jxb/erj123

Bruchmann, A., Roller, C., Walther, T. V., Schafer, G., Lehmusvaara, S., Visakorpi, T., et al. (2013). Bcl-2 associated athanogene 5 (Bag5) is overexpressed in prostate cancer and inhibits ER-stress induced apoptosis. *BMC Cancer* 13, 96. doi: 10.1186/1471-2407-13-96

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

Chung, K. K., and Dawson, T. M. (2004). Parkin and Hsp70 sacked by BAG5. *Neuron* 44, 899–901. doi: 10.1016/j.neuron.2004.12.007

Cui, B., Fang, S., Xing, Y., Shen, Y., and Yang, X. (2015). Crystallographic analysis of the arabidopsis thaliana BAG5-calmodulin protein complex. *Acta Crystallogr. F Struct. Biol. Commun.* 71, 870–875. doi: 10.1107/S2053230X15005956

Cui, M. H., Ok, S. H., Yoo, K. S., Jung, K. W., Yoo, S. D., and Shin, J. S. (2013). An arabidopsis cell growth defect factor-related protein, CRS, promotes plant senescence by increasing the production of hydrogen peroxide. *Plant Cell Physiol.* 54, 155–167. doi: 10.1093/pcp/pcs161

Dash, A., and Ghag, S. B. (2022). Genome-wide in silico characterization and stress induced expression analysis of BcL-2 associated athanogene (BAG) family in musa spp. *Sci. Rep.* 12, 625. doi: 10.1038/s41598-021-04707-5

Davidson, B., Valborg Reinertsen, K., Trinh, D., Reed, W., and Bohler, P. J. (2016). BAG-1/SODD, HSP70, and HSP90 are potential prognostic markers of poor survival in node-negative breast carcinoma. *Hum. Pathol.* 54, 64–73. doi: 10.1016/j.humpath.2016.02.023

Doong, H., Price, J., Kim, Y. S., Gasbarre, C., Probst, J., Liotta, L. A., et al. (2000). CAIR-1/BAG-3 forms an EGF-regulated ternary complex with phospholipase c-gamma and Hsp70/Hsc70. *Oncogene* 19, 4385–4395. doi: 10.1038/sj.onc.1203797

Doong, H., Vrailas, A., and Kohn, E. C. (2002). What's in the 'BAG'?–a functional domain analysis of the BAG-family proteins. *Cancer Lett.* 188, 25–32. doi: 10.1016/S0304-3835(02)00456-1

Doukhanina, E. V., Chen, S., van der Zalm, E., Godzik, A., Reed, J., and Dickman, M. B. (2006). Identification and functional characterization of the BAG protein family in arabidopsis thaliana. *J. Biol. Chem.* 281, 18793–18801. doi: 10.1074/jbc.M511794200

Edwards, K. D., Fernandez-Pozo, N., Drake-Stowe, K., Humphry, M., Evans, A. D., Bombarely, A., et al. (2017). A reference genome for nicotiana tabacum enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genomics* 18, 448. doi: 10.1186/s12864-017-3791-6

Fang, S., Li, L., Cui, B., Men, S., Shen, Y., and Yang, X. (2013). Structural insight into plant programmed cell death mediated by BAG proteins in arabidopsis thaliana. *Acta Crystallogr. D Biol. Crystallogr.* 69, 934–945. doi: 10.1107/S0907444913003624

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367

Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., et al. (2010). The pfam protein families database. *Nucleic Acids Res.* 38, D211–D222. doi: 10.1093/nar/gkp985

Ge, S., Kang, Z., Li, Y., Zhang, F., Shen, Y., Ge, R., et al. (2016). Cloning and function analysis of BAG family genes in wheat. *Funct. Plant Biol.* 43, 393–402. doi: 10.1071/FP15317

Grbić, V. (2002). Spatial expression pattern of SAG12:GUS transgene in tobacco (Nicotiana tabacum). *Physiologia Plantarum* 116, 416–422. doi: 10.1034/j.1399-3054.2002.1160318.x

Grbić, V. (2003). SAG2 and SAG12 protein expression in senescing arabidopsis plants. *Physiologia Plantarum* 119, 263–269. doi: 10.1034/j.1399-3054.2003.00168.x

Gutle, D. D., Roret, T., Muller, S. J., Couturier, J., Lemaire, S. D., Hecker, A., et al. (2016). Chloroplast FBPase and SBPase are thioredoxin-linked enzymes with similar architecture but different evolutionary histories. *Proc. Natl. Acad. Sci. U.S.A.* 113, 6779–6784. doi: 10.1073/pnas.1606241113

He, M., Wang, Y., Jahan, M. S., Liu, W., Raziq, A., Sun, J., et al. (2021). Characterization of SlBAG genes from solanum lycopersicum and its function in response to dark-induced leaf senescence. *Plants (Basel)* 10, 947. doi: 10.3390/plants10050947

Hortensteiner, S. (2006). Chlorophyll degradation during senescence. *Annu. Rev. Plant Biol.* 57, 55–77. doi: 10.1146/annurev.arplant.57.032905.105212

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

Jahan, M. S., Shu, S., Wang, Y., Hasan, M. M., El-Yazied, A. A., Alabdallah, N. M., et al. (2021). Melatonin pretreatment confers heat tolerance and repression of heat-induced senescence in tomato through the modulation of ABA- and GA-mediated pathways. *Front. Plant Sci.* 12, 650955. doi: 10.3389/fpls.2021.650955

Jiang, H., Ji, Y., Sheng, J., Wang, Y., Liu, X., Xiao, P., et al. (2022). Genome-wide identification of the bcl-2 associated athanogene (BAG) gene family in solanum lycopersicum and the functional role of SlBAG9 in response to osmotic stress. *Antioxidants (Basel)* 11, 589. doi: 10.3390/antiox11030598

Jiang, Y., Liang, G., Yang, S., and Yu, D. (2014). Arabidopsis WRKY57 functions as a node of convergence for jasmonic acid- and auxin-mediated signaling in jasmonic acid-induced leaf senescence. *Plant Cell* 26, 230–245. doi: 10.1105/tpc.113.117838

Jiao, S., Hazebroek, J. P., Chamberlin, M. A., Perkins, M., Sandhu, A. S., Gupta, R., et al. (2019). Chitinase-like1 plays a role in stalk tensile strength in maize. *Plant Physiol.* 181, 1127–1147. doi: 10.1104/pp.19.00615

Kabbage, M., and Dickman, M. B. (2008). The BAG proteins: a ubiquitous family of chaperone regulators. *Cell Mol. Life Sci.* 65, 1390–1402. doi: 10.1007/s00018-008-7535-2

Kabbage, M., Kessens, R., Bartholomay, L. C., and Williams, B. (2017). The life and death of a plant cell. *Annu. Rev. Plant Biol.* 68, 375–404. doi: 10.1146/annurev-arplant-043015-111655

Kabbage, M., Kessens, R., and Dickman, M. B. (2016). A plant bcl-2-associated athanogene is proteolytically activated to confer fungal resistance. *Microb. Cell* 3, 224–226. doi: 10.15698/mic2016.05.501

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Koyama, T. (2018). A hidden link between leaf development and senescence. *Plant Sci.* 276, 105–110. doi: 10.1016/j.plantsci.2018.08.006

Lee, D. W., Kim, S. J., Oh, Y. J., Choi, B., Lee, J., and Hwang, I. (2016). Arabidopsis BAG1 functions as a cofactor in Hsc70-mediated proteasomal degradation of unimported plastid proteins. *Mol. Plant* 9, 1428–1431. doi: 10.1016/j.molp.2016.06.005

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

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

Letunic, I., Doerks, T., and Bork, P. (2015). SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.* 43, D257–D260. doi: 10.1093/nar/gku949

Li, L., Xing, Y., Chang, D., Fang, S., Cui, B., Li, Q., et al. (2016a). CaM/BAG5/Hsc70 signaling complex dynamically regulates leaf senescence. *Sci. Rep.* 6, 31889. doi: 10.1038/srep31889

Li, Y., Kabbage, M., Liu, W., and Dickman, M. B. (2016b). Aspartyl protease-mediated cleavage of BAG6 is necessary for autophagy and fungal resistance in plants. *Plant Cell* 28, 233–247. doi: 10.1105/tpc.15.00626

Li, M., Li, Y., Guo, L., Gong, N., Pang, Y., Jiang, W., et al. (2017a). Functional characterization of tea (Camellia sinensis) MYB4a transcription factor using an integrative approach. *Front. Plant Sci.* 8, 943. doi: 10.3389/fpls.2017.00943

Li, Y., Williams, B., and Dickman, M. (2017b). Arabidopsis b-cell lymphoma2 (Bcl-2)-associated athanogene 7 (BAG7)-mediated heat tolerance requires translocation, sumoylation and binding to WRKY29. *New Phytol.* 214, 695–705. doi: 10.1111/nph.14388

Li, W. X., Wang, L., He, Z. C., Lu, Z. G., Cui, J. W., Xu, N. T., et al. (2020). Physiological and transcriptomic changes during autumn coloration and senescence in ginkgo biloba leaves. *Hortic. Plant J.* 6, 396–408. doi: 10.1016/j.hpj.2020.11.002

Liang, R., Zhao, J., Li, B., Cai, P., Loh, X. J., Xu, C., et al. (2020). Implantable and degradable antioxidant poly(epsilon-caprolactone)-lignin nanofiber membrane for effective osteoarthritis treatment. *Biomaterials* 230, 119601. doi: 10.1016/j.biomaterials.2019.119601

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

Nam, H. G. (1997). The molecular genetic analysis of leaf senescence. *Curr. Opin. Biotechnol.* 8, 200–207. doi: 10.1016/S0958-1669(97)80103-6

Nawkar, G. M., Maibam, P., Park, J. H., Woo, S. G., Kim, C. Y., Lee, S. Y., et al. (2017). In silico study on arabidopsis BAG gene expression in response to environmental stresses. *Protoplasma* 254, 409–421. doi: 10.1007/s00709-016-0961-3

Park, J. H., Oh, S. A., Kim, Y. H., Woo, H. R., and Nam, H. G. (1998). Differential expression of senescence-associated mRNAs during leaf senescence induced by different senescence-inducing factors in arabidopsis. *Plant Mol. Biol.* 37, 445–454. doi: 10.1023/A:1005958300951

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

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

Rashid Mehmood, R. (2012). Identification and characterization of the bcl-2-associated athanogene (BAG) protein family in rice. *Afr. J. Biotechnol.* 11, 88–98. doi: 10.5897/AJB11.3474

Sierro, N., Battey, J. N., Ouadi, S., Bakaher, N., Bovet, L., Willig, A., et al. (2014). The tobacco genome sequence and its comparison with those of tomato and potato. *Nat. Commun.* 5, 3833. doi: 10.1038/ncomms4833

Song, J., Takeda, M., and Morimoto, R. I. (2001). Bag1-Hsp70 mediates a physiological stress signalling pathway that regulates raf-1/ERK and cell growth. *Nat. Cell Biol.* 3, 276–282. doi: 10.1038/35060068

Takayama, S., and Reed, J. C. (2001). Molecular chaperone targeting and regulation by BAG family proteins. *Nat. Cell Biol.* 3, E237–E241. doi: 10.1038/ncb1001-e237

Takayama, S., Sato, T., Krajewski, S., Kochel, K., Irie, S., Millan, J. A., et al. (1995). Cloning and functional analysis of BAG-1: a novel bcl-2-binding protein with anti-cell death activity. *Cell* 80, 279–284. doi: 10.1016/0092-8674(95)90410-7

Takayama, S., Xie, Z., and Reed, J. C. (1999). An evolutionarily conserved family of Hsp70/Hsc70 molecular chaperone regulators. *J. Biol. Chem.* 274, 781–786. doi: 10.1074/jbc.274.2.781

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Thanthrige, N., Jain, S., Bhowmik, S. D., Ferguson, B. J., Kabbage, M., Mundree, S., et al. (2020). Centrality of BAGs in plant PCD, stress responses, and host defense. *Trends Plant Sci.* 25, 1131–1140. doi: 10.1016/j.tplants.2020.04.012

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882. doi: 10.1093/nar/25.24.4876

Vivancos, P. D., Dong, Y., Ziegler, K., Markovic, J., Pallardo, F. V., Pellny, T. K., et al. (2010). Recruitment of glutathione into the nucleus during cell proliferation adjusts whole-cell redox homeostasis in arabidopsis thaliana and lowers the oxidative defence shield. *Plant J.* 64, 825–838. doi: 10.1111/j.1365-313X.2010.04371.x

Wang, F., Liu, J., Zhou, L., Pan, G., Li, Z., Zaidi, S. H., et al. (2016). Senescence-specific change in ROS scavenging enzyme activities and regulation of various SOD isozymes to ROS levels in psf mutant rice leaves. *Plant Physiol. Biochem.* 109, 248–261. doi: 10.1016/j.plaphy.2016.10.005

Wang, M., Oppedijk, B. J., Lu, X., Van Duijn, B., and Schilperoort, R. A. (1996). Apoptosis in barley aleurone during germination and its inhibition by abscisic acid. *Plant Mol. Biol.* 32, 1125–1134. doi: 10.1007/BF00041396

Williams, B., Kabbage, M., Britt, R., and Dickman, M. B. (2010). AtBAG7, an arabidopsis bcl-2-associated athanogene, resides in the endoplasmic reticulum and is involved in the unfolded protein response. *Proc. Natl. Acad. Sci. U.S.A.* 107, 6088–6093. doi: 10.1073/pnas.0912670107

Wu, A., Allu, A. D., Garapati, P., Siddiqui, H., Dortay, H., Zanor, M. I., et al. (2012). JUNGBRUNNEN1, a reactive oxygen species-responsive NAC transcription factor, regulates longevity in arabidopsis. *Plant Cell* 24, 482–506. doi: 10.1105/tpc.111.090894

Yan, J. Q., He, C. X., and Zhang, H. (2003). The BAG-family proteins in arabidopsis thaliana. *Plant Sci.* 165, 1–7. doi: 10.1016/S0168-9452(03)00121-3

Yang, J., and Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* 43, W174–W181. doi: 10.1093/nar/gkv342

Zhang, L., Liu, J., Cheng, J., Sun, Q., Zhang, Y., Liu, J., et al. (2022). lncRNA7 and lncRNA2 modulate cell wall defense genes to regulate cotton resistance to verticillium wilt. *Plant Physiol.* 189, 264–284. doi: 10.1093/plphys/kiac041

Zhou, H., Li, J., Liu, X., Wei, X., He, Z., Hu, L., et al. (2021). The divergent roles of the rice bcl-2 associated athanogene (BAG) genes in plant development and environmental responses. *Plants (Basel)* 10, 2169. doi: 10.3390/plants10102169

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 09 February 2023
DOI 10.3389/fpls.2023.1078299

OPEN ACCESS

EDITED BY
Mehanathan Muthamilarasan,
University of Hyderabad, India

REVIEWED BY
Parviz Heidari,
Shahrood University of Technology, Iran
Haifeng Liu,
Chonnam National University, Republic of Korea
Huan Luo,
Chungnam National University, Republic of Korea
Zeguang Liu,
University of Geneva, Switzerland

*CORRESPONDENCE
Wang Chen
chenwangchw@163.com
Dongfang Ma
madf@yangtzeu.edu.cn
Zhichuang Lü
lvzhichuang@caas.cn

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

RECEIVED 24 October 2022
ACCEPTED 23 January 2023
PUBLISHED 09 February 2023

CITATION
Tang Y, Yang X, Li H, Shuai Y, Chen W, Ma D and Lü Z (2023) Uncovering the role of wheat magnesium transporter family genes in abiotic responses.
Front. Plant Sci. 14:1078299.
doi: 10.3389/fpls.2023.1078299

Uncovering the role of wheat magnesium transporter family genes in abiotic responses

Yanhong Tang[1,2], Xiaoyue Yang[1,2], Han Li[1,2], Yating Shuai[1], Wang Chen[1]*, Dongfang Ma[1,2]* and Zhichuang Lü[2]*

[1]MARA Key Laboratory of Sustainable Crop Production in the Middle Reaches of the Yangtze River (Co-construction by Ministry and Province)/Engineering Research Center of Ecology and Agricultural Use of Wetland, Ministry of Education/College of Agriculture, Yangtze University, Jingzhou, China, [2]State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China

Background: The CorA / MGT / MRS2 family proteins are an important group of magnesium transporter proteins that maintain magnesium ion homeostasis in plant cells. However, little is known about the MGT functions in wheat.

Methods: The known MGT sequences were used as queries to BlastP against wheat genome IWGSC RefSeq v2.1 assembly (E-value <10−5). Chromosome localization information for each *TaMGT* gene was obtained from the GFF3 file of the wheat genome data (IWGSCv2.1).The sequence of 1500 bp upstream of the *TaMGT* genes was extracted from the wheat genome data. The cis-elements were analyzed using PlantCARE online tool.

Result: A total of 24 *MGT* genes were identified on 18 chromosomes of wheat. After functional domain analysis, only *TaMGT1A*, *TaMGT1B*, and *TaMGT1D* had GMN mutations to AMN, while all the other genes had conserved GMN tripeptide motifs. Expression profiling showed that the *TaMGT* genes were differentially expressed under different stresses and at different growth and development stages. The expression levels of *TaMGT4B* and *TaMGT4A* were significantly up-regulated in cold damage. In addition, qRT-PCR results also confirmed that these *TaMGT* genes are involved in the wheat abiotic stress responses.

Conclusion: In conclusion, The results of our research provide a theoretical basis for further research on the function of *TaMGT* gene family in wheat.

KEYWORDS

TaMGT, gene structure, abiotic stresses, gene expression, fluorescence quantitative PCR, climate-resilience

1 Introduction

Magnesium (Mg^{2+}) is an important macronutrient (Williams and Salt, 2009). As the most abundant divalent cation in living plant cells, Mg^{2+} is involved in chloroplast synthesis, regulation of osmotic pressure, and intracellular enzyme activity (Shaul, 2002; Williams and Salt, 2009; Guo et al., 2016). In addition, Mg^{2+} is essential for the synthesis of proteins and

nucleic acids and maintains the cation-anion balance in cell (Marschner and Marschner, 2012). Moreover, magnesium deficiency can adversely affect plant cells, like reducing macromolecular synthesis, photosynthetic capacity, and plant growth (Hermans et al., 2005; Tang et al., 2012; Yang et al., 2012; Peng et al., 2015; Farhat et al., 2016; Li et al., 2016). Therefore, maintaining the balance and stability of Mg is crucial in plants (Shaul, 2002; Kobayashi and Tanoi, 2015).

Magnesium transporter proteins (MGTs), also known as Magnesium Transporter MRS2, play a critical role in maintaining Mg homeostasis in plant cells (Li et al., 2017). *ZmMGT10* has the ability to transport Mg under conditions of Mg deficiency in Maize (Li et al., 2016). The transgenic Arabidopsis plants with *ZmMGT10* overexpression exhibited vigorous growth, such as larger plant size, longer root length, higher fresh weight and increased chlorophyll content compared with the wild-type plants (Li et al., 2017). Moreover, *AtMGT7* can maintain normal growth and development in Arabidopsis under low Mg conditions (Mao et al., 2008).

Among the magnesium transporters, the CorA-type magnesium transporter is the most widely studied magnesium transporter protein. This type of magnesium transporter was originally identified in *Salmonella typhimurium*, and its mutants exhibit resistance to Co^{2+} growth inhibition (Silver, 1969). CorA is a funnel-shaped homopentamer with two transmembrane structural domains per monomer, the first of which forms the ion conduction pathway. Mg^{2+} transport firstly involves the binding of a cation to an extracellular binding loop connecting the transmembrane structural domains (Maguire, 2006). These proteins have a variable conserved N-terminal hydrophilic domain of about 260 amino acids and a fairly conserved hydrophobic domain of 55 amino acids (Maguire, 2006). The MPEL sequence and the YGMNF sequence are the most characteristic features of CorA proteins, both of which are located in the loop between the transmembrane helices. Among them, the (Gly-Met-Asn) GMN sequence is essential for the protein function (Palombo et al., 2013).

CorA-type Mg transport proteins mediate the influx and efflux of Mg^{2+} (Franken et al., 2022; Papp-Wallace and Maguire, 2008). CorA homologous proteins have been identified in yeast, animals and plants (Knoop et al., 2005). In yeast, the ALR1 family confers Al^{3+} tolerance and encodes the major plasma membrane Mg^{2+} uptake system (Liu et al., 2002). Overexpression of *AtMGT1* in tobacco plants increases Mg concentration and confers tolerance to low Mg environments (Deng et al., 2006).

Mutations in the GMN motif may eliminate Mg^{2+} transport, but the naturally occurring variants GVN and GIN may be associated with the transport of other divalent cations (Eshaghi et al., 2006). The GMN motif was changed to AMN in ZmMGT6 in maize, and in functional complementation experiments using *Salmonella typhimurium* mutant MM281, *ZmMGT6* was shown to rescue the susceptibility of Mg^{2+} in *Salmonella typhimurium mutant* MM281, but its complementation efficacy was lower than that of other ZmMGTs containing GMN motifs (Li et al., 2016). In rice, the GMN motif was changed to AMN in *OsMRS2-4* and *OsMRS2-5* and to GIN in *OsMRS2-8*, and in functional complementation experiments using Yeast mutant CM66, *OsMRS2-4*, *-5*, and *-8* did not observe complementation ability, but other *OsMRS2* genes had obvious complementation ability (Saito et al., 2013). These phenomena suggest that GMN motif mutations affected gene functions. Overall, these features are important markers of Mg^{2+} transporters. (Szegedy and Maguire, 1999; Knoop et al., 2005).

In plants, MGT proteins have been reported in *Arabidopsis thaliana* (Schock et al., 2000), maize (*Zea mays*) (Li et al., 2016), and rice (*Oryza sativa*) (Saito et al., 2013). Recently, it has been studied in tomato (*Solanum lycopersicum*) (Regon et al., 2019), sugarcane (*Saccharum spontaneum*) (Wang et al., 2019), pear (*Pyrus bretschneideri*) (Zhao et al., 2018), citrus (*Poncirus trifoliata*) (Liu et al., 2019), *Vitis vinifera* (Ge et al, 2022), *Citrullus lanatus* and *Cucumis sativus* (Heidari et al, 2022), *Theobroma cacao*, *Corchorus capsularis*, and *Gossypium hirsutum* (Heidari et al, 2021), and *Triticum turgidum* and *Camelina sativa* (Faraji et al, 2021). Studies have shown that Mg transporter proteins have different functions. The pear *PbrMGT7* gene mediates Mg transport between mitochondria and the cytoplasmic matrix (Zhao et al., 2018). The sugarcane *MTG6* gene is the main MGT that maintains the concentration of Mg in chlorophyll and transporting Mg ions into the chloroplast stroma, and *MGT10* transports Mg from roots to leaves over long distances (Wang et al., 2019). Banana *MaMRS2-5* and *MaMRS2-7* are involved in the uptake and transport of Mg and its partitioning between different tissues (Tong et al., 2020). The Arabidopsis *AtMGT5* and *AtMGT9* genes play critical roles in Mg supply during pollen mitosis and pollen in formation (Drummond et al., 2006; Chen et al., 2009; Li et al., 2015). In the Malvaceae family, MGTs appear to be involved in various pathways that control plant growth and development and respond to adverse conditions (Heidari et al, 2021). At the same time, MGTs can also affect the biosynthesis of phytohormones related to stresses by regulating Mg^{2+} concentration, thereby enhancing plant stress resistance (Guo et al., 2014).

Wheat (*Triticum aestivum*) is an important food crop worldwide, and Mg^{2+} deficiency may affect chlorophyll synthesis, multiple enzymes activation, photosynthesis, and partitioning and utilization of photoassimilates, which may inhibit the growth and development of wheat and further lead to yield loss (Shaul, 2002; Cakmak and Yazici, 2010). To date, studies on Mg^{2+} transporters in wheat have not been reported. In this study, the Mg^{2+} transporter genes in wheat were identified, their physicochemical properties, chromosomal location and gene structure were systematically analyzed, and their responses to Mg deficiency, Al stress, and abscisic acid treatment were investigated. The present study will provide a basis for further revealing the biological functions of *MGT* genes in wheat.

2 Materials and methods

2.1 Identification of *MGT* genes in wheat

In this experiment, 10 MGT protein sequences from Arabidopsis, 9 from rice, and 12 from maize were obtained from the Arabidopsis genome database (https://www.arabidopsis.org/), the rice genome database (http://Rice.plantbiology.msu.edu), and the maize genome database (https://www.maizegdb.org/), respectively. The known MGT sequences were used as queries to BlastP against wheat genome IWGSC RefSeq v2.1 assembly (E-value $<10^{-5}$). Duplicates and

mismatched candidate proteins were removed. The conserved protein domains were screened through the Pfam database (http://pfam.xfam.org/) (Finn et al., 2006) and the SMART database (http://smart.embl-heidelberg.de/) (Letunic et al., 2004). TaMGT family members were finally determined after excluding the sequences that did not contain the CorA type domain. Subcellular localization prediction of TaMGTs was performed *via* the Plant-mPLoc online tool (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/). Potential TM regions in each TaMGT protein were predicted using TMHM Server v2.0 (http://www.cbs.dtu.dk/services/TMHMM/) (Krogh et al., 2001). Protein length, average molecular weight, isoelectric point (pI), instability index, and mean hydrophilicity value (GRAVY) were predicted by ExPASy Server10 (SIB Bioinformatics Resource Portal, https://prosite.expasy.org/PS50011).

2.2 Construction of Phylogenetic tree, gene structure and protein motif analysis

The protein sequences of 10 AtMGTs, 9 OsMGTs, 12 ZmMGTs, and TaMGTs identified in present study were collected. Multiple sequence alignment was performed using ClustalW2 (Thompson et al., 1994). The TM structural domain and conserved GMN motifs were annotated using DNAMAN software (Zhao et al., 2012). The phylogenetic tree was constructed using MEGA7.0 software (version 7.0, Mega Limited, Auckland, New Zealand) using Neighbor-joining (NJ) method (Kumar et al., 2016) with the Bootstrap value setting as 1000. Furthermore, the phylogenetic tree was modified *via* the online tool iTOL (version 3.2.317, http://itol.embl.de) (Letunic and Bork, 2019).

To study the gene structures of *TaMGTs* genes, the genome annotation information of TaMGTs was obtained from wheat genome database IWGSC V2.1. And the gene structures were visualized by TBtools (Chen et al., 2020). The conserved protein motifs were analyzed by MEME Suite 5.1.1, with the motif number setting as 20 and other parameters with default values.

2.3 Chromosomal localization and gene duplication events

Chromosome localization information for each *TaMGT* gene was obtained from the GFF3 file of the wheat genome data (IWGSCv2.1). The chromosome distribution map of the *TaMGT* gene was subsequently generated using MapInspect software. MCScanX software was used to analyze *TaMGT* gene duplication events in wheat and the homology of *MGT* genes between wheat and other selected species(Wang et al., 2012). Duplicate gene pairs among *TaMGT* members were identified and visualized using the R package "circlize".

2.4 Cis-elements and transcriptome expression analysis

To identify cis-regulatory elements in the promoter regions, the sequence of 1500 bp upstream of the *TaMGT* genes was extracted from the wheat genome data. The cis-elements were analyzed using PlantCARE online tool (Lescot et al., 2002). The data was collected and visualized using the R package "pheatmap".

The RNA-seq data of *TaMGTs* was downloaded through the NCBI SRA database (The SRA numbers are detailed in Table S3) (Ramírez-González et al., 2018) and the raw data was aligned to the wheat reference genome by Hisat2 (Kim et al., 2015). The genes were then assembled by cufflinks to detect the expression levels of *TaMGTs* (Fragments per Kilobase Million of exon model per Million mapped fragments, FPKM) (Trapnell et al., 2012). Convert all the transcription data through log2, $P<0.05$. Heatmaps were created using the R package "pheatmap" to show the expression patterns of *TaMGT* genes under different conditions.

2.5 Plant materials and stress treatments

Seeds of the hexaploid wheat variety "Yangmai 158" were surface sterilized in 0.5% (w/v) sodium hypochlorite for 15 min. After germinated in a greenhouse at 25°C for 3 d, the seedlings were transferred to Hoagland solution (pH5.8) and were grown under standard greenhouse conditions. The parameters were set as follows: 16h/25°C and 8h/20°C diurnal cycle, 70% relative humidity, and 300 mmolm-2s-1 strong luminosity. The nutrient solution was replaced every two days.

For Mg^{2+} deficiency treatment, wheat seedlings at two leaves with a bud stage were transferred to Hogeland solution with $MgSO_4$-$7H_2O$ deficient. For aluminium chloride stress treatment, wheat seedlings were transferred to a Hoagland solution supplemented with 60uM aluminium chloride (pH4.5). The roots and leaves of treated seedlings were collected at different time points (6h, 12h, 24h, 48h, and 72h after treatment). For hormone treatment, seedlings were treated with 100 mmol/L abscisic acid. Roots and leaves were sampled at 6h, 12h, 24h, and 48h after treatment. Untreated wheat seedlings were used as control. Three biological replicates were set up for each treatment, with each replicate including three technical replicates. The collected samples were immediately frozen in liquid nitrogen and stored at -80° C for subsequent RNA extraction.

2.6 Real-time quantitative RT-PCR analysis

The relative expression levels of *TaMGT* genes in roots and leaves of wheat under Mg^{2+} deficiency (-Mg^{2+}), aluminum stress (+Al), and abscisic acid (+ABA) treatments were analyzed by qRT-PCR. Total RNA was extracted from roots and leaves using TRIzol reagent (Life, USA). cDNA was reverse transcribed using the HiScript II Reverse Transcriptase kit (Vazyme, Nanjing, China).Three biological replicates were performed for each sample, with three technical replicates repeated each. The qPCR primers used in present study were listed in Supplementary Table 1. The *Ta2291* gene, which was expressed stably under various conditions, was used as the internal reference gene (Paolacci et al., 2009). The expression levels of *TaMGT* genes were calculated using $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

3 Results

3.1 Identification of *MGT* genes in wheat

Twenty-four putative *MGT* genes were identified in wheat. The candidate genes were named according to their chromosomal location (Table 1). The predicted molecular weights of TaMGTs were ranged from 39688.7D to 55522.41D. The protein isoelectric points were ranged from 4.74 to 7.26, of which TaMGT1A, TaMGT1B, and TaMGT1D had isoelectric points greater than 6, and more interestingly, they were triplet homologous genes. The protein instability index showed that only TaMGT5D was less than 40, which indicated that it was a stable protein. Analysis of the predicted hydrophobicity of the proteins showed that TaMGT7A and TaMGT7B are hydrophobic proteins. The results of subcellular localization analysis of these 24 proteins showed that 20 of them were widely localized in chloroplasts, and might also be localized in mitochondria and cytoplasm. TaMGT3A.3, TaMGT4A, TaMGT5B and TaMGT7B were only localized in the nucleus (Table 1). The TM domain analysis of TaMGT showed that similar to CorA/MRS2/MGT transport proteins in other species (including fungi), all 24 members contain two hypothetical TM domains at the C-terminus (Supplementary Figure 1) and cora-type domain. Multiple sequence alignment of TaMGTs showed that TM1 was more conserved than TM2. Moreover, the GMN tripeptide motif was found to be present in all members except for TaMGT1A, TaMGT1B, and TaMGT1D, which had the mutated motif AMN (alanine-methionine-aspartate) (Figure 1).

3.2 Construction of phylogenetic tree, protein motif and gene structure analysis

To better understand the evolutionary relationships of TaMGT with other species, phylogenetic trees were constructed for 10 Arabidopsis, 9 rice, 12 maize, and 24 wheat MGTs. As shown in Figure 2, all MGTs were divided into four groups, each group

TABLE 1 TaMGTs protein features and prediction of subcellular location.

Gene name	Gene ID	Chr	Length	MW	TM	DM	pI	I. I	GRAVY	sub
TaMGT1A	*TraesCS1A03G0546100.1*	Chr1A:369306274-369308928	464	50775.05	2	cora	6.43	54.84	-0.167	Chl
TaMGT1B	*TraesCS1B03G0635400.1*	Chr1B:402375260-402377952	510	55522.41	2	cora	7.26	54.69	-0.163	Chl. Nuc
TaMGT1D	*TraesCS1D03G0526100.1*	Chr1D:297360401-297363124	465	50894.13	2	cora	6.23	53.96	-0.166	Chl.
TaMGT2A	*TraesCS2A03G0864500.1*	Chr2A:598758314-598762519	428	47298.31	2	cora	5.17	49.91	-0.048	Chl. Nuc
TaMGT2B	*TraesCS2B03G0965300.1*	Chr2B:544767428-544772321	428	47437.51	2	cora	5.24	49.93	-0.064	Chl. Nuc
TaMGT2D	*TraesCS2D03G0813900.1*	Chr2D:458190668-458194758	428	47407.42	2	cora	5.24	49.56	-0.082	Chl. Nuc
TaMGT3A.1	*TraesCS3A03G0899600.1*	Chr3A:631201551-631203919	408	45518.86	2	cora	5.33	45.03	-0.307	Chl. Nuc
TaMGT3A.2	*TraesCS3A03G0899700.1*	Chr3A:631206505-631210704	405	45290.65	2	cora	4.97	45.48	-0.351	Chl. Nuc.CM
TaMGT3A.3	*TraesCS3A03G0965700.1*	Chr3A:657786841-657791568	451	49332.86	2	cora	4.86	44.76	-0.22	Nuc
TaMGT3B.1	*TraesCS3B03G1023800.1*	Chr3B:665302551-665305027	407	45214.62	2	cora	5.58	43.68	-0.295	Chl. Nuc
TaMGT3B.2	*TraesCS3B03G1023900.1*	Chr3B:665306308-665311671	434	48638.26	2	cora	5.16	44.8	-0.41	Chl. Nuc .CM
TaMGT3B.3	*TraesCS3B03G1109400.1*	Chr3B:704250533-704254822	449	49127.73	2	cora	4.86	44.36	-0.191	Chl.
TaMGT3D.1	*TraesCS3D03G0826200.1*	Chr3D:487956862-487959377	407	45407.67	2	cora	5.26	42.69	-0.316	Chl.Nuc .CM .MN
TaMGT3D.2	*TraesCS3D03G0826300.1*	Chr3D:487960778-487965939	405	45395.67	2	cora	4.9	46.75	-0.358	Chl. Nuc .CM
TaMGT3D.3	*TraesCS3D03G0899300.1*	Chr3D:523593005-523597470	450	49246.81	2	cora	4.86	44.29	-0.208	Chl. Nuc
TaMGT4A	*TraesCS4A03G0733600.1*	Chr4A:590707040-590712507	448	49578.14	2	cora	5.43	57.92	-0.144	Nuc
TaMGT4B	*TraesCS4B03G0060700.1*	Chr4B:23432237-23437771	448	49639.31	2	cora	5.64	56.78	-0.151	Chl. Nuc
TaMGT4D	*TraesCS4D03G0040500.1*	Chr4D:11323563-11329187	447	49671.36	2	cora	5.71	57.78	-0.161	Chl. Nuc .CM
TaMGT5A	*TraesCS5A03G0916900.1*	Chr5A:582177688-582183695	392	43092.65	2	cora	4.85	40.17	-0.139	Chl. Nuc
TaMGT5B	*TraesCS5B03G0965000.1*	Chr5B:569002453-569007864	388	42905.33	2	cora	4.97	43.28	-0.218	Nuc
TaMGT5D	*TraesCS5D03G0876100.1*	Chr5D:463082622-463089037	388	42764.28	2	cora	4.85	39.98	-0.158	Chl. Nuc
TaMGT7A	*TraesCS7A03G1229000.1*	Chr7A:698716978-698721829	443	49161.28	2	cora	5.18	50.08	0.001	Chl
TaMGT7B	*TraesCS7B03G1115300.1*	Chr7B:689902980-689907578	348	39688.7	2	cora	4.74	50.65	0.082	Nuc
TaMGT7D	*TraesCS7D03G1169200.1*	Chr7D:604391025-604395163	445	49444.61	2	cora	5.18	50.82	-0.002	Chl

Chr, Chromosomal Location; Length, Amino acid length; MW, Molecular weight; pI, Isoelectric point; TM, Transmembranestructural domain; DM, cora domain; I.I, Instability index; GRAVY, Grand average of hydropathy; Sub, Subcellular localization; Nuc, Nucleus; Chl, Chloroplast; CM, Cytoplasm; MN, Mitochondrion.

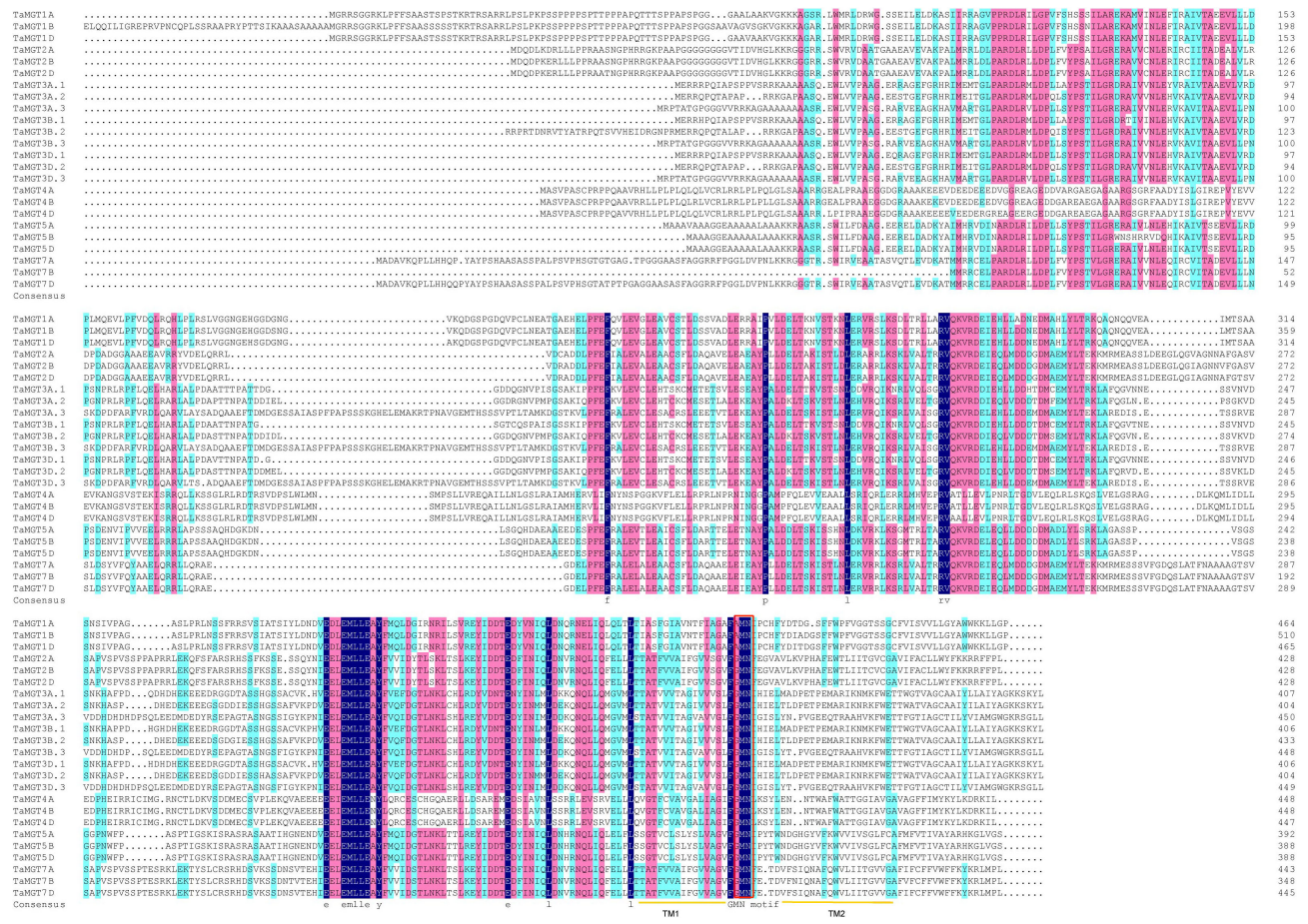


FIGURE 1
Multiple sequence alignment of TaMGTs. Multiple alignments were performed using DNAMAN software. TM domains are marked with orange lines. Conserved GMN motif is indicated by Red box.

contained members from different plant species. Among them, the second group with the most proteins also has the most members of TaMGTs. The phylogenetic tree (Figure 3) showed that the 24 *TaMGTs* were also divided into four groups, which is consistent with the phylogenetic tree results of the above four plant species. Moreover, members from the same subgroup have similar motif types and structures. It is worth noting that *TaMGT1A*, *TaMGT1B*, *TaMGT1D* in Group IV lacked Motif 11, *TaMGT7B* in Group I lacked Motif 12, and *TaMGT7A*, *TaMGT7B*, *TaMGT7D* in Group I contained two Motif 6. The position of the motif varies from subgroup to subgroup, for example, Motif 9 is located at the 5' end in group 3 and at the 3' end in other subgroups. Motif 1 and Motif 6 were located in the functional structural domain of TaMGTs.

The results of the gene structure analysis (Figure 3C) showed that there are differences in exons/introns among the members in different subgroups. Almost all members of subgroup I and subgroup II contain 4-6 exons/introns and 2 non-coding regions, such as *TaMGT2B*, *TaMGT3B.3*, and *TaMGT3D.2*. Individual *TaMGT* genes do not have non-coding regions, such as *TaMGT7D*, *TaMG1A*, and *TaMG1B*. While in the fourth subgroup, *TaMGT4D*, *TaMGT4B*, and *TaMGT4A* contain 13 exons, 12 introns and 2 non-coding regions.

3.3 Chromosomal localization and gene duplication events

Based on the GFF3 annotation file, the chromosomal location of *TaMGT* genes were mapped using MapInspect software (Figure 4A). 24 *TaMGT* genes were distributed on 18 chromosomes. *TaMGT* genes were evenly distributed among the three subgenomes of wheat (each subgenome contained 8 genes). However, the distribution of the genes is uneven among the different chromosomes, with three *TaMGT* genes distributed on chromosome 3 and only one on the other chromosomes.

Gene duplication is one of the most important mechanisms by which organisms acquire new genes and create gene novelty (Song et al., 2019). Gene duplication consists of both tandem duplication and segmental duplication (Vision et al., 2000). In order to obtain the

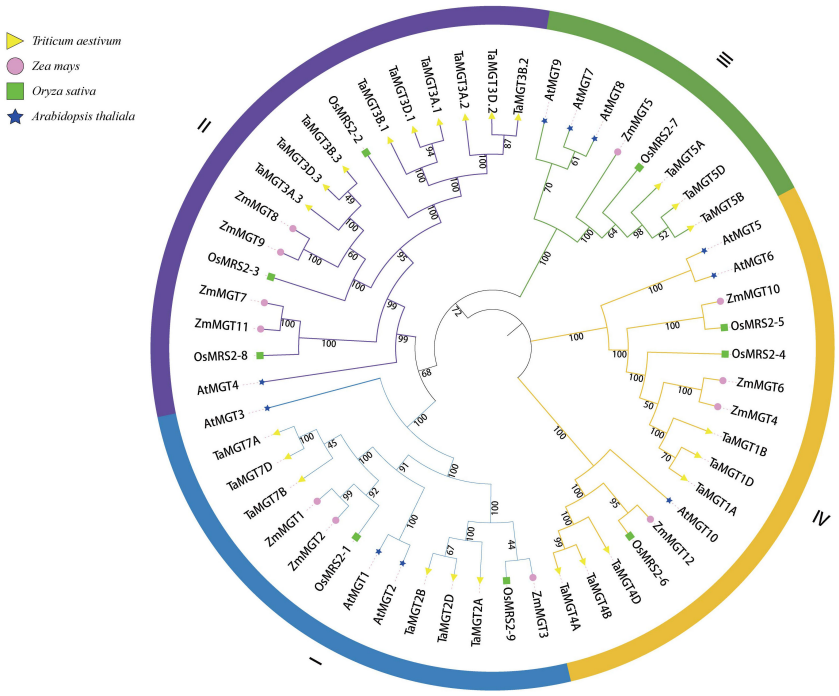


FIGURE 2
Phylogenetic analysis of Arabidopsis, rice, maize and wheat CorA/MRS2/MGT members. Phylogenetic trees were constructed by the neighbor-Joining method (NJ) using MEGA software for 24 TaMGT proteins, 11 Arabidopsis MRS2/MGT proteins, 9 rice proteins, and 12 maize proteins.

amplification mechanism of the *TaMGT* genes, the collinearity among *TaMGT* genes was analyzed by MCScanX software (Figure 4B). The results showed that there were twelve pairs of segmental duplication events among *TaMGT* genes.And most of the duplicated *TaMGT* genes were located on different chromosomes, and no tandem duplication relationships were found among *TaMGT* members on the same chromosome. These results suggest that the *TaMGT* genes may have undergone segmental replications during evolution.

To further analyze the evolutionary and homologous relationships of the MGT family in wheat, collinearity analysis of wheat with *Arabidopsis thaliana* and *Triticum urartu* were constructed. The results (Figure 4C) showed that there were 14 *MGT* homologous gene pairs in wheat and its close relative *Triticum urartu*, while there were only one *MGT* homologous gene pair in wheat and dicotyledonous *Arabidopsis*, which might be due to the short differentiation distance between wheat and its close relatives, with fewer events such as gene loss, insertion and transposition.

3.4 Analysis of Cis-acting elements and transcriptome expression analysis

Cis-acting regulatory elements were bound by appropriate transcription factors to control gene transcription (Liu et al., 2013). In this experiment, we identified and summarized the cis-acting elements in the promoter region of the *TaMGT* gene family related to growth and development, biotic stress and phytohormones (Figure 5 and Supplemental Table 2), and found that there were differences in the number and types of elements contained in the

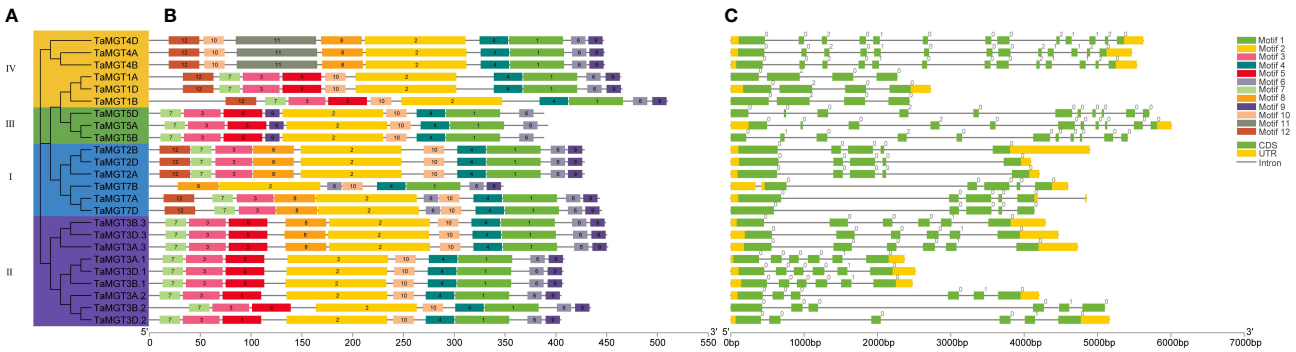


FIGURE 3
Phylogenetic relationships **(A)**, motif analysis **(B)** and gene structure **(C)** of TaMGT family members.

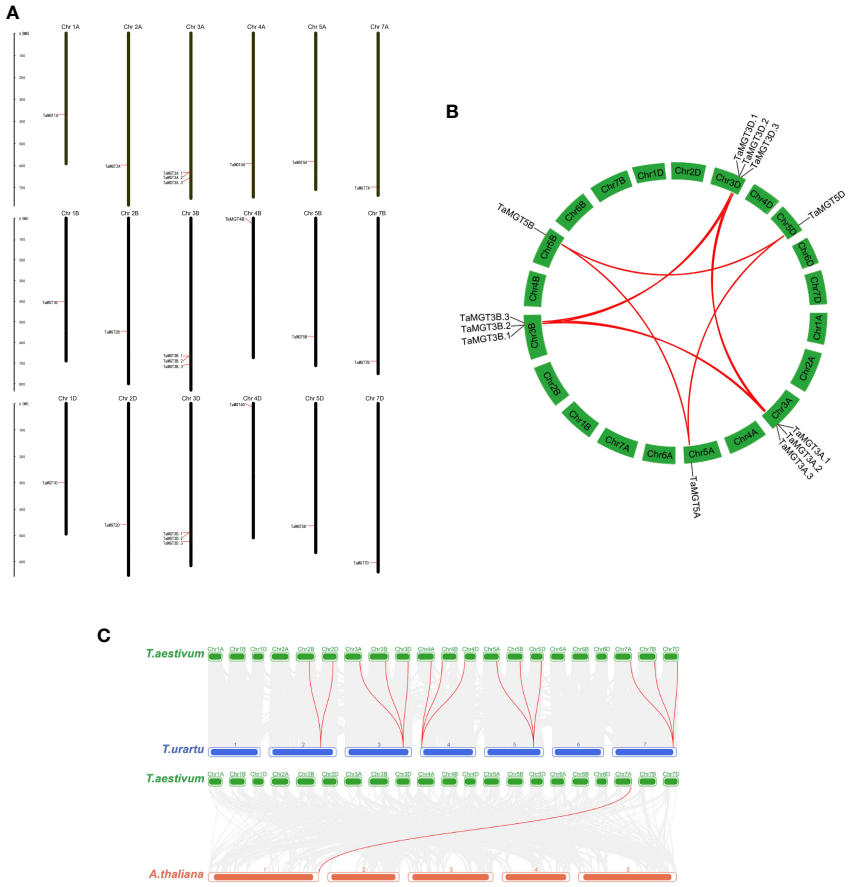


FIGURE 4
Chromosomal localization and gene duplication of TaMGTs genes. **(A)** Location of TaMGTs genes on chromosomes, with black representing chromosomes. **(B)** Gene duplication events in the wheat MGT gene family, where the curves represent gene duplication between chromosomes; the collinearity genes on chromosome 3A, 3B, 3C are highlighted with red lines. **(C)** Collinearity analysis of wheat with Arabidopsis thaliana and Triticum urartu. The gray line in the background indicates a collinear block in the genome of wheat and Arabidopsis thaliana and Triticum urartu, while the red line highlights the homologous MGT gene pair.

promoters of the *TaMGT* family. *TaMGT3D.3* promoter contained 69 cis-acting elements, while *TaMGT7B* just contained 20. Among the cis-elements associated with growth and development, the TATA-box was the most. Among the hormone-responsive cis-elements, TGACG-motif and CGTCA-motif were involved in the regulation of the methyl jasmonate response. Among the biotic and abiotic stress-related cis-elements, WUN-motif is a wound responsive element, MYB is a drought-inducibility element, ARE and GC-motif are an anaerobic induction element, and LTR is a low-temperature response element (Bang et al., 2013; Banerjee et al., 2013; Mittal et al., 2009; Kovalchuk et al., 2013; Li et al., 2022). These results suggest that this gene family may also play an important function in the response to adversity stresses.

We used available wheat RNA-seq data (Supplemental information: Tables S3, S4) to analyze the expression levels of the *TaMGT* gene at different growth and developmental stages of wheat, and the results were shown in Figure 6. These *TaMGTs* were expressed in grains, headings, stems, leaves, roots, and seeds. According to the clustering results of their expression levels, these *TaMGTs* can be divided into four groups, in which the genes of Groups 2 and 3 were expressed or lowly expressed in all tissues, while the expression levels of genes in Group 4 were much higher. For example, *TaMGT5D*, *TaMGT5A*, and *TaMGT5B* genes were highly expressed in all five tissues; *TaMGT3D.3* and *TaMGT3A.3* were abundantly expressed in stems, but not expressed or lowly expressed in other tissues, indicating that *TaMGT3D.3* and *TaMGT3A.3* may play important roles in stem development. In addition, the expression levels of these genes in the seeds also varied, with ten genes expressed in small amounts and the remaining genes hardly expressed. These results suggest that the differential expression of *TaMGTs* may play important but distinct roles in different tissues.

In addition to detecting the expression levels of each *TaMGTs* in different tissues, we also analyzed their expressions at different developmental stages. Based on the clustering results of their expression levels, *TaMGT4A* and *TaMGT4B* were highly expressed at the stem axis-seedling stage, the first leaf-seedling stage, the third leaf-three-leaf stage, the first leaf tillering stage, the root-three-leaf stage, and the flag leaf-30% Ear stage. While almost no expression was found in other periods. These two genes may be mainly involved in the regulation of these six tissue developmental stages in wheat.

Next, we analyzed the expression differences of *TaMGTs* under biotic and abiotic stresses. As shown in Figure 6C, all *TaMGTs* genes were not significantly induced under drought stress. In drought stress

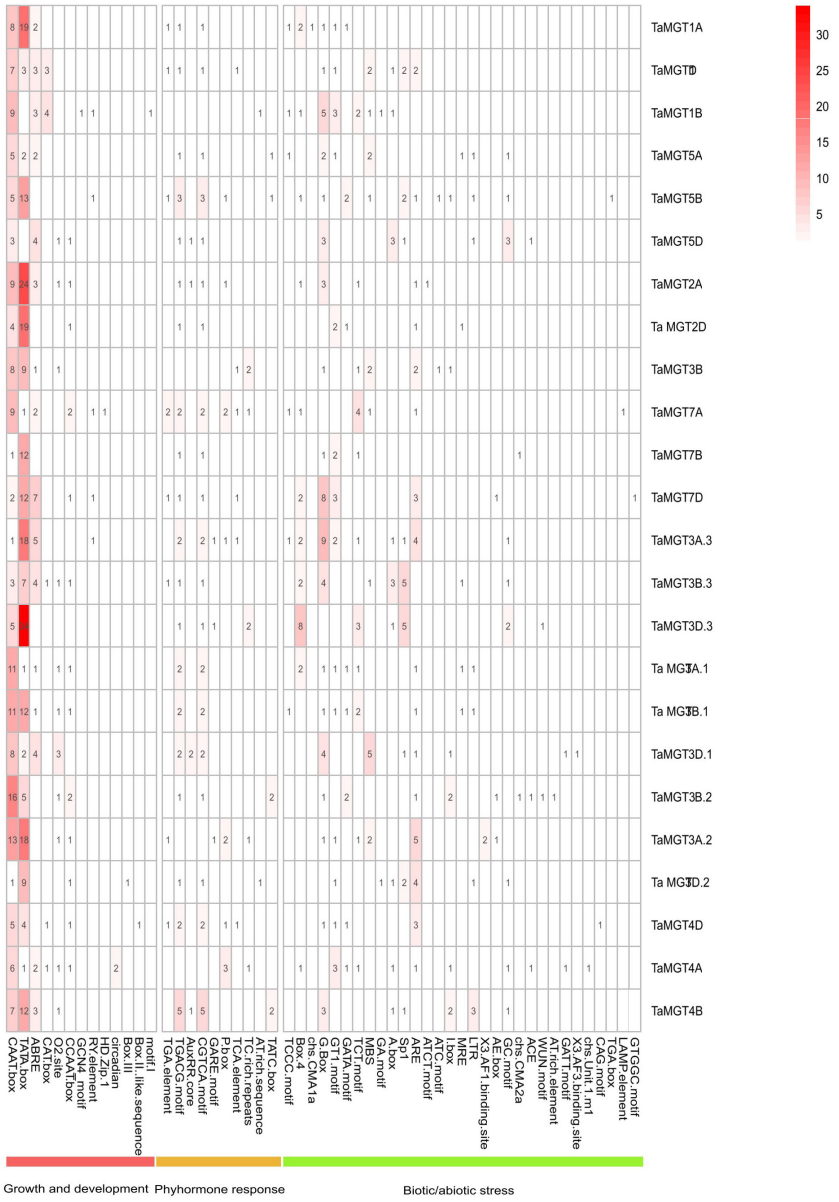


FIGURE 5
Cis-acting elements analysis of TaMGT genes.

treatment, the expression level of *TaMGT5A*, *TaMGT3A.3*, and *TaMGT3D.3* genes gradually increased with the increaseing of treatment time, while the expression of *TaMGT4A* and *TaMGT4B* genes gradually decreased. It can be seen that the genes in the 4 groups were affected to a certain extent by the increase of processing time. The expression levels of *TaMGT3D.3*, *TaMGT3A.3*, and *TaMGT3B.3* increased significantly with the prolongation of drought stress, heat stress, and drought and heat stress, indicating that these genes may be related to the degree of abiotic stress in plants.

After cold treatment, the expression of *TaMGT5D*, *TaMGT5B*, *TaMGT4A*, and *TaMGT4B* genes was significantly upregulated, which indicated that these genes might play positive roles against cold stress. Among them, *TaMGT4A* and *TaMGT4B* were barely expressed in drought stress, thermal stress, and drought and heat stress, and it is speculated that these two genes may be related to the cold resistance of plants.

In addition, under biotic stress, most *TaMGTs* had different gene expression patterns after *Magnaporthe oryzae* infection, indicating that these *TaMGTs* responded differently to rice blast (Figure 6D). The genes in Group 1 and 3 were unexpressed or low-expressed in all tissues, but those in Group 2 and 4 were expressed at much higher levels and responded significantly to different treatments. After the fifth day of infection, the gene expression level of *TaMGT3B.2*, *TaMGT3A.2*, *TaMGT3D.2*, *TaMGT3B.3*, *TaMGT5D*, and *TaMGT5B* were significantly down-regulated. However, with the increase of infection time, the expression levels of *TaMGT4A* and *TaMBT4B* were gradually increased. Therefore, we speculate that *TaMGT4A* and *TaMGT4B* may be involved in the response to *Magnaporthe oryzae* infection.

When plants were inoculated with powdery mildew and stripe rust, the expression patterns of *TaMGT* genes were similar in response to these two diseases. As the time of infection increases,

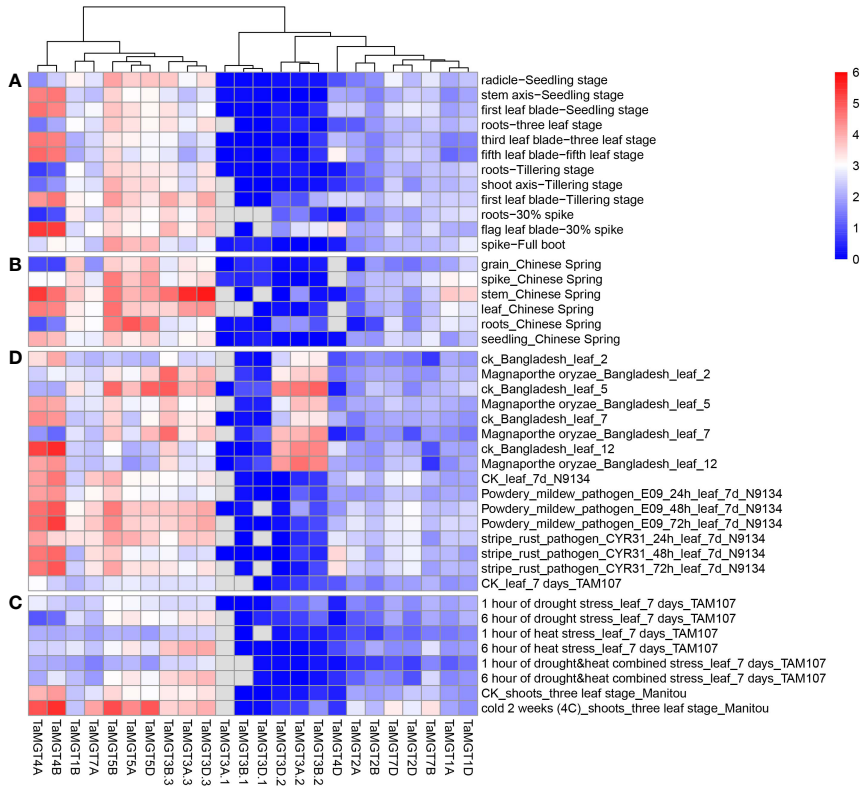


FIGURE 6
(A) different growth and developmental stages expression analysis **(B)** different tissues expression analysis **(C)** abiotic stresses expression analysis **(D)** biotic stresses expression analysis.

the expression level of some genes were increased significantly. Among them, *TaMGT4A* and *TaMGT4B* had significant responses to powdery mildew and stripe rust. Combined with the previous analysis, it is speculated that these two genes are mainly related to the regulation of plant disease resistance.

3.5 Response characteristics of *TaMGT* family genes to Mg^{2+} deficiency, AL stress, and ABA treatment

We analyzed the expression of *TaMGT* gene in roots and shoots under Mg^{2+} deficiency conditions using qRT-PCR. The results showed that under the condition of Mg^{2+} deficiency, the expression level of *TaMGT1A* was strongly upregulated and peaked at 48h in the leaf and peaked at 6h in the root, and *TaMGT5B* peaked at 24h in leaf. Compared with the untreated group, *TaMGT3B.3*, *TaMGT4B* and *TaMGT7A* showed a downward trend in leaves and roots. These results indicated that the *TaMGT* genes responded differently to magnesium deficiency.

Compared with the control group, the gene expression levels of *TaMGT1B* and *TaMGT3B* reached the maximum at 24h and 48h respectively (Figure 7) after the plant leaves were exposed to aluminum stress, which was about 10 and 2.75 times higher than those of the control group, respectively, indicating that these two genes may be involved in the regulation of aluminium stress. In addition, the gene expression levels of *TaMGT4B*, *TaMGT5B*, and *TaMGT7A* were significantly decreased during each period, and the expression levels of *TaMGT4B* and *TaMGT7A* were gradually down-regulated before 48h, which may be caused by the stress response of plants.

Compared with the control group, the expression of all genes was inhibited after the roots of the plant were subjected to aluminum stress. The expression levels of *TaMGT5B* and *TaMGT7A* showed a significantly downward trend and began to level-off after 48h. The expression levels of *TaMGT3B.3* and *TaMGT7A* showed the same trend at 6h, 12h, 24h and 48h, with *TaMGT7A* levelling off after 72h and *TaMGT3B.3* having a continued up-regulation trend with increasing treatment time. Gene expression in roots and leaves indicated that gene expression was inhibited under Al stress.

The expression patterns of *TaMGTs* showed (Figure 7) that after 24 h of abscisic acid treatment, *TaMGT5B* and *TaMGT7A* were down-regulated in the leaves and *TaMGT4B* was down-regulated in the roots. The expression levels of other *TaMGT* genes peaked in both leaves and roots 24 h after abscisic acid treatment. The expression levels of most *TaMGT* genes showed a tendency to descend and then rise under abscisic acid treatment. The above results show that the *TaMGT* family gene has a different role in wheat response to abscisic acid treatment.

4 Discussion

Although the magnesium transporter protein family genes have been extensively studied in a large number of plant species(Schock

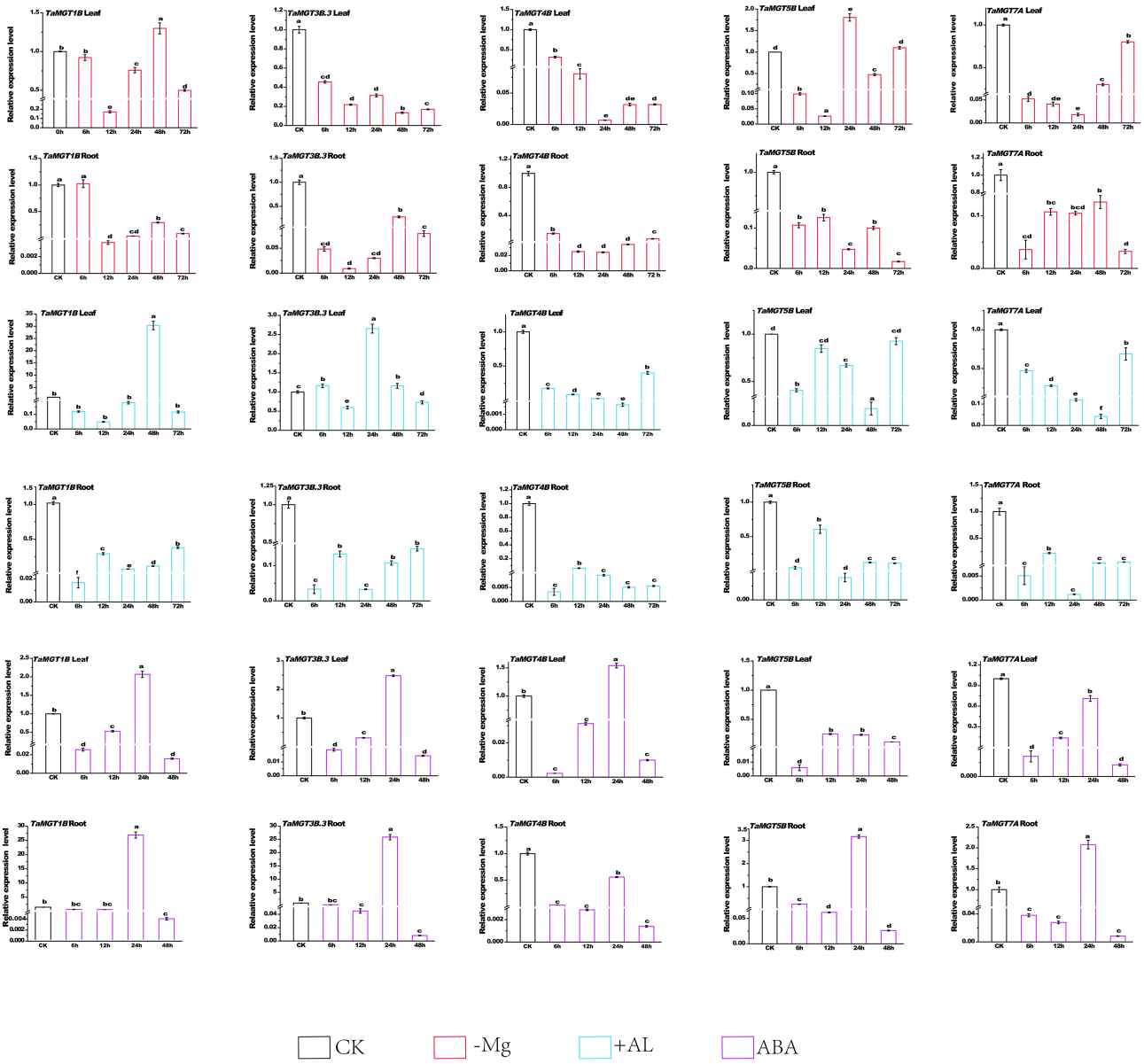


FIGURE 7
Expression patterns of five *TaMGT* genes under three different treatments. Three groups of treatments included 60uM aluminum chloride, 100 mmol/L ABA treatment, and magnesium deficiency treatment. The Y-axis represents the relative expression level, and the X-axis represents the time point of treatment. Significant differences at P<0.05 are indicated by letters. The expression levels of *TaMGT* genes were mapped using Origin software.

et al., 2000; Saito et al., 2013; Zhao et al., 2018; Liu et al., 2019; Heidari et al, 2022), no reports in wheat was published. In this study, 24 *TaMGT* genes were identified for the first time in the wheat genomeusing bioinformatics methods.

The results of gene structure and motif analysis showed that the deletion of Motif 11 in *TaMGT1A*, *TaMGT1B*, *TaMGT1D* and the deletion of Motif 12 in *TaMGT7B* may be caused by gene recombination (Melamed-Bessudo et al., 2016), while further experiments are needed to prove it. In addition, members of the MGT family had variations in the number of exons. The exon number can increase the diversity of gene-coding proteins by influencing post-transcriptional processes, such as alternative splicing (Koralewski and Krutovsky, 2011). Those with fewer exons can activate rapidly in response to stress, and these genes play a stronger role in adapting to adverse environmental conditions (Jeffares et al., 2008; Jiang et al., 2022).There were segmental duplication relationships between genes involved in gene duplication in *TaMGT* gene family, with no signs of tandem duplication relationships (Figure 4B), which is consistent with the previous report of interrelationships of genes in the *MGT* gene family in other species (Yan et al., 2019). Therefore, it is speculated that the expansion of MGT family caused by segmental duplication plays an important role in the adaptation of wheat to external environmental changes.

Cis-regulatory elements in promoter regions play important roles in the regulation of gene expression (Hernandez-Garcia and Finer, 2014). *MGT* genes have a high potential to respond to stresses, including hormones, growth and development, cis-related elements of biological stress and abiotic stress. It shows that this family plays a

major role in the growth and development of wheat and adapts to the external environments (such as changes in light signal, hormones and stress), and also indicates that there are multiple factors regulating the expression of *TaMGT* gene. Among them, the TaMGT promoter contained the growth and development response elements was the largest, indicating that *TaMGT* gene may affect the growth and development of wheat.

Differential analysis of the wheat transcriptome revealed significant differences in the *TaMGT* family of genes under diverse environmental stresses. *TaMGT4A* and *TaMGT4B* maintained high expression levels under different stress conditions. However, *TaMGT3B.2*, *TaMGT3A.2*, and *TaMGT3D.2* were only expressed under the stress of rice blast. There are few studies on magnesium transporters in response to disease stress. Watermelon *MSR2* gene were upregulated in response to mosaic virus stress(Heidari et al, 2022). *Arabidopsis MGT* gene, *AT2G21120*, which was recognized using a forward genetic screen, was involved in response to viral infection (Guo et al., 2018). Further studies of the functions of these genes are needed to elucidate the specific mechanisms involved in these phenomena.

Magnesium deficiency has been reported to have an effect on *MGT* gene expression (Li et al., 2016; Tong et al., 2020), so we analyzed the expression levels of the *MGT* gene in wheat under the condition of magnesium deficiency by qRT-PCR. The results showed that the expression levels of five *TaMGT* genes were significantly different in wheat leaves and roots. Studies have shown that *AtMGT6* and *ZmMGT10* are essential for maintaining Mg homeostasis and are highly expressed in roots under Mg deficiency (Mao et al., 2014; Li et al., 2017; Tang and Luan, 2017). Phylogenetic analysis showed that *TaMGT1B*, a homologue of *AtMGT6* and *ZmMGT10*, was significantly up-regulated in Mg-deficient roots and leaves. *AtMGT6* mediates Mg2+ uptake in roots and is required for plant adaptation to a low-Mg^{2+} environment (Mao et al., 2014). Whether *TaMGT1B* has the same function needs further experiments. *AtMGT7* is involved in Mg uptake and can maintain normal Arabidopsis growth and development under low Mg or Mg-deficient conditions (Mao et al., 2008). *TaMGT5B*, a homolog of *AtMGT7* and *OsMRS2-7*, was highly upregulated in leaves. It has been shown that branch A of *OSMRS2-6* and *AtMRS2-11* has been defined as chloroplast transporter group (Saito et al., 2013), *TaMGT4A*, *TaMGT4B*, *TaMGT4D*, *OSMRS2-6*, and *AtMRS2-11* were belonged to the same branch (Figure 2), and they may be related to chloroplast metabolism.

Studies have demonstrated that magnesium transporters can alleviate the toxic of Al in the absence of magnesium. Overexpression of Mg transporters (ALR1 or ALR2) in *Saccharomyces cerevisiae* confers Al tolerance in yeast (MacDiarmid and Gardner, 1998). The *OsMGT1* gene, a plasma membrane-localized transporter of Mg in rice, confers Al tolerance by increasing Mg uptake into cells (Chen et al., 2012). Most of the *MGT* genes in Arabidopsis were not up-regulated by Al treatment (Hoekenga et al., 2003; Sawaki et al., 2009; Zhao et al., 2009). In this study, it was found that most of the *MGT* genes in wheat were also not up-regulated by Al. This is consistent with related reports that rice is highly tolerant to Al toxicity (Chen et al., 2012).The transgenic lines for *AtMGT1* in *Nicotiana benthamiana* showed a reduction in Al toxicity (Deng et al., 2006). It seems that increasing MGT activity plays an important role in reducing the negative effects of some elements and ions. In the sugarcane MGT family, different genes were expressed differently under different ABA treatment times. The five *TaMGT* genes (*TaMGT1B*, *TaMGT3B.3*, *TaMGT4B*, *TaMGT5B* and *TaMGT7A*) in wheat have the same expression pattern at 12-48h under abscisic acid treatment, which may indicate that these five genes play some function in response to ABA. However, the exact functions performed need to be further verified.

In summary, this study identified 24 MGT-type magnesium transporter proteins and provided a detailed analysis of their classification, protein structure, evolutionary relationships, gene structure and localization. Changes in the CorA GMN motif in *Salmonella typhimurium* have been used as zinc transporter proteins, which may mediate the entry and exit of Zn but are unable to transport Mg^{2+} (Worlock and Smith, 2002). Given this situation, it remains to be demonstrated whether mutation of the GMN motif to AMN in wheat has a function in transporting Mg. Transcriptomic analysis and qRT-PCR analysis showed that *TaMGTs* play an important role in responding to diverse abiotic stresses. *TaMGT* genes are differentially expressed under Mg-deficient conditions, aluminium stress and abscisic acid treatment. Among them, *TaMGT1B* may play a prominent function in maintaining Mg homeostasis, Mg uptake and participation in hormone response in plants. The results of this study provide a theoretical basis for an in-depth study of the function of the *TaMGT* gene family in wheat.

5 Conclusions

In this study, the classification, protein structure, evolutionary relationship, gene structure and localization of 24 *TaMGT* genes in wheat were analyzed in detail. The quantitative PCR results suggested that *TaMGT* plays a role in maintaining magnesium homeostasis, magnesium absorption and hormone response in wheat, as well as in response to biological and non-organic important role in biotic stress. The results provide a theoretical basis for further research on the function of *TaMGT* gene family in wheat.

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

ZL, DM, and WC conceived the study and worked on the approval of the manuscript. YT, XY, and HL performed the experiments and wrote the first draft. WC, DM, and ZL revised the manuscript. YS contributed to data analysis. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Open Program of Engineering Research Center of Ecology and Agricultural Use of Wetland Ministry of Education (KFT202103) and the open project program of State Key Laboratory for Biology of Plant Diseases and Insect Pests (NO. SKLOF202008), and the National Natural Science Foundation of China (31672088).

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.2023.1078299/full#supplementary-material

References

Banerjee, J., Sahoo, D. K., Dey, N., Houtz, R. L., and Maiti, I. B. (2013). An intergenic region shared by At4g35985 and At4g35987 in arabidopsis thaliana is a tissue specific and stress inducible bidirectional promoter analyzed in transgenic arabidopsis and tobacco plants. *PLos One* 8, e79622. doi: 10.1371/journal.pone.0079622

Bang, S. W., Park, S. H., Jeong, J. S., Kim, Y. S., Jung, H., Ha, S. H., et al. (2013). Characterization of the stress-inducible OsNCED3 promoter in different transgenic rice organs and over three homozygous generations. *Planta* 237, 211–224. doi: 10.1007/s00425-012-1764-1

Cakmak, I., and Yazici, A. M. (2010). Magnesium: a forgotten element in crop production. *Better. Crops* 94, 23–25.

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, J., Li, L. G., Liu, Z. H., Yuan, Y. J., Guo, L. L., Mao, D. D., et al. (2009). Magnesium transporter AtMGT9 is essential for pollen development in arabidopsis. *Cell Res.* 19, 887–898. doi: 10.1038/cr.2009.58

Chen, Z. C., Yamaji, N., Motoyama, R., Nagamura, Y., and Ma, J. F. (2012). Up-regulation of a magnesium transporter gene OsMGT1 is required for conferring aluminum tolerance in rice. *Plant Physiol.* 159, 1624–1633. doi: 10.1104/pp.112.199778

Deng, W., Luo, K., Li, D., Zheng, X., Wei, X., Smith, W., et al. (2006). Overexpression of an arabidopsis magnesium transport gene, AtMGT1, in nicotiana benthamiana confers Al tolerance. *J. Exp. Bot.* 57, 4235–4243. doi: 10.1093/jxb/erl201

Drummond, R. S. M., Tutone, A. F., Li, Y.-C., and Gardner, R. C. (2006). A putative magnesium transporter AtMRS2-11 is localized to the plant chloroplast envelope membrane system. *Plant Sci.* 170, 78–89. doi: 10.1016/J.PLANTSCI.2005.08.018

Eshaghi, S., Niegowski, D., Kohl, A., Martinez Molina, D., Lesley, S. A., and Nordlund, P. (2006). Crystal structure of a divalent metal ion transporter CorA at 2.9 angstrom resolution. *Science* 313, 354–357. doi: 10.1126/science.1127121

Faraji, S., Ahmadizadeh, M., and Heidari, P. (2021). Genome-wide comparative analysis of mg transporter gene family between *Triticum turgidum* and *Camelina sativa*. *Biometals* 34, 639–660. doi: 10.1007/s10534-021-00301-4

Farhat, N., Elkhouni, A., Zorrig, W., Smaoui, A., Abdelly, C., and Rabhi, M. (2016). Effects of magnesium deficiency on photosynthesis and carbohydrate partitioning. *Acta Physiol. Plant* 38, 145. doi: 10.1007/s11738-016-2165-z

Finn, R. D., Mistry, J., Schuster-Böckler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., et al. (2006). Pfam: clans, web tools and services. *Nucleic Acids Res.* 34, D247–D251. doi: 10.1093/nar/gkj149

Franken, G. A. C., Huynen, M. A., Martínez-Cruz, L. A., Bindels, R. J. M., and de Baaij, J. H. F. (2022). Structural and functional comparison of magnesium transporters throughout evolution. *Cell. Mol. Life Sci.* 79, 418. doi: 10.1007/s00018-022-04442-8

Ge, M., Zhong, R., Sadeghnezhad, E., Hakeem, A., Xiao, X., Wang, P., et al. (2022). Genome-wide identification and expression analysis of magnesium transporter gene family in grape (Vitis vinifera). *BMC Plant Biol.* 22 (1), 217. doi: 10.1186/s12870-022-03599-5

Guo, W., Cong, Y., Hussain, N., Wang, Y., Liu, Z., Jiang, L., et al. (2014). The remodeling of seedling development in response to long-term magnesium toxicity and regulation by ABA-DELLA signaling in arabidopsis. *Plant Cell Physiol.* 55, 1713–1726. doi: 10.1093/pcp/pcu102

Guo, W., Nazim, H., Liang, Z., and Yang, D. (2016). Magnesium deficiency in plants: An urgent problem. *Crop J.* 4, 83–91. doi: 10.1016/J.CJ.2015.11.003

Guo, Z., Wang, X. B., Wang, Y., Li, W. X., Gal-On, A., and Ding, S. W. (2018). Identification of a new host factor required for antiviral RNAi and amplification of viral siRNAs. *Plant Physiol.* 176, 1587–1597. doi: 10.1104/pp.17.01370

Heidari, P., Abdullah,, Faraji, S., and Poczai, P. (2021). Magnesium transporter gene family: Genome-wide identification and characterization in *Theobroma cacao*, *Corchorus capsularis*, and *Gossypium hirsutum* of family malvaceae. *Agronomy* 11, 1651. doi: 10.3390/agronomy11081651

Heidari, P., Puresmaeli, F., and Mora-Poblete, F. (2022). Genome-wide identification and molecular evolution of the magnesium transporter (MGT) gene family in *Citrullus lanatus* and *Cucumis sativus*. *Agronomy* 12, 2253. doi: 10.3390/agronomy12102253

Hermans, C., Bourgis, F., Faucher, M., Strasser, R. J., Delrot, S., and Verbruggen, N. (2005). Magnesium deficiency in sugar beets alters sugar partitioning and phloem loading in young mature leaves. *Planta* 220, 541–549. doi: 10.1007/s00425-004-1376-5

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

Hoekenga, O. A., Vision, T. J., Shaff, J. E., Monforte, A. J., Lee, G. P., Howell, S. H., et al. (2003). Identification and characterization of aluminum tolerance loci in arabidopsis (Landsberg erecta × Columbia) by quantitative trait locus mapping. a physiologically simple but genetically complex trait. *Plant Physiol.* 132, 936–948. doi: 10.1104/pp.103.023085

Jeffares, D. C., Penkett, C. J., and Bähler, J. (2008). Rapidly regulated genes are intron poor. *Trends Genet.* 24, 375–378. doi: 10.1016/j.tig.2008.05.006

Jiang, B., Liu, Y., Niu, H., He, Y., Ma, D., and Li, Y. (2022). Mining the roles of wheat (*Triticum aestivum*) SnRK genes in biotic and abiotic responses. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.934226

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

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Knoop, V., Groth-Malonek, M., Gebert, M., Eifler, K., and Weyand, K. (2005). Transport of magnesium and other divalent cations: evolution of the 2-TM-GxN proteins in the MIT superfamily. *Mol. Genet. Genom.* 274, 205–216. doi: 10.1007/s00438-005-0011-x

Kobayashi, N. I., and Tanoi, K. (2015). Critical issues in the study of magnesium transport systems and magnesium deficiency symptoms in plants. *Int. J. Mol. Sci.* 16, 23076–23093. doi: 10.3390/ijms160923076

Koralewski, T. E., and Krutovsky, K. V. (2011). Evolution of exon-intron structure and alternative splicing. *PLos One* 6, e18055. doi: 10.1371/journal.pone.0018055

Kovalchuk, N., Jia, W., Eini, O., Morran, S., Pyvovarenko, T., Fletcher, S., et al. (2013). Optimization of TaDREB3 gene expression in transgenic barley using cold-inducible promoters. *Plant Biotechnol. J.* 11, 659–670. doi: 10.1111/pbi.12056

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315

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

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

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

Letunic, I., Copley, R. R., Schmidt, S., Ciccarelli, F. D., Doerks, T., Schultz, J., et al. (2004). SMART 4.0: towards genomic data integration. *Nucleic Acids Res.* 32, D142–D144. doi: 10.1093/nar/gkh088

Li, H., Du, H., Huang, K., Chen, X., Liu, T., Gao, S., et al. (2016). Identification, and functional and expression analyses of the CorA/MRS2/MGT-type magnesium transporter family in maize. *Plant Cell Physiol.* 57, 1153–1168. doi: 10.1093/pcp/pcw064

Li, J., Huang, Y., Tan, H., Yang, X., Tian, L., Luan, S., et al. (2015). An endoplasmic reticulum magnesium transporter is essential for pollen development in arabidopsis. *Plant Sci.* 231, 212–220. doi: 10.1016/j.plantsci.2014.12.008

Li, Y., Liu, X., Xiao, Y., Wen, Y., Li, K., Ma, Z., et al. (2022). Genome-wide characterization and function analysis uncovered roles of wheat LIMs in responding to adverse stresses and TaLIM8-4D function as a susceptible gene. *Plant Genome* 15, e20246. doi: 10.1002/tpg2.20246

Liu, X., Guo, L. X., Luo, L. J., Liu, Y. Z., and Peng, S. A. (2019). Identification of the magnesium transport (MGT) family in poncirus trifoliata and functional characterization of PtrMGT5 in magnesium deficiency stress. *Plant Mol. Biol.* 101, 551–560. doi: 10.1007/s11103-019-00924-9

Liu, G. J., Martin, D. K., Gardner, R. C., and Ryan, P. R. (2002). Large Mg(2+)-dependent currents are associated with the increased expression of ALR1 in saccharomyces cerevisiae. *FEMS Microbiol. Lett.* 213, 231–237. doi: 10.1111/j.1574-6968.2002.tb11311.x

Liu, Y., Wang, L., Xing, X., Sun, L., Pan, J., Kong, X., et al. (2013). ZmLEA3, a multifunctional group 3 LEA protein from maize (*Zea mays* l.), is involved in biotic and abiotic stresses. *Plant Cell Physiol.* 54, 944–959. doi: 10.1093/pcp/pct047

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

Li, H., Wang, N., Ding, J., Liu, C., Du, H., Huang, K., et al. (2017). The maize CorA/MRS2/MGT-type mg transporter, ZmMGT10, responses to magnesium deficiency and confers low magnesium tolerance in transgenic arabidopsis. *Plant Mol. Biol.* 95, 269–278. doi: 10.1007/s11103-017-0645-1

MacDiarmid, C. W., and Gardner, R. C. (1998). Overexpression of the saccharomyces cerevisiae magnesium transport system confers resistance to aluminum ion. *J. Biol. Chem.* 273, 1727–1732. doi: 10.1074/jbc.273.3.1727

Maguire, M. E. (2006). Magnesium transporters: properties, regulation and structure. *Front. Biosci.* 11, 3149–3163. doi: 10.2741/2039

Mao, D., Chen, J., Tian, L., Liu, Z., Yang, L., Tang, R., et al. (2014). Arabidopsis transporter MGT6 mediates magnesium uptake and is required for growth under magnesium limitation. *Plant Cell* 26, 2234–2248. doi: 10.1105/tpc.114.124628

Mao, D. D., Tian, L. F., Li, L. G., Chen, J., Deng, P. Y., Li, D. P., et al. (2008). AtMGT7: An arabidopsis gene encoding a low-affinity magnesium transporter. *J. Integr. Plant Biol.* 50, 1530–1538. doi: 10.1111/j.1744-7909.2008.00770.x

Marschner, H., and Marschner, P. (2012). *Marschner's mineral nutrition of higher plants* (Beijing: Science Press).

Melamed-Bessudo, C., Shilo, S., and Levy, A. A. (2016). Meiotic recombination and genome evolution in plants. *Curr. Opin. Plant Biol.* 30, 82–87. doi: 10.1016/j.pbi.2016.02.003

Mittal, D., Chakrabarti, S., Sarkar, A., Singh, A., and Grover, A. (2009). Heat shock factor gene family in rice: genomic organization and transcript expression profiling in response to high temperature, low temperature and oxidative stresses. *Plant Physiol. Biochem.* 47, 785–795. doi: 10.1016/j.plaphy.2009.05.003

Palombo, I., Daley, D. O., and Rapp, M. (2013). Why is the GMN motif conserved in the CorA/Mrs2/Alr1 superfamily of magnesium transport proteins? *Biochemistry* 52, 4842–4847. doi: 10.1021/bi4007397

Paolacci, A. R., Tanzarella, O. A., Porceddu, E., and Ciaffi, M. (2009). Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. *BMC Mol. Biol.* 10, 11. doi: 10.1186/1471-2199-10-11

Papp-Wallace, K. M., and Maguire, M. E. (2008). Magnesium transport and magnesium homeostasis. *EcoSal. Plus.* 3, 1–18. doi: 10.1128/ecosalplus.5.4.4.2

Peng, H. Y., Qi, Y. P., Lee, J., Yang, L. T., Guo, P., Jiang, H. X., et al. (2015). Proteomic analysis of citrus sinensis roots and leaves in response to long-term magnesium-deficiency. *BMC Genom.* 16, 253. doi: 10.1186/s12864-015-1462-z

Ramírez-González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., et al. (2018). The transcriptional landscape of polyploid wheat. *Science* 361, eaar6089. doi: 10.1126/science.aar6089

Regon, P., Chowra, U., Awasthi, J. P., Borgohain, P., and Panda, S. K. (2019). Genome-wide analysis of magnesium transporter genes in solanum lycopersicum. *Comput. Biol. Chem.* 80, 498–511. doi: 10.1016/j.compbiolchem.2019.05.014

Saito, T., Kobayashi, N. I., Tanoi, K., Iwata, N., Suzuki, H., Iwata, R., et al. (2013). Expression and functional analysis of the CorA-MRS2-ALR-type magnesium transporter family in rice. *Plant Cell Physiol.* 54, 1673–1683. doi: 10.1093/pcp/pct112

Sawaki, Y., Iuchi, S., Kobayashi, Y., Kobayashi, Y., Ikka, T., Sakurai, N., et al. (2009). STOP1 regulates multiple genes that protect arabidopsis from proton and aluminum toxicities. *Plant Physiol.* 150, 281–294. doi: 10.1104/pp.108.134700

Schock, I., Gregan, J., Steinhauser, S., Schweyen, R., Brennicke, A., and Knoop, V. (2000). A member of a novel arabidopsis thaliana gene family of candidate Mg^{2+} ion transporters complements a yeast mitochondrial group II intron-splicing mutant. *Plant J.* 24, 489–501. doi: 10.1046/j.1365-313x.2000.00895.x

Shaul, O. (2002). Magnesium transport and function in plants: the tip of the iceberg. *Biometals* 15, 309–323. doi: 10.1023/a:1016091118585

Silver, S. (1969). Active transport of magnesium in escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* 62, 764–771. doi: 10.1073/pnas.62.3.764

Song, S., Hao, L., Zhao, P., Xu, Y., Zhong, N., Zhang, H., et al. (2019). Genome-wide identification, expression profiling and evolutionary analysis of auxin response factor gene family in potato (Solanum tuberosum Group Phureja). *Sci. Rep.* 9, 1755. doi: 10.1038/s41598-018-37923-7

Szegedy, M. A., and Maguire, M. E. (1999). The CorA Mg(2+) transport protein of salmonella typhimurium. mutagenesis of conserved residues in the second membrane domain. *J. Biol. Chem.* 274, 36973–36979. doi: 10.1074/jbc.274.52.36973

Tang, N., Li, Y., and Chen, L. (2012). Magnesium deficiency–induced impairment of photosynthesis in leaves of fruiting citrus reticulata trees accompanied by up-regulation of antioxidant metabolism to avoid photo-oxidative damage. *J. Plant Nutr. Soil Sci.* 175, 784–793. doi: 10.1002/JPLN.201100329

Tang, R. J., and Luan, S. (2017). Regulation of calcium and magnesium homeostasis in plants: from transporters to signaling network. *Curr. Opin. Plant Biol.* 39, 97–105. doi: 10.1016/j.pbi.2017.06.009

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, 4673–4680. doi: 10.1093/nar/22.22.4673

Tong, M., Liu, W., He, H., Hu, H., Ding, Y., Li, X., et al. (2020). Identification and functional analysis of the CorA/MGT/MRS2-type magnesium transporter in banana. *PLos One* 15, e0239058. doi: 10.1371/journal.pone.0239058

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016

Vision, T.J., Brown, D.G., and Tanksley, S.D. (2000). The origins of genomic duplications in Arabidopsis. *Science* 290, 2114–2117. doi: 10.1126/science.290.5499.2114

Wang, Y., Hua, X., Xu, J., Chen, Z., Fan, T., Zeng, Z., et al. (2019). Comparative genomics revealed the gene evolution and functional divergence of magnesium transporter families in saccharum. *Curr. Opin. Plant Biol.* 12, 247–249. doi: 10.1186/s12864-019-5437-3

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. *Nucleic Acids Res.* 40, e49. doi: 10.1093/nar/gkr1293

Williams, L., and Salt, D. E. (2009). The plant ionome coming into focus. *Curr. Opin. Plant Biol.* 12 (3), 247. doi: 10.1016/j.pbi.2009.05.009

Worlock, A. J., and Smith, R. L. (2002). ZntB is a novel Zn^{2+} transporter in salmonella enterica serovar typhimurium. *J. Bacteriol.* 184, 4369–4373. doi: 10.1128/jb.184.16.4369-4373.2002

Yang, G. H., Yang, L. T., Jiang, H. X., Li, Y., Wang, P., and Chen, L. S. (2012). Physiological impacts of magnesium-deficiency in citrus seedlings: photosynthesis, antioxidant system and carbohydrates. *Trees* 26, 1237–1250. doi: 10.1007/s00468-012-0699-2

Yan, H., Wu, F., Jiang, G., Xiao, L., Li, Z., Duan, X., et al. (2019). Genome-wide identification, characterization and expression analysis of NF-y gene family in relation to fruit ripening in banana. *Postharvest. Biol. Technol.* 151, 98–110. doi: 10.1016/j.postharvbio.2019.02.002

Zhao, L. M. (2016). Bioinformatics analysis of two f-box genes in rice. *Hubei Agric. Sci.* 9, 2396–2399. doi: 10.14088/j.cnki.issn0439-8114.2016.09.063

Zhao, C. R., Ikka, T., Sawaki, Y., Kobayashi, Y., Suzuki, Y., Hibino, T., et al. (2009). Comparative transcriptomic characterization of aluminum, sodium chloride, cadmium and copper rhizotoxicities in arabidopsis thaliana. *BMC Plant Biol.* 9, 32. doi: 10.1186/1471-2229-9-32

Zhao, H., Ma, H., Yu, L., Wang, X., and Zhao, J. (2012). Genome-wide survey and expression analysis of amino acid transporter gene family in rice (*Oryza sativa* l.). *PLos One* 7, e49210. doi: 10.1371/journal.pone.0049210

Zhao, Z., Wang, P., Jiao, H., Tang, C., Liu, X., Jing, Y., et al. (2018). Phylogenetic and expression analysis of the magnesium transporter family in pear, and functional verification of PbrMGT7 in pear pollen. *J. Hortic. Sci. Biotech* 93, 51–63. doi: 10.1080/14620316.2017.1338923

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 21 June 2023
DOI 10.3389/fpls.2023.1192258

OPEN ACCESS

EDITED BY
Bhanu Prakash Petla,
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

REVIEWED BY
Dinakar Challabathula,
Central University of Tamil Nadu, India
Suleyman I Allakhverdiev,
Russian Academy of Sciences (RAS), Russia

*CORRESPONDENCE
Rajagopal Subramanyam
srgsl@uohyd.ac.in

RECEIVED 23 March 2023
ACCEPTED 16 May 2023
PUBLISHED 21 June 2023

CITATION
Ramachandran P, Pandey NK, Yadav RM, Suresh P, Kumar A and Subramanyam R (2023) Photosynthetic efficiency and transcriptome analysis of *Dunaliella salina* under hypersaline: a retrograde signaling mechanism in the chloroplast. *Front. Plant Sci.* 14:1192258. doi: 10.3389/fpls.2023.1192258

Photosynthetic efficiency and transcriptome analysis of *Dunaliella salina* under hypersaline: a retrograde signaling mechanism in the chloroplast

Pavithra Ramachandran[1], Naveen Kumar Pandey[2], Ranay Mohan Yadav[1], Praveena Suresh[1], Aman Kumar[2] and Rajagopal Subramanyam[1*]

[1]Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad, Telangana, India, [2]Novelegene Technologies Pvt. Ltd, Genomics division, Hyderabad, Telangana, India

Understanding the molecular mechanisms of environmental salinity stress tolerance and acclimation strategies by photosynthetic organisms facilitates accelerating the genetic improvement of tolerant economically important crops. In this study, we have chosen the marine algae *Dunaliella (D.) salina*, a high-potential and unique organism that shows superior tolerance against abiotic stresses, especially hypersaline conditions. We have grown the cells in three different salt concentrations 1.5M NaCl (control), 2M NaCl, and 3M NaCl (hypersaline). Fast chlorophyll fluorescence analysis showed increased initial fluorescence (Fo) and decreased photosynthetic efficiency, indicating hampered photosystem II utilization capacity under hypersaline conditions. Also, the reactive oxygen species (ROS) localization studies and quantification revealed elevated accumulation of ROS was observed in the chloroplast in the 3M condition. Pigment analysis shows a deficit in chlorophyll content and increased carotenoid accumulation, especially lutein and zeaxanthin content. This study majorly explored the chloroplast transcripts of the *D. salina* cell as it is the major environmental sensor. Even though most of the photosystem transcripts showed moderate upregulation in hypersaline conditions in the transcriptome study, the western blot analysis showed degradation of the core as well as antenna proteins of both the photosystems. Among the upregulated chloroplast transcripts, chloroplast Tidi, flavodoxin IsiB, and carotenoid biosynthesis-related protein transcripts strongly proposed photosynthetic apparatus remodeling. Also, the transcriptomic study revealed the upregulation of the tetrapyrrole biosynthesis pathway (TPB) and identified the presence of a negative regulator of this pathway, called the s-FLP splicing variant. These observations point towards the accumulation of TPB pathway intermediates PROTO-IX, Mg-PROTO-IX, and P-Chlide, those earlier reported as retrograde

signaling molecules. Our comparative transcriptomic approach along with biophysical and biochemical studies in *D. salina* grown under control (1.5 M NaCl) and hypersaline (3M NaCl) conditions, unveil an efficient retrograde signaling mechanism mediated remodeling of photosynthetic apparatus.

KEYWORDS

carotenoids, chloroplast, *Dunaliella salina*, photosystems, reactive oxygen species, transcriptome

Introduction

Dunaliella (D.) salina is a representative of highly salt-tolerant green algae and is well known to show superior tolerance against abiotic stresses. Hence it is an effective model organism to study various acclimatization mechanisms adapted by photosynthetic organisms under high salinity (Katz et al., 2007), which easily get acclimated to a range of salinities from 0.05M to 5M NaCl saturation. The salinity stress affects growth by impaired photosynthetic activity due to the accumulation of harmful oxidative species. Oxidative species are initially generated from the acceptor side of photosystem (PS) I. Hence, there is an emerging understanding of the mechanistic basis for changes in structure in response to environmental cues. *D. salina* is pertinent to our studies because as it grows in high salinity and thus its photosynthetic machinery should evolve to cope with its environment.

Apart from salt tolerance and photosynthetic efficiency, *D. salina* is a promising microalga worldwide for commercially producing an enormous amount of β-carotene. There is a high demand for β-carotene in various industries like pharmaceuticals, food, drugs, etc., The cells under stress turn orange due to the over-accumulation of β-carotene in plastidic oil globules (Ben-Amotz and Avron, 1983). As it has the property to accumulate β-carotene up to 8% of the dry cellular weight, it is a major crop for the production of natural β-carotene in the health and food industry (Borowitzka, 2013). Recently there has been a surge in unraveling the proteome, metabolome, and transcriptome of *D. salina* as it is a halotolerant with high photosynthetic efficiency and a promising crop for β-carotene production. This organism is high potential and unique to produce β-carotene, biofuel, and antioxidants on a commercial scale.

High salt levels impact about 20% of all cultivated land and nearly half of the irrigated land, significantly reducing crop yields to well below their genetic potential (Munns et al., 2006). One realistic strategy to cope with this problem is to increase the salt tolerance of crop plants, either *via* genetic introgression or transgenic technology for gene transfer (Koyama et al., 2001). The effects of salinity stress on crop growth are manifested by impaired photosynthetic capacity. High sodium concentrations in the soil solution impair cell metabolism and photosynthesis by imposing osmotic stress on cell–water relations and increasing sodium toxicity in the cytosol (Brini et al., 2007). Salt stress also impacts other physiological and biochemical processes. A variety of salt tolerance mechanisms have been observed in photosynthesizing organisms. Many plants and cyanobacteria exposed to salt stress produce co-solutes such as sucrose, trehalose, proline, glucosyl-glycerol, and glycine-glutamate-betaine that play essential roles in the salt tolerance of these organisms (Gupta and Huang, 2014). Exogenous proline addition protects plant growth and productivity by reducing free radical production and/or scavenging free radicals. In *Arabidopsis (A.) thaliana*, the Salt-Overly-Sensitive (SOS)1 gene encodes a plasma membrane Na^+/H^+ antiporter essential for salt tolerance (Shi et al., 2000). Transgenic *A. thaliana* showed substantial upregulation of SOS1 transcript levels upon NaCl treatment, suggesting post-transcriptional control of SOS1 transcript accumulation (Shi et al., 2003).

Other studies have demonstrated that glyoxalase II overexpression leads to salt tolerance in tobacco and rice plants, and choline oxidase expression plays a role in salt stress tolerance (Singla-Pareek et al., 2008). In the mangrove plant *Bruguiera parviflora*, a 23-kDa PSII-associated protein that plays a major role in oxygen evolution is getting dissociated under high NaCl conditions (Parida et al., 2003). In *Chlamydomonas (C.) reinhardtii*, salt stress-induced changes and acclimation mechanisms have been extensively studied, and morphological studies showed palmelloid formation as a stress-induced behavior (Neelam and Subramanyam, 2013; Devadasu et al., 2023). Also, abscisic acid is a key player in reducing oxidative damage during salt stress (Yoshida et al., 2004). In *C. reinhardtii*, a study on photosynthetic apparatus under salt stress showed oxidative damage to the two PSII core proteins CP43 and CP47 and light-harvesting complex (LHC)II proteins (Ji et al., 2018). A report states that maintenance of efficient photosynthesis by creating fewer and larger PS-I units by *D. salina* under iron deprivation (Varsano et al., 2007). Global transcriptome study in *D. salina* under short-term salinity stress revealed systematic changes in biological processes, including enhancement of photosynthesis, synthesis of glycerol, accelerated protein turnover, carbon fixation, etc (Gao et al., 2021).

Salt stress is a combination of ionic, oxidative, and nutrient deficiency stress, leading to the generation of an enormous amount of reactive oxygen species (ROS), especially in the chloroplast. Chloroplast is one of the major environmental sensors present in cells which helps in acclimation and adaptation under various stress by triggering redox and retrograde signals. Inside the chloroplast, there is an array of ROS-scavenging enzymes and nonenzymatic

compounds that work potentially to rebalance the redox homeostasis. But under extreme stress conditions, the balance between the accumulation of ROS molecules and the scavenging system gets disrupted, leading to the damage of biomolecules such as DNA, RNA, and protein. Recent research shows the evolution of a new concept where ROS molecules generated in the chloroplast in a coordinated and controlled manner act as signaling molecules under stress conditions. Once the redox state of the photosynthetic electron transfer (PET) apparatus gets imbalanced, ROS molecules get generated largely in different sites, including the donor and acceptor sides of both PSI and PSII (Foyer and Noctor, 2000; Asada, 2006). The regulation of the redox homeostasis mechanism in chloroplast under salt stress is still unclear. As per previous studies on ROS generation in photosynthetic apparatus, PS I generate hydrogen peroxide (H_2O_2), as well as superoxide (O_2^-), and PSII, generates O_2^- alone as signaling molecules. They make the chloroplast an environmental sensor for initiating acclimation mechanisms at transcriptional and post-translational levels (Foyer and Noctor, 2000). Salt stress imposition studies revealed more efficient PSII recovery than PSI. Recent studies on glutathione reductase (GR) knockout plants revealed the redox buffering of chloroplast enabled by a highly reduced glutathione pool to maintain efficient photosynthesis (Müller-Schüssele et al., 2020). The overexpression of glutathione transferase study in tobacco plants proved its enhancement in stress tolerance (Roxas et al., 2000).

Salt stress imposed Chlamydomonas transcriptome study showed impaired photosynthetic activity due to significant downregulation of several PS I proteins (Wang et al., 2018). A previous study on salinity tolerance in Chlorella also showed declined photosystem transcripts such as PsaA, PsaB, PsaC, PsaI, and PsaJ (PS I) and PsbA, PsbB, PsbC, PsbD, PsbE, PsbK, PsbH, PsbM, PsbQ, PsbR and PsbZ (PSII) as well as light-harvesting antenna proteins along with decreased pigment content in the system (Abdellaoui et al., 2019). Understanding metabolome and transcriptome will aid in improving knowledge of salt-induced mechanisms. Without understanding the fundamental mechanism, enhancing its large-scale production is not possible. Environmental pollution, the energy crisis, and the fossil fuel limitation increase the concern for large-scale production of biofuel. Our study aims to improve the fundamental understanding of such potential organism that can be used in several industries effectively.

RNA seq is a powerful and revolutionary tool designed to study the transcriptome of organisms, which helps in understanding the functional aspects and molecular elements at the RNA level. In various organisms, transcriptome study under salt stress unveiled many significant pathways which could have been unheeded (Qiao et al., 2013; Gao et al., 2021; Lv et al., 2021). A transcriptomic study of *D. salina* in Yuncheng Salt Lake, China, revealed several salt-stress responsive genes differentially expressed under different salt conditions, low, moderate, and high NaCl concentrations (Gao et al., 2021). A comparative transcriptomic approach on short-term glycerol and salt-stressed *D. salina* cells revealed unique pathways common to both the stress conditions as well as specific pathways in response to both conditions (Lv et al., 2021). In this study, we have focused on the long-term salt stress responses and acclimation mechanisms in *D. salina* cells by extensive analysis of differentially expressed genes (DEGs) fetched out from the high throughput RNA seq data. Our study explored the chloroplast transcripts and correlation of physiological studies of the *D. salina* cells as it is the primary environmental sensor. Gene expression profile studies revealed the cross-talk between redox and retrograde signaling in the chloroplast of *D. salina* and its significance in long-term acclimation under hypersalinity.

Materials and methods

Dunaliella salina growth and culture conditions

D. salina has been grown photoheterotrophically in different salt concentrations ranging from 1.5M (optimum concentration) and 3M (hypersaline concentration) saturation. Cells were grown in BG11 medium supplied with different NaCl concentrations at 25°C with a photon flux density of 50-60 µmol photons m^{-2} s^{-1} in a 1000 mL bottle shaken at 120 rpm in an orbital shaker. Optical density was measured at 730nm. The cell count was measured using a hemocytometer, followed by chlorophyll estimation based on standard protocol. Cells were harvested from 200 ml of culture in the mid-logarithmic growth phase after 72 h by centrifuging at 4000 rpm without damaging the cells. Immediately after harvesting, the cells have been frozen using liquid nitrogen and taken for RNA isolation. The chlorophyll and carotenoid pigment content in control and treated samples were measured as described earlier (Lichtenthaler, 1987; Porra et al., 1989).

Fast Chlorophyll *a* fluorescence measurement

Fast chlorophyll (Chl) measurements were measured using a Handy PEA fluorimeter to check the maximum quantum efficiency of PS II with incident light of high intensity (3000 µmol photons m^{-2} s^{-1}) in a very short period of 1 sec. The photosynthetic efficiency has been recorded as the fluorescence parameter Fv/Fm, whereas Fm is the maximum fluorescence and Fv is the variable fluorescence calculated by subtracting initial fluorescence Fo from Fm. The Chl fluorescence kinetics is represented as an induction curve in which peaks are denoted by four letters O, J, I, and P. O represents the initial fluorescence (Fo); O–J phase that reflects the reduction of the primary quinone acceptor of PSII, Q_A to Q_A^-; J–I and I–P phases involved in the reduction of PQ pool as well as electron acceptor side of PSI; at maximum fluorescence level (Fm), reduced electron carriers in between the PSII reaction center and NADP (Kodru et al., 2015).

Confocal imaging of *D. salina* cells

D. salina grown under different salt conditions has been harvested after 3 days of growth for confocal imaging. A Leica super-resolution microscope has been used for imaging the cells to

see morphological changes. Cells were immobilized using 0.025% iodine solution and observed in 100x magnification for live cell imaging and autofluorescence and images captured in 5nm scale in all conditions.

ROS localization and imaging

D. salina cells were grown under 1.5M, 2M, and 3M conditions and were harvested on the third day and pelleted by centrifugation at 3000 rpm for 10 min. H_2DCFDA dye is used to stain ROS, and 500 µM stock solution of dye was prepared in DMSO and prepared a dilution of 10µM in water. The cells were treated with 5µM dye and incubated in the dark for 1 h. Afterwards washed three times using its growth medium. H_2DCFDA fluorescence was detected at 500-530nm wavelength and an optical filter of 600nm. Samples were viewed with a magnification of 100x oil immersion. Also, the treated cells were used for fluorescence measurement using a plate reader, an excitation wavelength of 485 nm and an emission wavelength of 530nm.

Isolation of thylakoids

The cells were harvested after 72 h of growth by centrifugation at 4500 rpm for 5 min and washed once with buffer containing 30mM Tricine-NaOH (PH-8), 15mM NaCl, and 0.4 M sucrose. Discarded the supernatant after washing and resuspended the pellet within the above buffer. Sonicate the resuspended cells with an amplitude of 25% for 4 cycles of 10 sec on and 50 sec off. Centrifuge the suspension at 12000 rpm for 10 min and then take the supernatant for ultracentrifugation at 45000 rpm for 2h. after centrifugation, resuspends the pellet in buffer containing 30mM Tricine-NaOH (PH-8), 100mM NaCl, and 0.4 M sucrose buffer and incubate it for 30 min, then repeat the ultracentrifugation as previously. Collect the pellet and store it in the 30mM Tricine-NaOH (PH-8), 15mM NaCl, and 0.4 M sucrose buffer (Perez-Boerema et al., 2020).

Western blot analysis

The isolated thylakoid sample from control, 2M and 3 M grown cells are used to separate the photosystem proteins using SDS-PAGE (Schägger and Von Jagow, 1987). A 12% resolving gel is used for separating the denatured protein samples. Then the polypeptides separated in the gel are taken for semi-dry immunoblot transfer apparatus (Biorad), and the polypeptides are transferred to a nitrocellulose membrane. The membrane was kept for incubation primarily overnight, and the secondary antibody (antibody anti-rabbit HRP - conjugated antibody) was for 1h. The primary antibodies used in dilutions PSII: PsbA (1:10,000) PsbD (1:5000), PsbC (1: 3000), PsbB (1: 10000) PSI: PsaA (1:5000), PsaB (1: 1000) PsaD (1:10000) PSI LHCs: LhcA1 (1:5000), LhcA2 (1:5000) AtpC (1: 10000). The images were detected using ChemiDoc, Bio-Rad.

HPLC analysis of pigment composition

Pigments from the harvested *D. salina* cells grown under control and treated conditions have been identified using high-performance liquid chromatography (HPLC) as described (Croce et al., 2000), with some modifications. First, pigments were extracted using 100% acetone and centrifuged at 10000 rpm for 10 min to remove unwanted materials. The supernatant was thermal vacuum dried for 2 h. The dried pigments were solubilized with 400 µl of mobile phase solvent composed of acetonitrile: acetone: methanol (70: 20:10) and filtered using 0.4-micron filters. The analysis was carried out using HPLC (Shimadzu) system on the C18 column. Chl a and Chl b peaks have been extracted at 662 and 642 nm, respectively. Also, different carotenoid pigments, β-carotene, zeaxanthin, lutein, xanthophyll, and violaxanthin, were obtained at 453nm, 481nm, 446nm, 443nm, and 444nm, respectively.

RNA seq analysis

The frozen sample from three independent biological replicates was immediately taken for grinding using a motor and pestle without allowing the sample to thaw. The disrupted cells were isolated for RNA using an RNeasy plant mini kit (Qiagen). The isolated RNA sample has been taken for quantity and quality check using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, United States) and agarose gel electrophoresis to observe if any RNA degradation occurs. After quality and quantity checks of the RNA from three replicates, each from control and treated samples, have been submitted to Novelegen company for Illumina NovaSeq 6000 sequencing. The raw read sequences obtained from the sequencer have been processed using Fastq for quality control check. Approximately 75 million base pairs read from each sample have passed the Q20 and Q30 quality control checks with above 90% and 53% GC content; these reads have been separated as high-quality for further processing. All the high-quality reads were mapped against the reference generated by Trinity using the Trinity pipeline (bowtie2) genome assembler. After mapping the reads, the BAM file containing the mapping information used for quantifying reads was generated. The downstream analysis requires the number of reads mapping to each genomic feature, for example, each exon or each gene. Read summarization (Quantification) has been conducted using the RSEM program, and gene expression level (FPKM) was generated using this. Among each sample, approximately 40 million reads were mapped, and 25 million reads were aligned. Based on the global gene expression level principal component analysis (PCA analysis) and correlation analysis have been conducted (Supplementary Figures 1A, B). DEGs were generated among the mapped genes using the Dseq2 program with a fold change of ≥ 0.5 and FDR ≤ 0.05 (Supplementary Figure 2A, B). The transcriptomic data was uploaded to NCBI under the Bio project ID PRJNA946036 and submission ID SUB12963156.

qRT PCR analysis

The accuracy of differentially expressed genes has been validated using qRT-PCR analysis (Supplementary Figure 3). Among the most upregulated genes, 10 were randomly selected and primers were designed (Supplementary Table 1) using primer 3, and primer parameters were rechecked using oligocalc. cDNA synthesis has been done using Primescript^{TM} 1st strand synthesis kit. Ribosomal protein L2 served (RPL2) as the reference gene for qPCR analysis.

Results and discussion

Growth and photosynthetic efficiency of *D. salina*

D. salina cells were grown in various salt concentrations (1.5 to 3 M NaCl) to understand the remodeling of photosynthetic apparatus as a salt tolerance mechanism to operate photosynthesis. Here the 1.5M growth is considered a control, and this concentration is an optimal growth condition. In the 2M growth condition, even though cells are growing slowly, they are acclimatizing within 3 days, and in 3M, the growth is severely impaired (Supplementary Figures 4A–D). The culture color turns from light green to yellowish green from 1.5M to 3M, respectively (Supplementary Figure 4C). Pigment analysis showed a moderate decrease in chlorophyll pigments and an increase in the carotenoid pigments under hypersaline (Supplementary Figure 5); the same trend has also been observed in the isolated thylakoid membranes. Enhanced carotenoid pigments must be an efficient strategy in the acclimation process involving excess energy quenching and non-enzymatic ROS scavenging. The cell growth in the control and 2M treated conditions reached the stationary phase after 120 h of growth and 3M condition showed an actively growing phase of 164 h. When the cells reach the stationary phase, the control culture contains 1.73x10^6 cells. The 2M and 3M have 1.38 x10^6 and 0.97 x10^6 cells, respectively. The morphology has been observed to be a round shape with reduced size in hypersaline (Figures 1A, B).

The chlorophyll fluorescence induction curve is an efficient and fast method to analyze the photosynthetic efficiency of oxygen-evolving photosynthetic organisms. The Chl fluorescence induction kinetics shows the photochemical activity of PSII in a time course of 1 sec and gives OJIP transient data (Figure 2A). 'O' represents the initial fluorescence (F_O), and P represents the maximum fluorescence (F_m) when it reaches its maximum. J step represents the formation of $Q_A^-Q_B$, and I step reflects the formation of $Q_A^-Q_B^-$ form. Finally, the P reflects the accumulation of $Q_A^-Q_B^{2-}$ form (Strasser, 1992). Chl a fluorescence data has shown almost similar photosynthetic efficiency after three days of growth in optimum condition as well as in the hypersaline condition of 3M. Comparing the transient fluorescence recorded in hypersaline grown cells to optimum grown cells shows a slight increase in the F_O (Figure 2B) and a marginal decrease in the PS II efficiency (Fv)/Fm) (Figure 2D) without change in F_m,(Figure 2C). Perhaps, this could be because of the decrease in the overall utilization of trapped energy for photochemistry. Also, the slight dip observed in the I phase reflects the reduced rate of reoxidation of the PQ pool.

Transcriptomic analysis shows efficient acclimation strategies at the chloroplastic level

Among the obtained transcripts significant DEGs were separated, setting a fold change above 1.5. GO enrichment analysis of all the DEGs was performed using Blast2GO software. Both the upregulated (Figure 3) and downregulated (Figure 4) transcripts were categorized into three GO functional groups, cellular component, molecular function, and biological process. More than 50% of the transcript categorized under cellular component was localized in the chloroplast, especially components of photosystem and proteins located in the thylakoid membrane. The other major part of the transcriptome showed a significant uptrend localized in the nucleus, minichromosome maintainance complex (MCM), and nucleus indicating cellular reorganization. Transmembrane transport shows a significant role

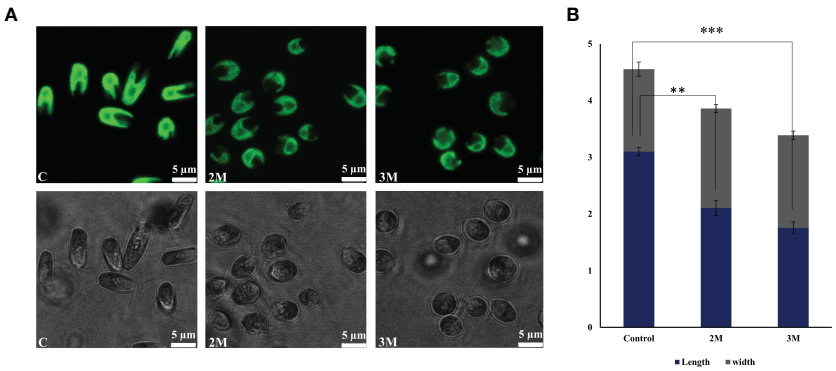


FIGURE 1
Morphological changes observed in *Dunaliella salina*. **(A)** Confocal images of cells were observed under control, 2M, and 3M hypersaline conditions. **(B)** Graph representing cell length and width change under control, 2M, and 3M hypersaline conditions. The means ± SD were calculated (NaCl) for three biological replicates. Statistical comparison was performed using one- way analysis of variance (ANOVA) followed by the Tukey multiple comparison tests, and p-values obtained are indicated asterisks (***p < 0.001, **p <0.01).

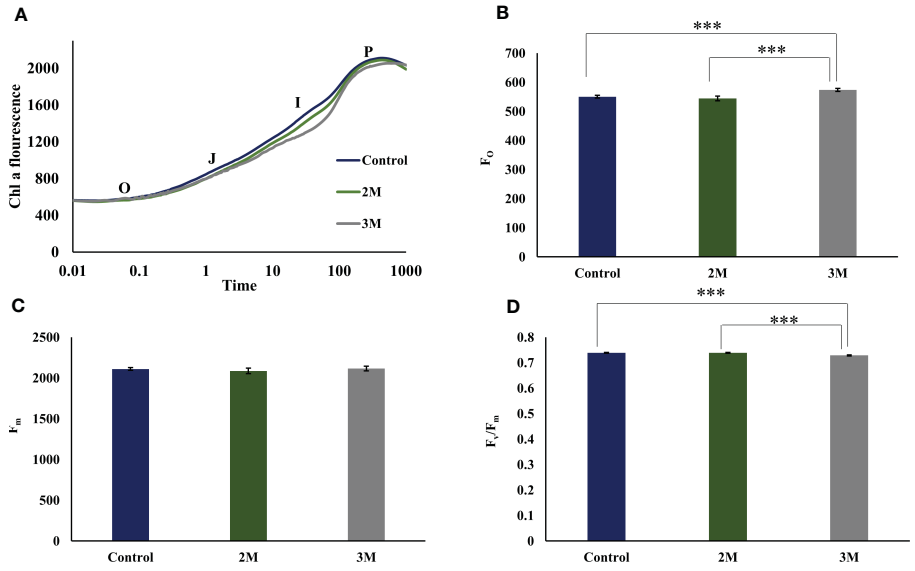


FIGURE 2
Fast chlorophyll a fluorescence analysis of control, 2M and 3M NaCl. **(A)** OJIP transient data was obtained from the three conditions. O represents the initial fluorescence O, initial fluorescence (Fo); OJ phase that reflects the reduction of the primary quinone acceptor of PSII, Q_A to Q_A^-; JI and IP phases involved in the reduction of PQ pool as well as electron acceptor side of PSI; at maximum fluorescence level (Fm). **(B)** Initial fluorescence (Fo) from control, 2M, and 3M conditions. **(C)** Maximum fluorescence (Fm)from control, 2M, and 3M conditions. **(D)** Maximum photosynthetic efficiency (Fv/ Fm) in control, 2M and 3M conditions. The means ±SD were calculated for 3 biological replicates. Statistical comparison was performed using one- way analysis of variance (ANOVA) followed by the Tukey multiple comparison tests, and p-values obtained are indicated asterisks (***p < 0.001)

under hypersaline conditions, especially the phosphate ion transporter and symporter activity. Pathway enrichment analysis of both upregulated (Figure 5A) and downregulated (Figure 5B) transcripts have been done using KOBAS software to select specific pathways showing significant changes. Among the pathways fetched out, carbon fixation, secondary metabolite synthesis, carotenoid biosynthesis, and fatty acid metabolism show substantial changes. Further, gene ontology (GO) analysis and pathway enrichment primarily indicate an efficient chloroplast signaling mechanism in response to the hypersaline condition as a significant environmental sensor. Especially photosystem subunit level changes showed in GO analysis must play a key role in salinity

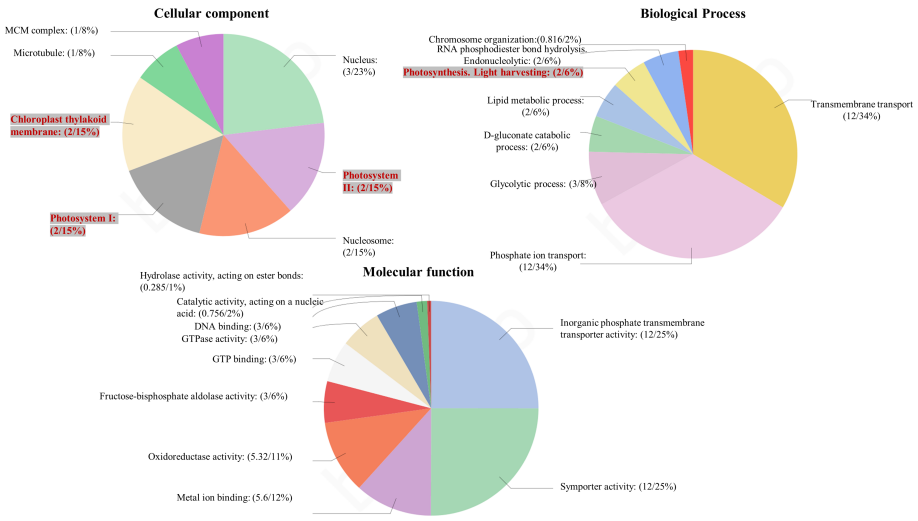


FIGURE 3
Gene ontology analysis of differentially expressed genes using BLAST2 GO software. The percentage of total upregulated transcripts in different GO categories biological processes, molecular function, and cellular components.

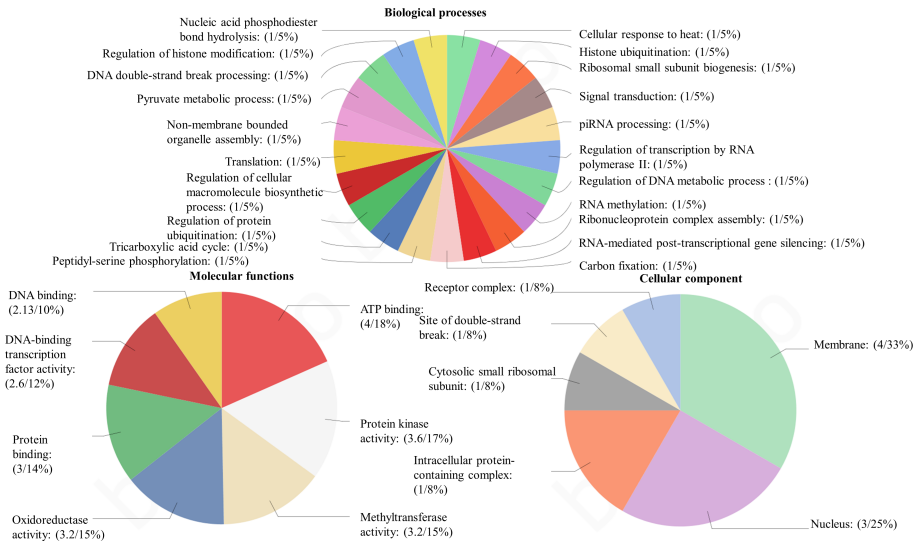


FIGURE 4
Gene ontology analysis of differentially expressed genes using BLAST2 GO software. The percentage of total downregulated transcripts in different GO categories biological processes, molecular function, and cellular components.

acclimation. Along with that, carotenoid biosynthesis and elevated secondary metabolite synthesis help in acclimation and repair processes.

Photosystem I remodeling and role of chloroplast Tidi

PS I function takes a short span of time for charge separation and conversion of excitation energy into radical pair (Wilhelm and Wilhelm., 1993). PSI is a multi-pigment-protein complex span over thylakoids with extrinsic, intrinsic, and transmembrane subunits, which together function as a strong redox system to convert $NADP^+$ to NADPH. The antenna molecules absorb photons and get excited, leading to charge separation between P700 and A_0 molecules. The excited electron subsequently reduces a chain of oxidized molecules A1, FeX, FeA, and FeB iron-sulfur clusters, and finally, it reduces the ferredoxin molecule. Ferredoxin transfer electron to NADP+ to generate NADPH (Golbeck, 1992). PSI is considered the most stable redox catalyst because of its shorter span of charge separation due to its tightly packed pigment molecules and symmetrical arrangement of Ao and RC Chl molecules. Being a fast redox system, PSI is less prone to photoinhibition under stress conditions when compared to PSII (Caffarri et al., 2014). Our transcript study also points towards PSI's potent and efficient remodeling to rebalance the excitation energy distribution in the photosynthetic apparatus under hypersaline conditions.

Analysis of DEGs reveals that the transcript level of PS I-related proteins has been showing a general trend of moderate upregulation of all proteins and a thylakoid transcript chloroplast Tidi was showing a significant increase (Figures 6A, B and Supplementary Table 2A). The transcriptional regulation of photosystem-related proteins varies in different photosynthetic organisms due to ecological and evolutionary reasons. In the model organism, *C. reinhardtii*, the transcriptional level of PS I light-harvesting proteins was reported as downregulated under salt stress conditions (Wang

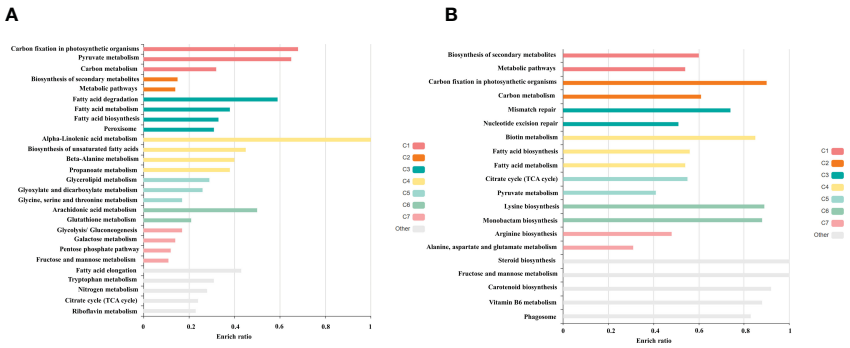


FIGURE 5
Gene enrichment analysis of differentially expressed gene using KOBAS 3.0 software **(A)** Upregulated **(B)** Downregulated transcripts.

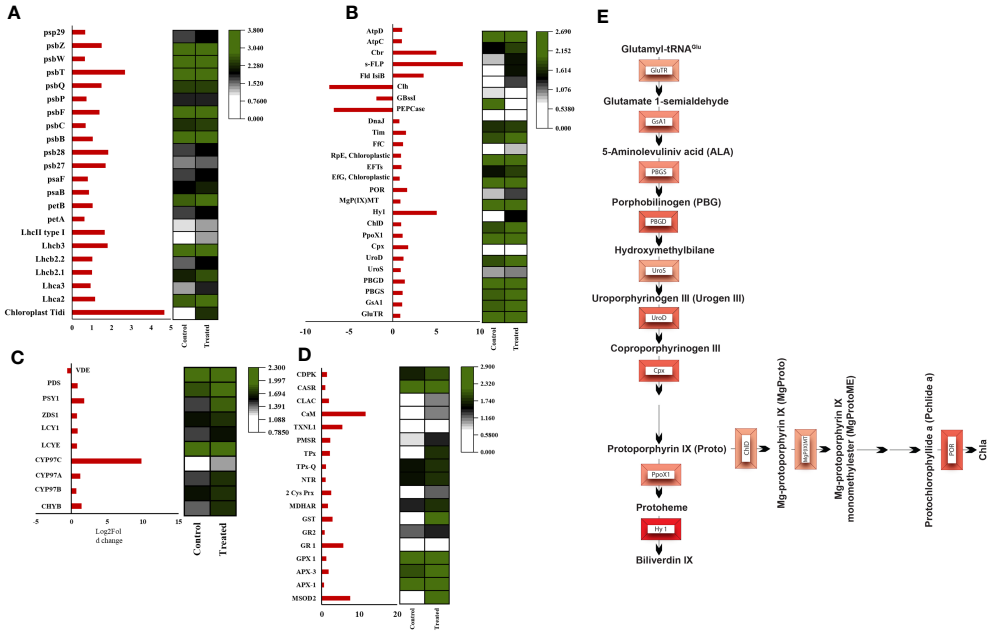


FIGURE 6
Heatmaps of significant differentially expressed genes. **(A)** Heatmap of photosystem-associated transcripts, **(B)** Heatmap of tetrapyrrole biosynthesis pathway-associated transcripts **(C)** Heatmap of carotenoid biosynthesis associated transcripts, **(D)** ROS and calcium signaling associated transcripts, **(E)** Pathway representation of all the upregulated (intensity of red representing the foldchange) transcripts involved in the tetrapyrrole biosynthesis pathway.

et al., 2018). Further, in Ulva, the subunits of PS I and II as well as all the LHC transcripts found to be significantly downregulated in salt stress.

Among the photosynthetic transcripts analyzed in both PSI and PSII from DEGs, the most upregulated transcript observed was chloroplast Tidi (TRINITY_DN12221_c0_g1_i9), showing a log2 fold change of 4.6. Also, most of the PSI light-harvesting proteins showed a moderate up-regulation. Chloroplast Tidi is a 45kD chlorophyll-binding protein localized in the thylakoid membrane reported earlier, specifically iron deficit inducible in *D. salina* (Varsano et al., 2007). The recent transcriptome analysis of short-term hypersaline-treated *D. salina* reported downregulation of the LHC proteins, and there was no report of chloroplast Tidi induction in shock treatment (Lv et al., 2021). When comparing this scenario with our study, chloroplast Tidi upregulation must play a key role in the long-term hypersaline acclimation of photosynthetic apparatus in *D. salina* and moderate upregulation of all other subunits. Recently the *D. salina* PSI structure has been elucidated as a minimal PSI structure with 11 subunits, including PsaA-PsaF, PsaI, and the LHC proteins Lhca1- Lhca4, without PsaG, PsaH, PsaI, PsaK, PsaL, and PsaO (Perez-Boerema et al., 2020; Caspy et al., 2023). From the above studies, the missing subunits of PSI structure can associate with additional Lhc proteins under certain physiological conditions. They suspect its functional implication as it could be for reducing the association in particular conditions or promoting specific remodeling and distinct interactions under certain conditions. In line with this hypothesis, chloroplast Tidi is associated with the PSI-LHCI complex with its unique N terminal domain under iron deficit conditions. Its accumulation is in correlation with decreased PSI subunits. Low-temperature fluorescence studies and BN-PAGE analysis of photosystem supercomplexes from iron deficit cells revealed an increase in PSI supercomplex size and Chl a fluorescence. They suspect that this is contributed by the CAB-like protein chloroplast Tidi by its association with PSI to rebalance the energy distribution between PSII and PSI in stress conditions (Varsano et al., 2006).

Our immunoblot studies of PS I subunits showed that PsaA undergoes more degradation than PsaB, and the PsaF subunit was unaffected. Among LHCs, Lhca2 and Lhca1 showed decreased expression levels, probably due to oxidative damage (Figure 7A). Also, the ROS localization study and elevated level of total ROS was observed in the hypersaline conditions compared to the control condition (Figures 7B, C). These elevated ROS levels must lead to the degradation of photosystem subunits. But both the western blot and transcript level analysis of ATP synthase subunits was showing an upregulated trend in hypersaline conditions. In hypersaline conditions, *D. salina* requires a high amount of energy to execute many acclimations mechanisms such as a high amount of glycerol production, cellular rearrangement, pigment production, etc., for its survival. So uninterrupted ATP synthase activity is essential to meet the energy requirements of the organism. The system must be protecting ATP synthase from degradation for better photosynthetic performance by uninterrupted electron transport through the photosynthetic apparatus. Compared with an earlier report of enhancement in PSI level in short-term salt treatment, our study shows the degradation of a few PSI subunits in the long term, which could be promoting the chloroplast Tidi association and enlargement in PSI particle size similar to iron deficit condition. The change in antenna complexes, the remodeling of the PSI occurred due to acclimation by over-expression of TiDi. Also,

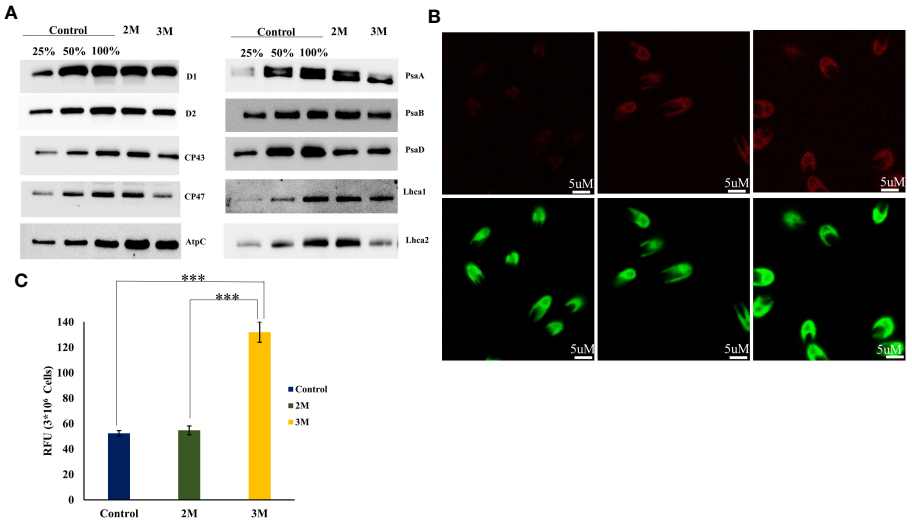


FIGURE 7
ROS accumulation in chloroplast and oxidative degradation of photosystem-associated proteins. **(A)** Western blot analysis of photosystem-associated proteins isolated from the thylakoid sample of control and hypersaline condition. **(B)** Confocal localization of total ROS accumulated in control, 2M, and 3M NaCl using Leica high-resolution microscopy technique, the cells incubated in H_2DCFDA for 1 h. **(C)** Total ROS measured using UV-visible spectrophotometry. Statistical comparison was performed using one- way analysis of variance (ANOVA) followed by the Tukey multiple comparison tests, and p-values obtained are indicated asterisks (***p < 0.001).

rebalancing the redox state of PSI and PSII as a long-term acclimation of the photosynthetic apparatus of *D. salina* under hypersaline conditions. PSI structure of *D. salina* is suspected to be evolved in a way to adapt to extreme saline conditions as well as other abiotic stress as it has an oceanic origin; structural plasticity of PSI structure must be playing a pivotal role in such acclimation processes under various physiological conditions specifically in *D. salina*.

Photosystem II assembly and repair regulation

Analysis of photosystem transcripts' expression level proclaims a general trend of upregulation among reviewed DEGs (Figure 6A, B and Supplementary Table 2A). PS II is a large transmembrane protein complex composed of several protein subunits and pigment molecules localized in the thylakoid membrane's stromal and luminal sides. According to the recent structure of PSII of C2S2M2 of *C. reinhardtii* comprises 3 LHCII trimers, 2 major antennae CP26, and CP29 surrounding a dimer core and peripheral protein subunits (Shen et al., 2019). Algal PSII is elucidated in cryo-EM structures as having extra light-harvesting antenna molecules in the photosystem. This could be due to its oceanic origin and underwater ecological niche. Photosystems undergo various types of repairs and remodeling under unfavorable conditions to acclimate and adapt photosynthetic apparatus. The protein-pigment complexes undergo degradation, repair, and remodeling depending on the type and intensity of stress. Photooxidative damage of PSII is an unavoidable evil that also helps in protecting the other part of the photosynthetic apparatus from further damage. PSII is more prone to photooxidative damage because of its monodirectional electron transfer, weaker pigment molecule interactions, and distorted arrangement of the porphyrin plane of all molecules bound to the PSII reaction center (Wilhelm and Wilhelm., 1993).

Several PSII-associated proteins showed transcriptional upregulation in our data, especially PSII repair and assembly proteins. One interesting transcript showing a significant upregulation with a log2 fold change of 2.6 was carotenoid biosynthesis-related protein (CBR) (TRINITY_DN2118_c0_g1_i13). CBR is an algal homolog of the ELIP family protein in plants, which was reported first in *D. bardawil* as high light-induced protein functions in photoprotection (Adamska et al., 1992). It is reported that CBR accumulation is in parallel with carotenogenesis. The earlier report proposed a physical association between the pigments and CBR protein to protect disassembled PSII under light stress (Levy et al., 1993). Our HPLC data and carotenogenesis-related transcripts data align with the upregulation of CBR transcript. Also, the disassembly and degradation of PSII subunits D1, D2, CP43, and CP47 proteins in 3M salt conditions were observed through western blot analysis (Figure 7A). This is probably due to oxidative degradation by the excess accumulated ROS in the chloroplast. It is tempting to suggest that CBR accumulation and upregulated carotenoid synthesis point towards the protective role of CBR in hypersaline-grown *D. salina* photosynthetic apparatus.

Along with this, the carotenoid hydroxylating enzymes, which play a role in the generation of xanthophyll molecules from α and β carotenoids, show marked upregulation in 3 M salt. Also, the transcript data support the protective function of CBR binding with xanthophyll molecules in long-term hypersaline conditions. Another transcript showing a significant upregulation was

flavodoxin IsiB, which consists of electron shuttles in the acceptor side of PSI complex functions similar to ferredoxin molecules in prokaryotes and some oceanic algae (Lodeyro et al., 2012). These molecules contain FMN prosthetic groups instead of Fe-sulfur clusters to participate in oxidation-reduction reactions in photosynthetic apparatus. The presence of flavodoxin molecule is an adaptive trait with the remodeling in photosynthetic apparatus to acclimate under iron deficit as well as some other abiotic stress (Lodeyro et al., 2012). In *D. salina* these molecules must be playing a pivotal role in remodeling and balancing the redox reactions replacing the ferredoxin molecules in hypersaline conditions.

Further analysis of PSII-associated protein transcripts revealed upregulation of several repair and assembly factors such as Psb27 (TRINITY_DN2037_c0_g1_i6), Psb28 (TRINITY_DN2123_c0_g1_i8), PsbZ (TRINITY_DN7535_c0_g1_i1), etc., Basically Psb28 and Psb27 are two major PSII biogenesis factors function in different steps in protecting the assembly intermediates which are more prone to damage in stress conditions (Dobáková et al., 2009; Boehm et al., 2012). This gives insight into the efficient biogenesis of the PSII complex and the protection of the intermediate complex to acclimate under hypersaline-triggered photooxidative damage of photosynthetic apparatus. Psb28 plays a major role in the association of CP47 with the reaction center intermediate forming the primary PSII assembly intermediate RCP47 complex. Psb27 protects free unbound CP43 from damage before incorporating it into the CP47 complex. It also suggested a possible function of Psb27 in restraining the flexibility of the CP43 E loop, which has an indispensable role in OEC assembly, for conformational changes for further OEC assembly (Komenda et al., 2012). PsbZ was also showing a moderate upregulation in our system, and a key role of Psbz as a small peripheral subunit that connects the core is proposed. The other subunits play a role in NPQ by inducing elicitation mechanisms in the core to dissipate excess excitation energy. This mechanism has not yet been well studied in terms of Psbz's contribution to the NPQ mechanism. This must be vital in stabilizing hypersaline-triggered damages in the *D. salina* photosynthetic apparatus. PsbT and PsbM are two important subunits engaged in assembly and protection, which showed moderate upregulation in high saline conditions. PsbM functions in the assembly of CP47 monomer complexes during the biogenesis of PSII. PsbM does not bind with this intermediate without PsbT (Iwai et al., 2004). Mutant studies in a double mutant of PsbM and PsbT showed increased photodamage and decreased recovery rate of photosynthetic apparatus, suggesting their inevitable role in the assembly and repair process (Bentley et al., 2008). Therefore, a moderate increase in PsbM and PsbT subunits involved in photoprotection against hypersaline in *D. salina*.

Oxygen evolving complex is an irreplaceable component of PSII which take part in the photolysis of water. In our study, both OEE2 (PsbQ) (TRINITY_DN19192_c0_g3_i1) and OEE3 (PsbP) (TRINITY_DN5262_c0_g1_i21) showed moderate upregulation. These are the two extrinsic subunits of the oxygen-evolving complex, which are involved in the binding of the cofactor's chloride (Cl^-) and calcium (Ca^-) (Mayfield et al., 1989). In other halophytic plants such as *Bruguiera gymnorrhiza* and *Salicornia europaea* (Momonoki et al., 2009), it has been reported that instead of OEE2 and OEE3, OEE1 was showing significant upregulation. So, among halophytes, the strategies differ in acclimation to protect and functionalize the photosynthetic apparatus. Apart from core subunits and assembly factors, light-harvesting complex proteins were also showing a general trend of upregulation, Lhcb3 (TRINITY_DN1444_c0_g1_i22), LhcII type I (TRINITY_DN1444_c0_g1_i14), Lhcb2.1 (TRINITY_DN1444_c0_g1_i1), Lhcb 2.2 (TRINITY_DN3207_c1_g1_i1) all lhcs and core. This indicates that long-term growth in the high salt of *D. salina* has different kinds of acclimation than freshwater algae and land plants. High salt is often has a deleterious effect on freshwater algae or plants, especially when the water oxidation complex (WOC) is damaged drastically (Gupta, 2020).

Carotenoid biogenesis

D. salina is known for its escalated accumulation of β-carotene under different abiotic stress conditions such as high light, UV exposure, nitrogen starvation, glycerol stress, and salinity stress. Carotenoids are the indispensable components of photosynthetic apparatus from bacterial photosystems to plant photosystems. Carotenoids are an integral component of light-harvesting complexes in the multi-protein photosystem complex. They play a crucial role in light harvesting and photoprotection by quenching the triplet chlorophyll, dissipating excess excitation energy in the system, and scavenging ROS. It has elucidated that specific xanthophyll molecules have a peculiar and precise function in photoprotection. α-carotene branch-derived lutein, β- carotene branch-derived zeaxanthin, and antheraxanthin have an irreplaceable role in dissipating excess excitation energy to avoid photodamage (Niyogi et al., 1997).

Interestingly, analysis of transcripts from carotenoid biogenesis revealed upregulation of most of the enzymes involved in α-carotene and β-carotene branch-derived pigments synthesis (Figure 6C and Supplementary Table 2C), which aligns with the upregulation of the chloroplastic Cbr gene. Both these point towards an efficient acclimation of photosynthetic apparatus. Cytochrome P-450 hydroxylase (CYP97C1) (TRINITY_DN4283_c0_g1_i7) was observed as the most abundant transcript, and it has shown a fold change of 9.7. This enzyme has carotenoid β- ring hydroxylating activity responsible for lutein accumulation (Inoue, 2004). Along with CYP97C1, two more hydroxylating enzymes show a moderate upregulation, CYP97A, and CYP97B. These three enzymes are well studied in *D. bardawil* and have been reported for their hydroxylating activity and conversion of α-carotene and β-carotene into lutein (Liang et al., 2020). CYP97C can hydroxylate β- carotene and α-carotene; however, CYP97A can hydroxylate only the β-carotene ring. In fact, CYP97B has only minor activity in the hydroxylation of α-carotene. Also, our HPLC data from the thylakoids from *D. salina* grown under 3M NaCl condition shows increased lutein production (Figure 8). These results reveal an essential role of lutein in long-term acclimation under hypersalinity conditions. The enzyme

considered to be the rate-limiting enzyme in carotenoid biogenesis, phytoene synthase (TRINITY_DN43_c1_g2_i1) was also observed to be moderately upregulated in long-term treatment. This might be the reason for the system's upregulation of total carotenoid content. β-carotene 3-hydroxylase (BCH) (TRINITY_DN10949_c0_g1_i2), which converts β-carotene to zeaxanthin by hydroxylation of its 3rd position in both the rings (Cazzonelli and Pogson, 2010) and this will be the precursor of many other important xanthophylls plays a crucial role in nonradiative energy dissipation under certain conditions showed a moderate upregulation. Also, in *D. bardawil* BCH has explained the hydroxylation of β-rings of α- carotene as well as β- carotene and leads to the accumulation of lutein. Considering the upregulated activity of BCH, CYP97A, CYP97B, CYP97C, and increased accumulation of lutein in HPLC analysis gives an insight into the presence of multiple pathways involved in lutein accumulation. It is involved in nonradiative energy dissipation under long-term hypersaline acclimation of the *D. salina* photosynthetic apparatus.

Cross-talk between ROS and calcium signaling mechanism

The proper functioning of chloroplast needs coordinated action of both the nucleus and plastid genome signaling. When the redox poise in the chloroplast is harnessed, it sends messages to the nucleus to initiate signaling mechanisms that rebalance the system, called retrograde signaling (Nott et al., 2006). The regulative mechanisms executed by retrograde signaling can be divided into two categories biogenic control (chloroplast biogenesis associated) and operational control (triggered by environmental stimuli) (Pogson et al., 2008). Retrograde signaling has been well explored in Arabidopsis by isolating and characterizing genome uncoupled (GUN) mutants, which are still nascent in the algal system. Analyzing the *D. salina* transcriptome under hypersalinity tempts to propose strong and efficient retrograde signaling to reestablish the chloroplast's redox state. The uptrend pattern shown in the expression level of ROS scavenging and calcium signaling transcripts promulgates the idea of cross-talk

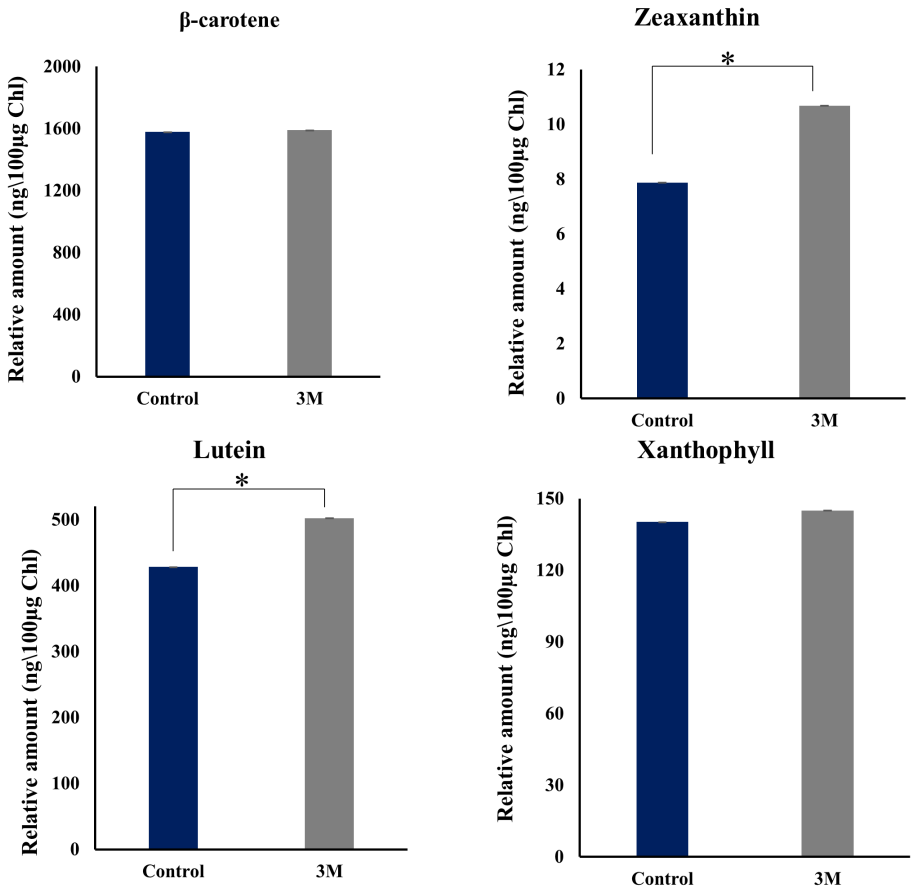


FIGURE 8
HPLC analysis of change in pigment composition from the thylakoid sample isolated from control, 2M, and 3M NaCl. Different carotenoid pigments are β-carotene, zeaxanthin, lutein, xanthophyll, and violaxanthin at 453nm, 481nm, 446nm, 443nm, and 444nm, respectively. The means ± SD were calculated for 3 biological replicates. Statistical comparison was performed using one- way analysis of variance (ANOVA) followed by the Tukey multiple comparison tests, and p-values obtained are indicated asterisks (*p < 0.001).

between both the signaling mechanisms and the establishment of chloroplast retrograde signaling.

Analyzing the transcriptome, non-enzymatic, and enzymatic ROS scavenging system shows a significant uptrend pattern (Figure 6D and Supplementary Table 2B). Many studies unveiled the importance of different antioxidant enzymes such as catalase (CAT), superoxide dismutase (SOD), ascorbate peroxidase (APD), and glutathione reductase (GR) in salt tolerance (Cejudo et al., 2021). Among those enzymes, manganese superoxide dismutase, which acts as the major ROS scavenging enzyme in the mitochondria (TRINITY_DN4937_c1_g1_i8), showed a maximum upregulation of 7-fold. Our analysis detected different isoforms of ascorbate peroxidase, and APX3, a peroxisomal enzyme, showed (TRINITY_DN3502_c0_g1_i122) an increase of 1.8 log2 fold change. From previous mutant studies in Arabidopsis, APX3 was proposed to be insignificant in oxidative stress (Narendra et al., 2006). However, our transcriptome data propose an important role of APX3 under oxidative stress. Enzymes involved in the glutathione metabolism, including Glutathione S transferase (TRINITY_DN11514_c0_g1_i24), glutathione peroxidase (TRINITY_DN16267_c0_g1_i9), glutathione reductase (TRINITY_DN6033_c0_g2_i13), and monodehydroascorbate reductase (TRINITY_DN131_c0_g1_i6) are significantly upregulated which reveals an efficient ROS detoxification system switch by glutathione to impart salt tolerance in *D. salina*. Chloroplastic 2-cys peroxiredoxin is an inevitable member of the photosynthetic antioxidant network, which was highly upregulated in high salt conditions. Observing the upraised antioxidant enzymes, both the organelles' mitochondria and chloroplast accommodate an efficient antioxidant network to minimize the deleterious effect of excess accumulated ROS and to rebalance the redox state of the internal environment. Redox imbalance leads to oxidative damage of proteins, especially by the generation of disulfide bonds in the thiol group of cysteine residues which leads to the misfolding and loss of function of proteins. These oxidized proteins got repaired by a system of reductases and proteins, which provide reducing power to the reductases for their activity. Several transcripts show upregulation associated with repairing oxidized proteins, such as NADPH-dependent thioredoxin reductase (NTR), peptide methionine sulfoxide reductase (PMSR), thioredoxin-dependent peroxidase, thioredoxin-like protein. PMSR catalyzes the reduction of the methionine sulfoxide back to methionine residue in damaged proteins (Romero et al., 2004), showing a 2.2-fold upregulation in *D. salina* under salinity stress. The active participation of thioredoxins and thioredoxin-like proteins in thiol-disulfide interchange is well studied; these proteins activate the reductase activity of disulfide reductases. This system relies on ferredoxin as the reducing power; NADPH also acts as the source of reducing power for NTR-dependent redox regulation (Schürmann and Buchanan, 2008; Balsera and Buchanan, 2019). NTR dependent model revealed the active participation of 2-cys peroxiredoxins in modulating TRX-regulated enzymes (Awad et al., 2015).

ROS generation in the biological system is necessary for normal growth as well as for acclimation under different biotic and abiotic stress conditions. Excess ROS accumulation, such as singlet oxygen, hydrogen peroxide, superoxide, and hydroxyl radical accumulation, leads to the damage of biological molecules in the system (Mittler, 2017). To avoid this deleterious situation, a powerful ROS scavenging system and compartmentalization of ROS occur in organisms to balance the redox system. Chloroplast is a major site of ROS generation. The major sites that generate ROS include 1) the semiquinone radical generated at Cyt b_6 complex transfer e- to O_2 (Shochat et al., 1990), 2) ferredoxin can directly transfer an electron to O_2 (Morehouse and Mason, 1988), 3) electron donation can happen from Cytb 559 cofactor of PSII to O_2, 4) from reduced plastoquinone pool, 5) incomplete oxidation of H_2O in water oxidation complex, 6) at PSI iron-sulfur clusters and phylloquinone donates an electron to O_2, 7) presence of metal centers such as Mn, Fe convert H_2O_2 into hydroxyl radical (Macpherson et al., 1993). An imbalance in the redox state between PSI and PSII leads to the overexcitation of electrons in the apparatus, which turns on ROS generation. As per previous studies on ROS generation in photosynthetic apparatus, PSI generates H_2O_2 as well as O_2^- and PSII generates O_2^- alone as signaling molecules and makes the chloroplast an environmental sensor for initiating acclimation mechanisms at transcriptional as well as in the post-translational level (Asada, 1987).

There are reports stating cross-talk between ROS signaling and scavenging mechanisms with calcium signaling in cells under biotic and abiotic stress conditions. In plants, the intracellular calcium level-activated antioxidant system has been studied in several mutants, but the algal system is under-studied. Increased calmodulin (CAD) (TRINITY_DN6782_c0_g1_i12), calcium-dependent kinase (CDK) (TRINITY_DN4313_c0_g1_i45), calcium load activated calcium channel (TRINITY_DN878_c0_g2_i8), calcium-sensing protein (TRINITY_DN9636_c0_g1_i11), transcript accumulation brings the idea of intracellular elevated calcium level in *D. salina* under hypersalinity. Calmodulin is a ubiquitous calcium-sensing protein that functions as a signaling molecule in several physiological mechanisms, including abiotic stress tolerance in plants showing upregulation of 11-fold change in transcript data. The calcium load-activated calcium channel showed a 1.9-fold change, which usually helps release calcium ions from the endoplasmic reticulum to avoid overfilling. It shows the cells' highly elevated calcium level under long-term hypersalinity conditions. CDPKs are an inevitable member of calcium-mediated signaling mechanisms to undertake several downstream processes, also offering moderate upregulation. ROS molecules are essential as signaling molecules, but their homeostasis is essential for the wellness of the intracellular environment. Calcium-calmodulin signaling plays an irreplaceable role in this regard. *In vitro* regulation of CAT3 expression by calcium-calmodulin signaling has been unveiled (Yang and Poovaiah, 2002). There are reports of peroxisomal Ca^{2+} mediated activation of enzymatic and non-enzymatic ROS scavenging mechanisms which control H_2O_2 levels (Costa et al., 2010). Also, in *A. thaliana*, calcium level elevation, and associated signaling mechanisms play a significant role in drought and salinity stress conditions (Knight et al., 1997). In *U. compressa*, marine algae, the activation of GR and GST is triggered by an elevation in both intracellular calcium levels and ROS (Gonzalez et al., 2010). Detailed analysis in Blast2GO of transcriptome shows many hypothetical and non-annotated transcripts with calcium-binding domains with significant fold change values; this clearly shows the indubitable calcium signaling mechanism in *D. salina*, and exhaustive studies in

D. salina can lead to the revealing of new members and pathways in calcium signaling in salt tolerance.

Retrograde signaling molecules involved in salinity tolerance

Research on chloroplast redox signaling molecules is limited in algal systems compared to that of plants, especially under abiotic stress tolerance. From the transcriptome data, we have fetched out the retrograde signaling molecules homologous to higher plants and *C. reinhardtii*, which showed significant upregulation in correlation with the ROS and calcium signaling molecule. Our transcriptome data show significant upregulation of most tetrapyrrole biosynthetic pathway enzymes (Figures 6B, E). But pigment analysis also showed decreased chlorophyll molecule (Supplementary Figure 5). So there is a possibility of inhibition of this pathway and accumulation of TPB pathway intermediates. An interesting chloroplastic transcript showing highly significant upregulation was s-FLP, an alternative splicing variant of the FLU-like gene (FLP). In *C. reinhardtii* there is a report of two transcripts of FLU-like gene (FLP) formed by alternative splicing, s-FLP, and l-FLP variants. Also, the alternative splicing of FLP has been reported as a blue light-induced mechanism, and mutant studies revealed chlorophyll biosynthesis intermediates PROTO-IX, Mg-PROTO-IX, and P-Chlide regulate the level of s-FLP and l-FLP (Falciatore et al., 2005).

Further studies have proven the protective role of these splicing variants against oxidative damage under varying light conditions by acting as retrograde signals from the plastid to the nucleus. Among these unique forms, s-FLP shows an upregulation of 8-fold change under hypersalinity conditions. It strongly indicates an FLP variant-mediated complex retrograde regulatory mechanism acting on chloroplast to acclimate under salt stress. The GUN5 mutant study revealed the function of the Chl H subunit of Mg-chelatase in inserting the Mg^{2+} into protoporphyrin IX in chlorophyll biosynthesis (Mochizuki et al., 2001) and also showed a moderate upregulation. Significant upregulation was also observed in heme oxygenase, mediates the cleavage of heme molecules under high salt. Oxidative breakdown of heme generates bilin molecules in algae, and they lack photosensors such as phytochrome. In Chlamydomonas, a bilin molecule-mediated retrograde signaling mechanism has been unveiled, which induces PhANGs expression as photoprotection as well as efficiently triggers metabolic pathways for the detoxification of ROS molecules (Duanmu et al., 2013). In our case, we expect bilin molecules as strong retrograde signals in PhANGs expression and the ROS detoxification system (Figure 9).

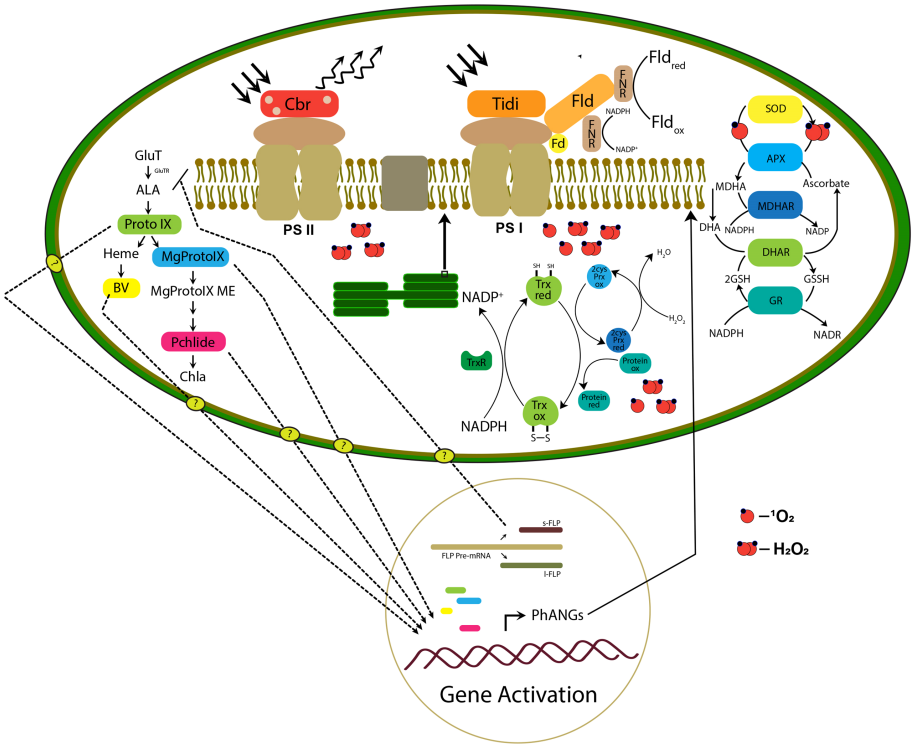


FIGURE 9
The hypothetical model proposed is based on the transcriptomic study representing the retrograde signaling and remodeling of photosynthetic apparatus in the chloroplast. The right panel of the figure represents the TPB pathway intermediate accumulation and retrograde signaling mechanism in which the transport of molecules to the nucleus is still unknown. The photosynthetic apparatus represents possible remodeling initiated by the photosystem-associated gene expression (PhANGs) with the involvement of chloroplast Tidi, CBR, and flavodoxin IsiB. Also, the efficient chloroplast ROS scavenging and protein repair mechanism is represented.

Conclusion

D. salina, as highly salt-tolerant green algae that grow in extremely saline environments, exhibits efficient acclimation mechanisms, especially at the chloroplastic level. Our transcriptomic analysis of chloroplastic genes and experimental approach support an establishment of efficient chloroplast retrograde signaling and possible remodeling of photosynthetic apparatus in *D. salina* under long-term hypersaline conditions. The direct and indirect effect of hypersaline conditions imposes a robust retrograde signaling mechanism by a possible cross-talk between the ROS and calcium signaling mechanism. The transcripts involved in both mechanisms show an uptrend pattern and positive correlation. Chloroplast acts as a primary environmental sensor. ROS accumulation in chloroplast was observed and the accumulated ROS leads to proportionate damage to photosystem proteins. The transcript analysis has proposed an efficient PSII repair and assembly mechanism such as Psb27, Psb28, PsbZ PsbT, and PsbM photosystem subunits. The major two transcripts upregulated in the PSI system chloroplast Tidi and flavodoxin IsiB, tempting to suggest a possible remodeling at a super complex level to rebalance the energy distribution and electron transport in the apparatus. The expression of CBR protein which was earlier reported in *D. bardawil*, having a photoprotection function in high light showed a correlation with our carotenoid biosynthesis transcripts upregulation; this suggests a protective role of CBR under the hypersaline condition to the PS II supercomplex. Expression pattern of ROS scavenging associated and protein refolding related transcripts such as Glutathione S-transferase, glutathione reductase, Chloroplastic 2-cys peroxiredoxin, NADPH-dependent thioredoxin reductase (NTR), peptide methionine sulfoxide reductase (PMSR), thioredoxin-dependent peroxidase, thioredoxin-like protein supports an efficient retrograde signaling mechanism. Tetrapyrrole biosynthesis pathway intermediates such as PROTO-IX, Mg-PROTO-IX, and P-Chlide are reported to act as retrograde signals in several studies. Also, the oxidation product of heme, the biliverdin molecules, suggested acting as retrograde signals. Our transcript analysis of TPB pathway enzymes and expression of the splicing variant of s-FLP, which was earlier reported as a negative regulator of the TPB pathway, supports the contribution of the TPB pathway in possible retrograde signaling mechanism and PhANGs expression. All these mechanisms play synergistically to rebalance the photosynthetic apparatus's redox state and help cope with the hypersaline environment in *D. salina*. We hope that the genes expressed in *D. salina* to the hypersaline condition could be transferred to the land plants/fresh algae to withstand the high salinity conditions and improve crop yield.

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

RS procured the grants and designed the experiments. PR performed most of the experiments while RY, PS and AK did some experiments. PR and NP together analyzed the transcriptome data. RS and PR analyzed the results and wrote the first draft. All authors contributed to the article and approved the submitted version.

Acknowledgments

RS was supported by the Joint UGC-ISF Research Grant - File No. 6-8/2018 (IC), Council of Scientific and Industrial Research (No. 38(1504)/21/EMR-II), Science & Engineering Research Board (CRG/2020/000489), and, DST-FIST and UGC-SAP, DBT-Builder, Govt. of India, for financial support. PR acknowledged the Prime minister Research Fellowship. We thank Prof. Nathan Nelson, Department of Biochemistry and Molecular Biology, Tel Aviv University, Israel, for his constructive suggestions.

Conflict of interest

Authors NP and AK were employed by Novelgene Technologies 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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1192258/full#supplementary-material

SUPPLEMENTARY FIGURE 1
(A) Principal component analysis of transcript data obtained from control and hypersaline condition to see how replicates within the sample cluster together in one group and how samples differ in two different conditions. **(B)** sample correlation among the samples and within the replicates represented in heatmap form.

SUPPLEMENTARY FIGURE 2
Differentially expressed gene expression pattern. **(A)** A volcano plot of differentially expressed genes was generated among the mapped genes using the Dseq2 program with a fold change of $\geq$ 0.5 and FDR $\leq$ 0.05. **(B)** Heat map generated from the differentially expressed genes from the triplicates of each sample.

SUPPLEMENTARY FIGURE 3
Real-time PCR analysis of 10 significant differentially expressed transcripts selected. RPL2 ribosomal protein was selected as the internal control. Each gene was tested with three biological replicates and was calculated.

SUPPLEMENTARY FIGURE 4
Growth kinetics and biomass measurement. **(A)** Cell density and **(B)** cell number of D. salina cells grown under control, 2M, and 3M conditions were analyzed throughout 24 hrs. **(C)** Morphological observation of change in color of culture from green (control), yellowish-green (hypersaline). **(D)** Biomass obtained after 72 hrs. of growth under control, 2M, and 3M conditions were quantified after lyophilization for 24 hrs. Statistical analysis was done using one-way ANOVA (Bonferroni t test) was performed to measure the P-values (ns, not significant). Asterisks indicate the level of significance of the drought effect compared to the respective control: ***P < 0.001.

SUPPLEMENTARY FIGURE 5
The chlorophyll and carotenoid pigment content in control and treated samples were measured per 1 ml of cells after 73 h of growth. UV-Visible spectrophotometer measurements in different wavelengths 480, 520, 645, and 663. The concentration was measured using the equations described in Lichtenthal (Lichtenthaler, 1987) and Porra et al. (1989).

SUPPLEMENTARY TABLE 1
Primers designed for real-time PCR analysis of 10 significant differentially expressed transcripts.

SUPPLEMENTARY TABLE 2
List of transcripts in **(A)** Photosystem-associated transcripts, **(B)** ROS and calcium signaling associated transcripts, **(C)** Carotenoid biosynthesis associated pathway, along with trinity IDs, log2 fold change and FPKM values.

References

Abdellaoui, N., Kim, M. J., and Choi, T. J. (2019). Transcriptome analysis of gene expression in chlorella vulgaris under salt stress. *World J. Microbiol. Biotechnol.* 35 (9), 1–11. doi: 10.1007/s11274-019-2718-6

Adamska, I., Kloppstech, K., and Ohad, I. (1992). UV Light stress induces the synthesis of the early light-inducible protein and prevents its degradation. *J. Biol. Chem.* 267 (34), 24732–24737. doi: 10.1016/S0021-9258(18)35825-3

Asada, K. (1987). Production and scavenging of active oxygen in photosynthesis. *Photoinhibition*, 227–287.

Asada, K. (2006). Production and scavenging of reactive oxygen species in chloroplasts and their functions. *Plant Physiol.* 141 (2), 391–396. doi: 10.1104/pp.106.082040

Awad, J., Stotz, H. U., Fekete, A., Krischke, M., Engert, C., Havaux, M., et al. (2015). 2-cysteine peroxiredoxins and thylakoid ascorbate peroxidase create a water-water cycle that is essential to protect the photosynthetic apparatus under high light stress conditions. *Plant Physiol.* 167 (4), 1592–1603. doi: 10.1104/pp.114.255356

Balsera, M., and Buchanan, B. B. (2019). Evolution of the thioredoxin system as a step enabling adaptation to oxidative stress. Free radical biology and medicine. 140, 28–35. doi: 10.1016/j.freeradbiomed.2019.03.003

Ben-Amotz, A., and Avron, M. (1983). On the factors which determine massive fl-carotene accumulation in the halotolerant alga dunaliella bardawil. *Plant Physiol.* 72, 593–597. doi: 10.1104/pp.72.3.593

Bentley, F. K., Luo, H., Dilbeck, P., Burnap, R. L., and Eaton-Rye, J. J. (2008). Effects of inactivating psbM and psbT on photodamage and assembly of photosystem II in synechocystis sp. PCC 6803. *Biochemistry* 47 (44), 11637–11646. doi: 10.1021/bi800804h

Boehm, M., Yu, J., Reisinger, V., Beckova, M., Eichacker, L. A., Schlodder, E., et al. (2012). Subunit composition of CP43-less photosystem II complexes of synechocystis sp. PCC 6803: implications for the assembly and repair of photosystem II. *Philos. Trans. R. Soc. B: Biol. Sci.* 367 (1608), 3444–3454.

Borowitzka, M. A. (2013). "Dunaliella: biology, production, and markets," in *The handbook of microalgal culture: applied phycology and biotechnology*, *2nd ed*. Eds. A. Richmond and Q. Hu (Hoboken, NJ: Wiley-Blackwell), p 359–p 368.

Brini, F. B., Hanin, M., Mezghani, I., Berkowitz, G. A., and Masmoudi, K. (2007). Overexpression of wheat Na+/H+ antiporter TNHX1 and h+-pyrophosphatase TVP1 improve salt- and droughtstress tolerance in arabidopsis thaliana plants. *J. Expt Bot.* 58, 301–308. doi: 10.1093/jxb/erl251

Caffarri, S., Tibiletti,, Jennings., C., and Santabarbara, (2014). A comparison between plant photosystem I and photosystem II architecture and functioning. *Curr. Protein Pept. Sci.* 15, 296–331. doi: 10.2174/1389203715666140327102218

Caspy, I., Fadeeva, M., Mazor, Y., and Nelson, N. (2023). Structure of dunaliella photosystem II reveals conformational flexibility of stacked and unstacked supercomplexes. *Elife* 12, e81150. doi: 10.7554/eLife.81150

Cazzonelli, C. I., and Pogson, B. J. (2010). Source to sink: regulation of carotenoid biosynthesis in plants. *Trends Plant Sci.* 15 (5), 266–274. doi: 10.1016/j.tplants.2010.02.003

Cejudo, F. J., González, M. C., and Pérez-Ruiz, J. M. (2021). Redox regulation of chloroplast metabolism. *Plant Physiol.* 186 (1), 9–21. doi: 10.1093/plphys/kiaa062

Costa, A., Drago, I., Behera, S., Zottini, M., Pizzo, P., Schroeder, J. I., et al. (2010). H2O2 in plant peroxisomes: an *in vivo* analysis uncovers a Ca2+-dependent scavenging system. *Plant J.* 62 (5), 760–772. doi: 10.1111/j.1365-313X.2010.04190.x

Croce, R., Cinque, G., Holzwarth, A. R., and Bassi, R. (2000). The soret absorption properties of carotenoids and chlorophylls in antenna complexes of higher plants. Photosynthesis research. 64, 221–231. doi: 10.1023/A:1006455230379

Devadasu, E., Kanna, S. D., Neelam, S., Yadav RM, R., Nama, S., Akhtar, P., et al. (2023). Long-and short-term acclimation of the photosynthetic apparatus to salinity in chlamydomonas reinhardtii. the role of Stt7 protein kinase. *Front. Plant Sci.* 14, 948. doi: 10.3389/fpls.2023.1051711

Dobáková, M., Sobotka, R., Tichy, M., and Komenda, J. (2009). Psb28 protein is involved in the biogenesis of the photosystem II inner antenna CP47 (PsbB) in the cyanobacterium synechocystis sp. PCC 6803. *Plant Physiol.* 149 (2), 1076–1086. doi: 10.1104/pp.108.130039

Duanmu, D., Casero, D., Dent, R. M., Gallaher, S., Yang, W., Rockwell, N. C., et al. (2013). Retrograde bilin signaling enables chlamydomonas greening and phototrophic survival. *Proc. Natl. Acad. Sci.* 110 (9), 3621–3626. doi: 10.1073/pnas.1222375110

Falciatore, A., Merendino, L., Barneche, F., Ceol, M., Meskauskiene, R., Apel, K., et al. (2005). The FLP proteins act as regulators of chlorophyll synthesis in response to light and plastid signals in chlamydomonas. *Genes Dev.* 19 (1), 176–187. doi: 10.1101/gad.321305

Foyer, C. H., and Noctor, G. (2000). Oxygen processing in photosynthesis: regulation and signalling. *New Phytol.* 146, 359–388. doi: 10.1046/j.1469-8137.2000.00667.x

Gao, F., Nan, F., Feng, J., Lü, J., Liu, Q., Liu, X., et al. (2021). Transcriptome profile of dunaliella salina in yuncheng salt lake reveals salt-stress-related genes under different salinity stresses. *J. Oceanology Limnology* 39 (6), 2336–2362. doi: 10.1007/s00343-021-0164-4

Golbeck, J. H. (1992). Structure and function of photosystem i. *Annu. Rev. Plant Physiol. Plant Mol. Bioi* 43, 293–324. doi: 10.1146/annurev.pp.43.060192.001453

Gonzalez, A., Vera, J., Castro, J., Dennett, G., Mellado, M., Morales, B., et al. (2010). Co-Occurring increases of calcium and organellar reactive oxygen species determine differential activation of antioxidant and defense enzymes in ulva compressa (Chlorophyta) exposed to copper excess. *Plant Cell Environ.* 33 (10), 1627–1640. doi: 10.1111/j.1365-3040.2010.02169.x

Gupta, R. (2020). The oxygen-evolving complex: a super catalyst for life on earth, in response to abiotic stresses. *Plant Signaling Behav.* 15 (12), 1824721. doi: 10.1080/15592324.2020.1824721

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

Inoue, K. (2004). Carotenoid hydroxylation–P450 finally! *Trends Plant Sci.* 9 (11), 515–517. doi: 10.1016/j.tplants.2004.09.001

Iwai, M., Katoh, H., Katayama, M., and Ikeuchi, M. (2004). PSII-Tc protein plays an important role in dimerization of photosystem II. *Plant Cell Physiol.* 45 (12), 1809–1816. doi: 10.1093/pcp/pch207

Ji, C., Mao, X., Hao, J., Wang, X., Xue, J., Cui, H., et al. (2018). Analysis of bZIP transcription factor family and their expressions under salt stress in chlamydomonas reinhardtii. *Int. J. Mol. Sci.* 19 (9), 2800. doi: 10.3390/ijms19092800

Katz, A., Waridel, P., Shevchenko, A., and Pick, U. (2007). Salt-induced changes intheplasmamembraneproteomeofthehalotolerant alga dunaliella salina as revealed by blue native gel electrophoresis and nano-LC-MS/MS analysis. *Mol. Cell. Proteomics* 6, 1459–1472. doi: 10.1074/mcp.m700002-mcp200

Knight, H., Trewavas, A. J., and Knight, M. R. (1997). Calcium signalling in arabidopsis thaliana responding to drought and salinity. *Plant J.* 12 (5), 1067–1078. doi: 10.1046/j.1365-313X.1997.12051067.x

Kodru, S., Malavath, T., Devadasu, E., Nellaepalli, S., Stirbet, A., and Subramanyam, R. (2015). The slow s to m rise of chlorophyll a fluorescence reflects transition from state 2 to state 1 in the green alga chlamydomonas reinhardtii. *Photosynthesis Res.* 125, 219–231. doi: 10.1007/s11120-015-0084-2

Komenda, J., Knoppová, J., Kopečná, J., Sobotka, R., Halada, P., Yu, J., et al. (2012). The Psb27 assembly factor binds to the CP43 complex of photosystem II in the cyanobacterium synechocystis sp. PCC 6803. *Plant Physiol.* 158 (1), 476–486. doi: 10.1104/pp.111.184184

Koyama, M. L., Levesley, A., Koebner, R. M. D., Flowers, T. J., and Yeo, A. R. (2001). Quantitative trait loci for component physiological traits determining salt tolerance in rice. *Plant Physiol.* 125, 406–422. doi: 10.1104/pp.125.1.406

Levy, H., Tal, T., Shaish, A., and Zamir, A. (1993). Cbr, an algal homolog of plant early light-induced proteins, is a putative zeaxanthin binding protein. *J. Biol. Chem.* 268 (28), 20892–20896. doi: 10.1016/S0021-9258(19)36870-X

Liang, M. H., Xie, H., Chen, H. H., Liang, Z. C., and Jiang, J. G. (2020). Functional identification of two types of carotene hydroxylases from the green alga dunaliella bardawil rich in lutein. *ACS Synthetic Biol.* 9 (6), 1246–1253. doi: 10.1021/acssynbio.0c00070

Lichtenthaler, H. K. (1987). "Chlorophylls and carotenoids: pigments of photosynthetic biomembranes," in *Methods in enzymology*, vol. 148. (Academic Press), 350–382. doi: 10.1016/0076-6879(87)48036-1

Lodeyro, A. F., Ceccoli, R. D., Karlusich, J. J. P., and Carrillo, N. (2012). The importance of flavodoxin for environmental stress tolerance in photosynthetic microorganisms and transgenic plants. mechanism, evolution and biotechnological potential. *FEBS Lett.* 586 (18), 2917–2924. doi: 10.1016/j.febslet.2012.07.026

Lv, H., Kim,, Park,, Baek,, Hyeonjun,, Polle, E. W., et al. (2021). Comparative transcriptome analysis of short-term responses to salt and glycerol hyperosmotic stress in the green alga dunaliella salina. *Algal Res.* 53 (2021), 102147. doi: 10.1016/j.algal.2020.102147

Macpherson, A. N., Telfer, A., Barber, J., and Truscott, T. G. (1993). Direct detection of singlet oxygen from isolated photosystem II reaction centres. *Biochim. Biophys. Acta (BBA)-Bioenergetics* 1143 (3), 301–309. doi: 10.1016/0005-2728(93)90201-P

Mayfield, S. P., Schirmer-Rahire, M., Frank, G., Zuber, H., and Rochaix, J. D. (1989). Analysis of the genes of the OEE1 and OEE3 proteins of the photosystem II complex from chlamydomonas reinhardtii. *Plant Mol. Biol.* 12 (6), 683–693. doi: 10.1007/BF00044159

Mittler, R. (2017). ROS are good. *Trends Plant Sci.* 22 (1), 11–19. doi: 10.1016/j.tplants.2016.08.002

Mochizuki, N., Brusslan, J. A., Larkin, R., Nagatani, A., and Chory, J.. (2001). Arabidopsis genomes uncoupled 5 (GUN5) mutant reveals the involvement of mg-chelatase h subunit in plastidto-nucleus signal transduction. *Proc. Natl. Acad. Sci. U.S.A* 98, 2053–2058. doi: 10.1073/pnas.98.4.2053

Momonoki, Y. S., Yamamoto, K., and Oguri, S. (2009). Molecular cloning of oxygen-evolving enhancer genes induced by salt treatment in a halophyte, salicornia europaea l. *Plant production Sci.* 12 (2), 193–198. doi: 10.1626/pps.12.193

Morehouse, K. M., and Mason, R. P. (1988). The transition metal-mediated formation of the hydroxyl free radical during the reduction of molecular oxygen by ferredoxin-ferredoxin: NADP+ oxidoreductase. *J. Biol. Chem.* 263 (3), 1204–1210. doi: 10.1016/S0021-9258(19)57287-8

Müller-Schüssele, S. J., Wang, R., Gütle, D. D., Romer, J., Rodriguez-Franco, M., Scholz, M., et al. (2020). Chloroplasts require glutathione reductase to balance reactive oxygen species and maintain efficient photosynthesis. *Plant J.* 103 (3), 1140–1154. doi: 10.1111/tpj.14791

Munns, R., James, R. A., and Lauchli, A. (2006). Approaches to increasing the salt tolerance of wheat and other cereals. *J. Exp. Bot.* 57, 1025–1043. doi: 10.1093/jxb/erj100

Narendra, S., Venkataramani, S., Shen, G., Wang, J., Pasapula, V., Lin, Y., et al. (2006). The arabidopsis ascorbate peroxidase 3 is a peroxisomal membrane-bound antioxidant enzyme and is dispensable for arabidopsis growth and development. *J. Exp. Bot.* 57 (12), 3033–3042. doi: 10.1093/jxb/erl060

Neelam, S., and Subramanyam, R. (2013). Alteration of photochemistry and protein degradation of photosystem II from chlamydomonas reinhardtii under high salt grown cells. Journal of photochemistry and photobiology b: biology. 124, 63–70. doi: 10.1016/j.jphotobiol.2013.04.007

Niyogi, K. K., Björkman, O., and Grossman, A. R. (1997). The roles of specific xanthophylls in photoprotection. *Proc. Natl. Acad. Sci.* 94 (25), 14162–14167. doi: 10.1073/pnas.94.25.14162

Nott, A., Jung, H. S., Koussevitzky, S., and Chory, J. (2006). Plastid-to-nucleus retrograde signaling. *Annu. Rev. Plant Biol.* 57, 739–759. doi: 10.1146/annurev.arplant.57.032905.105310

Parida, A. K., Das, A. B., and Mittra, B. (2003). Effects of NaCl stress on the structure, pigment complex composition, and photosynthetic activity of mangrove bruguiera parviflora chloroplasts. *Photosynthetica* 41, 191–200. doi: 10.1023/B:PHOT.0000011951.37231.69

Perez-Boerema, A., Klaiman, D., Caspy, I., Netzer-El, S. Y., Amunts, A., and Nelson, N. (2020). Structure of a minimal photosystem I from the green alga dunaliella salina. *Nat. Plants* 6 (3), 321–327. doi: 10.1038/s41477-020-0611-9

Pogson, B. J., Woo, N. S., Forster, B., and Small, I. D. (2008). Plastid signalling to the nucleus and beyond. *Trends Plant Sci.* 13, 602–609. doi: 10.1016/j.tplants.2008.08.008

Porra, R. J., Thompson, W. A. A., and Kriedemann, P. E. (1989). Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochim. Biophys. Acta (BBA)-Bioenergetics* 975 (3), 384–394. doi: 10.1016/S0005-2728(89)80347-0

Qiao, J., Huang, S., Te, R., Wang, J., Chen, L., and Zhang, W. (2013). Integrated proteomic and transcriptomic analysis reveals novel genes and regulatory mechanisms involved in salt stress responses in synechocystis sp. PCC 6803. *Appl. Microbiol. Biotechnol.* 97 (18), 8253–8264. doi: 10.1007/s00253-013-5139-8

Romero, H. M., Berlett, B. S., Jensen, P. J., Pell, E. J., and Tien, M. (2004). Investigations into the role of the plastidial peptide methionine sulfoxide reductase in response to oxidative stress in arabidopsis. *Plant Physiol.* 136 (3), 3784–3794. doi: 10.1104/pp.104.046656

Roxas, V. P., Lodhi, S. A., Garrett, D. K., Mahan, J. R., and Allen, R. D. (2000). Stress tolerance in transgenic tobacco seedlings that overexpress glutathione s-transferase/glutathione peroxidase. *Plant Cell Physiol.* 41 (11), 1229–1234. doi: 10.1093/pcp/pcd051

Schägger, H., and Von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical Biochem.* 166 (2), 368–379. doi: 10.1016/0003-2697(87)90587-2

Schürmann, P., and Buchanan, B. B. (2008). The ferredoxin/thioredoxin system of oxygenic photosynthesis. *Antioxidants Redox Signaling* 10 (7), 1235–1274. doi: 10.1089/ars.2007.1931

Shen, L., Huang, Z., Chang, S., Wang, W., Wang, J., Kuang, T., et al. (2019). Structure of a C2S2M2N2-type PSII–LHCII supercomplex from the green alga chlamydomonas reinhardtii. *Proc. Natl. Acad. Sci.* 116 (42), 21246–21255. doi: 10.1073/pnas.1912462116

Shi, H., Ishitani, M., Kim, C., and Zhu, J. K. (2000). The arabidopsis thaliana salt tolerance gen SOS1 encodes a putative Na+/H+ antiporter. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6896–9601. doi: 10.1073/pnas.120170197

Shi, H., Lee, B., Wu, S. J., and Zhu, J. K. (2003). Overexpression of a plasma membrane Na+/H+ antiporter gen improves salt tolerance in arabidopsis thaliana. *Nat. Biotech.* 21, 81–85. doi: 10.1038/nbt766

Shochat, S., Adir, N., Gal, A., Inoue, Y., Mets, L., and Ohad, I. (1990). Photoinactivation of photosystem II and degradation of the D1 protein are reduced in a cytochrome b6/f-less mutant of chlamydomonas reinhardtii. *Z. für Naturforschung C* 45 (5), 395–401. doi: 10.1515/znc-1990-0514

Singla-Pareek, S. L., Yadav, S. K., Pareek, A., and Reddy. M.K. and Sopory, S. K. (2008). Enhancing salt tolerance in a crop plant by overexpression of glyoxalase II. *Transgenic Res.* 17, 171–180. doi: 10.1007/s11248-007-9082-2

Strasser, R. J. (1992). "Govindjee," in *Research in photosynthesis*, vol. 2 . Ed. M. Murata (Dordrecht: Kluwer), 29^32.

Varsano, T., Kaftan, D., and Pick, U. (2007). Effects of iron deficiency on thylakoid membrane structure and composition in the alga dunaliella salina. *J. Plant Nutr.* 26, 2197–2210. doi: 10.1081/PLN-120024275

Varsano, T., Kaftan,, and Pick, (2007). Effects of iron deficiency on thylakoid membrane structure and composition in the alga dunaliella salina. *J. Plant Nutr.* 26, 10–11. doi: 10.1081/PLN-120024275

Varsano, T., Wolf., G., and Pick, (2006). A chlorophyll a/b-binding protein homolog that is induced by iron deficiency is associated with enlarged photosystem I units in the eucaryotic alga dunaliella salina. *THE J. OF Biol. Chem.* 281 (15), 10305–10315. doi: 10.1074/jbc.M511057200

Wang, N., Qian, Z., Luo, M., Fan, S., Zhang, X., and Zhang, L. (2018). Identification of salt stress responding genes using transcriptome analysis in green alga chlamydomonas reinhardtii. *Int. J. Mol. Sci.* 19 (11), 3359. doi: 10.3390/ijms19113359

Wang, N., Qian,, Luo,, Fan,, Zhang,, and Zhang, (2018). Identification of salt stress responding genes using transcriptome analysis in green alga chlamydomonas reinhardtii. *Int. J. Mol. Sci.* 19 (11), 3359. doi: 10.3390/ijms19113359

Wilhelm, H., and Wilhelm., T. (1993). Why do thylakoid membranes from higher plants form grana stacks. *TIBs* 18, 415–419. doi: 10.1016/0968-0004(93)90136-B

Yang, T. P. B. W., and Poovaiah, B. W. (2002). Hydrogen peroxide homeostasis: activation of plant catalase by calcium/calmodulin. *Proc. Natl. Acad. Sci.* 99 (6), 4097–4102. doi: 10.1073/pnas.052564899

Yoshida, K., Igarashi, E., Wakatsuki, E., Miyamoto, K., and Hirata, K. (2004). Mitigation of osmotic and salt stresses by abscisic acid through reduction of stress-derived oxidative damage in chlamydomonas reinhardtii. *Plant Sci.* 167 (6), 1335–1341. doi: 10.1016/j.plantsci.2004.07.002

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 07 July 2023
DOI 10.3389/fpls.2023.1198474

OPEN ACCESS

EDITED BY
Bhanu Prakash Petla,
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

REVIEWED BY
Suleyman I. Allakhverdiev,
Russian Academy of Sciences (RAS), Russia
Dinakar Challabathula,
Central University of Tamil Nadu, India

*CORRESPONDENCE
Rajagopal Subramanyam
srgsl@uohyd.ernet.in

†PRESENT ADDRESS
Sureshbabu Marriboina,
The French Associates Institute for Agriculture and Biotechnology of Drylands, The J. Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion, Israel

RECEIVED 01 April 2023
ACCEPTED 11 May 2023
PUBLISHED 07 July 2023

CITATION
Yadav RM, Marriboina S, Zamal MY, Pandey J and Subramanyam R (2023) High light-induced changes in whole-cell proteomic profile and its correlation with the organization of thylakoid super-complex in cyclic electron transport mutants of *Chlamydomonas reinhardtii*. *Front. Plant Sci.* 14:1198474. doi: 10.3389/fpls.2023.1198474

High light-induced changes in whole-cell proteomic profile and its correlation with the organization of thylakoid super-complex in cyclic electron transport mutants of *Chlamydomonas reinhardtii*

Ranay Mohan Yadav, Sureshbabu Marriboina†, Mohammad Yusuf Zamal, Jayendra Pandey and Rajagopal Subramanyam*

Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad, India

Light and nutrients are essential components of photosynthesis. Activating the signaling cascades is critical in starting adaptive processes in response to high light. In this study, we have used wild-type (WT), cyclic electron transport (CET) mutants like Proton Gradient Regulation (PGR) (*PGRL1*), and *PGR5* to elucidate the actual role in regulation and assembly of photosynthetic pigment–protein complexes under high light. Here, we have correlated the biophysical, biochemical, and proteomic approaches to understand the targeted proteins and the organization of thylakoid pigment–protein complexes in the photoacclimation. The proteomic analysis showed that 320 proteins were significantly affected under high light compared to the control and are mainly involved in the photosynthetic electron transport chain, protein synthesis, metabolic process, glycolysis, and proteins involved in cytoskeleton assembly. Additionally, we observed that the cytochrome (*Cyt*) b_6 expression is increased in the *pgr5* mutant to regulate proton motive force and ATPase across the thylakoid membrane. The increased Cyt b_6 function in *pgr5* could be due to the compromised function of chloroplast (cp) ATP synthase subunits for energy generation and photoprotection under high light. Moreover, our proteome data show that the photosystem subunit II (PSBS) protein isoforms (PSBS1 and PSBS2) expressed more than the Light-Harvesting Complex Stress-Related (LHCSR) protein in *pgr5* compared to WT and *pgrl1* under high light. The immunoblot data shows the photosystem II proteins D1 and D2 accumulated more in *pgrl1* and *pgr5* than WT under high light. In high light, CP43 and CP47 showed a reduced amount in *pgr5* under high light due to changes in chlorophyll and carotenoid content around the PSII protein, which coordinates as a cofactor for efficient energy transfer from the light-harvesting antenna to the photosystem core. BN-PAGE and circular dichroism studies indicate changes in macromolecular assembly and thylakoid super-complexes destacking in *pgrl1*

and *pgr5* due to changes in the pigment–protein complexes under high light. Based on this study, we emphasize that this is an excellent aid in understanding the role of CET mutants in thylakoid protein abundances and super-complex organization under high light.

KEYWORDS

Chlamydomonas reinhardtii, cyclic electron transport, differentially abundant proteins, high light, LHCSR, PSBS, photosynthesis, thylakoid super-complex

Introduction

Light is a predominant need for photosynthesis, and its quantity can influence photosynthetic efficiency. Optimum photosynthesis requires a subtle balance between photoreception and utilization of ideal light intensity and a constant supply of nutrients. Photosynthetic organisms that grow under natural conditions do not always receive the optimum light and nutrients, leading to photosynthetic imbalance. Photosynthetic organisms have been adapted to various photoprotective mechanisms and changes in core and antenna protein to acclimate under abiotic stress conditions. High light is particularly problematic because excess light affects the charge separation and damages the pigment–protein complexes involved in photosynthetic electron transport, which causes an imbalance in the reducing powers of NADPH and ATP for carbon fixation and photoacclimation. Absorption of light energy and its conversion into chemical form are carried out by the coordinated function of specialized thylakoid membrane protein complexes like photosystems (PS) I and PSII along with their light-harvesting (LHC) antenna LHC I and II, the cytochrome (Cyt) b_6f, and the ATP–synthase complex. These protein complexes are well-organized as super-complexes in the appressed sections of the thylakoid membrane (Shen et al., 2019; Wietrzynski et al., 2020). In *Chlamydomonas (C.) reinhardtii*, these super-complexes consist of PSII–LHCII, PSI–LHCI, LHC trimers, monomers, and other small subunits like CP26 and CP29 (Croce and van Amerongen, 2020).

The antenna proteins associate with reaction centers (RCs), allowing overexcitation of energy from the membrane until it drives stable photochemistry (Bennett et al., 2018). PSII has been modified to minimize ROS generation due to the RC complex strategy, redox tuning (Rutherford et al., 2012; Brinkert et al., 2016), and supporting mechanism dissipating excess energy and inhibiting the damage of photosystems (Allakhverdiev et al., 2008; Peers et al., 2009; Roach et al., 2020). In photosynthetic organisms, photoinactivation of PSII is rapidly repaired and reflects the balance between the light-induced damage to PSII and the repair of PSII (Allakhverdiev and Murata, 2004). Photosynthetic organisms have evolved with different agencies to dissipate extra energy absorbed by photosynthetic machinery. These mechanisms are collectively called non-photochemical quenching (NPQ). Energy quenching, qE, is mediated by the light harvesting complex stress related 3 (LHCSR3) antenna after the protonation of its lumen residues (Peers et al., 2009; Tian et al., 2019). In green algae, *C. reinhardtii*, the induction of qE requires gene products of LHCSR, i.e., LHCSR1, LHCSR3.1/3.2, and PSBS (Peers et al., 2009; Correa-Galvis et al., 2016). The PSBS protein does not form a complex with pigments (Bonente et al., 2008). More likely, its function depends on the protonation of the luminal-exposed glutamine residues, which can further activate the zeaxanthin and lutein-dependent quenching process in LHCs through chlorophyll–carotenoid charge transfer (Niyogi et al., 2001; Ahn et al., 2008; Bonente et al., 2008). PSI is particularly resistant to photoinhibition under oxidative stress conditions due to the high efficiency of protective mechanisms, which regulate the flow of electrons to the PSI donor side, including NPQ, lumen pH-dependent regulation of Cyt b_6f activity, and even PSII photoinhibition (Tikkanen and Aro, 2014). Electron consumption at the PSI acceptor side through the Calvin–Benson cycle, photorespiration, and cyclic and pseudo-cyclic electron flow are also protective factors that avoid PSI over-reduction (Yamori, 2016). In algae, the cyclic electron transport is operated by *pgrl1* and *pgr5* proteins, but the role of these proteins in high light photoprotection and the metabolic process is still unknown.

Proteomics must be an essential match to biochemical and biophysical data to improve the biological perspective on plant biology. This approach is applied to discover the dynamic changes in protein expression to a specific abiotic stress in various organisms. Several studies have investigated protein expression using a perspective method called whole proteome analysis. This technique can determine the relative levels of peptides in abiotic stress and normal physiological condition. The difference or similarity in expression levels under different light conditions can be revealed by differential expression in the peptide levels of specific sets of proteins. The impact or influence of high light on a particular biological process can be reported by changes in the proteome levels of a specific protein that has a functional role in the performance of photosynthesis. A detailed knowledge of biological pathways and processes is required to determine the reasons for microalgae acclimation and biomass production under high light. Hence, a new straightforward proteome analysis platform is necessary to elucidate these advantages and identify potential targets for higher plants to improve crop production.

Proteomics is a promising approach that could deliver insights into and knowledge of the protein, which helps improve and utilize

conventional breeding and genetic engineering to make new traits or recombinant plants with better-quality agronomical characteristics (Gao et al., 2019). Notably, several efforts have been accomplished to achieve gene expression in alfalfa in salt conditions (Jin et al., 2010; Postnikova et al., 2013). High-throughput techniques have rapidly been discovered and helped us better understand plants' transcriptome and proteome changes during light and dark (Vogel et al., 2014; Liang et al., 2016; Crisp et al., 2017; Huang et al., 2019). The present study characterized a bio-physiological, biochemical study integrated with nLCMS/MS-based proteomic analysis to quantify the global database resource by the eukaryotic microalgae *C. reinhardtii*. It also determined how *C. reinhardtii* globally allocates resources in response to high light stress.

The applied strategies can quantitatively and temporally account for all apparent behaviors, including changes in physiological characteristics, biochemical parameters, and whole-cell proteome content change in WT, *pgrl1*, and *pgr5* mutants. The cyclic electron transport *pgrl1* and *pgr5* proteins are extensively involved in energy generation, which is utilized for metabolic processes. These mutants showed reduced photosynthetic activity and non-photochemical quenching (Yadav et al., 2020; Chouhan et al., 2023). Recently, our findings showed that more biomass and lipid production are fundamental mechanisms of photoprotection in response to light stress (Chouhan et al., 2022). We have grown *C. reinhardtii* cells in Tris-acetate phosphate (TAP) to address these questions in high light conditions. No reports have been documented in the proteomic responses of microalgae *C. reinhardtii* when the cells were grown under high light intensities, mimicking natural conditions. The present study examined the effect of high light at different light intensities on green microalga *C. reinhardtii* through proteomic analysis combined with a biochemical study.

Materials and methods

Growth condition

C. reinhardtii wild-type strain 137*c* and mutants *pgrl1* and *pgr5* are a gift from Prof. Gilles Peltier, CEA-CNRS-Aix Marseille Université, France, and Prof. Michael Hippler, University of Munster. The cells were grown in a TAP medium in an Algaetron growth chamber (AG 230-ECO, Czech Republic) with white fluorescent LED light shaken at 120 rpm (Yadav et al., 2020). The WT, *pgrl1*, and *pgr5* cells were grown in 50 µmol photons $m^{-2}s^{-1}$ (optimal growth), moderate (250 µmol photons $m^{-2}s^{-1}$), and high light (500 µmol photons $m^{-2}s^{-1}$). The chlorophyll content was analyzed using 80% acetone. We have used three or four biological replicates for all the experiments.

Measurement of photosynthetic O_2 evolution and respiratory O_2 consumption

The light-saturated oxygen evolution and consumption rate (under ~600 µmol photons m^{-2} s^{-1}) were measured from 2-ml samples of intact cells with a Clark-type oxygen electrode (Hansatech Instruments Ltd, Norfolk, United Kingdom) at 25°C in a TAP medium. 20 µl of the artificial electron acceptor 2,5-dichloro1,4-benzoquinone (DCBQ, 0.5 mM) and 20 µl of 100 mM NH_4Cl were injected into the electrode reaction. The cell mixture was kept in the dark for 2 min, and then the light was turned on (or O_2 evolution rate measurement) for 6-min continuous light illumination (600 µmol photons m^{-2} s^{-1}) until the straight bend line no longer appeared. Respiratory oxygen consumption was measured by adding 100 mM sodium azide, 50 mM methyl viologen, 10 mM DCPIC, 50 mM DCMU, and 500 mM ascorbate and the cell is kept in the dark until the chamber oxygen is not entirely consumed. The oxygen electrode was calibrated with air-saturated water (100% oxygen) by adding sodium dithionite to achieve 0% oxygen. The final chlorophyll concentration used for each replicate was 20 µg ml^{-1}. For better activity, cells were mixed with buffer 5 mM Tris-HCl (pH 7.5), 10 M $MgCl_2$, 5 mM $CaCl_2$, 30 mM KCl, and 0.25 M sorbitol (Nakalai, Japan). Experiments were the average of three biological replicates.

Room temperature fluorescence

Room temperature spectra of chlorophyll fluorescence were recorded by a fluorescence spectrophotometer (Perkin Elmer, LS-55). 10 µg chlorophyll concentration was taken for each measurement. Fluorescence spectra were recorded from 800 to 600 nm, exciting at 435 nm with a scan speed of 100 nm/min. Each spectrum is an accumulation of two scans.

Measurement of circular dichroism under high light

The circular dichroism (CD) spectra of thylakoid membranes isolated from control and high light-treated cells were measured by a J-810 spectropolarimeter (Jasco Inc., Easton, MD, USA). The spectra were recorded within a visible wavelength range (400–800 nm) by a quartz cell with an optical path length of 1 cm and three bandwidths. The sample thylakoid was dissolved into a thylakoid suspension buffer of 20 mM Tris (pH 7.5), 0.3 M sorbitol, 10 mM $MgCl_2$, 10 mM NaF, and 5 mM $CaCl_2$ with a concentration of 20 µg/ml. The three scans were repeated with a 100-nm/min continuous scan speed for each replicate. Three independent experiments were taken for the measurement.

Blue native PAGE analysis of high light-treated thylakoid membranes

The thylakoid membranes were isolated from cells grown in high light dissolved in thylakoid resuspension buffer containing 0.2 M sorbitol, 5 mM Tris–HCl (pH 7.5), and 5 mM $CaCl_2$. Blue native PAGE was performed with a 4% stacking and 4%–12.5% gradient resolving gel polymerized from a 32/1 bis-acrylamide/acrylamide mixture (Järvi et al., 2011). Thylakoid was isolated from the control

and the light-treated sample containing 1 mM PMSF, 1 mM ACA, and 1 mM benzamidine hydrochloride as a protease inhibitor. A total of 10 µg of Chl was loaded per lane, with a final Chl concentration of 0.5 mg ml^{-1} and a final detergent concentration of 1% β-DM. Native protein markers were from GE Healthcare (UK).

Gel electrophoresis and immunoblotting

The proteins from cells were separated under denaturing conditions on SDS-PAGE. The cells were grown in optimal growth light and high light conditions, and cells were centrifuged, and the pellet was stored in 20% glycerol. Cells were solubilized into a 2× sample buffer containing 1 M Tris-HCl (pH 6.8), 10% 1 mM DTT, 100% glycerol, and 10% SDS slowly mixed with pipetting. The sample was heated for 1 min at 100°C and then centrifuged at 10,000 rpm for 10 min to collect the supernatant and loaded the sample on an equal Chl basis. SDS-PAGE was performed using an SDS-PAGE system with a 4% stacking and 12% or 15% resolving gel. Three different quantities [0.25 (25%) or 0.5 (50%) and 1.0 (100%) µg] of Chl were loaded for each lane to compare the quantification of the protein as reported earlier (Devadasu and Subramanyam, 2021). The nitrocellulose membrane was incubated with primary antibodies (LHCII, I, PSII, and PSI) raised in rabbit and the antibody dilutions were as follows: For the PSII–LHCII complex: PsbA and PsbB (1:5,000); CP47 (1:2,000); CP43 (AS04038; 1:10,000); Lhcb1 (AS01004; 1:10,000); LhcB2 (1:5,000); Lhcb4 (1:10,000); Lhcb5 (1:10,000); PsbO (1:5,000); For the PSI–LHCI complex: PsaA, PsaD, PsaH, PsaG, and PsaF (1:10,000) (all antibodies were purchased from Agrisera, Sweden). The LHCI antibodies followed the following dilutions: Lhca1 (1:5,000); Lhca2 (1:5,000); Lhca3 (1:5,000); Lhca4 (1:3,000); Lhca5 (1:3,000); Lhca6 (1:3,000); Lhca7 (1:2,500); and Lhca9 (1:3,500) (these antibodies were raised in our laboratory). Subsequently, secondary antibody (HRP-conjugated anti-rabbit, 1:20,000 dilution; Agrisera) embellishment was performed for 1 h at RT at 10 rpm. Chemiluminescence signal detection was performed using ECL Western Blotting solution (Bio-Rad) and the ChemiDoc Imaging System (Bio-Rad).

Whole proteome analysis of *C. reinhardtii* cells under high light stress: protein extraction and quantification

C. reinhardtii control and light-treated cells were grown and harvested at 3,000 rpm for 5 min. Cell pellets were finely mixed in 10% glycerol, dipped in liquid nitrogen, and stored at −80°C for further analysis. The cell pellet was taken into a 15-ml Falcon tube (Genaxy, India) and suspended in 4 ml of extraction buffer containing 0.5 M Tris-HCl (pH 7.5), 0.7 M sucrose, 0.1 M KCl, 50 mM EDTA, 2% β-mercaptoethanol, and 1 mM PMSF. After thoroughly mixing, Tris-saturated phenol (pH 7.5) was added to the extracted suspension. The whole suspension was mixed for 30 min at 4°C in a rotor spin cyclomixer. Tris-saturated phenol was prepared by mixing an equal volume of Tris-HCl (pH 7.5) and phenol with continuous stirring for 3–4 h. The lower phenolic layer was separated, and an equal volume of Tris-HCl (pH 7.5) was added with constant stirring for 2–3 h. The lower phenolic layer was collected and stored in an amber glass bottle at 4°C. The sample mixture was centrifuged at 5,000×*g* for 30 min at 4°C. The upper phenolic phase was collected carefully, and an equal volume of extraction buffer was added. This step was performed repeatedly, and the phenolic stage was re-extracted. Four volumes of ice-cold 0.1 M ammonium bicarbonate in methanol were added to the final collected phenolic phase and incubated overnight at −20°C for protein precipitation. The next day, the samples were centrifuged at 10,000×*g* for 30 min at 4°C. The pellet was washed thrice with ice-cold methanol, twice with acetone, and air-dried for a few minutes. The final pellet was dissolved in 200 µl of the rehydration solution containing 8 M urea, 2 M thiourea, 30 mM DTT, 4% CHAPS, and 0.8% IPG buffer of pH range 4–7 (GE Healthcare). The protein concentration was determined using Bradford reagent (Bio-Rad) with BSA as standard (standard curve of 0–100 mg ml^{-1} concentration).

nLC-MS/MS analysis

100 µg of the final pellet was treated with 100 mM DTT for 1 h at 95°C, followed by 250 mM iminodiacetic acid (IDA) for 45 min at room temperature in the dark. The sample suspension was incubated with trypsin at 37°C for overnight digestion. The trypsin-digested peptides were extracted in 0.1% formic acid solution at 37°C for 45 min. The solution was centrifuged at 10,000×*g*, and the supernatant was collected in the fresh tube for vacuum drying. The final sample was solubilized in 20 µl of 0.1% formic acid. For the separation of peptides, 10 µl of injection volume was loaded on the C18 UPLC column, and peptides were separated by Waters Synapt G2 Q-TOF for MS and MS/MS analysis. For LC-MS analysis, 10 µl of the sample was injected into the ACQUITY UPLC system (Waters, UK) equipped with an ACQUITY UPLC BEH C18 column (Waters, UK) (150 mm × 2.1 mm × 1.7 µm), a SYNAPT G2 QTOF (Waters, UK), and an electrospray ionization (ESI) source. The sample analysis was run on the positive mode by applying 3,500 V capillary voltage and 30 L of cone gas flow per hour. The source and desolvation glass flow were maintained at 1.8 and 800 L/h, and the temperatures of source and desolvation were 150°C and 350°C, respectively. The protein range was from 50 Da to 150 Da. The trap and the transfer collision energy were continuously maintained at 6 V. The ramp collision energy was set at 20 V and increased up to 45 V. The total acquisition time was 60 min, and the solution flow rate was 300 nl/min. The mobile phase consisted of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). A linear 60-min gradient consisted of solvent A 98% and solvent B 2% for 1 min, solvent A and B 50% for 29 min, solvent A 20% and solvent B 80% for 15 min, followed by 15 min solvent A 98% and solvent B 2%. A wash solution was used at the end of each program to reduce carryover between samples.

Protein identification

The raw data acquired from the above analysis were processed using PLGS software 3.0.2 (Waters, India; identification and expression algorithm), within which data processing and database search was performed. The samples—*C. reinhardtii* proteins—for two sample sequences in FASTA format were downloaded from Swiss-Prot and used to search peptides present in samples. On each run, the sample was processed using the following search parameters in the software: peptide tolerance, 50 ppm; fragment tolerance, 100 ppm; minimum number of fragment matches for peptide, 2; minimum number of fragment matches for proteins, 5; and carbamidomethylating of cysteine and oxidation of methionine were selected as fixed and variable modifications, respectively. UniProt (*C. reinhardtii*, reviewed protein) was used as the database against which the search was done.

Gene ontology and bioinformatics analysis

This study's identified protein species were annotated based on their molecular function, biological process, and cellular component with Gene Ontology (GO) annotation using UniProt. Hierarchical cluster analysis was performed using the R statistical package based on correlation values. Network analysis was performed by using Cytoscape bioinformatics software version (3.7.2).

Statistical analyses

The experiments conducted in the current study were carried out with at least three biological replications (n = 3). The R statistical package constructed hierarchical cluster matrices and protein–protein networks based on correlation values and p-values less than 0.05.

Results

High light changes oxygen evolution/consumption in *pgrl1* and *pgr5*

Factually, oxygen evolution, spectrometry, and biochemical measurements were used to quantify PSII damage under abiotic stress (Chow and Aro, 2005; Tyystjärvi, 2013). The oxygen evolution results demonstrate that the WT cells acclimate when grown under high light. Their oxygen-evolving capacity is increased by approximately 47% (194 µmol/mg Chl/h) in moderate light and 36% (180.33 µmol/mg Chl/h) in high light (Figure 1A). A significant decrease of 51.58% and 66.35% in oxygen evolution in *pgr5* has been observed in moderate and high light, respectively, than in WT growth light conditions. On the other hand, in *pgrl1*, a minor decrease is observed in high light conditions. The Chl fluorescence has mainly decreased in *pgrl1* and *pgr5* in moderate and high light than in WT growth light conditions. When cells were dark-adapted, the WT and *pgrl1* observed an almost similar O_2 consumption under high light (Figure 1B). There is a significant decrease in dark-induced O_2 consumption in moderate and high light, corresponding to nearly 15% and 24% in *pgrl1* and 32% and 38% in *pgr5* compared to WT growth light conditions. Furthermore, we correlate this result with biochemical changes; the oxygen-evolving complex subunit of PsbO slightly decreased in *pgrl1* and *pgr5* in moderate and high light than WT (Figure 1C), as shown in the O_2 uptake experiment

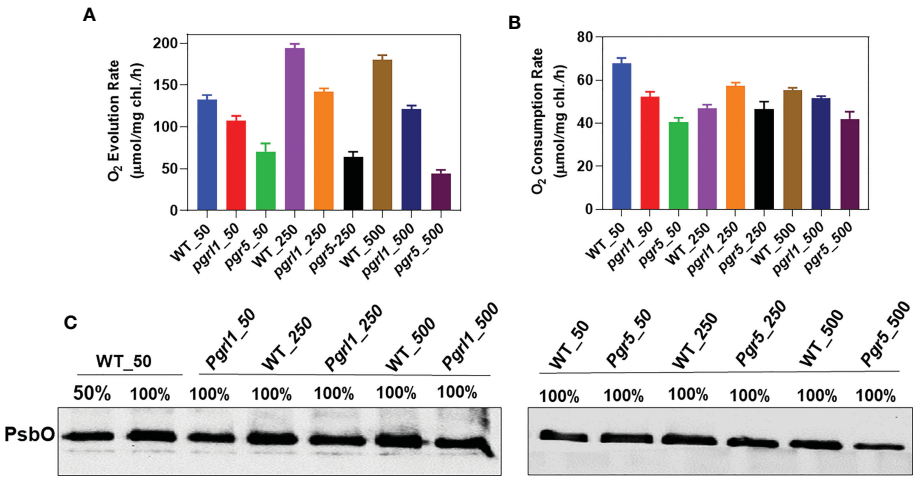


FIGURE 1
Measurement of oxygen evolution/consumption. Comparison of oxygen evolution and consumption rates of the wild type (WT), and knockdown strains, (*pgrl1* and *pgr5*). Cells were grown in different light conditions. The oxygen-evolving activity of cells was measured by taking an equal concentration of chlorophyll, 10 µg/ml, supplemented with the artificial electron acceptor 1 mM DCBQ and 1 mM NH_4Cl for PSII activity **(A)**. Oxygen evolution recordings were measured at 25°C under saturating light (600 µmol photons $m^{-2}s^{-1}$) using a Clark-type oxygen microelectrode (Hansateck Instrument Ltd, Norfolk, England) stuck inside the oxygen electrode chamber. Oxygen consumption measurements were carried out with cells incubated in a TAP medium containing NaN_3, ascorbate, and DCPIP at a final concentration of 1 mM **(B)**. In high light, the oxygen-evolving complex subunit of PsbO decreased in *pgrl1* and *pgr5* **(C)**. The electrode was calibrated in a medium flushed with air (O_2 saturated) in the presence of sodium hydrosulfite (anoxia). Standard deviations were estimated from three biological replicates.

(Supplementary Figures 1A–C). These results propose that WT cells adjust to high light stress by adaptative mechanism, but *pgrl1* and *pgr5* drastically damaged PSII, which agrees with the fluorescence result (Supplementary Figure 2A).

Structural changes in thylakoid membrane under high light

Thylakoid membrane structure and composition were analyzed using CD spectrometry and BN-PAGE. The thylakoid membrane is a dynamic structure; it can undergo numerous structural changes to acclimate under different environmental conditions (Nagy et al., 2014). CD determines the structural change of the macro-organization of super-complexes and pigment–pigment interactions. The WT and mutant CD spectra are presented in Figures 2A, B. The CD spectra of the thylakoid membrane are divided into the Qy region (600 to 700 nm or towards the red region) and the Soret region (400 to 550 nm or blue region). The CD spectra of the thylakoid membrane revealed the features called *psi*-type CD bands. The magnitude of the *psi*-type band is proportional to the size of the ordered array at a given set of conditions.

The visible region of the CD spectrum of *C. reinhardtii* cells consists of two intense *psi*-type bands at approximately (−) 676 and (+) 690 nm (Nagy et al., 2014). The positive band at approximately 690 nm (Chl *a*) and the negative band at approximately 674 nm (Chl *b*) originated from Chl molecules associated with macrodomains of photosynthetic proteins. These *psi*-type bands arose due to the presence of large well-ordered arrangements (size ranging from 200 to >400 nm) of PSII–LHCII complexes in the appressed region of the thylakoid membrane. In WT, there are marginal changes in the red psi-type at approximately (+) 692-nm signal amplitude, especially in moderate compared to high light. In contrast, we detected a significant increase in the *psi*-type signal from *pgrl1* and *pgr5* grown under high light compared to WT growth light (Figure 2A). On the other hand, another *psi*-type band at approximately (−) 678 is significantly decreased in *pgrl1* and *pgr5* in moderate and high light compared to WT control. This indicates that *pgrl1* and *pgr5* are crucial for maintaining the thylakoid dynamics and structural arrangements under high light.

The blue (Soret) region *psi*-type band at (+) 506 nm is attributed to carotenoids found in a long-range ordered structure, and the (−) 650-nm negative band originated from LHCII trimers (Garab and van Amerongen, 2009). Also, as reported earlier, the Soret region 506 nm originated from the β-carotene of the core complex (Tóth et al., 2016). Our results show that this peak is decreased drastically, especially in *pgr5*, which might be due to a decrease in the protein content of the core complex in cells subjected to high light (Figure 2B). The CD spectra of LHCII trimers are similar under normal and high light but slightly increased in moderate and high light in WT. At the same time, *pgrl1* LHCII trimers are increased, but in *pgr5*, peak intensity is drastically reduced compared to WT control, suggesting that the LHCII trimers were significantly affected in CET mutants under high light.

The spectra have a more complex structure in the Soret region (400 to 450 nm) because of an abundance of Chl *a* (+) 436 nm and Chl *b* (+) 446 nm, and the dominant peak at (−) 472 nm for LHCII. There is no such difference in the peak intensity of (−) 436 nm in WT and *pgrl1* in moderate and high light, and it decreased in *pgrl1* and *pgr5* in high light. On the other hand, the peak intensity of (+) 446 nm increased in *pgrl1* in moderate and high light but reduced

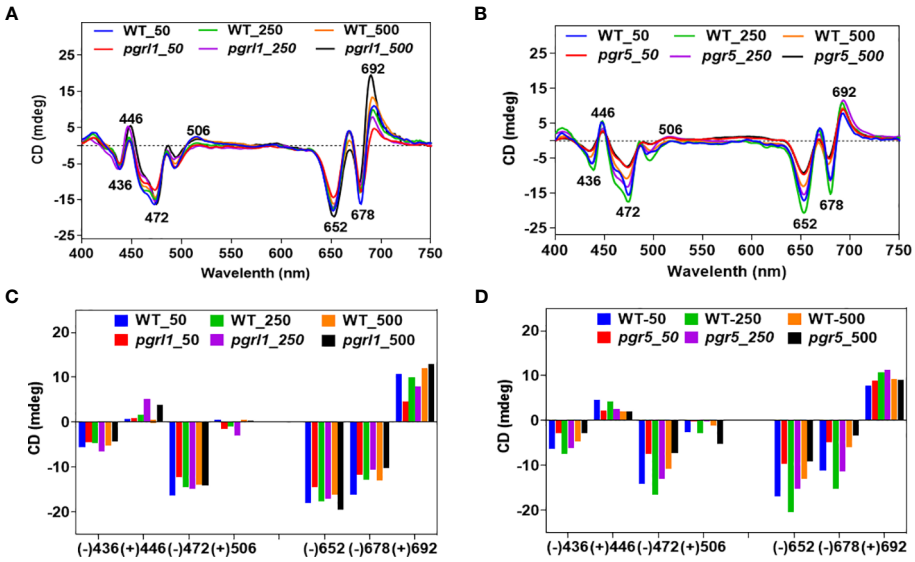


FIGURE 2
CD spectra of wild-type and mutants **(A, B)**. Spectra were recorded at room temperature, baseline-corrected, and normalized to equal chlorophyll (10 µg/ml). The total amplitude of the psi-type signal [i.e., the difference between the positive and negative amplitudes is shown as a bar **(C, D)** diagram] of WT, *pgrl1*, and *pgr5* cells grown under growth light and high light conditions, which received the same parameter and three accumulations with the scan speed of 100 from three independent experiments, on different batches. Values represent the mean and three replicates.

significantly in *pgr5* in high light compared to WT control (Figures 2A, B).

The LHCII band shows a (−) 472-nm similar rise in peak WT and *pgrl1* in moderate and high light but decreased dramatically under *pgr5* high light conditions as shown in Figures 2C, D. It is clear that though there is acclimation, the LHCs to RC assembly changed, describing that the LHCs are detached from the RC. The data agree with blue native gels that disturbed super-complexes and mega-complexes. However, these disturbed mega- and super-complexes were restored as the monomer and trimer content significantly reduced in *pgr5*.

Protein pigment interaction and changes in macro-organization of super-complexes

Thylakoid membranes comprise several multiprotein membrane complexes that run photosynthetic light reactions. Their composition, distribution, and stoichiometry may vary according to environmental conditions. Thylakoids include three structures: grana, stroma, and grana margins. The appressed regions of thylakoids, grana, are enriched with PSII–LHCII super-complexes, non-appressed parts, and stroma with PSI–LHCI and ATPase complexes (Garab and van Amerongen, 2009). Grana margins have been suggested to have both PS I and II form mega-complexes (Järvi et al., 2011). Figures 3A, B show typical patterns of thylakoid membrane protein complexes separated in BN-PAGE. Four bands of PSII–LHCII super-complexes between 669 and 440 kDa, PSII and PSI monomers, and Cyt b_6f complex devoid of LHCs, near 440–230 kDa, and the lower three bands consisting of different hierarchal combinations of LHC trimer and monomer have been resolved from the thylakoids under high light. The localization of the Cyt b_6f complex is under debate, but many studies reported its occurrence near PSII dimer (Anderson, 1982; Albertsson et al., 1991). The largest PSII super-complexes isolated from *C. reinhardtii* are classified as a strongly bound trimer (S-trimer), a moderately bound trimer (M-trimer), and a loosely bound trimer (L-trimer) depending on their binding strength with the PSII core complex as $C_2S_2M_2L_2$. From BN-PAGE, the four high-molecular-weight bands as a super-complex (~669 kDa) consist of different combinations of PSII core with S and M trimers, and the largest super-complex observed is $C_2S_2M_2$.

The WT under moderate and high light stress shows a marginal decrease of PSII–LHCII super-complex (Figure 3A). However, in *pgrl*1, the PSII–LHCII super-complexes decrease even more severely in *pgr5* under moderate and high light (Figures 3A, B). The reduction of PSII–LHCII super-complexes reflects in the form of LHCII trimer and monomer. In WT and *pgr5*, the fraction of LHCs is stored as a monomer than *pgrl1*; trimer is restored more under moderate and high light conditions (Figure 3A). The loosely bound LHCII complexes in mega-complexes and super-complexes (top bands) were dissociated and further accumulated as LHCII trimers. These changes are less noticeable in WT and mutants under high light.

Western blot analysis of proteins from *C. reinhardtii* WT and *pgrl1* and *pgr5* mutants

Immunoblot analysis was performed on WT, *pgrl1*, and *pgr5* grown in optimal (50 µmol photons m^{-2} s^{-1}), moderate (250 µmol photons m^{-2} s^{-1}), and high light (500 µmol photons m^{-2} s^{-1}) regimes. We have identified the proteins of PSII (D1, D2, CP43, and CP47) and PSI (PsaA, PsaD, PsaG, PsaH, and PsaF) core proteins along with both photosystems (PSII and PSI) and LHCs (light-harvesting complex) by Western blot. The changes in the stromal and luminal subunits (PsaG, PsaH, and PsaD) can distract the assembly of PSI. The proteins representing PSII core PsbA (D1) and PsbD (D2) accumulated more in *pgrl1* and *pgr5* than in WT under high light stress. However, D1 and D2 did not differ significantly in *pgr5* grown under moderate light compared to the WT control condition (Figures 4A, B). The PSII major antenna protein CP43 is decreased noticeably in *pgr5* while slightly decreased in *pgrl1* than WT control under moderate and high light. However, another antenna protein, CP47, significantly reduced in high light in *pgr5* but increased dramatically in *pgrl1* and *pgr5* in moderate light than WT control. The PsbA (D1) protein content slightly increased in *pgrl1* and *pgr5* in high light than WT control (Figure 4A). In high

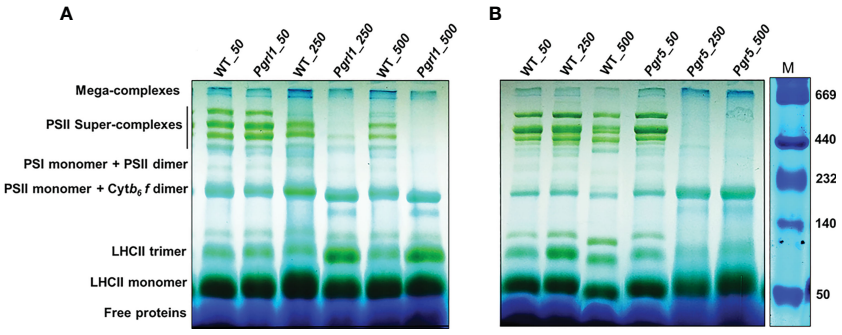


FIGURE 3
Pigment–protein complexes and super-complexes were seperated using BN-PAGE gel. Thylakoid membranes from WT, *pgrl1*, and *pgr5* samples were solubilized with 1% β-DM and separated using native PAGE. WT, *pgrl1*, and *pgr5* containing equal amounts of chlorophyll (10 µg) were loaded in each lane shown in **(A, B)** panel with a Native marker.

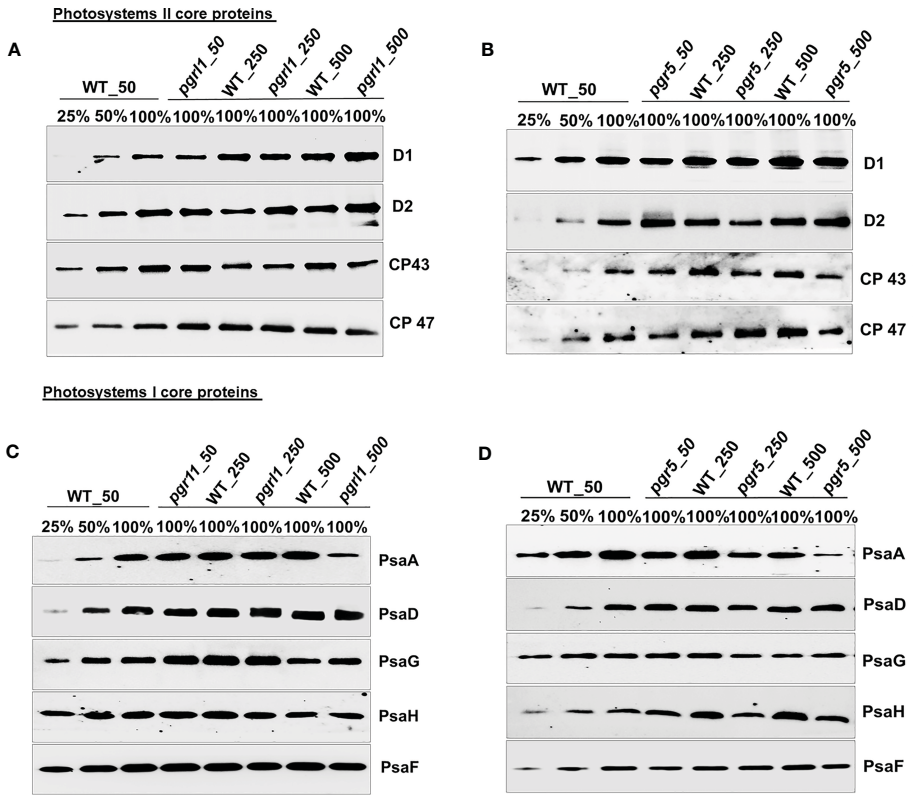


FIGURE 4
Immunoblot analysis of PSII **(A, B)** and PSI **(C, D)** core proteins selected from wild-type (WT) and *pgrl1* and *pgr5* mutants. The WT and *pgrl1* and *pgr5* mutants were grown under constant growth light (50 µmol photons m^{-2} s^{-1}) and moderate (250 µmol photons m^{-2} s^{-1}) and high light (500 µmol photons m^{-2} s^{-1}) conditions. The Western blots shown here are representative of at least three biological replicates.

light stress conditions, CP43 and CP47 showed a reduced amount in *pgr5* due to changes in chlorophyll and carotenoid molecules around the PSII protein, which coordinates the cofactors to transfer electrons to PSI (Figure 4B).

The PGRL1 and PGR5 proteins are part of the PSI complexes and are extensively involved in CET for energy generation. The absence of this protein affects the culture growth and other photosynthetic parameters related to both PSI and PSII in high light. Additionally, we checked the outcome of high light on *pgrl1* and *pgr5* strains, and how the absence of these proteins affects other PSI proteins under high light is studied. The PsaA subunit of the PSI showed a more than 50% decrease under moderate light while a 75% decrease in *pgr5* under high light (Figure 4D). However, PsaA protein content was unaffected in *pgrl1* under moderate light but significantly decreased in high light compared to WT control. At the same time, there was a marginal decrease of PsaD and PsaG protein content in *pgrl1* and WT under high light, but these proteins are unaffected under moderate light (Figure 4C). The stromal subunits PsaC and PsaD act as a binding site for ferredoxin, which leads to an efficient electron transport chain (Fischer et al., 1998). *pgrl1* the PsaD is marginally decreased in moderate light but increased slightly in *pgr5* under high light. At the same time, PsaG accumulates in WT and *pgr5* under moderate and high light.

On the other hand, another subunit PsaD is noticeably accumulated in WT and *pgr5* under moderate and high light than WT. Together, mutants confirmed that the protein content of PsaF did not change from WT under moderate and high light conditions, showing that the arbitrary presence in the PSI complexes was unaffected (Figures 4C, D). Parallel results have been described earlier in *A. thaliana pgr5* and *pgrl1* (DalCorso et al., 2008). The immunoblot analyses of the LHCII proteins were relatively stable in WT and *pgrl1* compared to *pgr5* under high light.

The Lhcb1 protein increased marginally in *pgrl1* and *pgr5* in high light, while the mutation did not affect the accumulation of Lhcb2 in *pgrl1* but slightly reduced in *pgr5* in high light compared to WT control (Figures 5A, B). The minor light-harvesting complexes Lhcb4 (CP29) and Lhcb5 (CP26) have been proposed to play a crucial role in the zeaxanthin-dependent (qZ) high light-induced regulation of NPQ in *C. reinhardtii*. The Lhcb4 and Lhcb5 proteins are almost unaffected in WT but increased in *pgrl1* under moderate and high light than in WT (Figure 5A). However, in *pgr5*, the Lhcb4 protein is decreased slightly but significantly in moderate and high light compared to WT, indicating that *pgr5* was affected more than *pgrl1* (Figure 5B).

Similarly, we can see that high light-induced photoprotective protein LHCSR3 accumulated more in WT and *pgrl1* but is absent in *pgr5* under moderate and high light. Furthermore, we examined the protein content of the light-harvesting complex of PSI that is arranged in two layers towards the stromal side of the PSI core and is essential for efficient energy transfer and PSI photoprotection. The four PSI light-harvesting proteins, Lhca1, Lhca8, Lhca7, and Lhca3, are situated

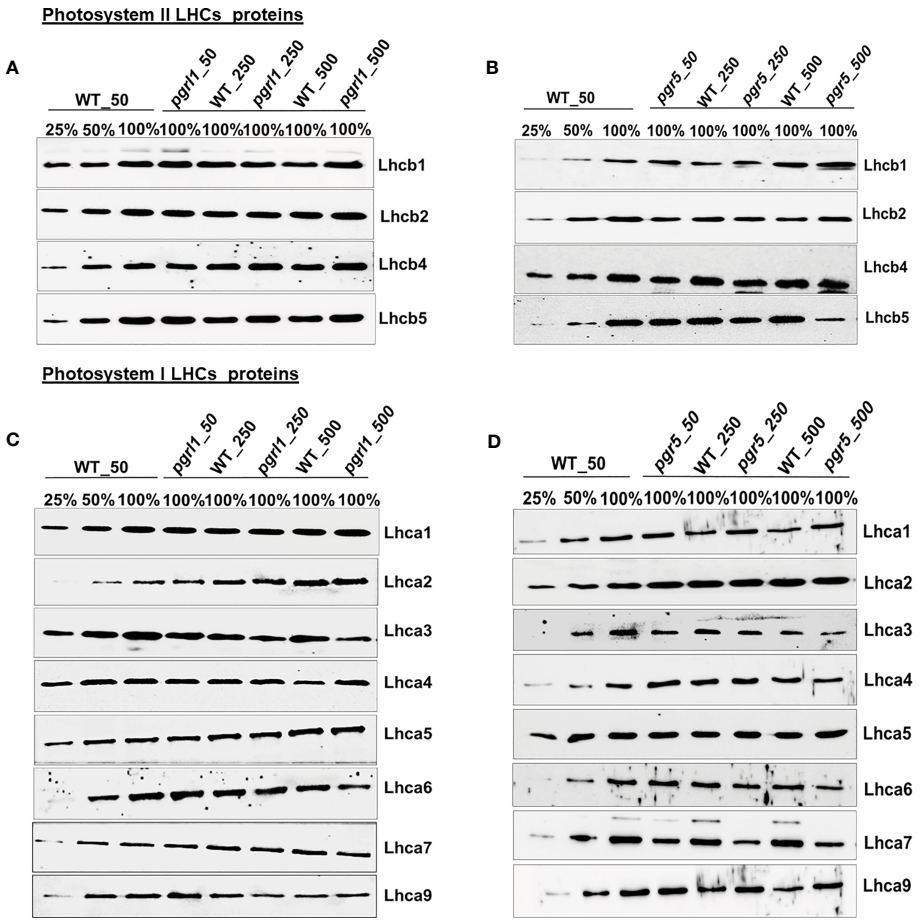


FIGURE 5
Western blotting analysis of all the LHCs of photosystem II and I subunit. Immunoblot analysis of the selected light-harvesting complex protein of PSII **(A, B)** and PSI **(C, D)** from wild-type (WT) and *pgrl1* and *pgr5* mutants that were grown under constant growth light (50 µmol photons m^{-2} s^{-1}) and moderate (250 µmol photons m^{-2} s^{-1}) and high light (500 µmol photons m^{-2} s^{-1}) conditions. The Western blots shown here are representative of at least three biological replicates.

in the internal layer, and another four, Lhca1, Lhca4, Lhca6, and Lhca5, are positioned in the outside layer towards the PSAF, and the two subunits, Lhca2 and Lhca9, associated with PSAB core (Ozawa et al., 2018). Interestingly, the light-harvesting complex of PSI—Lhca1, Lhca4, Lhca5, and Lhca7—is relatively stable in *pgrl1*, whereas Lhca2, Lhca3, Lhca6, and Lhca9 are extensively affected in *pgrl1* and *pgr5* under high light (Figures 5C, D).

Unusually, the association of Lhca1, Lhca7, Lhca8, and Lhca3 is more steady than Lhca4, Lhca5, and Lhca6 light-harvesting proteins, which are positioned in the innermost layer (Drop et al., 2011). The Lhca1 is almost unaffected in *pgrl1* and *pgr5* in moderate and high light compared to the WT control (Figures 5C, D). The Lhca2 and Lhca9 present near the PsaG subunit increased in *pgrl1* and *pgr5*, except that the Lhca9 in *pgrl1* decreased under moderate and high light WT. The location of lhca3 close to PSAK is consistent with previously reported results (Moseley et al., 2002). The Lhca3 is present in the outer side of the edge in the inner ring, and Lhca6 at the outer ring is reduced significantly in *pgrl1* and *pgr5* under moderate and high light compared to WT control. Two other LHCs, Lhca4 and Lhca5, are likely located in the outer ring of PSI polypeptide and unaffected under moderate and high light in *pgrl1* and *pgr5* compared to WT (Figures 5C, D). The accumulation of the inner subunit of Lhca7 is unchanged in *pgrl1*, whereas it decreased significantly in *pgr5* under moderate and high light than WT.

High light changes protein coverage and up-/downregulated and differentially expressed proteins between treated and control samples

The current study combined the *n*LC-MS/MS technique and targeted proteome to unravel the essential proteins accountable for resistance to high light from *C. reinhardtii*. We would like to understand the importance of the CET mutants photosynthetic electron transport chain (PETC) proteins and downstream metabolism in *C. reinhardtii*. Total cell proteomes were taken from cultures grown under optimal growth and high light conditions to address the following question (Supplementary Figure 3A). A Venn diagram of each protein was established and observed almost 295 proteins in WT, 306 in *pgrl1*, and 299 in *pgr5* under growth light while 306 in WT, 306 in *pgrl1*, and 286 in *pgr5* in

10.3389/fpls.2023.1198474

high light. We have placed almost 249 DEPs in WT_50 and 284 DEPs in high light stress from the entire proteome. Of the total, 284 proteins showed common expression in growth light WT_50 and high light W_500 (Supplementary Figure 3B). Seventeen DEPs in WT_50 while 52 DEPs in WT_500 showed unique expression under low and high light conditions.

Furthermore, we have annotated and enriched the *C. reinhardtii* proteome with the sample-identified peptide after nLC-MS/MS analysis categorized them based on molecular weight and protein sequence coverage (Figures 6A, B). We obtained 6,654 peptides in WT under control light conditions (denoted as WT_50, here 50 µmol photons m^{-2} s^{-1} is light intensity in growth light) and 14,822 peptides under high light-treated (WT_500, here 500 µmol photons m^{-2} s^{-1} is light intensity at high light) conditions. The *pgrl1* showed 23,133 peptides under growth light (*pgrl1*_50) and 20,515 peptides under high light (*pgrl1*_500) conditions. On the other hand, *pgr5* showed 12,733 peptides under growth light (*pgr5*_50) but only 4,362 peptides under high light (*pgr5*_500) conditions. Furthermore, we enriched all peptides into proteins, which share common and exclusively expressed proteins in treated versus control samples, as shown in Figure 6C. Overall, 302 proteins from several functional groups were significantly affected together in treated and control samples, out of which 14 (4.64%) were associated with PSII, 12 (3.97%) were involved in PSI, 11 (3.64%) were associated with Cyt b_6f complex, 10 (3.31%) were engaged in ATPase assembly, and 17 (5.62%) were involved in photosystem (PSI and PSII) assembly.

There were 64 (22%) proteins associated with photosynthesis, 29 (9.60%) proteins related to mitochondria, 28 (9.27%) proteins related to chloroplast/cytosol, 34 (11.25%) proteins related to translational machinery, 34 (11.26%) proteins associated with Cilia/Flageller assembly, and 113 (37.42%) proteins associated with primary and secondary metabolism (Figure 6D). These up- and downregulated proteins are primarily engaged in photosynthesis, metabolic pathway, glycolysis, protein synthesis, cilia, flagella-associated proteins, and proteins involved in cytoskeleton assembly.

We have also compared *Chlamydomonas* proteome-treated versus control samples to understand them much better. In WT_500 vs. WT_50 (Supplementary Table 1), we identified 274 as DEPs; 16 proteins are upregulated, 21 are downregulated, and 14 and 22 proteins are exclusively expressed in WT_50 and WT_500, respectively (Figure 6E). In *pgrl1*_500 vs. WT_50 (Supplementary

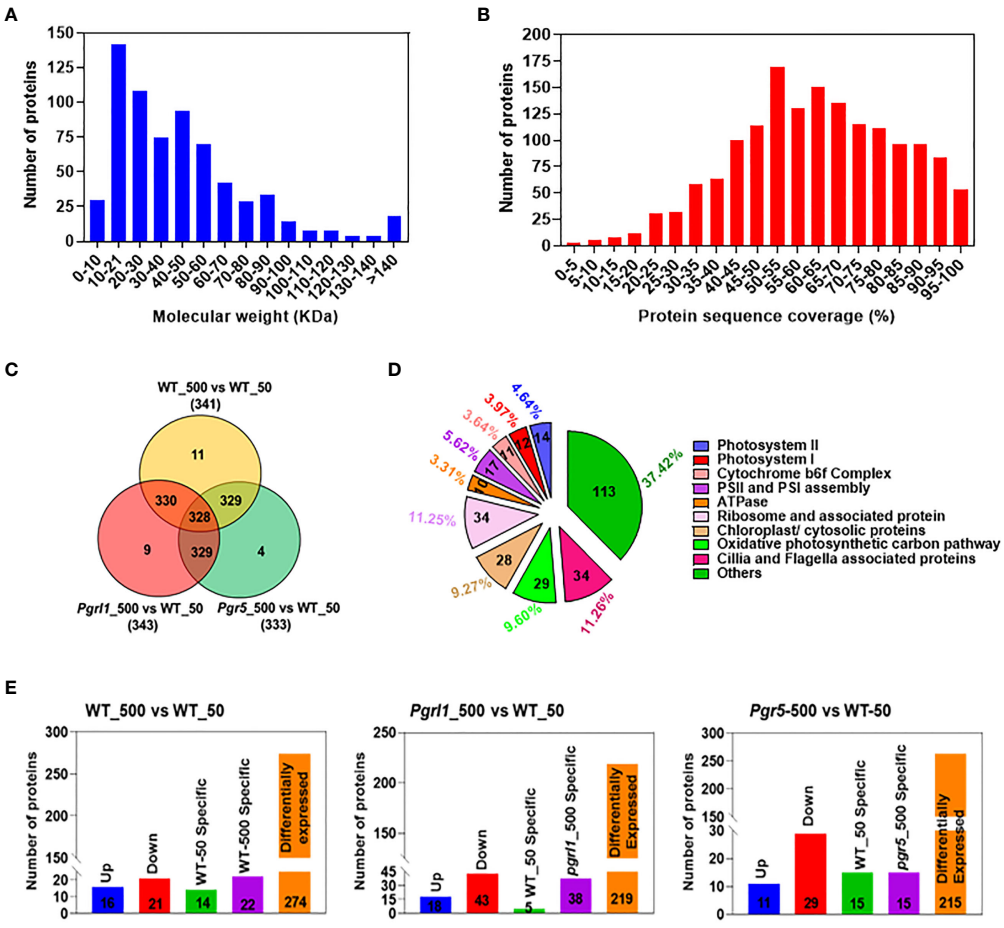


FIGURE 6
Whole-cell proteome analysis between treated and control cells. Total numbers of proteins against different molecular weights **(A)**. The number of proteins against additional protein sequence coverages **(B)**. Venn diagrams showing the significantly different protein levels in light treated vs. control cells **(C)**. The total number of proteins that share unique expression levels is represented as a percentage **(D)**. The comparative proteome of the total number of upregulated, downregulated, and differentially expressed proteins (DEPs) **(E)** under growth and high light conditions.

Table 2), we identified 219 as DEPs; 18 proteins are upregulated, 43 are downregulated, and 5 and 38 proteins are exclusively expressed in WT_50 and *pgrl1*_500, respectively. In *pgr5*_500 vs. W_50 (Supplementary Table 3), we have identified 215 as DEPs; 40 proteins are upregulated, 12 are downregulated, and 21 and 14 proteins are exclusively expressed in WT_50 and *pgr5*_500, respectively (Figure 6E).

Changes in photosynthetic and high light-inducible proteins of *C. reinhardtii* under high light

DEPs contributing to photosynthesis were defined under high light. The water-oxidizing complex, PSBO_P12853 (Oxygen-evolving enhancer protein 1_chloroplastic) and PSBP_P11471 (Oxygen-evolving enhancer protein 2_chloroplastic), are upregulating (≥2FC) in *pgrl1*_500 vs. WT_50 and *pgr5*_500, and PsbQ_P12852 (Oxygen-evolving enhancer protein 3_chloroplastic) was differentially regulated in all the variants (Figures 7A). Only water-oxidizing protein PSBO_P12853 is downregulated (≤0.5FC) in *pgr5* (*pgr5*_500 vs. WT_50) under high light.

The PSII RC proteins PSBA_P07753 (D1 protein), PSBD_P06007 (D2 protein), and PSBC_P10898 (PS II CP43 RC protein) are differentially regulated (≤2FC and ≥0.5FC) in WT, *pgrl1*, and *pgr5* under high light compared to growth light. On the other hand, the major antenna proteins PSBB_P37255 (PS II CP47 reaction center protein) are downregulated in *pgrl1* (*pgrl1*_500 vs. WT_50) under high light compared to growth light (Figure 7A).

Three PSI RC proteins, PSAB_P09144 (PSI P700 chlorophyll an apoprotein A2), PSAE_P12352 (PSI reaction center subunit III chloroplastic), and PSAF_P12356 (PSI reaction center subunit III_chloroplastic), have a more than twofold increase in expression in WT (WT_500 vs. WT_50) under high light. The two PSI core proteins with a downregulated < 0.5FC expression in *pgr5* are PSAA_P12154 (PSI P700 chlorophyll, an apoprotein A1) and PSAG_P14224 (PS I reaction center subunit V_chloroplastic), as they decrease in protein blot under high light conditions. Other PSI proteins like PSAC_Q00914 (PSI iron-sulfur center), PSAD_Q39615 (PSI reaction center subunit II_chloroplastic), and PSAH_P13352 (PSI reaction center subunit VI_chloroplastic) are differentially regulated (≤2FC and ≥0.5FC) under WT-, *pgrl1*-, and *pgr5*-treated vs. control light conditions (Figure 7A). The possible role of PSAH_P13352 could be provided with the docking site of the LHC I antenna protein to the core complex.

The CYC_P15451 was more downregulated in *pgr5*_500 vs. WT_50 (~one-third of log2FC). The log2FC expression of CYF_P23577 ($p < 0.05$) did not significantly changed in *pgrl1* and *pgr5*. Another essential Cyt complex protein is CYB6_Q00471 (Cyt b_6), whose log2FC expression is increased 281 times in *pgr5* (*pgr5*_500 vs. *pgr5*_50) compared to WT high light (WT_500 vs. WT_50, Figure 7A). The PETD_P23230 (Cyt complex subunit 4), another type of Cyt complex protein, was similarly expressed in WT, *pgrl1*, and *pgr5* in all high light conditions, except that a slight downregulation of PETD_P23230 was observed in *pgrl1* (*pgrl1*_500 vs. *pgrl1*_50) under the HL condition (Figure 7A). PETM_Q42496 (Cyt b_6f complex subunit 7_chloroplastic) and PETO_Q9LLC6 (Cyt b_6f complex subunit 5_chloroplastic) have been differentially

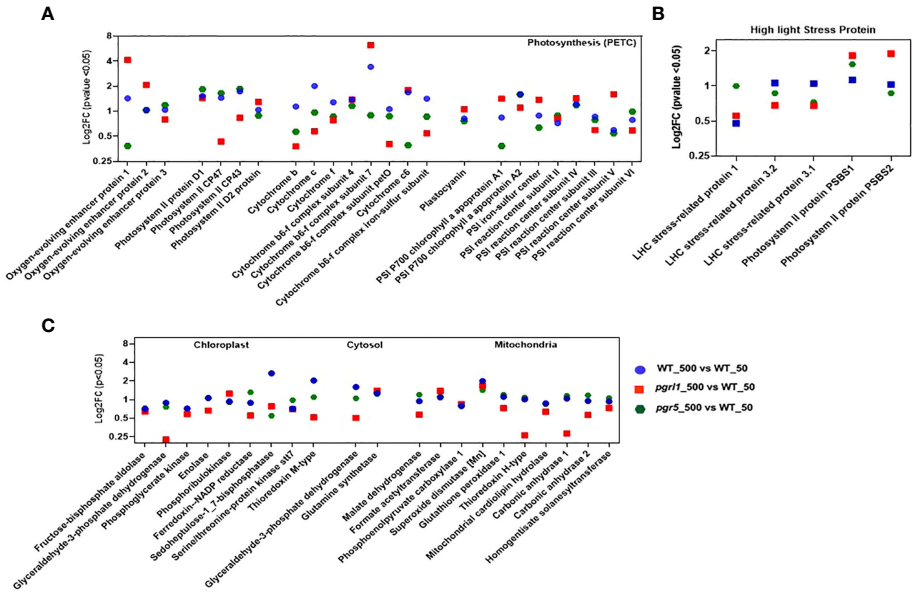


FIGURE 7
Proteins participating in photosynthetic machinery (PETC: photosynthetic electron transport chain) involved in OEC, PSII core, *Cyt* b_6f complex, and PSI core **(A)** proteins are significantly affected under high light stress and are shown as log2FC ($n = 3$). Proteins expressed in high light stress belonging to NPQ generation **(B)** are mainly involved under high light. Proteins participating in the chloroplast, cytosol, and mitochondria **(C)** that are involved considerably under high light are shown. Proteins with fold change and p-value < 0.05 are shown ($n = 3$).

expressed (≤2FC and ≥ 0.5FC) in *pgrl1* and *pgr5*, except for the PETM_Q42496, Cyt b_6f subunit encoded by the nucleus, which is apparently effective in the signaling pathway (Schneider et al., 2001) and whose expression is upregulated (≥2FC) in *pgrl1* (*pgrl1*_500 vs. WT_50) and WT (WT_500 vs. *pgr5*_50 and WT_500 vs. W_50) under high light. Similar results have been described earlier in *A. thaliana pgr5* and *pgrl1* (DalCorso et al., 2008).

The ATP synthase subunit ATPA_P26526 (ATP synthase subunit alpha_chloroplastic), ATPB_P06541 (ATP synthase subunit beta_chloroplastic), and ATBG_P12113 (ATP synthase gamma chain_chloroplastic) are essential for proton uptake from the lumen to balance the proton conductivity on the thylakoid membrane and are differentially regulated in *pgrl1* (WT_500 vs. *pgrl1*_50) under high light (Supplementary Figure 4A). The ATPD_Q42687 (ATP synthase delta chain_chloroplastic) and ATPE_P07891 (ATP synthase epsilon chain_chloroplastic) are upregulated in WT (WT_500 vs. WT_50), but ATPD_Q42687 is increased exceptionally 64 times in *pgr5* (*pgr5*_500 vs. WT_50), showing the importance of CET for energy generation. However, proteins ATPD_Q42687 and ATPE_P07891 are downregulated in *pgrl1* (*pgrl1*_500 vs. W_50) compared to WT, indicating the significance of CET under high light stress. *pgrl1* and *pgr5* substantially reduced CYF_P23577 (Cyt f) and CYC_P15451 (Cytc) of the Cyt b_6f complex compared to WT under high light.

The stress-related proteins LHCSR3 and PSBS were differently expressed (≤2FC and ≥0.5FC) in *pgrl1* and *pgr5* under high light. The LHC stress-related protein 1 (LHSR1_P93664) is mainly involved in zeaxanthin-dependent quenching and increased more than >2FC expression in *pgr5* (*pgr5*_500 vs. WT_50) than WT (WT_500 vs. WT_50) under high light. The LHC stress-related protein 3.1 (LHSR3.1_P0DO19) and LHCSR3.2_P0DO19, induced by lumen acidification, were decreased (log2FC expression) in *pgrl1* and *pgr5* mutants compared to WT in high light. The LHCSR3.2_P0DO19 has increased log2FC expression under WT high light (WT_500 vs. WT_50) compared to under WT control conditions (Figure 7B). However, *pgrl1* has decreased log2FC expression of LHSR3.2_P0DO19, whereas its expression was lacking in *pgr5* under high light conditions. This finding shows that *pgr5* can generate the least NPQ, while *pgrl1* has a smaller NPQ than WT (Yadav et al., 2020).

Another high light-expressing protein, PSBS, is present in two isoforms, the first one being PSBS1_A8HPM2 (PSII protein PSBS1), whose log2FC expression is higher in *pgr5* (*pgr5*_500 vs. WT_50) and *pgrl1* (*pgrl1*_500 vs. WT_50) than in WT under high light. Conversely, in PSBS2_A8HPM5 (PSII protein PSBS2), the log2FC expression increased mainly into *pgrl1* (*pgrl1*_500 vs. WT_50) under high light. The data show that the LHCSR3.1_P0DO19 and LHCSR3.2_P0DO19 are mainly expressed proteins in WT. In contrast, PSBS2_A8HPM5 and PSBS1_A8HPM2 are significantly expressed in *pgr5* and *pgrl1* and is the major effector protein under high light photoprotection in *C. reinhardtii* (Figure 7B). The PSBR_A0A2K3DMP5 (PSII protein PSBR chloroplastic) protein is essential for the steady binding of LHCSR3 to PSII–LHCII and is crucial for the efficient quenching and for maintaining the stability of the PSII–LHCII–LHCSR3 super-complex under high light.

High light changes the metabolism of chloroplast, cytosol, and mitochondria

C. reinhardtii is a mixotroph that can act as an autotroph or heterotroph, depending on medium conditions. The control light-grown cultures have a robust carbon fixation potential, but downstream metabolites were changed when grown in high light, where the growth was affected. The RubisCO subunit RUBA_Q42694 (RuBisCO large subunit-binding protein subunit alpha_chloroplastic) and RUBB_Q42693 (RuBisCO large subunit-binding protein subunit beta-1), a chaperone protein involved in RubisCO assembly, were deferentially regulated in WT (WT_500 vs. WT_50), but RUBA_Q42694 is downregulated in *pgrl1* (*pgrl1*_500 vs. WT_50) and RUBB_Q42693 in *pgr5* (*pgr5*-500 vs. W-50) under high light. Another subunit of RuBisCO assembly, RUBC_Q42695 (RuBisCO large subunit-binding protein subunit beta-2), binds RuBisCO's small and large subunits, is occupied in the enzyme assembly, and is differentially regulated in WT and *pgrl1* but downregulated in *pgr5* (*pgr5*_500 vs. W_50) under high light conditions (Supplementary Figure 4B).

The glycolytic enzyme ALFC_Q42690 (Fructose-bisphosphate aldolase 1_chloroplastic) is an essential enzyme in the energy pathways of algae; it was differentially regulated (≤2FC and ≥0.5FC) in WT, *pgrl1*, and *pgr5* under high light (Figure 7C). The enzyme G3PA_P50362 (Glyceraldehyde-3-phosphate dehydrogenase A_chloroplastic) involved in the glucose metabolic process was downregulated in *pgrl1* (*pgrl1*_500 vs. WT_50) under high light. Similarly, it is differentially regulated in WT (WT_500 vs. WT_50) under high light. Simultaneously, the protein KPPR_P19824 (Phosphoribulokinase_chloroplastic) is associated with the Calvin cycle; part of carbohydrate biosynthesis was differentially regulated in *pgrl1* and *pgr5* (*pgrl1*_500 vs. WT_50) and *pgr5* (*pgr5*_500 vs. WT_50) under high light.

The enzyme FENR_P53991 (Ferredoxin-NADP reductase_chloroplastic) is vital for adjusting the cyclic and non-cyclic electron transport to meet the requirement of ATP, and reducing power was differentially regulated in WT (WT_500 vs. WT_50) and *pgrl1* (*pgrl1*_500 vs. WT_50). TRXM_P23400 (Thioredoxin M-type_chloroplastic) is a vital regulatory player receiving information about the redox state of the PETC and participates in the reversible oxidation of the active center dithiol to a disulfide, which was highly upregulated by more than threefold (>3 log2FC) in WT (WT_500 vs. WT_50) and was differentially regulated in *pgrl1* (*pgrl1*_500 vs. W_50) and *pgr5* (*pgr5*_500 vs. WT_50) under high light.

The M type of TRX is known to activate NADP-malate dehydrogenase to hold the CET indirectly. The enzyme ENO_P31683 (Enolase) was differentially regulated in WT (WT_500 vs. WT_50), *pgrl1* (*pgrl1*_500 vs. WT_50), and *pgr5* (*pgr5*_500 vs. WT_50) under high light. The key enzyme STT7_Q84V18 (Serine/threonine-protein kinase stt7_chloroplastic) is required for state transition by phosphorylating the LHC II external antenna differentially regulated (≤2FC and ≥0.5FC) in WT, *pgrl1*, and *pgr5* in high light (Figure 7C). The S17P_P46284 (Sedoheptulose-1,7-bisphosphatase) chloroplastic protein is highly upregulated (more

than >2FC) in WT (WT_500 vs. WT_50) and was differentially regulated in *pgrl1* (*pgrl1*_500 vs. W_50) and *pgr5* (*pgr5*_500 vs. WT_50) in high light.

Comparing the whole cell proteome of *C. reinhardtii*, high light-grown cultures suggested that metabolic enzyme was strongly affected in *pgrl1* and *pgr5*. The tricarboxylic acid (TCA) cycle of MDHM_Q42686 (Malate dehydrogenase_mitochondrial) was differentially regulated in WT (WT_500 vs. WT_50), *pgrl1* (*pgrl1*_500 vs. WT_50), and *pgr5* (*pgr5*_500 vs. WT_50) under high light (Figure 7C). MDHM_Q42686, a mitochondria enzyme, synthesizes extra reducing power for chloroplast by converting oxaloacetate to malate by a malate shuttle shunt. This way, an additional NADPH pool is synthesized and regenerated (Scheibe, 2004).

The mitochondrial protein GPX1_P83564 (Glutathione peroxidase 1) constitutes a glutathione peroxidase-like protective system against oxidative stresses, synthesizing PUFA through arachidonic acid metabolic pathways, and was differentially regulated in *pgrl1* and *pgr5* in high light. The protein PAPP1_P81831 (phosphoenolpyruvate carboxylase 1) is involved in the carboxylation of phosphoenolpyruvate (PEP), forms oxaloacetate, and was differentially regulated (≤2FC and ≥0.5FC) in WT, *pgrl1*, and *pgr5* in high light. This enzyme is activated by glutamine and dihydroxyacetone phosphate and inhibited by glutamate and malate.

Superoxide dismutase (Mn), mitochondrial (SODA), destroys superoxide anion radicals within the cells. It is toxic to biological systems with more than twofold change (>2FC) upregulated in *pgrl1* (*pgrl1*_500 vs. W_50) and more than the threefold differentially (>3FC) upregulated in *pgr5* (*pgr5*_500 vs. WT_50) under high light, although its expression was flexible and statistically significant ($p < 0.05$), whereas TRXH_P80028 (thioredoxin H-type) protein contributes to several redox responses through the reversible oxidation of the dithiol into disulfide by taking the electron from ferredoxin, with the activity of enzyme FTR (ferredoxin: thioredoxin reductase) differentially regulated (≤2FC and ≥0.5FC) in WT, *pgrl1*, and *pgr5* and its H form is known to activate several cytosolic enzymes (Figure 7C).

CAH1_P20507 and CAH2_P24258 (Carbonic anhydrase 1 and 2) are present in the mitochondria and responsible for the reversible carbonate dehydratase activity because the formation of CO_2 was downregulated (<0.2 log2FC) in *pgrl1* (*pgrl1*_500 vs. WT_50) under high light. The metabolic protein homogentisate solanesyltransferase (HSTC_A1JHN0) not only is involved in PQ synthesis but also participates in carotenoids, and abscisic acid (ABA) biosynthesis pathways were differentially regulated (≤2FC and ≥0.5FC) in WT, *pgrl1*, and *pgr5* under high light.

Heatmap representing the comparative expression of the differentially abundant proteins

To elucidate the heatmap analysis, we employed the expression value of WT, *pgrl1*, and *pgr5*. Most of the proteins were upregulated in WT, and few were in the *pgrl1*, but the majority of proteins are downregulated in *pgr5* in high light conditions. The heatmap study categorized the identified DEPs into four distinctive categories: photosynthetic electron transport protein, Calvin cycle and metabolism protein, light-harvesting pigment, photoprotection proteins, and protein involved in translational machinery. The photosynthetic electron transport proteins were divided into 42 sub-categories with three major clusters: PS II and PS I protein subunits, Cyt b_6f complex, ATP synthase, RuBisCO subunits, and electron carrier proteins (Figure 8A).

During the light reaction of the photosynthesis, we identified 10 DEPs from Cluster 1 [PSBB_P37255, PSBD_P06007, PSAA_P12154, PSAB_P09144, CYB6_Q00471, CYB_P23662 (*Cyt* b), ATPF_P12356, CCSA_P48269 (*Cyt* c biogenesis protein CcsA), RBL_P00877 (Ribulose bisphosphate carboxylase large chain), and PLAS_P18068 (Plastocyanin chloroplastic)], in which most of the proteins are in *pgrl1*, but few of them (PSBB_P37255, ATPF_P12356, and CYB6_Q00471) in WT are upregulated compared to *pgr5* in high light. Cluster 2 includes 17 proteins [PSBA_P07753, PSBC_P10898, PSAH_P13352, PETD_P23230, UCRIA, PHOT_Q8LPD9 (Phototropin), CYC_P15451, PETO_Q9LLC6, ATPE_P07891, RUBC_Q42695, PSAD_Q39615, ATPG_P12113, ATPD_Q42687, PETM_Q42496, RUBB_Q42693, RBS1_P00873 (Ribulose bisphosphate carboxylase small chain 1_chloroplastic), and RBS2_P08475 (Ribulose bisphosphate carboxylase small chain 2_chloroplastic)] that are upregulated in WT high light (Figure 8A). However, 13 proteins upregulated in *pgr5* were associated with PSII chlorophyll-binding protein D1, PSI protein, ATP synthase subunit, Rubisco subunit, *Cyt* c subunit, and phototropin protein in growth light conditions.

Cluster 3 includes 15 proteins [CALM_P04352 (Calmodulin), PSAG_P14224, PSBQ_P12852, PSAF_P12356, PSAC_Q00914, CYF_P23577, RUBA_Q42694, PSAE_P12352, PSBO_P12853, PSBP_P11471, ATPBM_P38482, ATPX_A8J785 (ATP synthase subunit b'_chloroplastic), ATPA_P26526, ATPB_P06541, and RCA_P23489 (Ribulose bisphosphate carboxylase/oxygenase activase_chloroplastic)] exclusively involved in oxygen-evolving proteins, Cyt f protein, and ATP synthase subunit upregulated in WT high light conditions.

A total of 250 DEPs were divided into 49 sub-categories of the Calvin cycle and metabolic proteins into three significant clusters. Cluster 1 includes CAO_Q9ZWM5 (Chlorophyllide a oxygenase_chloroplastic), D4FAD_I2CYZ4 (Acyl-lipid (7-3)-desaturase_chloroplastic), G3PA_P50362, SODM_Q42684 (Superoxide dismutase (Mn)_mitochondrial), HSTC_A1JHN0 (Homogentisate solanesyltransferase_chloroplastic), HEM2_Q42682 (Delta-aminolaevulinic acid dehydratase_chloroplastic), THI4_A8J9T5 (Thiamine thiazole synthase_chloroplastic), SULP1_Q8RVC7, SULP2_Q6QJE2 (Sulfate permease 1 and 2_chloroplastic), ALFC_Q42690, LISC_A8I2V9 (Lipoyl synthase_chloroplastic), DES_Q2HWK7 (Acyl-lipid omega-13 desaturase), TRXH_P80028 (), SRGT1_H3JU05 (Peptidyl serine alpha-galactosyltransferase), METK_A8HYU5 (S-adenosylmethionine synthase), and 15 DEPs with the maximum number observed in the dark reaction and are upregulated in *pgrl1* in growth and high light conditions (Figure 8B).

Cluster 2 includes CP12_A6Q0K5 (Calvin cycle protein CP12_chloroplastic), GLNA2_Q42689 (Glutamine synthetase_

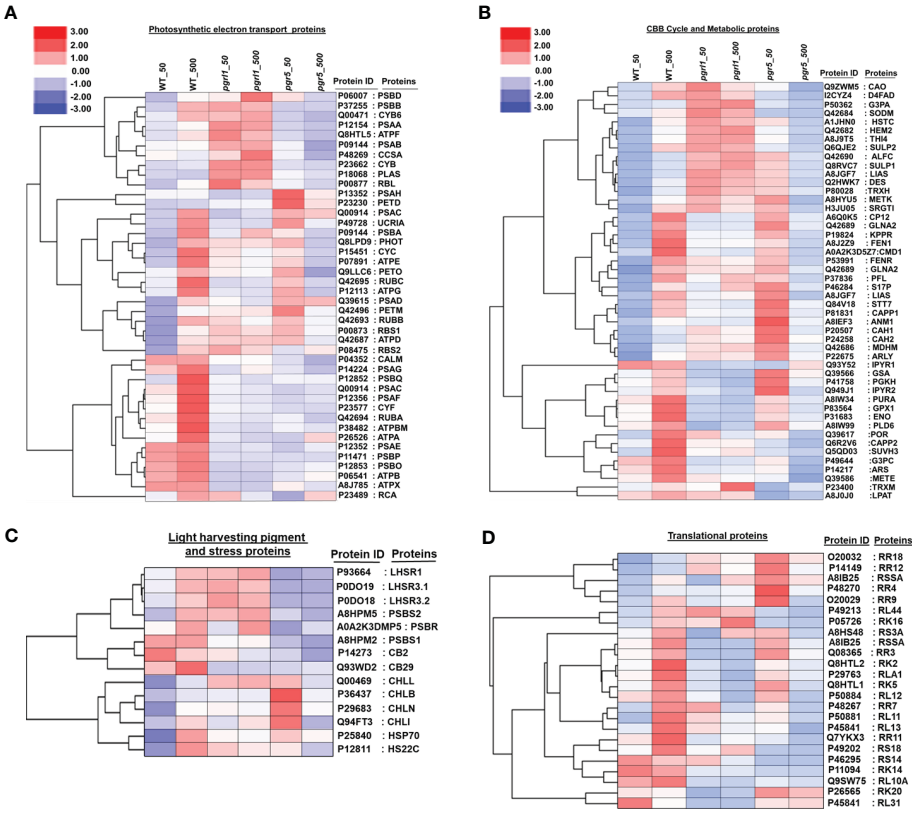


FIGURE 8
Global protein expression levels in *C reinhardtii* in response to high light stress. Hierarchical clustering heatmap representing the relative expression of the DAPs between light-treated versus control cells. Hierarchical clustering based on the abundant levels of protein identified in photosynthesis electron transport **(A)**, CBB and metabolic proteins **(B)**, light-harvesting pigment and stress proteins **(C)**, and translational proteins **(D)** was performed using Gene Cluster 3.0 software with the similarity metric and linkage method. The resulting clusters were visualized using Java Tree View software.

chloroplastic), KPPR_P19824 (Phosphoribulokinase_chloroplastic), FEN1_A8J2Z9 (Flap endonuclease 1), CMD1_A0A2K3D5Z7 (5-methylcytosine-modifying enzyme), FENR_P53991, GLNA1_Q42688 (Glutamine synthetase cytosolic isozyme), PFL_P37836 (Formate acetyltransferase), S17P_P46284, LIAS_A8JGF7 (Lipoyl synthase_mitochondrial), STT7_Q84V18 (Serine/threonine-protein kinase stt7_chloroplastic), CAPP1_P81831 (Phosphoenolpyruvate carboxylase 1), and ANM1_A8IEF3 (Protein arginine N-methyltransferase 1). It also contains CAH1_(P20507), CAH2_(P24258), MDHM_Q42686 (Malate dehydrogenase_mitochondrial), ARLY_P22675 (Argininosuccinate lyase), and 17 DEPs in which 7 are upregulated in WT high light and 15 are upregulated in *pgr5* growth light compared to *pgrl1*. These proteins are involved in carbon fixation, protein lipoylation, modifiable cyclic, and non-cyclic electron transport, supplying additional ATP and reducing power and CO_2 uptake.

Cluster 3 includes TRXM_P23400 (Thioredoxin M-type_chloroplastic), G3PC_P49644 (Glyceraldehyde-3-phosphate dehydrogenase_cytosolic), GPX1_P83564 (Glutathione peroxidase 1_mitochondrial), GSA_Q39566 (Glutamate-1-semialdehyde 2_1-aminomutase_chloroplastic), IPYR1_Q93Y52 (Soluble inorganic pyrophosphatase 1_chloroplastic), PURA_A8IW34 (Adenyl succinate synthetase_chloroplastic), POR_Q39617 (Protochlorophyllide reductase_chloroplastic), PGKH_P41758 (Phosphoglycerate kinase_chloroplastic), CAPP2_Q6R2V6 (Phosphoenolpyruvate carboxylase 2), ARS_P14217 (Arylsulfatase), ENO_P31683 (Enolase), IPYR2_Q949J1 (Solubleinorganicpyrophosphatase2), METE_Q39586 (5-methyl tetra hydropteroyl triglutamate-homocysteine methyltransferase), SUVH3_Q5QD03 (histone-lysine N-methyltransferase_H3 lysine-9 specific SUVH3), PLD6_A8IW99 (Mitochondrial cardiolipin hydrolase), LPAT_A8J0J0 (1-acyl-sn-glycerol-3-phosphate acyltransferase), and 16 DEPs that commonly upregulate in WT high light. Five proteins are exclusively upregulated in *pgr5* growth light and are involved in the glycolytic process (G3PC_P49644, PGKH_P41758, GSA_Q39566, and ENO_P31683), purine nucleotide biosynthesis (PURA_A8IW34), light-independent chlorophyll biosynthesis (POR_Q39617), and lipid biosynthesis (LPAT_A8J0J0) (Figure 8B). The light-harvesting pigments and photoprotection proteins were subdivided into 14 sub-categories with two major clusters.

Cluster 1 includes LHR3.1_P0DO18, LHSR1_P93664, LHSR3.2_P0DO19, PSBS1_A8HPM2, PSBS2_A8HPM5, PSBR_A0A2K3DMP5, CB29_Q93WD2, and CB2_P14273; two photoprotective proteins (LHCSR and PSBS); and two Chl a/b binding proteins (CB29 and CB2) upregulated in WT and *pgrl1* but downregulated in *pgr5* high light (Figure 8C). We have found that Cluster 2 includes CHLB_P37824 (Light-independent protochlorophyllide reductase subunit B), CHLL_Q00469 (Light-

independent protochlorophyllide reductase iron-sulfur ATP-binding protein), CHLN_P29683 (Light-independent protochlorophyllide reductase subunit N), CHLI_Q94FT3 (Magnesium-chelatase subunit ChlI_chloroplastic), HSP70_P25840 (Heat shock 70 kDa protein), HS22C_P12811 (Heat shock 22 kDa protein_chloroplastic), and H2A_P50567 (Histone protein), and seven DEPs associated with chlorophyll pigment-binding protein upregulated in WT high light.

The protein involved in translational machinery was further subdivided into 24 sub-categories with two clusters. Among them, minor clusters containing five DEPs [RR18_O20032, RR4_P48270, RR12_P14149 (30S ribosomal protein S7_chloroplastic), RSSA_A8IB25 (40S ribosomal protein SA), and RL44_P49213 (60S ribosomal protein L44)] were upregulated in *pgr5* growth light. In contrast, significant clusters included 19 DEPs [RR2A_O47027, RR3_Q08365, RR7_P48267, RR9_O20029, RR11_P14149 (30S ribosomal proteins_chloroplastic), RS3A_A8HS48, RS14_P46295, RS18_P49202 (40S ribosomal protein), RK2_Q8HTL2, RK5_Q8HTL1, RK14_P11094, RK16_P05726, RK20_P26565 (50S ribosomal proteins_chloroplastic), RLA1_P29763, RL10A_Q9SW75, RL11_P50881, RL12_P50884, RL13_O48513, and RL31_P45841 (60S ribosomal protein L13)] in which 17 were upregulated in WT and the remaining 2 (RK20_P26565 and RL31_P45841) were upregulated in *pgr5* high light condition (Figure 8D).

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analysis

The Gene Ontology (GO) enrichment analysis includes a biological process (BP), cellular component (CC), and molecular function (MF) of upregulated, downregulated, and DEPs (p-value is <0.05) under high light stress. Gene ontology analysis in WT_500 vs. WT_50 high light proteomics reveals a special enrichment for PSII. We detected considerable changes in proteins growing in high light WT samples. Out of 350 DEPs, 18 were involved in 18 BP, 15 proteins in 14 CC, and 23 were engaged in 35 MF categories. We have also separated some core proteins from the groups WT_500 vs. WT_50, mainly involved in ETR, CET, and water photolysis. KEGG pathway showed that 16 DEPs are primarily interested in photosynthesis and metabolic pathway proteins (Figure 9A). GO enrichment analysis of the global proteomic dataset of lightly treated *pgrl1*_500 vs. WT_50 cultures revealed that *pgrl1* was enriched over WT in which 39 proteins were involved in 83 biological functions, 32 types of proteins in 88 MF, and 45 types of proteins in 58 CC, including protein, unfolding, PSII repairing, and oxidative responses related to ROS.

KEGG pathway showed that 15 DEPs are mainly involved in three major categories, i.e., photosynthesis, oxidative phosphorylation, and metabolic pathway (Figure 9B). Quantitative proteomics was performed on *pgr5_500* vs. WT_50 under high light conditions, where 294 proteins were quantified in all functional categories, in which 15 proteins were involved in 15 different biological functions, 11 proteins in 12 MF, and 16 proteins in 31 CC categories. In contrast, *pgr5* lacked GO terms related to carbohydrate metabolism, TCA cycle, and PSII electron transport (Figure 9C). KEGG pathway showed that nine DEPs are mainly engaged in three major categories, i.e., photosynthesis, oxidative phosphorylation, and metabolic pathway. These proteins show differential protein expression in high light and play an essential role in photoprotection. In contrast, *pgr5* proteins remarkably decrease in high light, including assembly and stabilization of PSII, OEC complex assembly, and oxygen evolution, which is crucial for the regulation of photosynthetic-related protein that has not been reported yet in algae so far. Notably, few photosynthesis-related proteins were not detected in WT high light.

Volcano plot and protein network analysis of DEP-regulated protein under high light stress

Volcano plots represent the proteins with substantial alterations made by high light in WT, *pgrl1*, and *pgr5*. Following light treatment, 274 DEPs significantly changed in WT_500 vs. WT_50, with 34 upregulated and 19 downregulated proteins (Figure 10A), while 219 DEPs significantly changed in the *pgrl1* treated vs. control samples (*pgrl1*_500 vs. WT_50), with 18 upregulated and 124 downregulated (Figure 10C). Comparing the *pgr5*_500 vs. WT_50, a total of 215 DEPs are significantly identified, of which 18 are upregulated and 25 are downregulated after light treatment (Figure 10E). Based on the volcano plot, we have taken the DEPs and elucidated the protein–protein interaction network of selected fold values of ≤0.86 (downregulated—low abundant protein) and ≥1.2 (upregulated—high abundant protein) in treated vs. control samples in WT_50 vs. WT_50, *pgrl1*_500 vs. WT_50 and *pgr5*_500 vs. WT_50. The nodes were shown in circles with two colors (red—upregulated protein, green—downregulated) to simplify further the correlation network based on the degree of interactions.

For the WT_500 vs. WT_50 DEPs correlation network, every data point was recognized as a node, and a total of 28 nodes (each node represents a single protein) formed 76 edges (neighboring interactions); on average, each node shared two to three edges with adjacent nodes (Figure 10B). The red nodes belonging to proteins associated with PETC PSII (PSBA_P07753, PSBB_P37255, PSBC_P10898, and PSBD_P06007), OEC (PSBO_P12853), PSI (PsaA_P12154), ATP synthase (ATPF_P12356 and ATPBM_P38482), high light-induced protein (LHSR3.2_P0DO19), and proteins involved in interflagellar assembly [IFT57_Q2XQY7, IFT81_Q68RJ5, and IFT74_Q6RCE1 (intraflagellar transport protein)] are upregulated.

The green nodes belonging to heat shock protein HSP70_P25840, PSI PSAH_P13352, L-serine amino-acid biosynthesis PSP1_A8JEM3 (O-phosphoserine phosphohydrolase), and protein involved in interflagellar assembly (IFT27_A8HN58 (Intraflagellar transport protein 27) are downregulated (Figure 10B). To simplify the *pgrl1*_500 vs. WT_50 correlation network, the 52 nodes were categorized in circles by two colors (red and green) based on the degree of interactions. Green nodes comprise a dense network containing 172 interaction edges with neighboring nodes. Interestingly, the green nodes belonging to proteins related to PETC (PSBB_P37255, PSBC_P10898, PSBQ_P12852, PSAF_P12356, and PSAK_P14225), plastocyanin (PLAS_P18068), ATP synthase

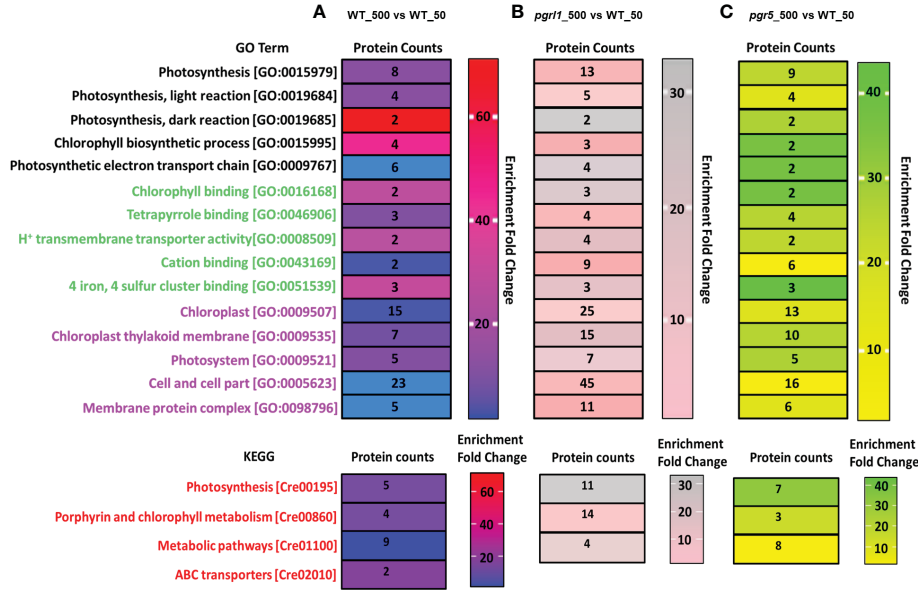


FIGURE 9
Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis on light stress-induced up-/downregulation of expressed proteins is divided into three categories of Go Term, including GO function and ID, Protein count, and Enrichment fold change. Panels **(A–C)** represent the up-/downregulation of proteins induced by WT-500 vs. W-50 **(A)**, *pgrl1*-500 vs. W-50 **(B)**, and *pgr5*-500 vs. W-50 **(C)** treatment versus control. BP1 -5, biological process; MF1 -5, molecular function; CC1 -5, cellular components The top three biological pathways were presented as KEGG pathways as follows for upregulated, downregulated, and differentially regulated proteins: (1) BP1 (biological process), Photosynthesis; BP2, Photosynthesis, light reaction; BP3, Photosynthesis, dark reaction; BP4, chlorophyll biosynthetic process; BP5, photosynthetic electron transport chain; (2) MF1 (molecular function), chlorophyll-binding; MF2, tetrapyrrole binding; MF3, H^+ ion transmembrane transporter activity; MF4, cation binding; MF5, four iron, four sulfur cluster binding; (3) CC1-5 (cellular components)—chloroplast, chloroplast thylakoid membrane, photosystems, cell and cell part, and membrane protein complex. However, the KEGG pathway represents the top three pathways in differentially abundant proteins involved in photosynthesis, porphyrin and chlorophyll metabolism, and metabolic pathways.

subunits (ATPD_Q42687 and ATPF_P12356), Cyt b_6f subunit for CET (PETO_Q9LLC6 and PETM_Q42496) interflagellar assembly (IFT81_Q68RJ5, IFT57_Q2XQY7, and IFT74_Q6RCE1), inorganic carbon uptake (CEMA_Q37050), light-independent chlorophyll biosynthesis (CHLI_Q94FT3), CCM proteins (CAH1_P20507 and CAH2_P24258), flagellar radial spoke protein or flagellar bending (RSP2_Q6UBQ3, RSP4_Q01656, and RSP5_Q27YU7), and PsbB mRNA maturation factor (MBB1_Q9FNS4_chloroplastic) are downregulated.

For the DEPs correlation network, every data point was reported as a node (each node represents a single protein), and each forms edges (neighboring interactions) with adjacent nodes (Figures 10B, D, E). Furthermore, red nodes that formed a dense network containing an average of ~15 interaction edges with neighboring nodes, which correspond to protein species related to PSII (PSBA_P07753 and PSBD_P06007), PSI (PSAG_P14224), chlorophyll-binding/light-harvesting (CB29_Q93WD2 and CB2_P14273), light-independent chlorophyll biosynthesis (CHLB_P37824), translational protein (RR2B_Q8HUH1 (putative 30S ribosomal S2-like protein), RR3_Q08365), and histone protein for DNA binding or nucleosome remodeling [H2B3_P54346 (Histone H2B.3) and H2B4_P54347 (Histone H2B.4)], are upregulated (Figure 10D). Similarly, in *pgr5*_500 vs. WT_50, 24 nodes formed a solid interaction network with 47 edges. To simplify the correlation matrix, each node was represented with a specific color (red—upregulated and green—downregulated) based on the degree of interaction. Furthermore, green nodes that formed a network containing an average of two interaction edges with neighboring nodes, which correspond to proteins related to PSII (PSBI_Q8DJZ6 and PSBO_P12853), PSI (PSAC_Q00914), heat shock proteins [HSP70_P25840 and H32_Q6LCW8 (Histone H3 type 2)], chlorophyll-binding/light-harvesting (CB29_Q93WD2), flagellar bending/calmodulin-binding [RSP2_Q6UBQ3 and RSP3_P12759 (Flagellar radial spoke protein 2 and 3)], and metabolic reaction (IPYR2_Q949J1 and S17P_P46284), are downregulated. Furthermore, red nodes that formed a weak network with neighboring nodes, which correspond to proteins related to PSII (PSBA_P07753 and PSBC_P10898), PSI (PSAB_P09144), ATP Synthesis (ATPD_Q42687), flagellar bending/calmodulin-binding RSP4_Q01656 (Flagellar radial spoke protein 4), RSP14_A8HNV0 (Radial spoke protein 14), heat shock protein [HS22C_P12811 (Heat shock 22 kDa protein_chloroplastic)], and Cyt b_6f subunit for CET (PETD_P23230), are upregulated (Figure 10F).

Discussion

Light stress has developed an accurate risk to *Chlamydomonas* growth and biomass development. *Chlamydomonas*' regulating adaptation mechanisms have not been identified adequately under high light. Our recent report shows that cyclic electron transport is crucial in photoprotection, where we showed that *pgrl1* and *pgr5* play an essential role under high light stress

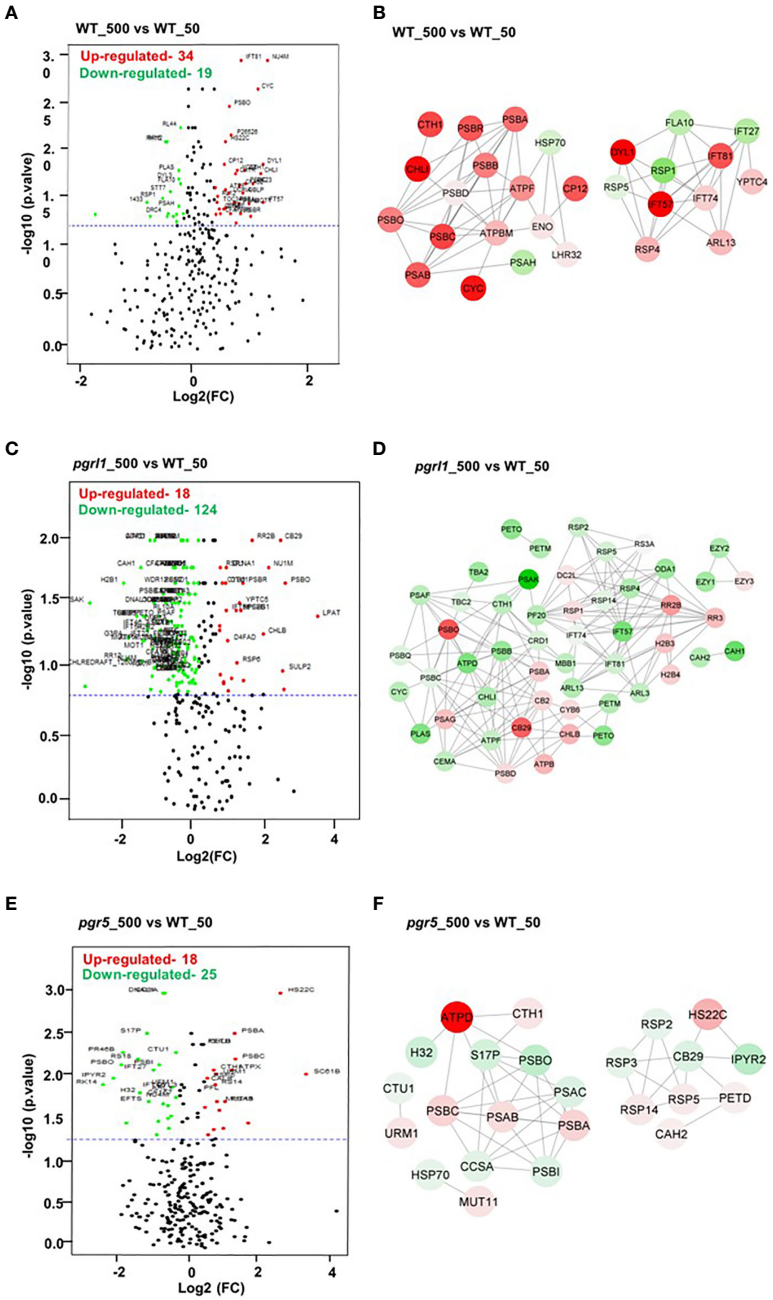


FIGURE 10
Volcano plot and correlation network analysis of DEPs in WT, *pgrl1*, and *pgr5* mutants. Volcano plots **(A, C, E)** and protein–protein interaction networks **(B, D, F)** represent the proteins with significant differences induced by WT-500 vs. WT-50, *pgrl1*-500 vs. WT-50, and *pgr5*-500 vs. WT-50. Proteins with significant differences and large fold changes are shown in red scatters as upregulated and green as downregulated. *X*- and *Y*-axis indicate a considerable difference in terms of both the two-sided Student's *t*-test (*p*-value < 0.05) and the fold cutoff (>1.2 and <0.86).

(Yadav et al., 2020; Chouhan et al., 2023). In this study, we showed that NPQ is significantly increased in WT, whereas it is reduced in *pgrl1* and *pgr5*. However, the role of total proteins and the organization of thylakoid super-complexes are not yet understood under high light stress. This study comprehensively analyzed crucial PETC and metabolic proteins directly involved under high light. We have addressed the entire proteome profile and its relation to the cell's function of photosynthesis efficiency under high light from WT, *pgrl1*, and *pgr5* mutants of *C. reinhardtii*.

CET mutation reduced the capacity of oxygen uptake under high light stress

A high oxygen evolution has been observed for cells grown under high light in WT compared to growth light, indicating that the impact of high light was not drastic on PSII damage. The light-induced O_2 uptake in *pgrl1* and *pgr5* was significantly reduced than in WT, demonstrating that the mutants' lack of CEF affected acclimation to high light (Figure 1A; Supplementary Figure 2B).

This result correlates with Chl fluorescence, which is similarly reduced in *pgrl1* and *pgr5* (Supplementary Figure 2A). Therefore, O_2 evolved by chlororespiration, Mehler reaction, or the malate shunt, stimulating the recurrence of PSII photochemistry (Schreiber and Vidaver, 1974). These results suggest that cells adjust under high light stress by different adaptive mechanisms and show agreements with our biochemical data; i.e., oxygen-evolving complex subunit proteins PsbO and PsbP decreased in *pgrl1* and *pgr5* (Figure 1C; Supplementary Figure 1D). Moreover, proteome data showed that all the subunits of the oxygen-evolving complex are downregulated in *pgrl1* and *pgr5* than WT under high light (Figure 7A). These results indicated that WT acclimates better under high light, whereas *pgrl1* and *pgr5* mutants are more prone to photosystem damage.

The pigment–protein composition and membrane stoichiometry were affected under high light in *pgrl1* and *pgr5*

In WT cells, the LHCSR3 and PSBS are defined under high light; therefore, the thylakoid membrane architecture and composition of PSII–LHCII super-complexes are not affected much (Figure 3A). The weaker *psi*-type CD signal in *pgrl1* and *pgr5* cells grown under high light indicates significant rearrangement in the macromolecular arrays of the PSII–LHCII complex (Figures 2A, B). The positive band at 506 nm in the Soret region originates from the β-carotene of the core complex (Tóth et al., 2016). It exhibits no or weak signal due to a decreased protein-containing core complex in cells subjected to high light stress (Figures 2C, D). The same result was observed from the composition of thylakoid membrane protein complexes in BN-PAGE analysis (Figure 3). In *C. reinhardtii*, PSII–LHCII comprises two reaction center cores and six LHCII trimers (Järvi et al., 2011). In *pgrl1* and *pgr5*, most of the LHCs of super-complex are associated with higher-order LHCs (LHC II trimers) or LHCII monomers under high light (Figures 3A, B). From BN-PAGE, owing to solubilization with β-DM, loosely bound trimers dissociated from PSII–LHCII super-complexes, leaving different core combinations with trimers and monomers of *pgrl1* and *pgr5* in high light (Figure 3). The abundance of super-complexes was not altered significantly in the *pgrl1*, but a dramatic change was observed in *pgr5* since it lacks LHCSR3 expression. The BN-PAGE result complements the CD results, where higher-order structures in the thylakoid membrane are affected under high light conditions. These observations conclude that the expression of LHCSR3 in high light could photo-protect the photosystems to hinder the over-excitation energy transfer from LHCII to RC.

The protein profile analysis from immunoblotting provides insight into the *pgrl1* and *pgr5* under high light

The D1 and D2 proteins are the reaction center-binding proteins of PSII that carry out essential redox components for charge separation. The protein level changes in the cells grown in high light showed a marginal increase of D1 and D2 in WT (Figures 4A, B). In high light-grown cells, the D1 and D2 proteins accumulated more in *pgrl1* and *pgr5*; the continuous repair mechanisms change their abundance. PSII core protein synthesis and its photodamage repair depend on many factors that could affect the process of their acclimation resulting in minimal change in its content. The photodamage of PSII repair was inhibited by ROS production (Murata et al., 2007), and the translational initiation of PsbA transcripts was activated through a redox signal associated with electron transport in PSI and reduction of the plastoquinone pool (Trebitsh and Danon, 2001).

Our recent reports show that in *pgr5*, significant ROS production was observed under high light, which could be one of the reasons for the change in photosynthetic efficiency and proteins (Chouhan et al., 2022). Closely associated with these core proteins are core antennae chlorophyll-binding proteins such as CP43 and CP47. Its association with the PSII super-complex transfer excitation energy from light-harvesting complexes to the chlorophyll-binding protein CP43 is selectively inhibited compared to CP47, preventing excess excitation energy from the overloading reaction center to change the core antenna proteins' content (Kim et al., 2017). The PsaA is significantly decreased in *pgr5* than in *pgrl1* and shows reduced protein content due to changes in chlorophyll and carotenoid molecules around the protein, which coordinates the cofactors to transfer electrons to ferredoxin (Figures 4C, D). The reduction of stromal subunits PsaD can distract the assembly of PSI. The associated stromal subunits PsaD and PsaE act as a ferredoxin binding site, leading to an efficient electron transport chain (Fischer et al., 1998).

The LHCII chlorophyll a/b-binding proteins were relatively consistent between WT and the *pgrl1* except for *pgr5* under high light. In the LHCII trimer, Lhcb1 and Lhcb3, peripherally arranged, depicted a drastic change in high light, while Lhcb2 near the PSII core showed no difference in protein content. This result indicates that high light mainly affects LHCs arranged peripherally in the super-complexes. The LHC monomeric protein, Lhcb5, reveals the decreased content, while there was no change in Lhcbm5, indicating that light stress mainly affects outermost arranged light-harvesting antennas. Light-harvesting complex proteins Lhcb3 and Lhcb5 were more tightly associated with the PSII core than Lhcb1 and Lhcb2 (Iwai et al., 2008). This might be the reason for the inhibition of outer antennae proteins.

An isoform of Lhcb4 connects trimeric LHCII with the PSII core; it stabilizes the PSII–LHCII dimer and binds to the monomeric core with the help of LHCII, suggesting its primary role in regulating the excitation energy flow and non-photochemical quenching process (Caffarri et al., 2009). Interestingly, the Lhcb4 has no significant change; thus, Lhcb4 might increase the rate of non-photochemical quenching, which is usually higher in high light conditions (Figures 5A, C, D). Furthermore, we examined the content of light-harvesting complex proteins of PSI that are arranged in two layers towards the stromal side of the PSI core. The outer half-ring of the super-complex includes Lhca4, Lhca5, and Lhca6, whereas the inner half-ring of the super-complex is composed of Lhca1, Lhca3, Lhca7, and

Lhca8 with Lhca3 located near PsaK, as previously suggested (Drop et al., 2011; Yadavalli et al., 2011a; Ozawa et al., 2018). In assumption, eight LHCI subunits bind to PSAF in two layers from PsaG to the PsaK subunit. The innermost layer contains Lhca 1, 4, 6, and 5, and the outside layer comprises Lhca 3, 4, 6, and 7, which decreased severely in *pgr5* and might be the distracting assembly of PSI-LHCI under high light conditions.

Earlier, our group reported that under extreme high light conditions (1,000 μmol photons m^{-2} s^{-1}), most of the inner LHCI proteins were affected by the ROS production in WT (Nama et al., 2019). Interestingly, under moderate or high light, we can say that the light-harvesting complexes of PSI are stable. However, their composition might be affected by a significant decrease in their binding site of PsaG and PsaH (Figures 5B–D). We recently reported that some of the core proteins were aggregated in moderately high light conditions, especially in *pgr5*, due to the induction of ROS (Chouhan et al., 2022). This could be the reason the efficiency of photosynthesis is reduced in *pgr5*. However, the stoichiometry of the PSI and PSII LHCs is maintained, indicating that *C. reinhardtii* acclimates rapidly by changing the antenna proteins under high light conditions.

The proteome data supports the role of CET in maintenance of photosynthesis and photoprotection

Here, we have discussed the comparative analysis of Western blot and proteome data of the PETC protein. The phosphorylation and repair cycle of PSII prevents D1 degradation (Koivuniemi et al., 1995). Our data show significant downregulation of PSBA_P07753 (D1) and PSBD_P06007 (D2) in *pgrl1* and *pgr5* compared to WT high light, indicating that D1 incurs more damage under high light conditions, and it is in agreement with the Western blot data (Figures 4A, B). The primary antenna proteins PSBB_P37255 (CP43) and PSBC_P10898 (CP47) are downregulated, especially in *pgr5* than in *pgrl1*, as we can see in our Western blot data where the protein content was decreased in *pgr5* more than the *pgrl1* (Figure 7B). These outcomes designate that damaging the catalytic subunits D1 and D2 leads to the downregulation of the reaction centers associated with antenna protein PSBC_P10898 (CP43) and PSBB_P37255 (CP47) in the mutants under high light conditions (Figure 7A). This might be due to light stress, primarily phosphorylated D1, D2, and CP43 in threonine residues at N-terminal (Elich et al., 1992; Booij-James et al., 2002) by the serine and threonine kinase (Vainonen et al., 2005; Rochaix et al., 2012).

The PSI reaction center proteins, PSAA_P12154, PSAB_P09144, PSAC_Q00914, PSAD_Q39615, PSAE_P12352, PSAF_P12356, PSAG_P14224, and PSAH_P13352, are downregulated in *pgrl1* and *pgr5* compared to WT, reflecting the importance of these proteins in the maintenance of the PSI assembly and CET under high light (Figure 7A). The LHCSR3 protein regulated NPQ (Peers et al., 2009) and showed a substantial change in *pgrl1* and *pgr5* compared to WT, signifying that CET and high light changed NPQ at the post-translational level. The NPQ is activated in the PSII antenna and disperses extra energy as heat (Ruban, 2016). Similarly, *Chlamydomonas* exhibited a substantial downregulation on LHCSR3.1_P0DO19 and LHCSR3.2_P0DO19 in *pgrl1* and *pgr5* due to the absence of CET under high light conditions. Another report showed that the LHCSR3 expression was supplementary for NPQ initiation in *Chlamydomonas* (Bonente et al., 2011). LHCSR1_P93664 was upregulated in *pgrl1* and *pgr5* and might be the zeaxanthin cycle compensated without CET under high light, which is also in line with our recent report (Figure 7B) (Chouhan et al., 2022). This finding aligns with NPQ data, showing that *pgr5* cannot induce NPQ, but *pgrl1* has a lower NPQ than WT under high light, to agree with our previous report (Yadav et al., 2020).

The PSBS1_A8HPM2 and PSBS2_A8HPM5 were upregulated, especially in *pgr5*; limited proton motive force (*pmf*) and LHCSR3 expression might increase the abundance of another photoprotective protein in *Chlamydomonas* under high light stress (Figure 7B). The PSBR_A0A2K3DMP5 is highly upregulated in *pgr5*, which offers that PSBR is essential for the steady binding of LHCSR3 to PSII–LHCII super-complexes and required for effective quenching under high light (Xue et al., 2015). The *pgr5* induced PSBS protein under high light stress (Figure 7B). The ATP synthase and Cyt b_6f are two primary components of the PETC regulating the thylakoid lumen acidification or *pmf* under high light conditions. The ATP synthase subunits ATPA_P26526, ATPB_P06541, ATPE_P07891, and ATPG_P12113 are upregulated in WT than *pgrl1* and *pgr5* under high light. *pgr5* displays higher upregulation ATPD levels than WT, suggesting that most of the H^+ is utilized of energy generation in the absence of CET in high light (Supplementary Figure 4B). The Cyt b_6f complex highly regulates the thylakoid lumen acidification and CET pathway. The Cyt b_6f complex subunit CYF_P23577, CYC6_P08197 (Cyt C6_chloroplastic), PETD_P23230, PETO_Q9LLC6, and PETM_Q42496 are upregulated in WT than *pgrl1* and *pgr5* under high light (Figure 7A). Another subunit of the Cyt b_6f complex CYB6_Q00471 is extensively involved in the regulation of lumen acidification and is highly upregulated in *pgr5* mutant showing the importance of CET under high light stress.

Whole-cell revealed significant changes in metabolomic protein under high light

It is known that photosynthesis is sensitive to high light. The change in photosynthesis would affect the metabolism of the cell. Also, the absence of the CET pathway alters the whole metabolic flux, leading to reroutes of the metabolic pathway and compositions under high light conditions to acclimate to the high light. Here, the Fructose bis-phosphate aldolase (ALFC_Q42690) was differentially expressed in WT, *pgrl1*, and *pgr5* in high light in gluconeogenesis and starch biosynthesis (Figure 7C). In addition, G3PA_P50362, engaged in glycerol synthesis, was the most upregulated in *pgr5* under high light. The G3PA_P50362 is also related to the conversion of DHAP to sn-glycerol-3-phosphate, and finally, glycerol kinase (GK) or G3P phosphatase (GPP) produces glycerol (Driever et al., 2017). The G3P is a predecessor for TAG biosynthesis and increases lipid synthesis under salinity stress

(Herrera-Valencia et al., 2012). This could be the reason under high light, where a significant accumulation of lipids and TAG was observed in our recent report in order to photoprotect (Chouhan et al., 2022).

The Ferredoxin-NADP reductase (FENR_P53991), a chloroplastic protein involved in ATP and NADPH synthesis, was downregulated in *pgr5* under high light. These results suggested that, as in higher plants, reduction by the Fd-TRX system regulates G6PDH and PRK activity to optimize carbon reduction and oxidation in the chloroplast following the supply of electrons and response to the metabolic demands of ATP. The protein ENO_P31683 synthesized pyruvate from the D-glyceraldehyde 3-phosphate pathway for glycolysis, and carbohydrate degradation was differentially regulated in WT, *pgrl1*, and *pgr5* in high light. Thioredoxin M-type (TRXM_P23400, chloroplastic) are vital regulatory players receiving information about the redox state, and participating in reversible oxidation was highly upregulated in WT. The M type of TRX is known to activate NADP-malate dehydrogenase to hold the CET indirectly.

The TCA cycle protein malate dehydrogenase (MDHM_Q42686) was differentially regulated in WT, *pgrl1*, and *pgr5* under high light (Figure 7C). The MDHM converts oxaloacetate to malate, generating extra reducing power (Scheibe, 2004). Another mitochondrial enzyme, Glutathione peroxidase 1 (GPX1_P83564), was upregulated in WT and may constitute a glutathione peroxidase-like protective system against oxidative stresses synthesizing PUFA through arachidonic acid metabolic pathways and was differentially regulated in *pgrl1* and *pgr5* in high light. Phosphoenolpyruvate carboxylase 1 (Ppc1), the carboxylation of phosphoenolpyruvate (PEP) that forms oxaloacetate, was differentially regulated in WT, *pgrl1*, and *pgr5* in high light. It is activated by glutamine and dihydroxyacetone phosphate and inhibited by glutamate and malate. The enzymes of the CBB cycle are highly upregulated in WT, including the protein complexities involved in RuBisCO assembly (RUBA_Q42694 and RuBB_Q42693 proteins). The phosphoribokinase converts ribulose-5-phosphate to RuBP (ribulose-1,5-bisphosphate), the leading CO_2-accepting enzyme differentially regulated in *pgrl1* and *pgr5* under high light. Rubisco uses RuBP as a substrate, creating two molecules of 3-phosphoglycerate. Increased expression of fructose bisphosphate enhances fixing CO_2 fixation and RuBP renewal (Lefebvre et al., 2005) by refining photosynthesis under stress (Driever et al., 2017).

The reduced amount of P3GA_P50362 (Glyceraldehyde-3-phosphate dehydrogenase A_chloroplastic), synthesized by RuBisCO, activates high carbon sinks to fixed CO_2 in high light-grown cultures. Lastly, the overexpression of these necessary CBB enzymes maintained strong protein translation abilities, i.e., 24 ribosomal proteins in which 17 were upregulated in WT and 7 were upregulated in *pgr5* under high light. The heatmap data show that 250 DEPs were divided into 49 Calvin cycle and metabolism protein sub-categories. The Carbonic anhydrase 1 (CAH1_P20507) and 2 (CAH2_P24258) proteins that play a crucial role in CCM responsible for the reversible carbonate dehydratase activity for the formation of CO_2 were upregulated in WT, and remarkably downregulated in *pgrl1* and *pgr5* under high light stress (Figure 7C). The enhanced CBB protein in WT supports healthy photosynthetic growth. When photosynthesis is compromised, the starch converts into sugars and other metabolites to deliver carbon and energy under abiotic stress. Our recent laboratory reports show that *pgrl1* and *pgr5* accumulated significant lipids to acclimate to high light (Chouhan et al., 2022). The over-reduction of the PETC generates additional NADPH and NADH by mitochondria, which balances the equilibrium by cellular redox homeostasis under salt stress (Wingler, 2002).

Heatmap, volcano plot, and Gene Ontology analysis revealed the CET-involved enrichment of photosynthesis proteins but significantly lacks a light-harvesting complex

The Gene Ontology analysis (Figure 9) and volcano plot data (Figure 10) revealed an unexpected enrichment of photosynthetic targets not described in previous studies in *Chlamydomonas* under high light. Similar to our heatmap data, GO proteins mainly engage in chlorophyll biosynthesis, photosynthetic electron transport chain, chlorophyll-binding, heat shock protein, light-harvesting, translational protein, and protein involved in interflagellar assembly (Figure 9). We have identified only two light-harvesting pigment chlorophyll a/b binding proteins CB29_Q93WD2 (Chlorophyll a-b binding protein CP29) and CB2_P14273 (Chlorophyll a-b binding protein of LHCII type I chloroplastic) upregulated in WT and *pgrl1* but downregulated in *pgr5* in high light that shows agreement with our protein data (Figure 8C). The remaining PSII–LHCII and PSI–LHCI proteins are not identified during the analysis and might be lost during protein extraction or are underneath the nLC/MS/MS detection limit.

Conclusion

Targeted proteome analysis was conducted to compare the photosystem-related and central metabolic protein abundance in WT, *pgrl1*, and *pgr5*. The physiological and biochemical data show that the *pgr5* is prone to high light and observed compromised photosynthesis. Also, the thylakoid super-complexes and long-order array structure of LHCII were diminished in *pgr5* due to the induction of ROS. The proteomic data distinguish the possible molecular mechanism of CET on photosynthesis and metabolomic profile in response to high light. Our results specify a robust relationship between CET and Cyt b_6f that impacts lumen acidification and NPQ. We have shown that mutants share mutual targets of CET that control photosynthetic machinery and compensation mechanisms, including regulation of *pmf* by Cyt b_6f and ATP synthase activity that indirectly affect the photoprotective protein in high light. We have also identified that increased expression of PSBS protein in *pgr5* compared to WT acts independently, controlling NPQ in high light. In the GO and KEGG pathway studies, we recognized that high light affected photosynthesis, oxidative phosphorylation, and metabolic pathway proteins to prevent abolishing carbon assimilation; few of them play a

significant role in photoprotection. The limitation of metabolite appears to be accessible via the TCA cycle affecting the restriction of CET. Generating additional NADPH by LET directly supplies the energy to carbon metabolism to eliminate ROS and neutralize the deleterious inhibitory effects. Our study reveals that CET is crucial in acclimation to the high light stress. Since the proteomic results have given a lot of clues, future efforts would address the role of these CET proteins *pgrl1* and *pgr5* in affecting the photosynthetic electron transport and inducing the downregulation of cellular metabolism in chloroplast independently or in coordination with mitochondria over different targets to maintain cellular homeostasis.

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

RY and RS planned the experiments and evaluated the results. RY, SM, MZ, and JP performed the experiments. RY and RS wrote the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

RS was supported by the Institute of Eminence (UoH/IoE/RC1/RC1-20-019), Department of Biotechnology (BT/PR14964/BPA/118/137/2015), Council of Scientific and Industrial Research (38 (1504)/21/EMR-ll), Science & Engineering Research Board (CRG/2020/000489), Department of Biotechnology-Builder (BT/INF/22/SP41176/2020), DST-FIST, and UGC-SAP, Govt. of India. We thank Gilles Peltier, CEA-CNRS-Aix Marseille Université, France, and Michael Hippler, Institute of Plant Biotechnology, University of Münster, Germany, for providing the mutants *pgrl1* and *pgr5*. We thank Gyozo Garab, Biological Research Center, Szeged, Hungary, for his critical manuscript reading. We also thank Sandor Life Sciences Pvt Ltd, Hyderabad, for the Proteomic data analysis.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1198474/full#supplementary-material

References

Ahn, T. K., Avenson, T. J., Ballottari, M., Cheng, Y. C., Niyogi, K. K., Bassi, R., et al. (2008). Architecture of a charge-transfer state regulating light harvesting in a plant antenna protein. *Science* 320 (5877), 794–797. doi: 10.1126/science.1154800

Albertsson, P.Å., Andreasson, E., Svensson, P., and Yu, S. G. (1991). Localization of cytochrome f in the thylakoid membrane: evidence for multiple domains. *Biochim. Biophys. Acta (BBA)-Bioenergetics* 1098 (1), 90–94. doi: 10.1016/0005-2728(91)90012-D

Allakhverdiev, S. I., Kreslavski, V. D., Klimov, V. V., Los, D. A., Carpentier, R., and Mohanty, P. (2008). Heat stress: an overview of molecular responses in photosynthesis. *Photosynthesis Res.* 98, 541–550. doi: 10.1007/s11120-008-9331-0

Allakhverdiev, S. I., and Murata, N. (2004). Environmental stress inhibits the synthesis *de novo* of proteins involved in the photodamage–repair cycle of photosystem II in synechocystis sp. PCC 6803. *Biochim. Biophys. Acta (BBA)-Bioenergetics* 1657 (1), 23–32. doi: 10.1016/j.bbabio.2004.03.003

Anderson, J. M. (1982). Distribution of the cytochromes of spinach chloroplasts between the appressed membranes of grana stacks and stroma-exposed thylakoid regions. *FEBS Lett.* 138 (1), 62–66. doi: 10.1016/0014-5793(82)80395-5

Bennett, D. I., Fleming, G. R., and Amarnath, K. (2018). Energy-dependent quenching adjusts the excitation diffusion length to regulate photosynthetic light harvesting. *Proc. Natl. Acad. Sci.* 115 (41), E9523–E9531. doi: 10.1073/pnas.1806597115

Bonente, G., Ballottari, M., Truong, T. B., Morosinotto, T., Ahn, T. K., Fleming, G. R., et al. (2011). Analysis of LhcSR3, a protein essential for feedback de-excitation in the green alga chlamydomonas reinhardtii. *PLoS Biol.* 9 (1), e1000577. doi: 10.1371/journal.pbio.1000577

Bonente, G., Howes, B. D., Caffarri, S., Smulevich, G., and Bassi, R. (2008). Interactions between the photosystem II subunit PsbS and xanthophylls studied *in vivo* and *in vitro*. *J. Biol. Chem.* 283 (13), 8434–8445. doi: 10.1074/jbc.M708291200

Booij-James, I. S., Swegle, W. M., Edelman, M., and Mattoo, A. K. (2002). Phosphorylation of the D1 photosystem II reaction center protein is controlled by an endogenous circadian rhythm. *Plant Physiol.* 130 (4), 2069–2075. doi: 10.1104/pp.013441

Brinkert, K., De Causmaecker, S., Krieger-Liszkay, A., Fantuzzi, A., and Rutherford, A. W. (2016). Bicarbonate-induced redox tuning in photosystem II for regulation and protection. *Proc. Natl. Acad. Sci.* 113 (43), 12144–12149. doi: 10.1073/pnas.1608862113

Caffarri, S., Kouřil, R., Kereïche, S., Boekema, E. J., and Croce, R. (2009). Functional architecture of higher plant photosystem II supercomplexes. *EMBO J.* 28 (19), 3052–3063. doi: 10.1038/emboj.2009.232

Chouhan, N., Devadasu, E., Yadav, R. M., and Subramanyam, R. (2022). Autophagy induced accumulation of lipids in *pgrl1* and *pgr5* of *Chlamydomonas reinhardtii* under high light. *Front. Plant Sci.* 12, 3308. doi: 10.3389/fpls.2021.752634

Chouhan, N., Yadav, R. M., Pandey, J., and Subramanyam, R. (2023). High light-induced changes in thylakoid supercomplexes organization from cyclic electron transport mutants of chlamydomonas reinhardtii. *Biochim. Biophys. Acta (BBA)-Bioenergetics* 864(1), 148917. doi: 10.1016/j.bbabio.2022.148917

Chow, W. S., and Aro, E. M. (2005). *Photoinactivation and mechanisms of recovery. in photosystem II* (Springer, Dordrecht), 627–648.

Correa-Galvis, V., Redekop, P., Guan, K., Griess, A., Truong, T. B., Wakao, S., et al. (2016). Photosystem II subunit PsbS is involved in the induction of LHCSR protein-dependent energy dissipation in chlamydomonas reinhardtii. *J. Biol. Chem.* 291 (33), 17478–17487. doi: 10.1074/jbc.M116.737312

Crisp, P. A., Ganguly, D. R., Smith, A. B., Murray, K. D., Estavillo, G. M., Searle, I., et al. (2017). Rapid recovery gene downregulation during excess-light stress and recovery in arabidopsis. *Plant Cell* 29 (8), 1836–1863. doi: 10.1105/tpc.16.00828

Croce, R., and van Amerongen, H. (2020). Light harvesting in oxygenic photosynthesis: structural biology meets spectroscopy. *Science* 369 (6506), eaay2058. doi: 10.1126/science.aay2058

DalCorso, G., Pesaresi, P., Masiero, S., Aseeva, E., Schünemann, D., Finazzi, G., et al. (2008). A complex containing PGRL1 and PGR5 is involved in the switch between linear and cyclic electron flow in arabidopsis. *Cell* 132 (2), 273–285. doi: 10.1016/j.cell.2007.12.028

Devadasu, E., and Subramanyam, R. (2021). Enhanced lipid production in chlamydomonas reinhardtii caused by severe iron deficiency. *Front. Plant Sci.* 12, 615577. doi: 10.3389/fpls.2021.615577

Driever, S. M., Simkin, A. J., Alotaibi, S., Fisk, S. J., Madgwick, P. J., Sparks, C. A., et al. (2017). Increased SBPase activity improves photosynthesis and grain yield in wheat grown in greenhouse conditions. *Philos. Trans. R. Soc. B: Biol. Sci.* 372 (1730), 20160384. doi: 10.1098/rstb.2016.0384

Drop, B., Webber-Birungi, M., Fusetti, F., Kouřil, R., Redding, K. E., Boekema, E. J., et al. (2011). Photosystem I of chlamydomonas reinhardtii contains nine light-harvesting complexes (Lhca) located on one side of the core. *J. Biol. Chem.* 286 (52), 44878–44887. doi: 10.1074/jbc.M111.301101

Elich, T. D., Edelman, M., and Mattoo, A. K. (1992). Identification, characterization, and resolution of the *in vivo* phosphorylated form of the D1 photosystem II reaction center protein. *J. Biol. Chem.* 267 (5), 3523–3529. doi: 10.1016/S0021-9258(19)50761-X

Fischer, N., Hippler, M., Setif, P., Jacquot, J. P., and Rochaix, J. D. (1998). The PsaC subunit of photosystem I provides an essential lysine residue for fast electron transfer to ferredoxin. *EMBO J.* 17 (4), 849–858. doi: 10.1093/emboj/17.4.849

Gao, Y., Cui, Y., Long, R., Sun, Y., Zhang, T., Yang, Q., et al. (2019). Salt-stress induced proteomic changes of two contrasting alfalfa cultivars during germination stage. *J. Sci. Food Agric.* 99 (3), 1384–1396. doi: 10.1002/jsfa.9331

Garab, G., and van Amerongen, H. (2009). Linear dichroism and circular dichroism in photosynthesis research. *Photosynthesis Res.* 101 (2), 135–146. doi: 10.1007/s11120-009-9424-4

Herrera-Valencia, V. A., Macario-González, L. A., Casais-Molina, M. L., Beltran-Aguilar, A. G., and Peraza-Echeverria, S. (2012). In silico cloning and characterization of the glycerol-3-phosphate dehydrogenase (GPDH) gene family in the green microalga chlamydomonas reinhardtii. *Curr. Microbiol.* 64 (5), 477–485. doi: 10.1007/s00284-012-0095-6

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, 800. doi: 10.3389/fpls.2019.00800

Iwai, M., Takahashi, Y., and Minagawa, J. (2008). Molecular remodelling of photosystem II during state transitions in *Chlamydomonas reinhardtii*. *Plant Cell* 20 (8), 2177–2189. doi: 10.1105/tpc.108.059352

Järvi, S., Suorsa, M., Paakkarinen, V., and Aro, E. M. (2011). Optimized native gel systems for separating thylakoid protein complexes: novel super-and mega-complexes. *Biochem. J.* 439 (2), 207–214. doi: 10.1042/BJ20102155

Jin, H., Sun, Y., Yang, Q., Chao, Y., Kang, J., Jin, H., et al. (2010). Screening of genes induced by salt stress from alfalfa. *Mol. Biol. Rep.* 37 (2), 745–753. doi: 10.1007/s11033-009-9590-7

Kim, E., Akimoto, S., Tokutsu, R., Yokono, M., and Minagawa, J. (2017). Fluorescence lifetime analyses reveal how the high light-responsive protein LHCSR3 transforms PSII light-harvesting complexes into an energy-dissipative state. *J. Biol. Chem.* 292 (46), 18951–18960. doi: 10.1074/jbc.M117.805192

Koivuniemi, A., Aro, E. M., and Andersson, B. (1995). Degradation of the D1-and D2-proteins of photosystem II in higher plants is regulated by reversible phosphorylation. *Biochemistry* 34 (49), 16022–16029. doi: 10.1021/bi00049a016

Lefebvre, B. G., Liu, W., Peterson, R. W., Valentine, K. G., and Wand, A. J. (2005). NMR spectroscopy of proteins encapsulated in a positively charged surfactant. *J. Magn. Reson.* 175 (1), 158–162. doi: 10.1016/j.jmr.2005.03.008

Liang, C., Cheng, S., Zhang, Y., Sun, Y., Fernie, A. R., Kang, K., et al. (2016). Transcriptomic, proteomic and metabolic changes in arabidopsis thaliana leaves after the onset of illumination. *BMC Plant Biol.* 16 (1), 1–17. doi: 10.1186/s12870-016-0726-3

Moseley, J. L., Allinger, T., Herzog, S., Hoerth, P., Wehinger, E., Merchant, S., et al. (2002). Adaptation to fe-deficiency requires remodeling of the photosynthetic apparatus. *EMBO J.* 21 (24), 6709–6720. doi: 10.1093/emboj/cdf666

Murata, N., Takahashi, S., Nishiyama, Y., and Allakhverdiev, S. I. (2007). Photo inhibition of photosystem II under environmental stress. *Biochim. Biophys. Acta (BBA)-Bioenergetics* 1767 (6), 414–421. doi: 10.1016/j.bbabio.2006.11.019

Nagy, G., Ünnep, R., Zsiros, O., Tokutsu, R., Takizawa, K., Porcar, L., et al. (2014). Chloroplast remodelling during state transitions in chlamydomonas reinhardtii was revealed by noninvasive techniques *in vivo*. *Proc. Natl. Acad. Sci.* 111 (13), 5042–5047. doi: 10.1073/pnas.1322494111

Nama, S., Madireddi, S. K., Yadav, R. M., and Subramanyam, R. (2019). Non-photochemical quenching-dependent acclimation and thylakoid organization of chlamydomonas reinhardtii to high light stress. *Photosynthesis Res.* 139, 387–400. doi: 10.1007/s11120-018-0551-7

Niyogi, K. K., Shih, C., Soon Chow, W., Pogson, B. J., Della Penna, D., and Björkman, O. (2001). Photoprotection in a zeaxanthin-and lutein-deficient double mutant of arabidopsis. *Photosynthesis Res.* 67 (1), 139–145. doi: 10.1023/A:1010661102365

Ozawa, S. I., Bald, T., Onishi, T., Xue, H., Matsumura, T., Kubo, R., et al. (2018). Configuration of ten light-harvesting chlorophyll a/b complex I subunits in chlamydomonas reinhardtii photosystem I. *Plant Physiol.* 178 (2), 583–595. doi: 10.1104/pp.18.00749

Peers, G., Truong, T. B., Ostendorf, E., Busch, A., Elrad, D., Grossman, A. R., et al. (2009). An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. *Nature* 462 (7272), 518–521. doi: 10.1038/nature08587

Postnikova, O. A., Shao, J., and Nemchinov, L. G. (2013). Analysis of the alfalfa root transcriptome in response to salinity stress. *Plant Cell Physiol.* 54 (7), 1041–1055. doi: 10.1093/pcp/pct056

Roach, T., Na, C. S., Stöggl, W., and Krieger-Liszkay, A. (2020). The non-photochemical quenching protein LHCSR3 prevents oxygen-dependent photoinhibition in chlamydomonas reinhardtii. *J. Exp. Bot.* 71 (9), 2650–2660. doi: 10.1093/jxb/eraa022

Rochaix, J. D., Lemeille, S., Shapiguzov, A., Samol, I., Fucile, G., Willig, A., et al. (2012). Protein kinases and phosphatases involved in the acclimation of the photosynthetic apparatus to a changing light environment. *Philos. Trans. R. Soc. B: Biol. Sci.* 367 (1608), 3466–3474. doi: 10.1098/rstb.2012.0064

Ruban, A. V. (2016). Nonphotochemical chlorophyll fluorescence quenching: mechanism and effectiveness in protecting plants from photodamage. *Plant Physiol.* 170 (4), 1903–1916. doi: 10.1104/pp.15.01935

Rutherford, A. W., Osyczka, A., and Rappaport, F. (2012). Back-reactions, short-circuits, leaks and other energy wasteful reactions in biological electron transfer: redox tuning to survive life in O2. *FEBS Lett.* 586 (5), 603–616. doi: 10.1016/j.febslet.2011.12.039

Scheibe, R. (2004). Malate valves to balance cellular energy supply. *Physiologia plantarum* 120 (1), 21–26. doi: 10.1111/j.0031-9317.2004.0222.x

Schneider, D., Berry, S., Rich, P., Seidler, A., and Roügner, M. (2001). A regulatory role of the PetM subunit in a cyanobacterial cytochrome b6 f complex. *J. Biol. Chem.* 276 (20), 16780–16785. doi: 10.1074/jbc.M009503200

Schreiber, U., and Vidaver, W. (1974). Chlorophyll fluorescence induction in anaerobic scenedesmus obliquus. *Biochim. Biophys. Acta (BBA)-Bioenergetics* 368 (1), 97–112. doi: 10.1016/0005-2728(74)90100-5

Shen, L., Huang, Z., Chang, S., Wang, W., Wang, J., Kuang, T., et al. (2019). Structure of a C2S2M2N2-type PSII–LHCII super-complex from the green alga chlamydomonas reinhardtii. *Proc. Natl. Acad. Sci.* 116 (42), 21246–21255. doi: 10.1073/pnas.1912462116

Tian, L., Nawrocki, W. J., Liu, X., Polukhina, I., Van Stokkum, I. H., and Croce, R. (2019). pH dependence, kinetics and light-harvesting regulation of nonphotochemical quenching in chlamydomonas. *Proc. Natl. Acad. Sci.* 116 (17), 8320–8325. doi: 10.1073/pnas.1817796116

Tikkanen, M., and Aro, E. M. (2014). Integrative regulatory network of plant thylakoid energy transduction. *Trends Plant Sci.* 19 (1), 10–17. doi: 10.1016/j.tplants.2013.09.003

Tóth, T. N., Rai, N., Solymosi, K., Zsiros, O., Schröder, W. P., Garab, G., et al. (2016). Fingerprinting the macro-organisation of pigment-protein complexes in plant thylakoid membranes *in vivo* by circular-dichroism spectroscopy. *Biochim. Biophys. Acta (BBA)-Bioenergetics* 1857 (9), 1479–1489. doi: 10.1016/j.bbabio.2016.04.287

Trebitsh, T., and Danon, A. (2001). Translating chloroplast psbA mRNA is regulated by signals initiated by both photosystems II and I. *Proc. Natl. Acad. Sci.* 98 (21), 12289–12294. doi: 10.1073/pnas.211440698

Tyystjärvi, E. (2013). Photoinhibition of photosystem II. *Int. Rev. Cell Mol. Biol.* 300, 243–303. doi: 10.1016/B978-0-12-405210-9.00007-2

Vainonen, J. P., Hansson, M., and Vener, A. V. (2005). STN8 protein kinase in arabidopsis thaliana is specific in phosphorylation of photosystem II core proteins. *J. Biol. Chem.* 280 (39), 33679–33686. doi: 10.1074/jbc.M505729200

Vogel, M. O., Moore, M., König, K., Pecher, P., Alsharafa, K., Lee, J., et al. (2014). Fast retrograde signaling in response to high light involves metabolite export, MITOGEN-ACTIVATED PROTEIN KINASE6, and AP2/ERF transcription factors in arabidopsis. *Plant Cell* 26 (3), 1151–1165. doi: 10.1105/tpc.113.121061

Wietrzynski, W., Schaffer, M., Tegunov, D., Albert, S., Kanazawa, A., Plitzko, J. M., et al. (2020). Charting the native architecture of chlamydomonas thylakoid membranes with single-molecule precision. *Elife* 9, e53740. doi: 10.7554/eLife.53740.sa2

Wingler, A. (2002). The function of trehalose biosynthesis in plants. *Phytochemistry* 60 (5), 437–440. doi: 10.1016/S0031-9422(02)00137-1

Wingler, A., Lea, P. J., Quick, W. P., and Leegood, R. C. (2000). Photorespiration: metabolic pathways and their role in stress protection. *Philos. Trans. R. Soc. London. Ser. B: Biol. Sci.* 355 (1402), 1517–1529.

Xue, H., Tokutsu, R., Bergner, S. V., Scholz, M., Minagawa, J., and Hippler, M. (2015). PHOTOSYSTEM II SUBUNIT r is required for efficient binding of LIGHT-HARVESTING COMPLEX STRESS-RELATED PROTEIN3 to photosystem II-light-harvesting supercomplexes in. *Chlamydomonas reinhardtii*. *Plant Physiol.* 167 (4), 1566–1578. doi: 10.1104/pp.15.00094

Yadav, R. M., Aslam, S. M., Madireddi, S. K., Chouhan, N., and Subramanyam, R. (2020). Role of cyclic electron transport mutations pgrl1 and pgr5 in acclimation process to high light in chlamydomonas reinhardtii. *Photosynthesis Res.* 146 (1), 247–258. doi: 10.1007/s11120-020-00751-w

Yadavalli, V., Malleda, C., and Subramanyam, R. (2011a). Protein-protein interactions by molecular modelling and biochemical characterization of PSI-LHCI supercomplexes from chlamydomonas reinhardtii. *Mol. Biosyst.* 7 (11), 3143–3151. doi: 10.1039/c1mb05218g

Yamori, W. (2016). Photosynthetic response to fluctuating environments and photoprotective strategies under abiotic stress. *J. Plant Res.* 129 (3), 379–395. doi: 10.1007/s10265-016-0816-1

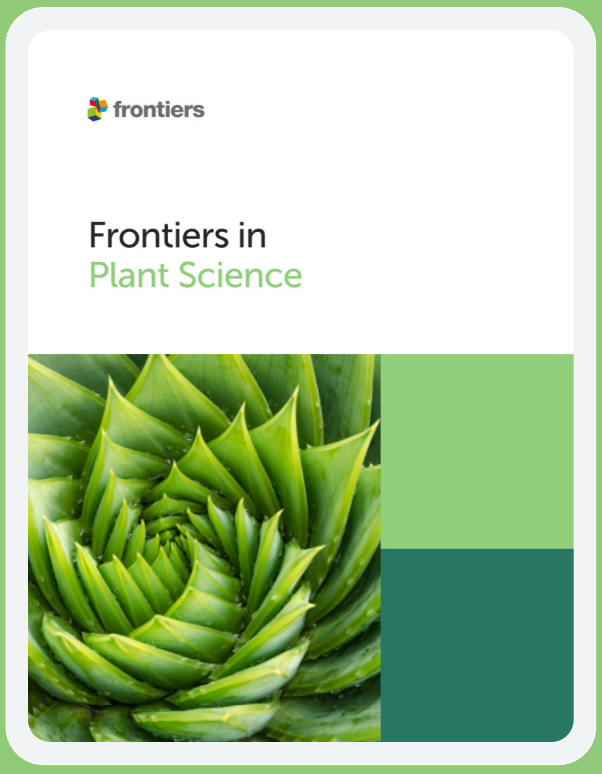
Frontiers in
Plant Science

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