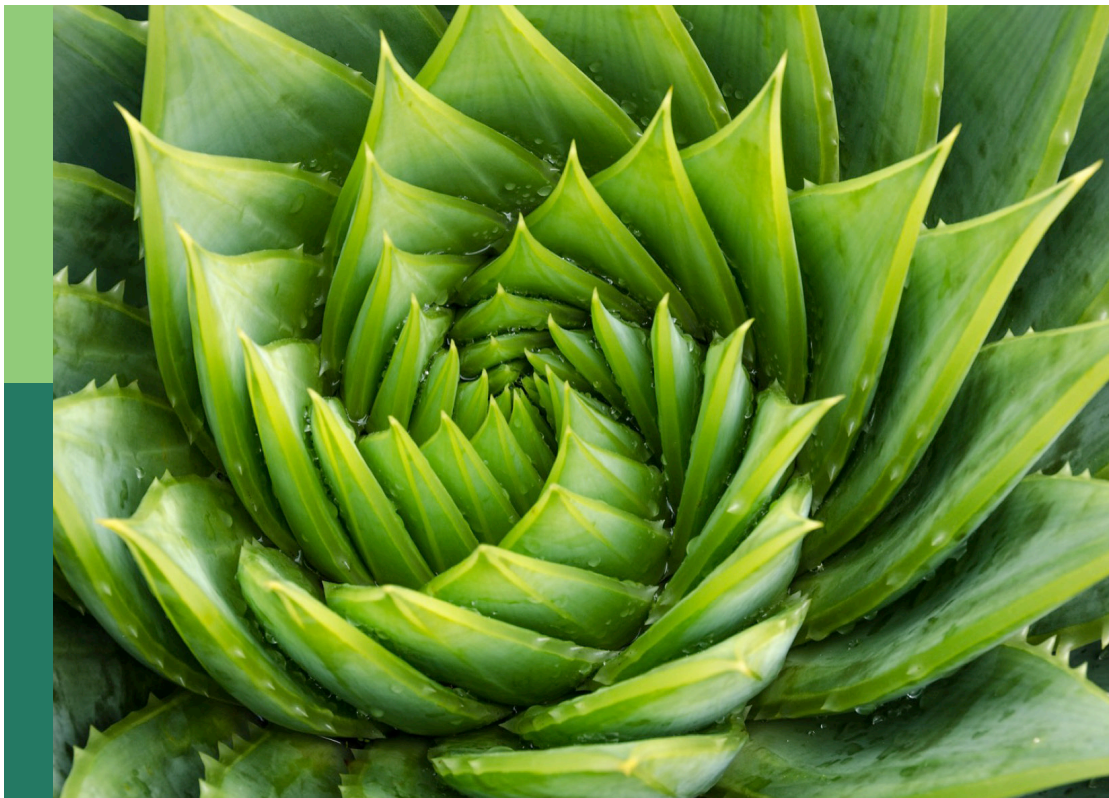
frontiers | Research Topics

Omics-driven crop improvement for stress tolerance, volume II

Edited by
Weicong Qi, Jian Chen, Yi Han and Zhen Li

Coordinated by
Freddy Kuok San Yeo and Xiaofeng Su

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-5405-0
DOI 10.3389/978-2-8325-5405-0

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

Omics-driven crop improvement for stress tolerance, volume II

Topic editors
Weicong Qi — Jiangsu Academy of Agricultural Sciences (JAAS), China
Jian Chen — Jiangsu University, China
Yi Han — Anhui Agricultural University, China
Zhen Li — Ghent University, Belgium

Topic coordinators
Freddy Kuok San Yeo — University of Malaysia Sarawak, Malaysia
Xiaofeng Su — Institute of Biotechnology, Chinese Academy of Agricultural Sciences, China

Citation
Qi, W., Chen, J., Han, Y., Li, Z., Yeo, F. K. S., Su, X., eds. (2024). *Omics-driven crop improvement for stress tolerance, volume II*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-5405-0

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 26 August 2024
DOI 10.3389/fpls.2024.1461217

OPEN ACCESS

EDITED AND REVIEWED BY
Nunzio D'Agostino,
University of Naples Federico II, Italy

*CORRESPONDENCE
Yi Han
yi.han@ahau.edu.cn

RECEIVED 08 July 2024
ACCEPTED 07 August 2024
PUBLISHED 26 August 2024

CITATION
Han Y, Qi W, Chen J, Li Z, Su X and Yeo FKS (2024) Editorial: Omics-driven crop improvement for stress tolerance, Volume II. *Front. Plant Sci.* 15:1461217. doi: 10.3389/fpls.2024.1461217

Editorial: Omics-driven crop improvement for stress tolerance, Volume II

Yi Han[1]*, Weicong Qi[2], Jian Chen[3], Zhen Li[4], Xiaofeng Su[5] and Freddy Kuok San Yeo[6]

[1]National Engineering Laboratory of Crop Stress Resistance Breeding, School of Life Sciences, Anhui Agricultural University, Hefei, China, [2]Excellence and Innovation Center, Jiangsu Academy of Agricultural Sciences, Nanjing, China, [3]International Genome Center, Jiangsu University, Zhenjiang, China, [4]Vlaams Instituut voor Biotechnologie (VIB) Center for Plant Systems Biology, Department of Plant Biotechnology and Bioinformatics, Ghent University, Gent, Belgium, [5]Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing, China, [6]Faculty of Resource Science and Technology, University of Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia

KEYWORDS

omics, crops, abiotic and biotic stresses, breeding, stress tolerance

Editorial on the Research Topic
Omics-driven crop improvement for stress tolerance, Volume II

Crops are vulnerable to biotic and abiotic stresses that lead to reduced yields. Biotic stresses, such as fungi and pests, cause crops to rot and develop diseases. In contrast, abiotic stresses—such as high temperatures, salinity and mineral toxicity, and water shortages—irreversibly affect crops at different developmental stages, such as flowering, grain filling, and maturation through signal transduction, gene expression, and protein modifications. An increase of 1°C in the global average temperature is projected to significantly reduce crop yields. Excessive soil salinity tends to inhibit plant growth, hinder photosynthesis, and require metabolic adjustments. However, some crop types and species can tolerate modest levels of salinity without affecting their growth and yield. Drought is one of the most damaging abiotic stresses affecting severely the productivity of cereal crops. Rice struggles to survive in water-deficient fields, while maize is highly susceptible to drought. Therefore, improving crop stress tolerance is crucial for yield stability and healthy growth.

Advancements in omics technologies, such as genomics, transcriptomics, proteomics, and metabolomics, have significantly increased the feasibility and depth of biological research. These technologies offer a more comprehensive and holistic view of crops from different perspectives, which is critical for breeders. Additionally, the application of multiple omics effectively addresses bottlenecks that single omics technologies struggle to overcome. The advent of high-throughput, large-scale, and highly sensitive sequencing technologies, such as next-generation sequencing (NGS), has made genome sequencing more affordable and accessible. The integrated approach of genome-wide association studies (GWAS) and NGS technologies has provided significant opportunities for predicting stress-related genes. Transcriptomics focuses on the study of gene transcription and transcriptional regulation in a specific tissue or cell at specific developmental stages or conditions. Conversely, proteomics is the study of the composition and changes in protein content within cells, tissues, or organisms at various stages. These two fields reflect gene expression in the organism at two different molecular levels and are

useful for gaining a deeper understanding of the intrinsic relationship between proteins and gene expression. They reveal gene expression and post-transcriptional regulatory states, providing more comprehensive information on the expression profile of the organism. Metabolomics, which is emerging after genomics and proteomics, significantly complements the understanding of cellular signaling, energy transfer, and intercellular communication after gene transcription and protein modification under genetic and environmental influences.

Currently, omics approaches play a crucial role in understanding crop stress resistance, disease resistance mechanisms, related-gene function research, germplasm resource identification, and assisted breeding. Recently, the application of omics strategies in crop stress tolerance research has significantly contributed to the advancement of the third "Green Revolution" in agriculture. Combining multiple omics approaches can reveal the molecular mechanisms underlying crop responses to stress and provide new strategies for developing stress-resistant varieties. Researchers have made significant progress in improving crop stress resistance using omics tools. Examples include enhancing stripe rust resistance in wheat (Lai et al.), necrotrophic bacterial disease resistance in solanum (Joshi et al.), drought tolerance in sugarcane (Li et al.), salinity stress resistance in cotton (Chen et al.), heat tolerance in grapes (Wu et al.), the response to grazing stress in *Phaeodactylum tricornutum* (Liu et al.) and antioxidant ability in postharvest crops (Wang et al.).

Stripe rust, caused by the fungus *Puccinia striiformis f.sp. tritici (Pst)* is a major threat to global wheat production. Lai et al. phenotyped the seedlings of 137 spring and 149 winter wheat varieties from Xinjiang, China, for resistance to six races of stripe rust. Subsequently, combining molecular markers, 10 stripe rust resistance genes that may correlate with the differences in resistance among wheat varieties were identified.

Pectobacterium species cause severe diseases, leading to rotting and wilting in cultivated potatoes, while wild potato relatives exhibit strong resistance. Studies have demonstrated that phytochemicals play a critical role in antivirulence traits. Therefore, Joshi et al. analyzed the metabolites in a population derived from a hybrid of *S. tuberosum* and *S. chacoense*. Approximately 30 metabolites, including alkaloids, terpenes, and several prenylated compounds, were found to be associated with resistance. Additionally, a QTL analysis was conducted to explore the genetic basis of antivirulence traits using recombinant inbred lines. Genetic mapping identified five QTLs associated with the inhibition of quorum sensing and two QTLs associated with the inhibition of protease activity. The metabolomics and QTL results demonstrated that quorum sensing inhibition and exo-protease inhibition could be efficient strategies for breeders to improve potato resistance to necrotrophic bacterial pathogens.

To identify drought resistance regulator genes and aid breeders in improving sugarcane (*Saccharum* spp.), Li et al. conducted a comparative transcriptomic study of two sugarcane cultivars with significantly different drought tolerance. The transcriptome results indicated that differential gene expression in various metabolic pathways, especially the photosynthesis pathway, may be the primary cause of the different drought tolerances in sugarcane varieties.

Cotton is vulnerable to abiotic stresses that often affect the production of fiber and cottonseed oil. A-galactosidases (AGALs), encoded by raffinose family oligosaccharide (RFO) catabolic genes, have been reported to be involved in plant stress tolerance. Chen et al. sequenced the whole genomes of *Gossypium hirsutum*, *Gossypium arboreum*, *Gossypium barbadense*, and *Gossypium raimondii*. Based on the sequencing results and genetic evidence, the *GhAGAL3* gene was shown to contribute to improved cotton tolerance to salinity stress by increasing the raffinose content.

With the increasing frequency of extreme weather events, high temperatures pose an unavoidable environmental threat to crops such as grapes, which are essential for fruit consumption and wine production. Wu et al. evaluated the heat tolerance of four grape varieties. Subsequently, RNA-seq was conducted on the four grape varieties under different temperature conditions. Based on bioinformatics analysis and differential expression comparisons, coexpression networks were constructed, identifying six genes associated with grape heat tolerance.

The quality of harvested crops is the basis for gaining economic benefits. Enzymatic browning reactions, triggered by oxidative stress, significantly compromise the quality of harvested crops. Therefore, a better understanding of the mechanism underlying browning is essential for the development of anti-browning treatments. Wang et al. unraveled enzymatic browning in crops by integrating omics strategies, including transcriptomic, proteomic, and metabolomic methods.

In conclusion, these Research Topics offer new insights into the application of omics tools for crop improvement in stress tolerance. Multi-dimensional omics studies have generated extensive datasets, revealing various molecular, physiological, and metabolic pathways related to abiotic and biotic stress tolerance, thereby opening new horizons for future investigations. The complex genome structure, variant epigenetic regulation and post-transcriptional modification and unpredictable environmental condition are still big challenges for the effective applications of omics strategies. Much more sensitive and higher resolution sequencing techniques, detective tools, and more scientific analysis methods are needed to advance research on polyploid crops.

Author contributions

YH: Writing – original draft. WQ: Writing – review & editing. JC: Writing – review & editing. ZL: Writing – review & editing. XS: Writing – review & editing. FY: Writing – review & editing.

Conflict of interest

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

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

Publisher's note

frontiers | Frontiers in Plant Science

TYPE Opinion
PUBLISHED 31 May 2023
DOI 10.3389/fpls.2023.1208111

OPEN ACCESS

EDITED BY
Yi Han,
Anhui Agricultural University, China

REVIEWED BY
Xi Chen,
Jiangsu Vocational College of Agriculture and Forestry, China
Liu Guoming,
Institute of Soil Science, Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE
Haoming Chen
chenhaoming89@hotmail.com

RECEIVED 18 April 2023
ACCEPTED 09 May 2023
PUBLISHED 31 May 2023

CITATION
Jiang J, Hu X, Ji X and Chen H (2023) High throughput sequencing technology facility research of genomic modification crop cultivation influencing soil microbe. *Front. Plant Sci.* 14:1208111. doi: 10.3389/fpls.2023.1208111

High throughput sequencing technology facility research of genomic modification crop cultivation influencing soil microbe

Jinyan Jiang, Xin Hu, Xincheng Ji and Haoming Chen*

School of Environmental and Biological Engineering, Nanjing University of Science and Technology, Nanjing, China

KEYWORDS

genetic modified crop, soil microbe, high throughput sequencing technology, microorganisms, crop

Introduction: the risk in the GM crop cultivation influencing soil microbe and new generation sequencing technologies

The safety of transgenic crops is a hot spot concerned deeply worldwide. Although transgenic crops have great advantages improving agricultural production, their potential impacts on soil safety and soil microorganisms need to be further studied. As an important place for material circulation and energy transformation in the ecosystem, soil is sensitive to changes in the external environment, and soil microbes has been recognized as one of the early warning indicators of ecosystem changes. Soil microorganisms are mainly composed of bacteria, fungi and actinomycetes. Last decades, with the development of modern biotechnology, the main research methods of soil microbial community include microbial plate culture method, Biolog microplate method, biomarker method and the emerging new technologies of high-throughput sequencing.

Technologies monitoring soil microbe and its transmission: from the classic tech to the NGS

Molecular detection technology is highly sensitive to changes in microbial community structure and can detect even small differences in bacterial composition. It has played a certain advantage in exploring microbial community diversity, particularly when identifying new microbial species. Classical methods include fluorescence *in situ* hybridization (FISH), DNA microarray, single strand conformation polymorphism (SSCP), restriction fragment length polymorphism (RFLP), amplified ribosomal DNA

restriction analysis (ARDRA), amplified fragment length polymorphism (AFLP), denaturation/temperature gradient gel electrophoresis (DGGE/TGGE), etc. Among them, DGGE technology is the most widely used method in the study of soil microbial community. These molecular biological methods have the advantages of high sensitivity and no need to cultivate microorganisms, which can better reflect the real situation of soil microorganisms. However, the disadvantages are also obvious. For example, the microbial communities that can be detected in DGGE are usually colonies with significant dominance, and the same band may be shared by multiple microbial species (Ansorge, 2009).

With the rapid development of molecular biology research, especially the wide application of next-generation sequencing (NGS) technology, the research scope and depth of soil microbiology have greatly improved (Nielsen and Wall, 2013; Ranjard et al., 2013). High throughput sequencing technologies mainly include 454 pyrophosphate, ABI SOLiD, Illumina Hiseq2000 and Illumina Miseq, Hecos Heliscope and PacBioSequel sequencing platforms. Up to date, these high-throughput sequencing technologies have gradually replaced DGGE technology, especially the third-generation sequencing technology characterized by single molecule and long reading length, and gradually become the main analytical methods of microbial community research (Fahner et al., 2016). For example, PacBio 16S amplification sequencing technology firstly corrects, filters and removes the chimeric sequences from the amplification and sequencing results (raw reads) of bacterial 16S rRNA genes to obtain valid data (clean reads), then bioinformatic analysis clusters the clean reads into OTUs, and annotates them with species to obtain taxonomic information. Next, Qiime and other software were used to calculate the observed OTUs, Chao1, Shannon, Simpson and other alpha diversity indexes, calculate the Unifrac distance, construct PCoA and UPGMA clustering trees, and analyze the differences between groups. Finally, Anosim, MRPP, ADONIS, t-test and other statistical analysis methods were used to test the difference significance of species composition and community structure among groups.

Macro genome sequencing is new research strategy based on NGS. It takes the entire microbial community in a specific habitat as the research object, and directly extract environmental sample DNA, and obtain the total amount of environmental microbial gene information via high throughput sequencing technology (Tringe et al., 2005). Unlike other high-throughput-sequencing strategy, clean reads need to be assembled by Metagenome to annotate common functional databases, resistance gene databases, and a series of advanced information analysis (such as CCA/RDA analysis, copy number variation analysis, etc.). At the same time, it can also be combined with environmental factors, pathological indicators or special phenotypes for in-depth correlation analysis. Macro genome sequencing technology can help avoid experimental errors due to microbial differences caused by the environment. This technology provides a comprehensive microbial profile in the samples, including their composition and interaction. The resulting data can be used for more in-depth research, such as analyzing metabolic pathways and gene function at the molecular level (Mende et al., 2012). Macro transcriptome sequencing refers to the study of the transcription and regulation of all genomes in a specific environment and period of life groups at all levels. Compared with macro genome sequencing, it takes all RNA in the ecological environment as the research object to study the changes of complex microbial communities at the transcription level, which is another strategy to tap potential new genes (Vieites et al., 2009). All of the cases mentioned above, and the others were listed in Table 1.

The bibliometric studies with the keywords "high-throughput sequencing technology", "microorganisms" and "crops" provide further evidence of the trend towards high-throughput sequencing technology, and the retrieval formula was TS= (“microbial” OR “microorganism*” OR “microbe*” OR “microbial community” OR “bacteria” OR “bacterial” OR “fungus” OR “fungi”) AND TS= (“NGS” OR “next generation sequencing” OR “high throughput sequencing technologies” OR “high throughput” OR “sequencing” OR “454 pyrophosphate” OR “ABI SOLiD” OR “Illumina Hiseq2000” OR “Illumina Miseq” OR “Hecos Heliscope” OR “PacBioSequel”) AND

TABLE 1 Apply high throughput sequencing technologies to analyze the microbe in GM crop cultivation field.

Crop	Exogenous gene	References
Maize	*Bt*	(Zeng et al., 2014)
Maize	*CrylAb*	(Ondreičková et al., 2014)
Maize	*Bt*	(Cheeke et al., 2015)
Cotton	*CrylAb/lAc*	(Xie et al., 2017)
Maize	*Bt*	(Zeng et al., 2019)
Maize	*Bt*	(van Wyk et al., 2017)
Rapeseed	*Glufosinate tolerant*	(Tang et al., 2019)
Soybean	*Glyphosate resistant*	(Girgan et al., 2020)
Soybean	*EPSPS*	(Lu et al., 2018)
Rice	*Bar*	(He et al., 2019)
Maize	*EPSPS*	(Wen et al., 2019)

TS= ("crop*"). Research exploring the mechanisms of crop-microbe interactions using high-throughput sequencing technologies has generally shown a steady growth trend over the past 22 years (2000.1.1-2022.11.30, 3411 research papers, reviews, conference papers, etc. from Web of Science). The total number of citations up to 77776 times also shows that this research direction has received high attention (Figure 1A). Meanwhile, this area of research has piqued the interest of scholars across a range of fields, with the highest interest in microbiology (22%), plant science (17%), soil science (12%) and biotechnology and applied microbiology (12%). The Vosviewer software analysis yielded a co-occurrence network and clustering distribution of the most popular keywords in the field (40 occurrences, 136 nodes) (Figure 1B).

The keyword (K) clustering atlas (LLR algorithm: log likelihood ratio, purple and yellow representing the first and latest time respectively, g-index K =15) found that High-throughput technology has gradually become the main means of studying soil microorganisms and plants, and has been widely used in crop variety improvement, plant stress resistance, disease and pest resistance, toxicity resistance, biological gene transformation and other aspects (Figure 1C). Moreover, the high-rate appearance of keywords in recent years indicated that transgenic crops affecting soil enzyme activity (e.g., cellobiose dehydrogenase, glucose-hydrogenase, cellobiose dehydrogenase, etc.), risk assessment and the study of soil microbial community structure and diversity have become hot spots between GM crops and microorganisms (Figure 1D).

High throughput sequencing technologies facilitating soil microbe identification in the land of GM crop cultivation

Insect resistant transgenic crops

Insect resistant genes that have been transferred into crops including: Bacillus thuringiensis toxin protein gene, protease inhibitor gene, amylase inhibitor gene, foreign lectin gene, ribosome inactivating protein gene and pea lipoxygenase gene. At present, the *Bt* and *CpT1* gene are the most popular in the research of evaluating the impact of insect resistant transgenic crops on non-target soil microorganisms (YJ et al., 2014). The high-throughput sequencing results of Liang et al. showed that there were no significant differences in rhizosphere soil bacterial richness and diversity index and community structure between *Cry1Ie* transgenic maize and non-transgenic maize, while principal component analysis (PCA) showed that there were significant differences between

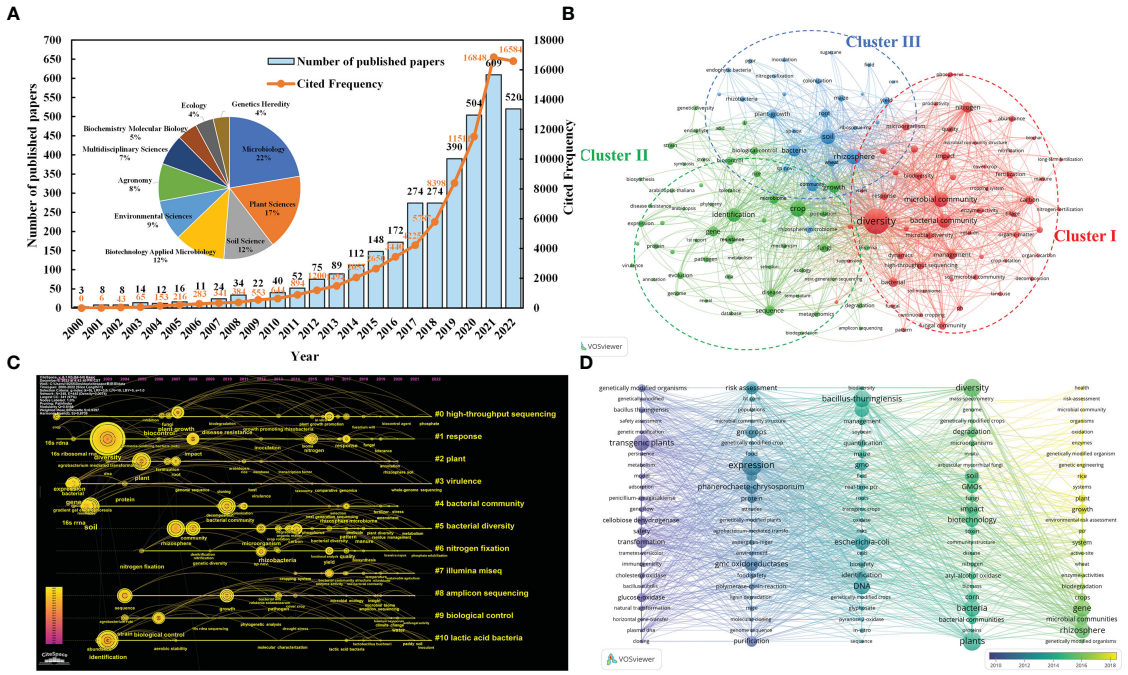


FIGURE 1
(A) Publications, citations performance and the top ten research fields in terms of number of documents. **(B)** Keywords co-occurring network. The main keywords of the three clusters can be divided into the research between soil ecological environment and microbial community/diversity (Cluster I red), the research between microorganisms and crop disease control based on gene level (Cluster II green), and the research between plant root soil microbial characteristics and plant growth (Cluster III blue). The key words "microbial communities" (425), "crop crop" (450) and "high-throughput sequencing" (151) are highly correlated. It can be seen that high-throughput sequencing technology has occupied a place in the study of the interaction mechanism between microorganisms and crops. **(C)** keyword clustering atlas of "High throughput sequencing technology", "microorganisms" and "crops". #0 (high-throughput sequencing) is closely connected with # 1 (response), # 2 (plant), # 3 (environment), and # 4 (bacteria community) respectively. Illumina MiSeq (# 7) and cluster # 8 (amplicon sequencing) are commonly used in high throughput technology. **(D)** keywords timeline visualization of "Microorganisms" and "Genetically Modified Crops" from WOS during 2000.01–2022.11. From 2014 to 2018, "GMOs", "expression", "plants", "diversity", "bacteria", "gene", "transgenic plants", "rhizosphere", "DNA" and "resistance" appeared frequently.

seedling, flowering and maturity stages or between different test years (Jingang et al., 2018). The bacterial species diversity and community structure in rhizosphere soil of Bt maize and non-Bt maize were studied by Illumina MiSeq sequencing technology (van Wyk et al., 2017). The results showed that there were significant differences in bacterial alpha diversity and community composition between the two groups. Proteobacteria and Acidobacteria were the dominant bacteria in Bt maize rhizosphere soil, while actinobacteria were the dominant bacteria in non-Bt maize rhizosphere soil. Further study showed that there were significant differences in the contents of organic carbon and nitrate and the activities of acid phosphatase and β-glucosidase in rhizosphere soil between Bt maize and control. These differences in soil physical and chemical properties and related enzyme activities may be the main reasons for the differences in bacterial community structure.

Herbicide-resistant genetically modified crops

Herbicide-resistant GM crops are the most widely grown crops in the world, even more than insect-resistant GM crops containing Bt genes (Kwon and Kim, 2001). Glyphosate and glyphosate resistance genes are the most commonly used genes in transgenic crops with herbicide resistance. Among them, there are 7 anti-glyphosate genes, which are *cp4 epsps*, *gat4601*, *goxv247*, *epsps*, *epsps grg23ace5*, *mepsps*, *2mepsps*, and anti-glyphosate genes *bar* and *pat*. Similar to the results of studies on insect-resistant GM crops, there are also two kinds of results in studies evaluating the effects of herbicide-resistant GM crops on soil microbial ecosystems. Most of the results showed that the introduction of herbicide resistance genes into crops had no specific effects on the soil microbial community, but was mainly influenced by the seasons and plant development. Tang et al. adopted 16S rRNA gene high-throughput sequencing technology and found that phosphonate resistant rapeseed had no significant effect on the alpha diversity of the rhizosphere bacterial community, and the dominant bacteria with high relative abundance were *Proteobacteria*, *Bacteroides*, *Acidobacteria*, *Blastomonas* and *Actinomycetes* (Tang et al., 2019). These dominant microbes were highly present in transgenic and non-transgenic recipient rapeseed treated with phosphate-oxyphosate, respectively, but there were significant differences in different growth stages (seedling, bolting, flowering and maturity). Moreover, some studies have shown that over a long period of time, some herbicide-resistant transgenic crops can affect the relative abundance of some microorganisms in soil, thus changing the composition of microbial community. Babujia et al. conducted 10 years of field monitoring on glyphosate-resistant transgenic soybeans (Babujia et al., 2016). Although the introduction of glyphosate-resistant genes did not affect the yield of soybeans, metagenomic sequencing analysis found that actinomycta and acidobacteria were abundant in the bacterial community of conventional soybean rhizosphere soil. The relative abundance of Proteus, Firmicutes and green algae was higher in transgenic soybean, and the abundance of sequences related to protein metabolism and cell division cycle were also higher. He et al. conducted Illumina MiSeq sequencing on the 16S rRNA gene of bacteria at maturity stage and found that there was no significant difference in soil bacterial diversity index between BAR-gene transgenic rice and conventional rice; but there were significant differences in the abundance of individual genera (such as *Cerococcus*) between transgenic japonica rice and conventional rice (He et al., 2019). The content of anaerobic bacteria in soil of transgenic japonica rice was significantly higher than that of conventional rice.

Prospect

Soil microorganisms are crucial to the soil ecosystem and agricultural production. The introduction of foreign genes may affect plant metabolism and root secretion production, thus changing soil microecological environment. Therefore, we need to carry out in-depth risk assessment of GM crops on soil ecosystems. The core of soil ecosystem security risks of GM crops is to assess whether the cultivation of GM crops will change the function of soil ecosystems. So far, a large number of studies have evaluated the potential impact of transgenic crops on the presence of soil microorganisms. However, due to the different objects (including exogenous genes and plant species), environment and experimental methods in these studies, the obtained data are analyzed in different ways. In addition, most of the research results are based on the cultivation period of 1-2 years, so the results are bound to show some differences. With the development of molecular breeding, the development of transgenic crops with multiple trait overlays has not only made research more complex, but also posed new challenges for environmental safety assessment and research. Due to the above limitations of the current research, we should pay more attention to the correlation between the composition and function of soil microbial community structure and its response to the natural changes of soil system (such as season, climate, crop rotation, heavy metal pollution and pesticide use) in the future, so as to assess the impact of GM crops on soil microorganisms more comprehensively and accurately.

Author contributions

JJ, HC, XH, and XJ conceived and wrote the paper. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of China (NO. 42007105) to HC.

Conflict of interest

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

Publisher's note

References

Ansorge, W. J. (2009). Next-generation DNA sequencing techniques. *N. Biotechnol.* 25, 195–203. doi: 10.1016/j.nbt.2008.12.009

Babujia, L. C., Silva, A. P., Nakatani, A. S., Cantão, M. E., Vasconcelos, A. T. R., Visentainer, J. V., et al. (2016). Impact of long-term cropping of glyphosate-resistant transgenic soybean [Glycine max (L.) merr.] on soil microbiome. *Transgenic Res.* 25, 425–440. doi: 10.1007/s11248-016-9938-4

Cheeke, T. E., Schütte, U. M., Hemmerich, C. M., Cruzan, M. B., Rosenstiel, T. N., and Bever, J. D. (2015). Spatial soil heterogeneity has a greater effect on symbiotic arbuscular mycorrhizal fungal communities and plant growth than genetic modification with *Bacillus thuringiensis* toxin genes. *Mol. Ecol.* 24, 2580–2593. doi: 10.1111/mec.13178

Fahner, N. A., Shokralla, S., Baird, D. J., and Hajibabaei, M. (2016). Large-Scale monitoring of plants through environmental DNA metabarcoding of soil: recovery, resolution, and annotation of four DNA markers. *PloS One* 11, e0157505. doi: 10.1371/journal.pone.0157505

Girgan, C., Claassens, S., and Fourie, H. (2020). Nematode assemblages and soil microbial communities in soils associated with glyphosate-resistant soybean. *South Afr. J. Plant Soil* 37, 11–22. doi: 10.1080/02571862.2019.1640297

He, M., Zhang, J., Shen, L., Xu, L., Luo, W., Li, D., et al. (2019). High-throughput sequencing analysis of microbial community diversity in response to indica and japonica bar-transgenic rice paddy soils. *PloS One* 14, e0222191. doi: 10.1371/journal.pone.0222191

Jingang, L., Ying, L., Yue, J., Longtao, X., Xinyuan, S., Xiaobo, Z., et al. (2018). No significant differences in rhizosphere bacterial communities between bt maize cultivar IE09S034 and the near-isogenic non-bt cultivar Zong31. *Plant Soil Environ.* 64, 427–434. doi: 10.17221/260/2018-PSE

Kwon, Y. W., and Kim, D.-S. (2001). Herbicide-resistant genetically-modified crop: its risks with an emphasis on gene flow. *Weed Biol. Manage.* 1, 42–52. doi: 10.1046/j.1445-6664.2001.00001.x

Lu, G.-H., Tang, C.-Y., Hua, X.-M., Cheng, J., Wang, G.-H., Zhu, Y.-L., et al. (2018). Effects of an EPSPS-transgenic soybean line ZUTS31 on root-associated bacterial communities during field growth. *PloS One* 13, e0192008. doi: 10.1371/journal.pone.0192008

Mende, D. R., Waller, A. S., Sunagawa, S., Järvelin, A. I., Chan, M. M., Arumugam, M., et al. (2012). Assessment of metagenomic assembly using simulated next generation sequencing data. *PloS One* 7, e31386. doi: 10.1371/journal.pone.0031386

Nielsen, U. N., and Wall, D. H. (2013). The future of soil invertebrate communities in polar regions: different climate change responses in the Arctic and Antarctic? *Ecol. Lett.* 16, 409–419. doi: 10.1111/ele.12058

Ondreičková, K., Mihálik, D., Ficek, A., Hudcovicová, M., Kraic, J., and Drahovská, H. (2014). Impact of genetically modified maize on the genetic diversity of rhizosphere bacteria: a two-year study in Slovakia. *Pol. J. Ecol.* 62, 67–76. doi: 10.3161/104.062.0107

Ranjard, L., Dequiedt, S., Chemidlin Prévost-Bouré, N., Thioulouse, J., Saby, N. P. A., Lelievre, M., et al. (2013). Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity. *Nat. Commun.* 4, 1434. doi: 10.1038/ncomms2431

Tang, T., Chen, G., Liu, F., Bu, C., Liu, L., and Zhao, X. (2019). Effects of transgenic glufosinate-tolerant rapeseed (Brassica napus l.) and the associated herbicide application on rhizospheric bacterial communities. *Physiol. Mol. Plant Pathol.* 106, 246–252. doi: 10.1016/j.pmpp.2019.03.004

Tringe, S. G., von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., et al. (2005). Comparative metagenomics of microbial communities. *Science (1979)* 308, 554–557. doi: 10.1126/science.1107851

van Wyk, D. A. B., Adeleke, R., Rhode, O. H. J., Bezuidenhout, C. C., and Mienie, C. (2017). Ecological guild and enzyme activities of rhizosphere soil microbial communities associated with bt-maize cultivation under field conditions in north West province of south Africa. *J. Basic Microbiol.* 57, 781–792. doi: 10.1002/jobm.201700043

Vieites, J. M., Guazzaroni, M.-E., Beloqui, A., Golyshin, P. N., and Ferrer, M. (2009). Metagenomics approaches in systems microbiology. *FEMS Microbiol. Rev.* 33, 236–255. doi: 10.1111/j.1574-6976.2008.00152.x

Wen, Z.-L., Yang, M.-K., Du, M.-H., Zhong, Z.-Z., Lu, Y.-T., Wang, G.-H., et al. (2019). Enrichments/Derichments of root-associated bacteria related to plant growth and nutrition caused by the growth of an EPSPS-transgenic maize line in the field. *Front. Microbiol.* 10. doi: 10.3389/fmicb.2019.01335

Xie, M., Zhang, Y., Peng, D., Li, Q., Hu, X., and Zhang, Z. (2017). No significant impact of transgenic Cry1Ab/1Ac cotton on rhizosphere-soil enzyme activities and bacterial communities. *Agron. J.* 109, 1271–1279. doi: 10.2134/agronj2016.10.0618

YJ, Z., Xie, M., and DL, P. (2014). Effects of the transgenic CrylAc and CpTI insect-resistant cotton SGK321 on rhizosphere soil microorganism populations in northern China. *Plant Soil Environ.* 60, 285–289. doi: 10.17221/192/2014-PSE

Zeng, H., Tan, F., Zhang, Y., Feng, Y., Shu, Y., and Wang, J. (2014). Effects of cultivation and return of bacillus thuringiensis (Bt) maize on the diversity of the arbuscular mycorrhizal community in soils and roots of subsequently cultivated conventional maize. *Soil Biol. Biochem.* 75, 254–263. doi: 10.1016/j.soilbio.2014.04.024

Zeng, H., Zhong, W., Tan, F., Shu, Y., Feng, Y., and Wang, J. (2019). The influence of bt maize cultivation on communities of arbuscular mycorrhizal fungi revealed by MiSeq sequencing. *Front. Microbiol.* 9. doi: 10.3389/fmicb.2018.03275

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 11 October 2023
DOI 10.3389/fpls.2023.1243664

OPEN ACCESS

EDITED BY
Weicong Qi,
Jiangsu Academy of Agricultural Sciences (JAAS), China

REVIEWED BY
Khushi Muhammad,
Hazara University, Pakistan
Xinlong Liu,
Yunnan Academy of Agricultural Sciences, China
Youxiong Que,
Fujian Agriculture and Forestry University, China

*CORRESPONDENCE
Xihui Liu
xihuiliu2006@126.com
Hui Zhou
zhouhui@gxaas.net

†These authors have contributed equally to this work

RECEIVED 21 June 2023
ACCEPTED 27 September 2023
PUBLISHED 11 October 2023

CITATION
Li H, Gui Y, Zhu K, Wei J, Zhang R, Yang R, Tang L, Zhou H and Liu X (2023) Comparative transcriptomic analyses of two sugarcane *Saccharum* L. cultivars differing in drought tolerance. *Front. Plant Sci.* 14:1243664. doi: 10.3389/fpls.2023.1243664

Comparative transcriptomic analyses of two sugarcane *Saccharum* L. cultivars differing in drought tolerance

Haibi Li[1,2,3†], Yiyun Gui[1,2†], Kai Zhu[1,2†], Jinju Wei[1,2], Ronghua Zhang[1,2], Rongzhong Yang[1,2], Liqiu Tang[3], Hui Zhou[1,2*] and Xihui Liu[1,2*]

[1]Sugarcane Research Center of Chinese Academy of Agricultural Sciences, Nanning, China, [2]Guangxi Key Laboratory of Sugarcane Genetic Improvement, Guangxi Academy of Agricultural Sciences, Nanning, China, [3]Guangxi South Subtropical Agricultural Science Research Institute, Guangxi Academy of Agricultural Sciences, Chongzuo, China

Sugarcane (*Saccharum* spp.) is an important cash crop, and drought is an important factors limiting its yield. To study the drought resistance mechanism of sugarcane, the transcriptomes of two sugarcane varieties with different levels of drought resistance were compared under different water shortage levels. The results showed that the transcriptomes of the two varieties were significantly different. The differentially expressed genes were enriched in starch and sucrose metabolism, linoleic acid metabolism, glycolysis/gluconeogenesis, and glyoxylate and dicarboxylate metabolic pathways. Unique trend genes of the variety with strong drought resistance (F172) were significantly enriched in photosynthesis, mitogen-activated protein kinases signaling pathway, biosynthesis of various plant secondary metabolites, and cyanoamino acid metabolism pathways. Weighted correlation network analysis indicated that the blue4 and plum1 modules correlated with drought conditions, whereas the tan and salmon4 modules correlated with variety. The unique trend genes expressed in F172 and mapped to the blue4 module were enriched in photosynthesis, purine metabolism, starch and sucrose metabolism, beta-alanine metabolism, photosynthesis-antenna proteins, and plant hormone signal transduction pathways. The expression of genes involved in the photosynthesis-antenna protein and photosynthesis pathways decreased in response to water deficit, indicating that reducing photosynthesis might be a means for sugarcane to respond to drought stress. The results of this study provide insights into drought resistance mechanisms in plants, and the related genes and metabolic pathways identified may be helpful for sugarcane breeding in the future.

KEYWORDS

sugarcane, drought, transcriptome, variety, metabolic pathways, photosynthesis

Introduction

Drought stress is a major factor limiting global agricultural production, and the development of drought-resistant crop varieties is of great significance in modern agriculture (Ozturk et al., 2021; Conti et al., 2023). The cultivation of drought-resistant crop varieties requires an understanding of the damage inflicted by drought and the mechanisms of crop drought resistance. Drought stress can affect the basic physiological activities of plants, such as enzymatic function, osmotic pressure, and energy supply, and inhibit mitosis and normal cell metabolism (Tardieu et al., 2018). In response, plants have evolved a series of mechanisms to overcome drought stress or drought-stress conditions, such as closing the stomata to reduce water loss from transpiration, regulating osmotic pressure, altering the expression of numerous genes, adjusting photosynthesis, modulating abscisic acid, and pigment levels, and altering sugar metabolism (Agurla et al., 2018; Conti et al., 2023).

Sugarcane (*Saccharum* spp.) is an economically important crop that can be used as food, feed, and fuel, and has strict water requirements for cultivation (Meena et al., 2020; Dinesh Babu et al., 2022). To adapt to water scarcity, sugarcane has evolved drought resistance mechanisms involving morphological and physiological responses, such as abscisic acid accumulation, ROS scavenging and antioxidant activity, lipid peroxidation and altered expression of certain genes (Ferreira et al., 2017). As the basis of life function, gene expression and its products play a central role in drought resistance of crops. For example, upon exposure to drought stress, dirigent proteins exhibit significant transcriptional responses and improve physiological and biochemical indices (Gentile et al., 2015). Studies have shown that miRNA-mediated post-transcriptional regulation plays an important role in drought resistance in sugarcane, particularly in regulating the production of transcription factors, transporters, senescence-related proteins, and proteins associated with flower development (Ferreira et al., 2017). The ScDREB2B-1 gene cloned from the *Saccharum* spp. hybrid ROC22 responds to drought stress by regulating the abscisic acid signaling pathway, ROS levels, and stress-related gene expression (Li et al., 2022). In addition, the expression of ShCBSD-PB1-5A and ShCBSD-PB1-7A-1 significantly decreased, whereas that of SsCBSDCBS-5A distinctly increased in ROC22 cells in response to drought stress (Gentile et al., 2013). Most of these studies have focused on one aspect of gene expression; however, to gain a comprehensive understanding of gene expression under water-deficit conditions, it is necessary to focus on the expression of all genes, and the rise of sequencing and transcriptomic technologies provides a technical means to solve this problem.

Since the publication of the whole-genome sequence of *Arabidopsis thaliana* in December 2000, research on crop plants has undergone significant advances, such as genome sequencing, and decoding of gene expression and function during development, and during the response to various environmental stimuli (Chen et al., 2022). With the development of sequencing and omics technologies, transcriptome analysis has been widely used to study the relationships between various factors and drought resistance in sugarcane, including those among varieties (Meena et al., 2020). A previous study showed that drought conditions can cause changes in the expression of many sugarcane genes. A total of 3,389 genes have been identified in wild sugarcane exposed to drought stress, including 1,772 upregulated and 1,617 downregulated genes (Belesini et al., 2017). Leaf transcriptomic analysis has shown that the expression of genes related to water retention, antioxidant secondary metabolite biosynthesis, oxidation, and osmotic stress responses is higher in the drought-tolerant sugarcane genotype, while the sensitive genotype has a higher number of downregulated genes, which include those involved in photosynthesis, carbon fixation, and the Calvin cycle (Nawae et al., 2020). A similar study showed that the drought-tolerant genotype Co-06022 expressed more genes than the drought-susceptible genotype Co-8021 under different degrees of drought stress. However, more genes are expressed in sensitive genotypes during the recovery period (Selvi et al., 2020). The results of these studies indicated that the relationship between drought resistance and sugarcane varieties is closely related gene expression at the transcriptome level under drought conditions. In addition, different parts of the sugarcane plant respond differently to drought stress. Fewer genes are upregulated and downregulated in the leaves, whereas more genes re upregulated and downregulated in the roots (Taheri et al., 2022). The organ heterogeneity of multiple gene expression is difficult to study using traditional methods and transcriptomic technology has helped to overcome this difficulty. In addition to the aspects mentioned above, the effects of biological factors such as disease, abiotic factors such as nutritional deficiencies, and extreme temperatures on the sugarcane transcriptome have also been studied (Li et al., 2023).

Although the relationships between the genic expression and drought resistance of sugarcane, as well as some cultivar-related studies, have been reported, these studies are insufficient; understanding the mechanism of drought resistance requires further exploration because of the complexity of the sugarcane genome as well as its source, and the development of modern sugarcane varieties (Pereira-Santana et al., 2017; Liu et al., 2018). To gain a more comprehensive understanding of the drought tolerance mechanisms for different sugarcane varieties, differences in transcriptomes of two sugarcane cultivar GT31, with weak drought tolerance, and F172 with strong drought tolerance were investigated in this study. Based on the transcriptome data, we further explored the differences in metabolic pathways and related gene expression between the different varieties under drought stress and confirmed that drought-resistant sugarcane responds to drought stress by regulating metabolic pathways and related gene expression, particularly the photosynthesis pathway. These results enrich our understanding of the molecular mechanisms underlying drought resistance in plants and provide a basis for sugarcane breeding.

Materials and methods

Plant material and sampling

In this study, plant materials from two sugarcane *Saccharum* L. cultivars, F172 and Guitang 31 (GT31), with strong and weak drought tolerance, developed by the Taiwan Sugar Research

Institute (Taiwan, China) and Sugarcane Research Institute, Guangxi Academy of Agricultural Sciences (Nanning China), respectively, were used (Yang and Li, 1992; Li et al., 2011).

The sugarcane seedlings were observed and photographed at 4, 5, and 7 d after water withdrawal and were categorized under different drought treatment conditions: mild drought stress (B), moderate drought stress (C), and severe drought stress (D) (Figure 1). Both varieties showed significant changes across the three time points, and cultivar F172 showed stronger drought tolerance than Guitang31 (Figure 1). Sugarcane leaf tissues from cultivars F172 and GT31 were obtained under B, C, and K drought conditions for subsequent transcriptomic sequencing and photosynthetic rate detection. In parallel, leaf tissues from the two cultivars under normal watering conditions were collected at the same time points as controls for subsequent transcriptomic sequencing and detection of the photosynthetic rate, which are abbreviated as BCK (on day 4), CCK (on day 5), and DCK (on day 7).

The experiment was performed in triplicates for each condition and cultivar. All the leaf samples were flash-frozen with liquid nitrogen and stored at −80°C until further use.

RNA extraction and RNA-seq analysis

RNA was extracted using the RNeasy Mini Kit (Qiagen, Beijing, China), followed by purification, fragmentation, and quality control using a NanoDrop 2000 (Thermo Scientific) and an Agilent 2100 Bioanalyzer (Agilent Technologies). Strand-specific libraries were obtained using dUTP for second-strand synthesis and subsequently sequenced on a BGISEQ-500 instrument (BGI, Shenzhen, China). The experiments were conducted in triplicates for each cultivar at each time point under drought (B, C, and K) and control (BCK, CCK, and DCK) conditions.

Preprocessing of the paired-end reads was performed using FASTP (v 0.23.1), and mapped to the sugarcane genome (NCBI accession: ASM2245720v1) using HISAT2 (v 2.20) (Kim et al., 2019). Raw read counts were quantified using the featureCounts software (SUBREAD v2.0.0) (Liao et al., 2014). Differential expression analysis was performed using the DESeq2 package (v1.38.3) (Love et al., 2014) for transcriptome comparisons between cultivars under the same conditions (GT31-B-vs-F172-B, GT31-C-vs-F172-C, GT31-D-vs-F172-D, GT31-BCK-vs-F172-BCK, GT31-CCK-vs-F172-CCK, and GT31-DCK-vs-F172-DCK) and between treatments and the corresponding controls for each cultivar (GT31-BCK-vs-GT31-B, F172-BCK-vs-F172-B, GT31-CCK-vs-GT31-C, F172-CCK-vs-F172-C, GT31-DCK-vs-GT31-D, and F172-DCK-vs-F172-D) to identify the differentially expressed genes (DEGs) with an absolute value of log2 FC > 1.0 and false discovery rate < 0.05. KEGG pathway analyses of the identified DEGs were conducted using ClusterProfiler (v4.3.1) (Yu et al., 2012).

Temporal analysis

Clusters of genes with the same expression profile over different time points were identified using the short time-series expression miner (v1.3.13) (Ernst and Bar-Joseph, 2006) for cultivar F172 under drought stress (F172), cultivar F172 controls (F172CK), cultivar GT31 under drought stress (GT31), and cultivar GT31 controls (GT31CK). The statistical significance of the number of genes for each profile compared to the expected number was computed using a permutation-based test. Unique and common trend genes for significant profile clusters ($p < 0.05$) from the above-mentioned four groups were selected and KEGG enrichment analysis was performed as described above.

Weighted correlation network analysis

Weighted correlation network analysis (v 1.69) (Langfelder and Horvath, 2008) was used to infer the network modules (parameters: softPower = 20, mergeCutHeight = 0.7, minModuleSize = 30)for 58,873 genes after filtering those with low expression levels (Fragments per kilobase of transcripts per million fragments mapped < 0.5). Module-trait associations were estimated using the correlation between module eigengenes and traits, including cultivars, drought conditions, and the strength of drought stress. Module–trait associations were considered statistically significant at $p < 0.05$. Trait-related genes with significant correlations were extracted from the module and subjected to KEGG enrichment analysis as described above. Hub genes in the modules were identified using the CytoHubba module in Cytoscape software.

Measurement of photosynthetic rate

A portable photosynthesissystem (Li-6800, Li-COR Biosciences, Lincoln, NE, USA) was used to observe the net photosynthetic rate for the functional topvisible dewlap leaf (leaf + 1) of sugarcane. as previously described (Verma et al., 2020). The photosynthesis rate was measured with three biological replicates for each cultivar at each time point under both drought treatment and control conditions.

Statistical analysis

The pheatmap package (v1.0.12) (https://CRAN.R-project.org/package=pheatmap) was used to plot the heatmaps. Differences were calculated using the t-test and were considered statistically significant at $p < 0.05$. Photosynthetic rate data were statistically analyzed using one-way ANOVA followed by Duncan's multiple range test.

Results

Transcriptome analysis of the sugarcane cultivars

RNA sequencing generated a total of 379.6 G of raw data for all 36 samples (each in the range of 7.8 G–13.9 G) (Table S1).

10.3389/fpls.2023.1243664

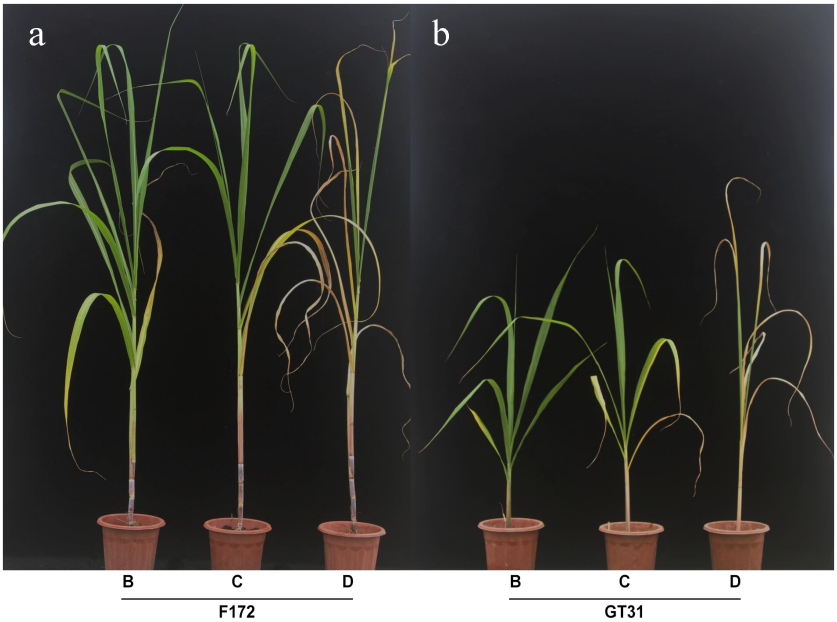


FIGURE 1
Appearance of two sugarcane varieties in drought tolerance. **(A)** Appearance of cultivar F172 under mild (left), moderate (middle), and severe (right) drought stress, respectively. **(B)** Appearance of cultivar GT31 under mild (left), moderate (middle), and severe (right) drought stress, respectively.

Approximately 373.8 G clean reads were obtained and passed through quality control; all samples were of high quality (Q20 ≥ 96.68% and Q30 ≥ 91.86%) (Table S1). The mapping rate of clean reads to the sugarcane reference genome ranged from 76.16% to 79.84% (Table S1). Principal component analysis showed low inter-replicate variability, and the samples in each group clustered together (Figure 2A). The principal component analysis clearly distinguished between the water deficit and control conditions, and the cultivars were also well separated (Figure 2A), indicating a large variability between the F172 and GT31 cultivars.

Differential gene expression analysis was performed between both cultivars grown under the same conditions, and between those grown under different conditions along with the corresponding controls (Figure 2B; Table S2). A total of 8,546, 10,036, 13,517, 11,776, 14,564, and 5,506 DEGs were identified between the GT31 and F172 cultivars under the same drought conditions, including B, BCK, C, CCK, D, and DCK, respectively (Figure 2B). On comparing intra-cultivar drought treatments and controls, we identified 33,058, 34,197, and 39,741 DEGs from the comparisons of three conditions (B vs. B-vs-BCK, C vs. C-vs-CCK, and D-vs-DCK) for strain GT31,

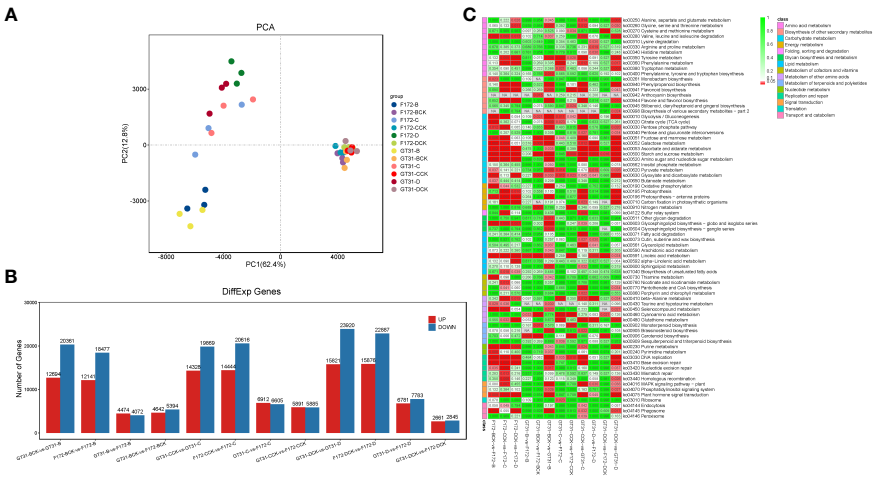


FIGURE 2
Transcriptome analysis of the sugarcane cultivars. **(A)** Principal component analysis (PCA) of transcriptome data. **(B)** Bar plot of the differentially expressed genes (DEGs) between cultivars in the same condition and between treatments and the corresponding controls for each cultivar, respectively. **(C)** Heatmap of relative enrichment qvalue of each pathway (rows) for each comparison (columns).

and 30,618, 35,060, and 38,763 DEGs for strain F172, respectively (Figure 2B). In general, the number of DEGs between the intra-cultivar drought treatments and controls was much greater than that between the inter-cultivar differences under the same conditions. These DEGs between cultivars grown under the same conditions were mainly enriched in starch and sucrose metabolism, linoleic acid metabolism, glycolysis/gluconeogenesis, glyoxylate and dicarboxylate metabolism, nitrogen metabolism, carotenoid biosynthesis, and ribosome pathways among others (Figure 2C). Compared with the DEGs between cultivars under normal watering conditions, those under drought conditions enriched in pyruvate metabolism, beta-alanine metabolism, glutathione metabolism specifically under moderate and severe drough stress (Figure 2C). In the comparison groups BCK-vs-GT31-B, CCK-vs-GT31-C, and DCK-vs-GT31-D, we identified 33,058 DEGs (12,697 up and 20,361 down), 34,197 DEGs (14,328 up and 19,869 down), and 39,741 DEGs (15,821 up and 23,920 down) for strain GT31, and 30,618 DEGs (12,141 up and 18,477 down), 35,060 DEGs (14,444 up and 20,616 down), and 38,763 DEGs (15,876 up and 22,887 down) for stain F172, respectively (Figure 2B). The DEGs between cultivars grown under different conditions along with the corresponding controls were mainly enriched in starch and sucrose metabolism, linoleic acid metabolism, glycolysis/gluconeogenesis, glyoxylate and dicarboxylate metabolism, valine, leucine and isoleucine degradation, phenylpropanoid biosynthesis, flavone and flavonol biosynthesis, photosynthesis-antenna proteins, glycosphingolipid biosynthesis (globo- and isoglobo- series), and phagosome pathways among others (Figure 2C). Many DEGs were identified between the drought stress and control groups for the same time points for both cultivars (Figure 2B; Table S2), indicating that drought stress has a significant impact on gene expression.

Comparison of trends between F172 and GT31 cultivars under in different drought stress conditions

Time-series expression analysis was performed to examine the dynamic transcriptomic differences in drought tolerance between cultivars. The gene expression patterns for cultivars grown under different conditions and their corresponding controls were analyzed for both cultivars across different drought time points, and the profiles with $p < 0.05$ in the permutation test were considered as significant (Figure 3). Profiles 0, 1, and 6 were considered to be significant for cultivar F172 in drought stress (Figure 3A; Table S3); profiles 4 and 3 were considered to be significant for cultivar F172 controls (Figure 3B; Table S4); profiles 1, 6, and 0 were considered to be significant for cultivar GT31 under drought stress (Figure 3C; Table S5); and profiles 1 and 6 were considered to be significant for cultivar GT31 controls (Figure 3D; Table S6).

To identify the trend genes specific to F172 in response to drought stress, an intersection analysis was conducted for the trend genes in different groups, including the cultivar F172 under drought stress (F172), cultivar F172 controls (F172CK), GT31 under drought stress (GT31), and GT31 controls (GT31CK) (Figure 4A). Among them, 5,103 genes exhibited a specific trend in the F172 group that was inconsistent with those in the GT31 and the F172 controls, potentially related to the drought tolerance of F172. The KEGG pathway enrichment results showed that these genes were significantly enriched in photosynthesis, mitogen-activated protein kinases (MAPK) signaling pathway, biosynthesis of various plant secondary metabolites, and cyanoamino acid metabolism pathways (Figure 4B).

Weighted gene co-expression network analysis of F172 and GT31 cultivars

We employed WGCNA to identify potential co-expression modules and key regulatory networks, and further elucidate their roles in the response of the F172 cultivar to drought stress. WGCNA categorized all genes into 15 modules (Figure 5A). Module-trait relationships were explored to extract significant associations between cultivars, drought conditions, strength of drought stress, and modules (Figure 5B). The blue4 and plum1 modules showed significant negative (–0.86, $p < 0.01$) and positive (0.96, $p < 0.01$) correlations, respectively, with drought conditions, (Figure 5B). The tan and salmon4 modules showed significant negative (-0.93, $p < 0.01$) and positive (0.8, $p < 0.01$) correlations, respectively, with the cultivars (Figure 5B).

All unique trend genes (Figure 4A) were mapped to weighted correlation network analysis modules (Table S7). Among the unique trend genes identified in F172 compared to GT31 and controls, most were in the blue4 (1,301) and plum1 (983) modules (Table S7); these were then used for KEGG enrichment analysis. Genes in the blue4 module were mainly enriched in phagosome, endocytosis, photosynthesis-antenna proteins, amino sugar and nucleotide sugar metabolism, and oxidative phosphorylation, among others (Figure 5C; Table S8). Those genes in the plum1 module were mainly enriched in spliceosome, RNA transport, aminoacyl-tRNA biosynthesis, basal transcription factors, peroxisome, and alanine, aspartate, and glutamate metabolism, among others (Figure 5D). Several enriched pathways, including those of photosynthesis, purine metabolism, starch and sucrose metabolism, beta-alanine metabolism, photosynthesis-antenna proteins, and plant hormone signal transduction, were shared between those genes in the blue4 module and the unique trend genes in F172 (Figure 5C; Table S8). The hub genes identified in the blue4 module included *ARF14* (auxin response factor 14) and *Os10g0147400* (similar to auxin influx carrier protein), which are associated with the plant hormone signal transduction pathway, and *Os02g0733300* (glycoside hydrolase gene), which is associated with the starch and sucrose metabolism pathway (Table S9). The hub genes of plum1 module included the spliceosome-related genes *RS31* (serine/arginine-rich splicing factor RS31) and *CLPC1* (chaperone protein ClpC1, chloroplastic), and peroxisome-related genes *PEX1* (peroxisomal biogenesis factor 1) and *PEX2* (peroxisomal biogenesis factor 2) (Table S10).

10.3389/fpls.2023.1243664

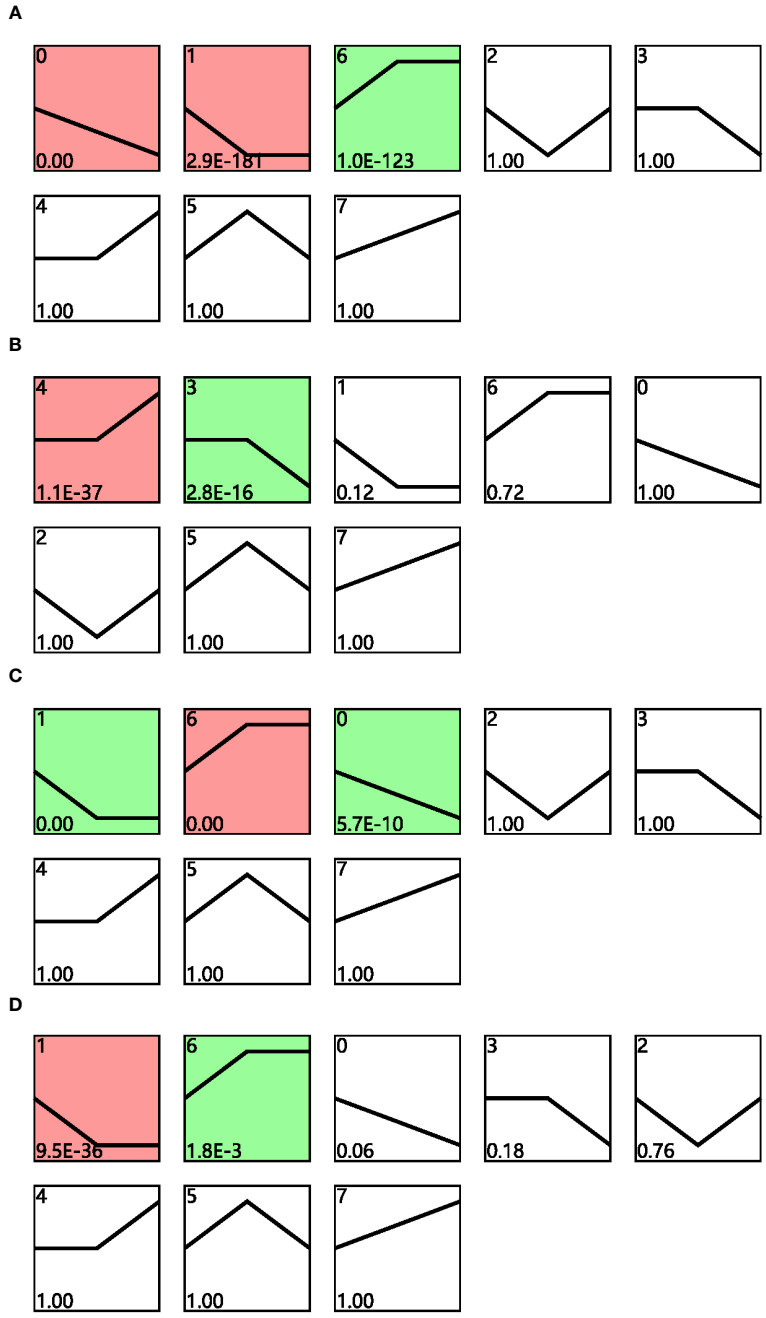


FIGURE 3
STEM analysis of gene expression profiles. Distinct expression profiles of cultivar F172 in drought stress **(A)**, cultivar F172 controls **(B)**, cultivar GT31 in drought stress **(C)**, and cultivar GT31 controls **(D)** were identified, respectively. The profiles were ordered by pvalue and those highlighted with colored background were significant profiles.

Differences in the photosynthesis pathway between F172 and GT31 cultivars

Photosynthesis is closely associated with drought stress in plants (Figure 6). In the photosynthesis-antenna protein pathway for F172, the expression levels of *Lhca1* (light-harvesting complex I chlorophyll a/b binding protein 1) and *Lhca4* (light-harvesting complex I chlorophyll a/b binding protein 4) were high at stage 1 and significantly decreased at later stages, exhibiting a distinct downward trend (Figure 6). Further, *PsbO* (photosystem II oxygen-evolving enhancer protein 1), *PsbP* (photosystem II oxygen-evolving enhancer protein 2), *PsbQ* (photosystem II oxygen-evolving enhancer protein 3), *PsbW* (photosystem II PsbW protein), *PsbY* (photosystem II PsbY protein), *Psb27* (photosystem II Psb27 protein), *PsaD* (photosystem I subunit II), *PsaE* (photosystem I subunit IV), *PsaF* (photosystem I subunit III), *PsaH* (photosystem I subunit VI), *PsaL* (photosystem I subunit XI), *PsaN* (photosystem I subunit PsaN), *PetF* (ferredoxin), and *petH* (ferredoxin–NADP+

10.3389/fpls.2023.1243664

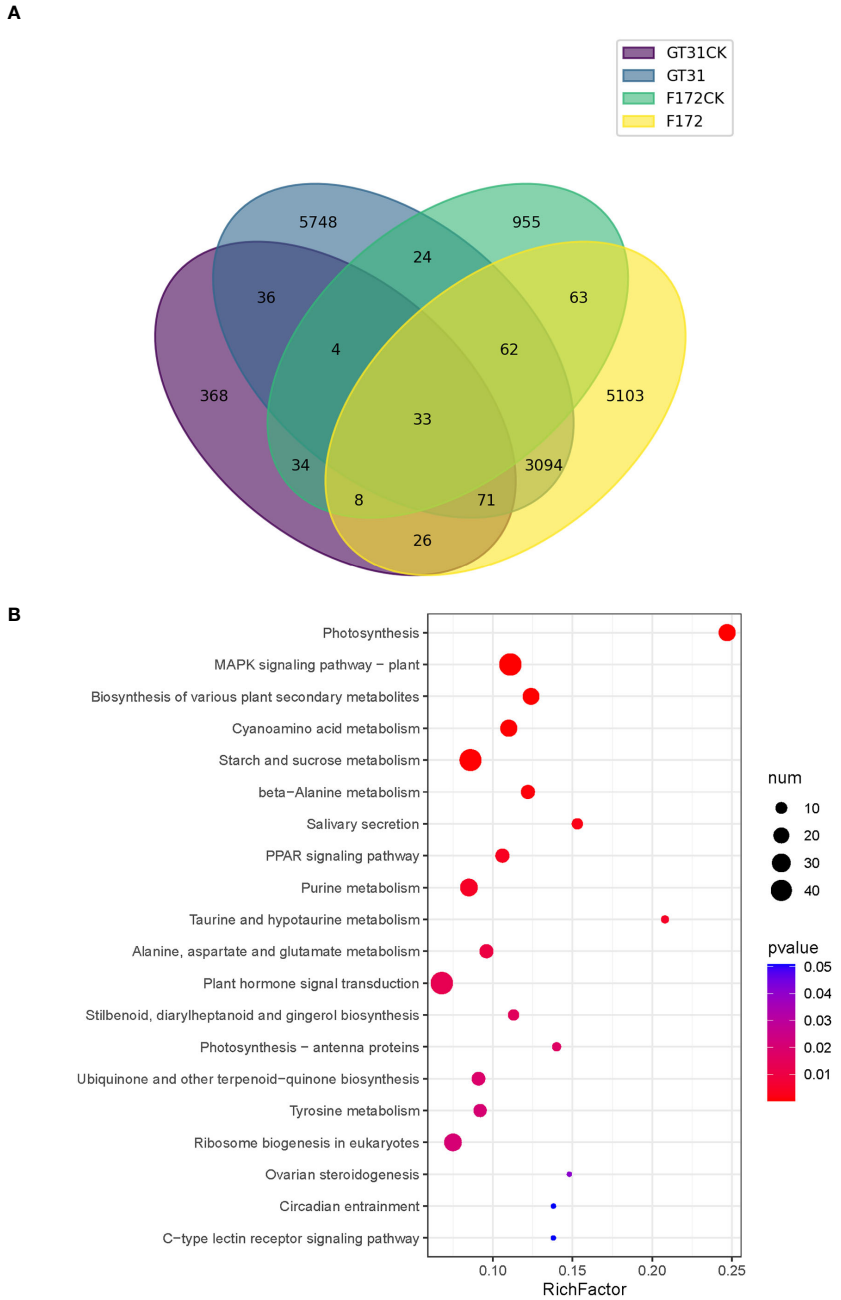


FIGURE 4
Comparison of trends between F172 and GT31 in different drought stress. **(A)** identification of unique trend genes in cultivar F172 in drought stress (F172) compared with cultivar F172 controls (F172CK), cultivar GT31 in drought stress (GT31), and cultivar GT31 controls (GT31CK). **(B)** Significantly enriched KEGG pathways of the unique trend genes in F172.

reductase), which participate the photosynthesis pathway, also showed a similar expression pattern for F172 (Figure 6).

The photosynthetic rates for two cultivars F172 and GT31 were significantly down-regulated under drought conditions. However, this decrease was more pronounced for the F172 variety, which was consistent with the declining trend observed in the expression of genes related to the photosynthetic pathway as described above (Figure 7). These results indicate that drought stress resulted in a significantly greater reduction in photosynthesis for cultivar F172, which is associated with stronger drought resistance.

Discussion

Sugarcane is an important commercial crop of global significance, more than 80% of the world's sugar production is derived from sugarcane, which is grown in more than 90 nations (Barnabas et al., 2015). Drought limits sugarcane production (Liu et al., 2021), and different sugarcane cultivars react differently to drought stress (da Silva et al., 2017). In this study, principal component analysis could discriminate between two different sugarcane cultivars under water deficit conditions, but not when

10.3389/fpls.2023.1243664

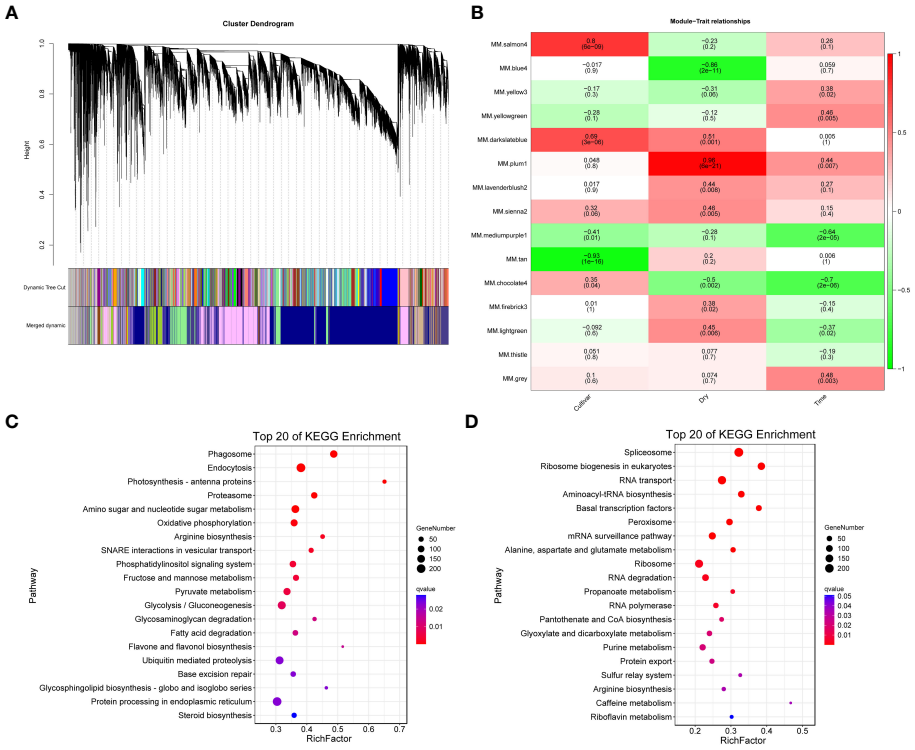


FIGURE 5
Weighted Gene Co-expression Network Analysis (WGCNA). **(A)** WGCNA module detection. **(B)** Heatmap of module - trait correlation, displaying the correlation values for each module (rows) and each trait (columns). **(C)** Significantly enriched KEGG pathways of the genes in blue4 module. **(D)** Significantly enriched KEGG pathways of the genes in plum1 module.

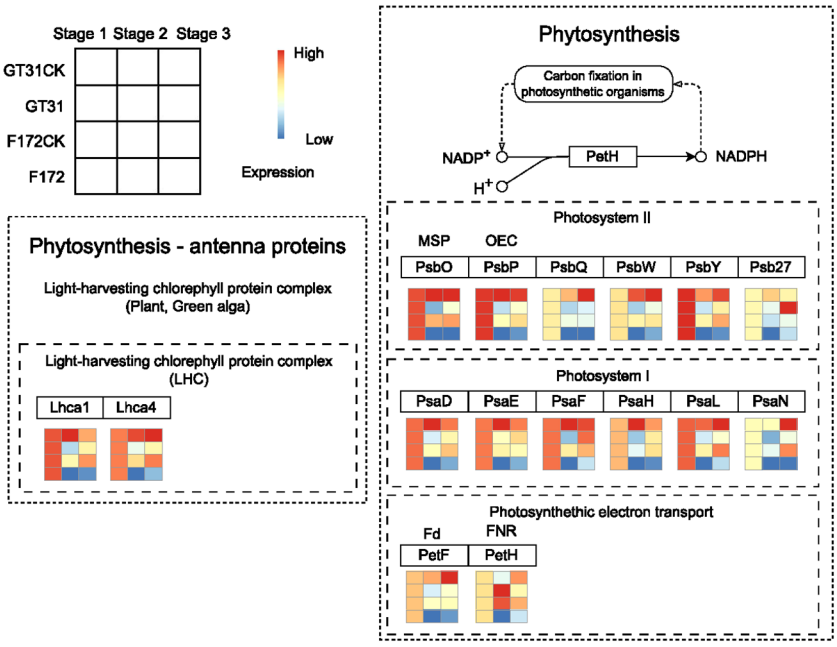


FIGURE 6
The expression pattern of genes in blue4 module which involved in Phytosynthesis - antenna proteins and Phytosynthesis pathways for cultivar F172 in drought stress (F172), cultivar F172 controls (F172CK), cultivar GT31 in drought stress (GT31), and cultivar GT31 controls (GT31CK), respectively.

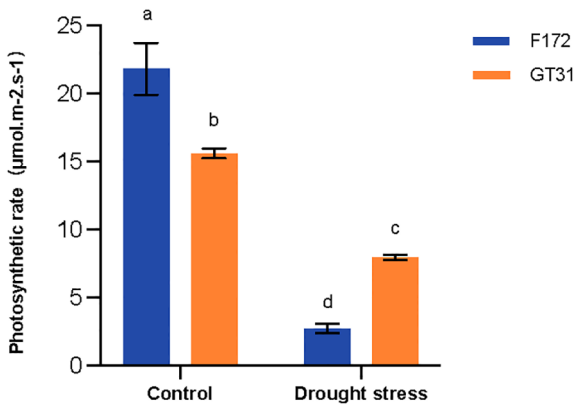


FIGURE 7
Photosynthetic rate of cultivars F172 and GT31 in drought stress and the corresponding controls. Different letters indicate significant differences (Duncan's test at $P < 0.05$).

the moisture level was normal (Figure 2A). Drought can also alter plant gene expression (Takahashi et al., 2018; Bashir et al., 2019). In this study, based on the DEGs between F172 and GT31 under drought stress, the results of the KEGG pathway analysis showed that starch and sucrose metabolism, and glycolysis/gluconeogenesis were extensively enriched for all comparisons (Figures 2B, C). Changes in these two metabolic pathways may be some of the most common responses to drought because these pathways are also significantly enriched for the IACSP97-7065, IACSP94-2094, and ROC22 sugarcane varieties under drought stress (Contiliani et al., 2022; Yang et al., 2022).

Many other physiological and metabolic processes in plants are involved in the drought stress response; Takahashi et al. (2018) reported that drought inhibits photosynthesis; further, MAPK cascades modulate plant tolerance to drought (Bashir et al., 2019). Drought can also affect the levels of secondary plant metabolites such as calcium, anthocyanins, flavonoids, phenolic acids, chlorophyll and saponins (Yu et al., 2020; Contiliani et al., 2022; Yang et al., 2022). Wan et al. (2021) found that changes in the cyanoamino acid metabolism pathway might be a key factor causing the difference in drought resistance between two cherry rootstocks. In this study, based on the unique trend of genes in F172 compared to that of GT31 and the controls, the results of the KEGG pathway analysis indicated that photosynthesis, MAPK signaling pathway, biosynthesis of various plant secondary metabolites, and cyanoamino acid metabolism pathways were significantly enriched in F172. In addition to these pathways, other enriched metabolic pathways, such as cyanoamino acid, starch, sucrose, urine, and ribosome biogenesis in eukaryotes (Figure 4B), have rarely been associated with drought tolerance in plants. The role of genes exhibiting specific trends in these pathways in the F172 group deserves further investigation.

Weighted correlation network analysis has been an important method used in previous studies on plant drought resistance (Yu et al., 2020; Cao et al., 2021). In *Arabidopsis thaliana*, both blue and salmon modules responded to drought (Sharma et al., 2018). Moreover, the blue module cprresponded to drought resistance traits in sunflower and Tartary buckwheat (Meng et al., 2022; Wu et al., 2022). In this study, the results of the weighted correlation network analysis demonstrated that the blue4 module was significantly negatively correlated with drought conditions, and the salmon4 module was significantly positively correlated with the cultivars (Figure 5B). In addition to the aforementioned metabolic pathways, several other pathways are involved in plant drought resistance, such as the plant hormone signal transduction pathways (Verma et al., 2016). The purine and phenylpropanoid metabolism pathways in *Dendrobium sinense* and Arabidopsis are also involved in the drought stress response (Watanabe et al., 2010; Zhang et al., 2021). Metabolome analysis showed that drought stress resulted in an increase in β-alanine in tomato fruit (Asakura et al., 2021). In addition, transcriptome analysis suggested that the photosynthesis-antenna protein pathway in Shanlan upland rice and peanuts is involved in the drought stress response (Ren et al., 2021; Zhou et al., 2022). In this study, KEGG pathway analysis showed that genes with unique trends in the blue module were enriched in photosynthesis, purine metabolism, starch and sucrose metabolism, alanine metabolism, photosynthesis-antenna proteins, and plant hormone signal transduction pathways (Figure 5C).

The most prominent effects of drought on crops are related to the germination and photosynthesis processes (Flexas et al., 2004; Nadeem et al., 2019). In the present study, both the photosynthesis-antenna protein and photosynthesis pathways were enriched for the drought-resistant variety, F172, under drought stress (Figures 4B, 5C). The proteins involved in the two pathways increased at normal moisture levels, but decreased under water-deficit conditions (Figure 6). These results indicated that under water-deficit conditions, drought-resistant sugarcane varieties might mitigate the effects of drought by adjusting the photosynthesis process. Reduced photosynthesis can save plenty of water for the plant itself, which may be the basis of drought resistance oin the F172 variety of sugarcane (Figure 7).

Conclusion

In the present study, we used comparative temporal analysis to unveil the differences in gene expression trends during drought stress between two sugarcane cultivars with differing drought tolerance levels. By performing WGCNA, we found that the two cultivars showed different trends in genes related to photosynthesis, MAPK signaling, biosynthesis of various plant secondary metabolites, and cyanoamino acid metabolism pathways in response to drought stress. The most notable change in this process was the reduction in expression of genes related to photosynthesis; the corresponding decrease in photosynthesis may be an important strategy adopted by plants to cope with drought. The bioinformatic analysis strategies used in this study may also be valuable for uncovering key factors related to important traits by comparing different cultivars.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA975299.

Author contributions

HL, RY and RZ: conceptualization. HL, XL and HZ: validation. KZ and JW: data curation. RZ, YG and HL: writing original draft preparation. LT, HZ and XL. writing review and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Foundation of Guangxi Province of China (No. GKZY20198005, 2020GXNSFAA297132), the National Natural Science Foundation of China (No. 32160486), the National Modern Agricultural Industry Technology System Construction Special Project (No. CARS-170105), and Guangxi Innovation Team Building Project of National Modern Agricultural Industrial Technology System of China(No. nycytxgxcxtd-2021-03-04).

Conflict of interest

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

The reviewer YQ declared a past co-authorship with the author XL to the handling editor at the time of review.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1243664/full#supplementary-material

References

Agurla, S., Gahir, S., Munemasa, S., Murata, Y., and Raghavendra, A. S. (2018). Mechanism of stomatal closure in plants exposed to drought and cold stress. *Adv. Exp. Med. Biol.* 1081, 215–232. doi: 10.1007/978-981-13-1244-1_12

Asakura, H., Yamakawa, T., Tamura, T., Ueda, R., Taira, S., Saito, Y., et al. (2021). Transcriptomic and Metabolomic Analyses Provide Insights into the Upregulation of Fatty Acid and Phospholipid Metabolism in Tomato Fruit under Drought Stress. *J. Agric. Food Chem.* 69, 2894–2905. doi: 10.1021/acs.jafc.0c06168

Barnabas, L., Ramadass, A., Amalraj, R. S., Palaniyandi, M., and Rasappa, V. (2015). Sugarcane proteomics: An update on current status, challenges, and future prospects. *Proteomics* 15, 1658–1670. doi: 10.1002/pmic.201400463

Bashir, K., Matsui, A., Rasheed, S., and Seki, M. (2019). Recent advances in the characterization of plant transcriptomes in response to drought, salinity, heat, and cold stress. *F1000Res* 8, 658. doi: 10.12688/f1000research.18424.1

Belesini, A. A., Carvalho, F. M. S., Telles, B. R., de Castro, G. M., Giachetto, P. F., Vantini, J. S., et al. (2017). *De novo* transcriptome assembly of sugarcane leaves submitted to prolonged water-deficit stress. *Genet. Mol. Res.* 16, gmr16028845. doi: 10.4238/gmr16028845

Cao, L., Lu, X., Wang, G., Zhang, P., Fu, J., Wang, Z., et al. (2021). Transcriptional regulatory networks in response to drought stress and rewatering in maize (*Zea mays* L.). *Mol. Genet. Genomics* 296, 1203–1219. doi: 10.1007/s00438-021-01820-y

Chen, Y., Li, Z., Sun, T., Wang, D., Wang, Z., Zhang, C., et al. (2022). Sugarcane scDREB2B-1 confers drought stress tolerance in transgenic nicotiana benthamiana by regulating the ABA signal, ROS level and stress-related gene expression. *Int. J. Mol. Sci.* 23, 9557. doi: 10.3390/ijms23179557

Conti, V., Parrotta, L., Romi, M., Del Duca, S., and Cai, G. (2023). Tomato biodiversity and drought tolerance: A multilevel review. *Int. J. Mol. Sci.* 24, 10044. doi: 10.3390/ijms241210044

Contiliani, D. F., de Oliveira Nebó, J. F. C., Ribeiro, R. V., Andrade, L. M., Peixoto Júnior, R. F., Lembke, C. G., et al. (2022). Leaf transcriptome profiling of contrasting sugarcane genotypes for drought tolerance under field conditions. *Sci. Rep.* 12, 9153. doi: 10.1038/s41598-022-13158-5

da Silva, M. D., de Oliveira Silva, R. L., Ferreira Neto, J. R. C., Benko-Iseppon, A. M., and Kido, E. A. (2017). Genotype-dependent regulation of drought-responsive genes in tolerant and sensitive sugarcane cultivars. *Gene* 633, 17–27. doi: 10.1016/j.gene.2017.08.022

Dinesh Babu, K. S., Janakiraman, V., Palaniswamy, H., Kasirajan, L., Gomathi, R., and Ramkumar, T. R. (2022). A short review on sugarcane: its domestication, molecular manipulations and future perspectives. *Genet. Resour. Crop Evol.* 69, 2623–2643. doi: 10.1007/s10722-022-01430-6

Ernst, J., and Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinf.* 7, 191. doi: 10.1186/1471-2105-7-191

Ferreira, T. H. S., Tsunada, M. S., Bassi, D., Araújo, P., Mattiello, L., Guidelli, G. V., et al. (2017). Sugarcane water stress tolerance mechanisms and its implications on developing biotechnology solutions. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01077

Flexas, J., Bota, J., Loreto, F., Cornic, G., and Sharkey, T. D. (2004). Diffusive and metabolic limitations to photosynthesis under drought and salinity in C(3) plants. *Plant Biol. (Stuttg).* 6, 269–279. doi: 10.1055/s-2004-820867

Gentile, A., Dias, L. I., Mattos, R. S., Ferreira, T. H., and Menossi, M. (2015). MicroRNAs and drought responses in sugarcane. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00058

Gentile, A., Ferreira, T. H., Mattos, R. S., Dias, L. I., Hoshino, A. A., Carneiro, M. S., et al. (2013). Effects of drought on the microtranscriptome of field-grown sugarcane plants. *Planta* 237, 783–798. doi: 10.1007/s00425-012-1795-7

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and. genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9, 559. doi: 10.1186/1471-2105-9-559

Li, A. M., Liao, F., Wang, M., Chen, Z. L., Qin, C. X., Huang, R. Q., et al. (2023). Transcriptomic and proteomic landscape of sugarcane response to biotic and abiotic stressors. *Int. J. Mol. Sci.* 24, 8913. doi: 10.3390/ijms24108913

Li, X., Liu, Z., Zhao, H., Deng, X., Su, Y., Li, R., et al. (2022). Overexpression of sugarcane scDIR genes enhances drought tolerance in nicotiana benthamiana. *Int. J. Mol. Sci.* 23, 5340. doi: 10.3390/ijms23105340

Li, H. G., Tan, Y. M., Tan, F., Wang, L. W., Yang, R. Z., and Liu, X. J. (2011). Breeding of new sugarcane variety guitang 31 with high productivity and sugar content and ratoon ability. *Seed* 30 (08), 116–118. doi: 10.16590/j.cnki.1001-4705.2011.08.061

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Liu, X., Zhang, R., Ou, H., Gui, Y., Wei, J., Zhou, H., et al. (2018). Comprehensive transcriptome analysis reveals genes in response to water deficit in the leaves of *Saccharum narenga* (Nees ex Steud.) hack. *BMC Plant Biol.* 18, 250. doi: 10.1186/s12870-018-1428-9

Liu, Q., Zhao, X., Liu, Y., Xie, S., Xing, Y., Dao, J., et al. (2021). Response of sugarcane rhizosphere bacterial community to drought stress. *Front Microbiol.* 12. doi: 10.3389/fmicb.2021.716196

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8

Meena, M. R., Kumar, R., Chinnaswamy, A., Karuppaiyan, R., Kulshreshtha, N., and Ram, B. (2020). Current breeding and genomic approaches to enhance the cane and sugar productivity under abiotic stress conditions. *3 Biotech.* 10, 440. doi: 10.1007/s13205-020-02416-w

Meng, H. L., Sun, P. Y., Wang, J. R., Sun, X. Q., Zheng, C. Z., Fan, T., et al. (2022). Comparative physiological, transcriptomic, and WGCNA analyses reveal the key genes and regulatory pathways associated with drought tolerance in Tartary buckwheat. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.985088

Nadeem, M., Li, J., Yahya, M., Sher, A., Ma, C., Wang, X., et al. (2019). Research progress and perspective on drought stress in legumes: A review. *Int. J. Mol. Sci.* 20 (10), 2541. doi: 10.3390/ijms20102541

Nawae, W., Shearman, J. R., Tangphatsornruang, S., Punpee, P., Yoocha, T., Sangsrakru, D., et al. (2020). Differential expression between drought-tolerant and drought-sensitive sugarcane under mild and moderate water stress as revealed by a comparative analysis of leaf transcriptome. *PeerJ* 8, e9608. doi: 10.7717/peerj.9608

Ozturk, M., Turkyilmaz Unal, B., García-Caparrós, P., Khursheed, A., Gul, A., and Hasanuzzaman, M. (2021). Osmoregulation and its actions during the drought stress in plants. *Physiol. Plant* 172, 1321–1335. doi: 10.1111/ppl.13297

Pereira-Santana, A., Alvarado-Robledo, E. J., Zamora-Briseño, J. A., Ayala-Sumuano, J. T., Gonzalez-Mendoza, V. M., Espadas-Gil, F., et al. (2017). Transcriptional profiling of sugarcane leaves and roots under progressive osmotic stress reveals a regulated coordination of gene expression in a spatiotemporal manner. *PLoS One* 12, e0189271. doi: 10.1371/journal.pone.0189271

Ren, J., Zhang, H., Shi, X., Ai, X., Dong, J., Zhao, X., et al. (2021). Genome-wide identification of key candidate microRNAs and target genes associated with peanut drought tolerance. *DNA Cell Biol.* 40, 373–383. doi: 10.1089/dna.2020.6245

Selvi, A., Devi, K., Manimekalai, R., and Prathima, P. T. (2020). Comparative analysis of drought-responsive transcriptomes of sugarcane genotypes with differential tolerance to drought. *3 Biotech.* 10, 236. doi: 10.1007/s13205-020-02226-0

Sharma, R., Singh, G., Bhattacharya, S., and Singh, A. (2018). Comparative transcriptome meta-analysis of Arabidopsis thaliana under drought and cold stress. *PLoS One* 13, e0203266. doi: 10.1371/journal.pone.0203266

Taheri, S., Gantait, S., Azizi, P., and Mazumdar, P. (2022). Drought tolerance improvement in Solanum lycopersicum: an insight into "OMICS" approaches and genome editing. *3 Biotech.* 12, 63. doi: 10.1007/s13205-022-03132-3

Takahashi, F., Kuromori, T., Sato, H., and Shinozaki, K. (2018). Regulatory gene networks in drought stress responses and resistance in plants. *Adv Exp Med Biol.* 1081, 189–214. doi: 10.1007/978-981-13-1244-1_11

Tardieu, F., Simonneau, T., and Muller, B. (2018). The physiological basis of drought tolerance in crop plants: A scenario-dependent probabilistic approach. *Annu. Rev. Plant Biol.* 69, 733–759. doi: 10.1111/ppl.13297

Verma, V., Ravindran, P., and Kumar, P. P. (2016). Plant hormone-mediated regulation of stress responses. *BMC Plant Biol.* 16, 86. doi: 10.1186/s12870-016-0771-y

Verma, K. K., Song, X. P., Zeng, Y., Li, D. M., Guo, D. J., Rajput, V. D., et al. (2020). Characteristics of leaf stomata and their relationship with photosynthesis in *Saccharum officinarum* under drought and silicon application. *ACS Omega.* 4, 5(37):24145–24153. doi: 10.1021/acsomega.0c03820

Wan, T., Feng, Y., Liang, C., Pan, L., He, L., and Cai, Y. (2021). Metabolomics and transcriptomics analyses of two contrasting cherry rootstocks in response to drought stress. *Biology* 10, 201. doi: 10.3390/biology10030201

Watanabe, S., Nakagawa, A., Izumi, S., Shimada, H., and Sakamoto, A. (2010). RNA interference-mediated suppression of xanthine dehydrogenase reveals the role of purine metabolism in drought tolerance in Arabidopsis. *FEBS Lett.* 584, 1181–1186. doi: 10.1016/j.febslet.2010.02.023

Wu, Y., Wang, Y., Shi, H., Hu, H., Yi, L., and Hou, J. (2022). Time-course transcriptome and WGCNA analysis revealed the drought response mechanism of two sunflower inbred lines. *PLoS One* 17, e0265447. doi: 10.1371/journal.pone.0265447

Yang, L., and Li, Y. (2022). Analysis of photosynthetic rate and sugar accumulation of Taitang 172. *Guangxi Agric. Sci.* 02, 61–62.

Yang, S., Chu, N., Zhou, H., Li, J., Feng, N., Su, J., et al. (2022). Integrated Analysis of Transcriptome and Metabolome Reveals the Regulation of Chitooligosaccharide on Drought Tolerance in Sugarcane (*Saccharum* spp. Hybrid) under Drought Stress. *Int. J. Mol. Sci.* 23, 9737. doi: 10.3390/ijms23179737

Yu, B., Liu, J., Wu, D., Liu, Y., Cen, W., Wang, S., et al. (2020). Weighted gene coexpression network analysis-based identification of key modules and hub genes associated with drought sensitivity in rice. *BMC Plant Biol.* 20, 478. doi: 10.21203/rs.3.rs-56040/v2

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118

Zhang, C., Chen, J., Huang, W., Song, X., and Niu, J. (2021). Transcriptomics and metabolomics reveal purine and phenylpropanoid metabolism response to drought stress in dendrobium sinense, an endemic orchid species in hainan island. *Front. Genet.* 12. doi: 10.3389/fgene.2021.692702

Zhou, S., He, L., Lin, W., Su, Y., Liu, Q., Qu, M., et al. (2022). Integrative analysis of transcriptome and metabolism reveals potential roles of carbon fixation and photorespiratory metabolism in response to drought in Shanlan upland rice. *BMC Genomics* 23, 862. doi: 10.1186/s12864-022-09094-3

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 25 October 2023
DOI 10.3389/fpls.2023.1270933

OPEN ACCESS

EDITED BY
Weicong Qi,
Jiangsu Academy of Agricultural Sciences (JAAS), China

REVIEWED BY
Tie Liu,
University of Florida, United States
Yihe Yu,
Henan University of Science and Technology, China

*CORRESPONDENCE
Qian Zha
zhaqian@saas.sh.cn
Haixia Zhong
zhonghaixia1@sina.cn

†These authors have contributed equally to this work

RECEIVED 01 August 2023
ACCEPTED 25 September 2023
PUBLISHED 25 October 2023

CITATION
Wu J, Zhang F, Liu G, Abudureheman R, Bai S, Wu X, Zhang C, Ma Y, Wang X, Zha Q and Zhong H (2023) Transcriptome and coexpression network analysis reveals properties and candidate genes associated with grape (*Vitis vinifera* L.) heat tolerance. *Front. Plant Sci.* 14:1270933. doi: 10.3389/fpls.2023.1270933

Transcriptome and coexpression network analysis reveals properties and candidate genes associated with grape (*Vitis vinifera* L.) heat tolerance

Jiuyun Wu[1,2†], Fuchun Zhang[1,2†], Guohong Liu[1,2†], Riziwangguli Abudureheman[1,2], Shijian Bai[1,3], Xinyu Wu[1,2], Chuan Zhang[1,2], Yaning Ma[1,2], Xiping Wang[1,4], Qian Zha[1,5*] and Haixia Zhong[1,2*]

[1]Turpan Research Institute of Agricultural Sciences, Xinjiang Academy of Agricultural Sciences, Xinjiang Grape Engineering Technology Research Center, Turpan, China, [2]The State Key Laboratory of Genetic Improvement and Germplasm Innovation of Crop Resistance in Arid Desert Regions (Preparation), Key Laboratory of Genome Research and Genetic Improvement of Xinjiang Characteristic Fruits and Vegetables, Institute of Horticultural Crops, Xinjiang Academy of Agricultural Sciences, Urumqi, China, [3]Xinjiang Uighur Autonomous Region of Grapes and Melons Research Institution, Turpan, China, [4]Colleges of Horticulture, Northwest A&F University, Xianyang, China, [5]Research Institute of Forestry and Pomology, Shanghai Academy of Agricultural Science, Shanghai, China

Temperature is one of the most important environmental factors affecting grape season growth and geographical distribution. With global warming and the increasing occurrence of extreme high-temperature weather, the impact of high temperatures on grape production has intensified. Therefore, identifying the molecular regulatory networks and key genes involved in grape heat tolerance is crucial for improving the resistance of grapes and promoting sustainable development in grape production. In this study, we observed the phenotypes and cellular structures of four grape varieties, namely, Thompson Seedless (TS), Brilliant Seedless (BS), Jumeigui (JMG), and Shine Muscat (SM), in the naturally high-temperature environment of Turpan. Heat tolerance evaluations were conducted. RNA-seq was performed on 36 samples of the four varieties under three temperature conditions (28°C, 35°C, and 42°C). Through differential expression analysis revealed the fewest differentially expressed genes (DEGs) between the heat-tolerant materials BS and JMG, and the DEGs common to 1890 were identified among the four varieties. The number of differentially expressed genes within the materials was similar, with a total of 3767 common DEGs identified among the four varieties. KEGG enrichment analysis revealed that fatty acid metabolism, starch and sucrose metabolism, plant hormone signal transduction, the MAPK signaling pathway, and plant-pathogen interactions were enriched in both between different temperatures of the same material, and between different materials of the same temperature. We also conducted statistical and expression pattern analyses of differentially expressed transcription factors. Based on Weighted correlation network analysis (WGCNA), four specific

modules highly correlated with grape heat tolerance were identified by constructing coexpression networks. By calculating the connectivity of genes within the modules and expression analysis, six candidate genes (*VIT_04s0044g01430*, *VIT_17s0000g09190*, *VIT_01s0011g01350*, *VIT_01s0011g03330*, *VIT_04s0008g05610*, and *VIT_16s0022g00540*) related to heat tolerance were discovered. These findings provide a theoretical foundation for further understanding the molecular mechanisms of grape heat tolerance and offer new gene resources for studying heat tolerance in grapes.

KEYWORDS

grape, heat stress, RNA-Seq, WGCNA, candidate genes

1 Introduction

Grapes (*Vitis vinifera* L.), as sessile organisms, inevitably encounter various biotic or abiotic stresses during their growth and development (Ju et al., 2020; Ren et al., 2022). Grapes ripen during the summer, coinciding with periods of high temperatures, which significantly impact grape-growing regions. With global climate change, high temperatures have had a severe impact on grape yield and quality (Gouot et al., 2019a; Gouot et al., 2019b). According to the Intergovernmental Panel on Climate Change's (IPCC) Sixth Assessment Report in 2021, the global average temperature is projected to continue rising. It is estimated that from 2021 to 2040, the global average temperature will increase by 1.5 to 1.6°C compared to the period of 1850-1900 (IPCC, 2021). By the end of this century, the global average temperature is expected to rise by 1.4 to 4.4°C (https://public.wmo.int/en/media/press-release/global-temperatures-set-reach-new-records-next-five-years). The frequency and extent of extreme heat events will continue to increase. High-temperature stress has become one of the major constraints on the health and sustainable development of the grape industry (Mori et al., 2007). Dealing with high-temperature stress will be an unavoidable challenge for the global grape industry (Sun et al., 2018).

High-temperature stress triggers cellular stress responses through signal transduction pathways (Lian et al., 2021). The most sensitive organ in cells affected by high temperatures is the cell membrane (Sadura and Janeczko, 2022). High temperatures can alter the fluidity and structural integrity of the cell phospholipid membrane, leading to protein denaturation and inducing oxidative reactions, stress gene expression, and protein response (Sadura and Janeczko, 2022; Kim et al., 2022). These processes contribute to the manifestation of heat-related phenotypes in plants. However, to adapt to high-temperature environments, plants have evolved ecological habits that enable them to respond and adapt to heat stress promptly. Field observations have shown that sustained high temperatures during the summer can cause grape leaves to curl, lose water, and experience severe sunburn (Mori et al., 2007; Sun et al., 2018). This damage often has a significant impact on normal grape growth and fruit quality (Mori et al., 2007; Sun et al., 2018). High temperatures can also restrict photosynthesis in grape plants, reducing nutrient synthesis and transportation and thereby affecting fruit formation and development (Correia et al., 2021), which can result in smaller and fewer fruits, leading to reduced yield. Additionally, high temperatures can increase the acidity of grape berries, resulting in poor taste and flavor (Mori et al., 2007; Sun et al., 2018). Furthermore, high temperatures may decrease the pigment content in fruits, leading to less vibrant colors (Sun et al., 2018).

In recent years, the widespread application of high-throughput technologies such as RNA-seq-based coexpression network analysis has provided powerful tools and methods for uncovering the molecular characteristics and candidate genes related to plant heat tolerance (Cao et al., 2021; Li et al., 2022b; Meng et al., 2022). Coexpression network analysis is a systems biology approach that constructs gene coexpression networks by analyzing the correlation of gene expression, thereby identifying functionally related gene modules (Langfelder and Horvath, 2008; Tian et al., 2020). In the study of plant heat tolerance, coexpression network analysis can help identify gene sets closely associated with heat tolerance and predict the interaction relationships and regulatory networks of these genes in physiological processes (Liu et al., 2019; Wang et al., 2022). By analyzing key genes in coexpression networks, we can uncover the molecular mechanisms of plants during high-temperature stress, providing potential candidate genes for breeding heat-tolerant varieties (Liu et al., 2019; Wang et al., 2022).

Because grapes are important fruits and wine-making materials, there are significant implications for improving grape yield and quality through the study of heat tolerance (Mori et al., 2007; Sun et al., 2018). Although some grape species have a certain degree of heat tolerance, there is significant variation in heat tolerance among different varieties, and the mechanisms underlying heat tolerance are not well understood. Moreover, prolonged extreme heat stress may permanently affect grape physiological metabolism and yield attributes. Therefore, we selected four grape varieties (detailed

information on the materials is provided in Table S1) with different levels of heat tolerance for our study and conducted leaf phenotypic and cellular structure analyses as well as RNA-seq sequencing under three temperature conditions during the summer in Turpan, Xinjiang, China. Through clustering analysis, differential expression analysis, GO and KEGG enrichment analysis, expression analysis of transcription factors (TFs), construction of weighted gene coexpression networks, and qRT-PCR, we identified candidate genes for grape heat tolerance. This study provides a theoretical foundation for further understanding the molecular mechanisms of grape heat tolerance and offers new genetic resources for studying heat tolerance in grapes.

2 Materials and methods

2.1 Plant materials and growth conditions

In this study, four grape varieties were used. They were planted in the Grape Germplasm Repository of Turpan Agricultural Research Institute, Xinjiang Academy of Agricultural Sciences (89°18′E, 42°53′N). The region has an average annual temperature of 17.6°C, annual precipitation of 12.5 mm, annual sunshine duration of 3109.2 hours, and a frost-free period of approximately 210 days. The grapevines were spaced 1.2 meters within rows and 2.5 meters between rows, arranged in a north-south direction, and were 4 years old. The cultivation employed a 'V'-shaped trellis system. All grapes received similar irrigation, soil management, pruning, and disease control methods. During the high-temperature period in Turpan in 2021 (June to August), there were 76 days with temperatures above 35°C, including 26 days with temperatures above 40°C (Figure 1; Table S2). On August 10, 2021, leaf samples were collected at air temperatures of 28°C (T1), 35°C (T2), and 42°C (T3). Leaf samples (the 5th to 9th fully expanded leaves from the top of the canopy, sampling is performed when the temperature is maintained at this temperature for 30 minutes) were collected from each material (four replicates per sample, three for RNA-seq and one for qRT-PCR), rapidly frozen in liquid nitrogen, and stored for subsequent experiments.

2.2 Transmission electron microscopy sample preparation

Grape leaf samples (the 5th to 9th fully expanded leaves from the top of the canopy) were collected, excluding the main veins, and cut into small pieces of approximately 1 mm^2. The leaf samples were then fixed overnight at 4°C in a 2.5% glutaraldehyde solution. Subsequently, the samples were sliced into sections of 70-90 nm using a LEICA EM UC7 ultramicrotome. The sections were stained with lead citrate solution and uranyl acetate-50% ethanol saturated solution for 5-10 minutes each. Finally, the sections were observed under a Hitachi H-7650 transmission electron microscope.

2.3 RNA extraction, cDNA library preparation, and sequencing

RNA was extracted using the TRIzol method, and the integrity of the RNA was assessed using 1% agarose gel electrophoresis (Fleige and Pfaffl, 2006). The extracted total RNA was stored at -80°C and transported on dry ice to the PARSUNO Company (Shanghai, China) for sequencing. Fragmentation of the extracted RNA was performed using a PCR plate with a magnetic plate holder (Landolt et al., 2016). The fragmented mRNA was reverse transcribed into cDNA using SuperScript II and random primers (Invitrogen, Carlsbad, California, USA) (Kusser et al., 2006). The obtained data were filtered and quality controlled using fastp software, and the clean data were used for subsequent analysis (Chen et al., 2018). The reads were aligned to the grape genome (http://plants.ensembl.org/Vitis_vinifera/Info/Index) using HISAT2, and StringTie was used for read quantification (Pertea et al., 2016; Kim et al., 2019).

2.4 Identification of differentially expressed genes

FPKM (fragments per kilobase of exon per million fragments mapped) was used to measure gene expression levels, which

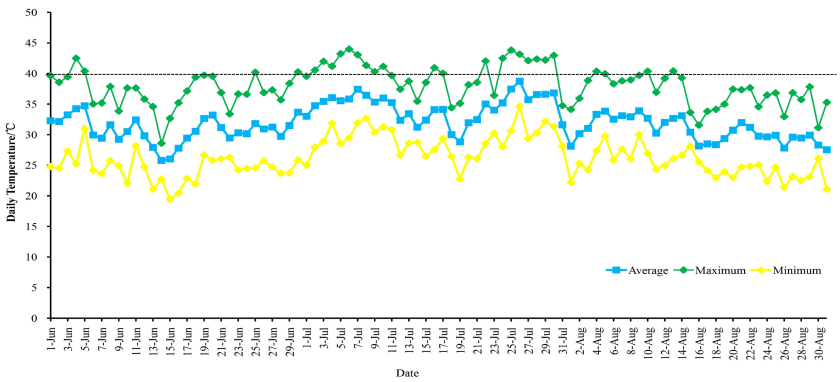


FIGURE 1
Temperature changes in Turpan from June to August 2021.

represents the number of reads mapped to exons per kilobase of exon length per million mapped reads. DESeq2 was used to calculate the fold change in gene expression between different samples based on gene expression levels (Liu et al., 2021). DEGs were selected based on the criteria of FDR ≤ 0.01 and absolute log2-fold change ≥ 1. The amino acid sequences of all DEGs were submitted to the KEGG database (https://www.kegg.jp/ghostkoala/) for identification of genes involved in hormone biosynthesis and signal transduction (Kanehisa and Goto, 2000). The grape whole-genome sequence was submitted to the PlantTFDB (http://planttfdb.cbi.pku.edu.cn/) for transcription factor prediction (Jin et al., 2017).

2.5 Construction of coexpression networks

The gene expression profiles of the DEGs were subjected to coexpression analysis using the dynamic branch cutting method in the R package WGCNA (Langfelder and Horvath, 2008). To ensure the scale-free distribution of the network, the weight coefficient β was set to 8, which should have a high correlation coefficient close to 0.8 and a certain degree of gene connectivity. The Blockwise Modules function was used to construct the network, resulting in multiple effective modules with varying numbers of genes. Modules with a minimum module size of 30 and a merge cut height of 0.25 were merged if their similarity exceeded 0.75. The module eigengene (ME) vectors and the correlation coefficients between hormone content and different treatment time points were calculated. Specific modules were selected based on a threshold of $r > 0.80$ and $P < 0.05$. The genes in the specific modules and predicted transcription factors were used to construct the coexpression network, which was visualized using Cytoscape software (version 3.10) (Shannon et al., 2003).

2.6 Quantitative real-time PCR

Total RNA was extracted using the E.Z.N.A. Plant RNA Kit (Omega Bio-Tek, Doraville, GA, USA). The concentration of each RNA sample was measured using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Then, 1 μg of isolated RNA was reverse transcribed into first-strand cDNA using the PrimeScript™ RT Reagent Kit with gDNA Eraser (Takara Bio Inc., Shiga, Japan). qRT-PCR analysis was performed using a Roche LC480 instrument (Roche Diagnostics GmbH, Mannheim, Germany) and SYBR Green (Takara Bio Inc.). Initially, a two-step PCR amplification program was used, with an initial denaturation at 95°C for 30 seconds, followed by 40 cycles of denaturation at 95°C for 5 seconds and annealing at 60°C for 34 seconds. Amplification, melting, and standard curves were generated using Roche LC480 software. geNorm software (https://genorm.cmgg.be/) was used to calculate the relative expression levels of target genes, with *VvGAPDH* as the reference gene. Each program was performed with three biological replicates. The primers used in this study are listed in Table S3.

3 Results

3.1 Effects of high temperature on grape leaf phenotype and cell structure

Leaf characteristics directly reflect the degree of high-temperature damage and the ability to resist high-temperature stress (Gong et al., 2023). High temperature stress can cause changes in the morphological characteristics of grape leaves, but different grape varieties exhibit inconsistent phenotypic responses to high-temperature conditions. When the temperature was 28°C, all leaf samples showed no obvious chlorosis (Figure 2A). At an air humidity of 35°C, TS exhibited leaf edge curling, SM showed leaf yellowing and chlorosis, and BS and JMG maintained normal leaf morphology without any evident heat damage symptoms (Figure 2A). At 42°C, BS and JMG showed marginal desiccation without yellowing, while TS and SM exhibited severe leaf edge curling, desiccation, and pronounced yellowing (Figure 2A). High-temperature stress has adverse effects on plant cell structure (Zhao et al., 2019). At 28°C, the cell structures, including the nucleus, vacuole, mitochondria, and chloroplasts, appeared normal in the grape leaf cells of all four materials. The chloroplast membranes and thylakoid membranes were clearly visible with intact structures, the vacuolar membranes had clear edges, and the grana and thylakoid stacks were dense and well-arranged (Figure 2B). At 35°C, the thylakoid arrangement was relatively orderly in JMG and TS, while the chloroplasts in BS and SM showed thylakoid swelling or

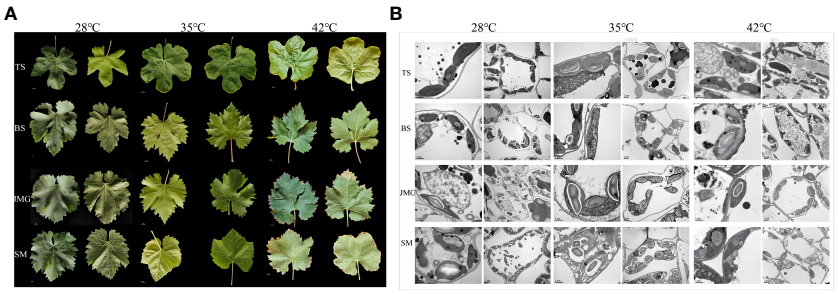


FIGURE 2
(A), Leaf morphology of four grape varieties at three temperatures. **(B)**, Leaf cell structure of 4 grape varieties at 3 temperatures.

disorganized arrangement with abundant osmiophilic particles (Figure 2B). At 42°C, the ultrastructure of leaf cells in all four varieties was damaged to varying degrees, but the chloroplast structure in JMG and BS remained stable with lower levels of damage. TS and SM exhibited an increase in plastoglobules within chloroplasts, and multiple organelle membranes showed fuzzy and damaged appearances (Figure 2B). Based on these results, we found that BS and JMG exhibited higher heat tolerance than TS and SM. To further determine the potential molecular mechanisms and candidate genes involved, RNA-seq analysis was conducted.

3.2 RNA-seq analysis

In total, 36 samples of RNA-seq data were obtained from 4 materials under 3 temperature conditions, resulting in 249.78 Gb of raw data. After filtering, a total of 228.05 Gb clean data were obtained. The effective data obtained from each sample were at least 5.35 Gb, with Q30 base percentages ranging from 91.15% to 93.92% and an average of 93.27%. The alignment rate with the reference genome ranged from 87.70% to 92.64%, with an average alignment rate of 91.31% (Table S4). Correlation analysis was initially performed among the samples, and it was found that the correlation among the three replicates of each material at different temperatures exceeded 0.96, indicating a high correlation between the replicates (Figure S1). PCA revealed that each replicate clustered together, and the differences between treatments were greater than the differences between materials (Figure 3). In summary, the high correlation among the replicates indicates their consistency across different time periods. To confirm the accuracy of the transcriptome expression profile, six randomly selected genes were subjected to qRT-PCR analysis with three independent replicates, and the results showed similar expression patterns to those observed in RNA-seq, confirming the reliability of the RNA-seq data for further analysis.

3.3 Differential expression analysis

First, differential expression analysis was performed among different materials under the same temperature. There were 5508 DEGs between BS and JMG, 8511 DEGs between TS and BS, 8756 DEGs between TS and JMG, 9952 DEGs between TS and SM, 8021 DEGs between SM and BS, and 8601 DEGs between SM and JMG (Figure 4A). Using the k-means clustering method, a total of 4 statistically significant clusters were identified among the 1890 commonly differentially expressed genes (Figures 4B, C). In BS, although Cluster 1 showed a trend of downregulation followed by upregulation, the overall expression trend change was not significant. Cluster 2 exhibited an increasing expression level with increasing temperature, while Cluster 3 showed the opposite expression trend compared to Cluster 2. Cluster 4 displayed an initial increase followed by a decrease in expression (Figure 4B). In JMG, Cluster 1 showed a trend of downregulation followed by a plateau, Cluster 2 exhibited an expression trend of initial increase followed by a decrease, Cluster 3 showed a gradual decrease in expression with increasing temperature, and Cluster 4 displayed increasing expression with increasing temperature (Figure 4B). In TS, Cluster 1 exhibited increasing expression with increasing temperature, Cluster 2 showed a trend of upregulation followed by downregulation, Cluster 3 displayed a downregulation trend, and Cluster 4 showed the opposite trend compared to Cluster 2 (Figure 4B). In SM, Cluster 1 exhibited increasing expression with increasing temperature, Cluster 2 showed a trend of upregulation followed by downregulation, Cluster 3 displayed a downregulation trend, and Cluster 4 showed a slight increase after an initial decrease (Figure 4B).

We also performed differential expression analysis on the same material under different temperatures. There were 9952 DEGs in BS, 9451 DEGs in JMG, 8266 DEGs in TS, and 9604 DEGs in SM. The DEGs between heat-tolerant and heat-sensitive materials did

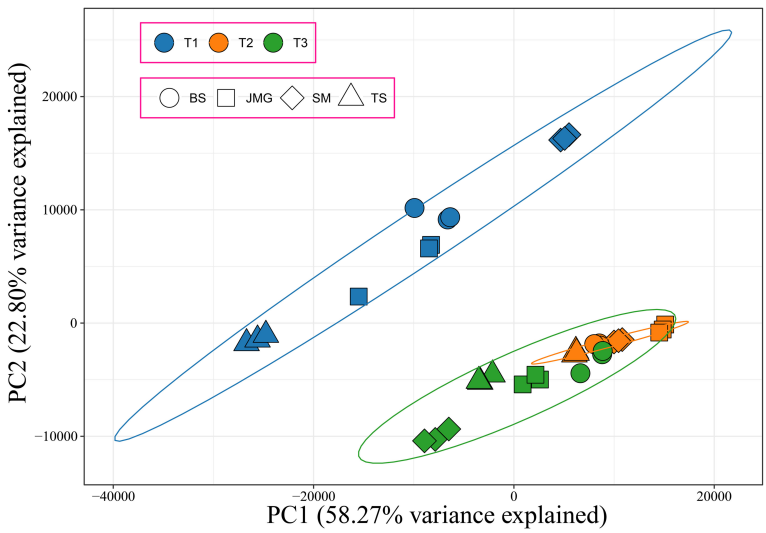


FIGURE 3
PCA of 36 RNA-seq samples, different colors represent different periods, and different shapes represent different materials.

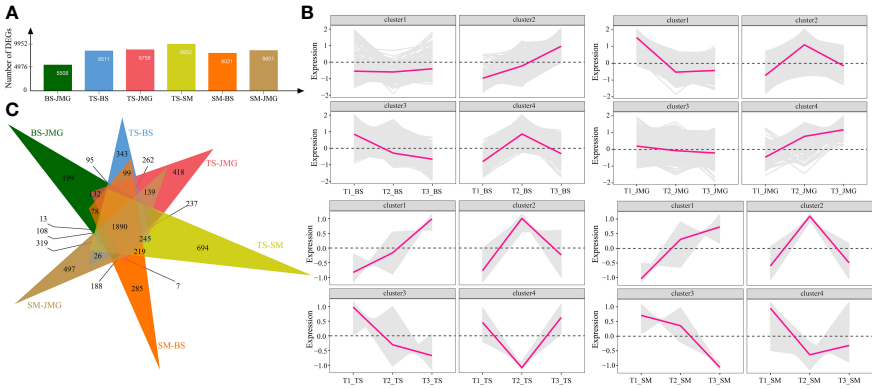


FIGURE 4
(A) Number of DEGs between different materials. **(B)** Line chart of DEG expression patterns between different materials. **(C)** DEG Wayne diagrams between different materials.

not differ significantly (Figure 5A). Using the k-means clustering method, a total of 4 statistically significant clusters were identified among the 3767 common DEGs (Figures 5B, C). In BS, although Cluster 1 showed a trend of upregulation followed by downregulation, the overall expression trend change was not significant. Cluster 2 exhibited an expression trend of an initial increase followed by a decrease, while Cluster 3 showed decreasing expression with increasing temperature, and Cluster 4 displayed increasing expression with increasing temperature (Figure 5B). In JMG, Cluster 1 showed an initial increase followed by a decrease in the expression trend, Cluster 2 exhibited a gradual decrease after an initial decrease, Cluster 3 showed a slight decrease with temperature change, and Cluster 4 displayed a trend of initial stability followed by a decrease (Figure 5B). In TS, Cluster 1 showed an initial increase followed by relatively stable expression, Cluster 2 exhibited a trend of downregulation followed by a plateau, Cluster 3 showed an initial increase followed by a decrease in the expression trend, and Cluster 4 exhibited increasing expression with increasing temperature (Figure 5B). In SM, Cluster 1 showed an initial increase followed by a slight decrease in expression, Cluster 2 exhibited a slight decrease in expression, Cluster 3 displayed a downregulation trend, and Cluster 4 showed a slight decrease followed by an increase in expression (Figure 5B).

3.4 GO and KEGG enrichment analysis of DEGs

Gene Ontology (GO) enrichment analysis was performed on the common DEGs between materials (Figure 4). The significantly enriched GO terms included defense response, response to oxygen-containing compound, response to light intensity, abscisic acid metabolic process, response to external stimulus, response to stress, response to hormone, response to water, response to water deprivation, and abscisic acid biosynthetic process (Figure 6A). KEGG pathway enrichment analysis of the common DEGs between materials revealed enrichment in plant-pathogen interaction, MAPK signaling pathway, plant hormone signal transduction, metabolism by cytochrome P450, DNA replication, flavone and flavonol biosynthesis, fatty acid biosynthesis, glycerolipid metabolism, fatty acid metabolism, and starch and sucrose metabolism (Figure 6B). GO enrichment analysis was also

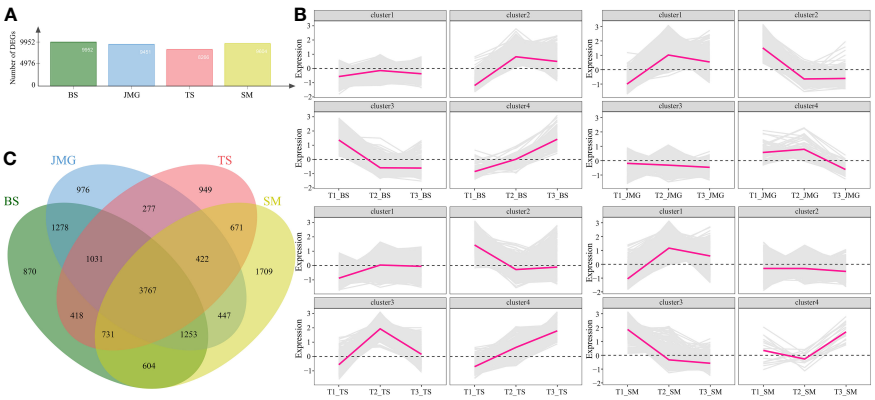


FIGURE 5
(A) Number of DEGs between different temperatures. **(B)** Line chart of DEG expression patterns between different temperatures. **(C)** DEG Wayne diagrams between different temperatures.

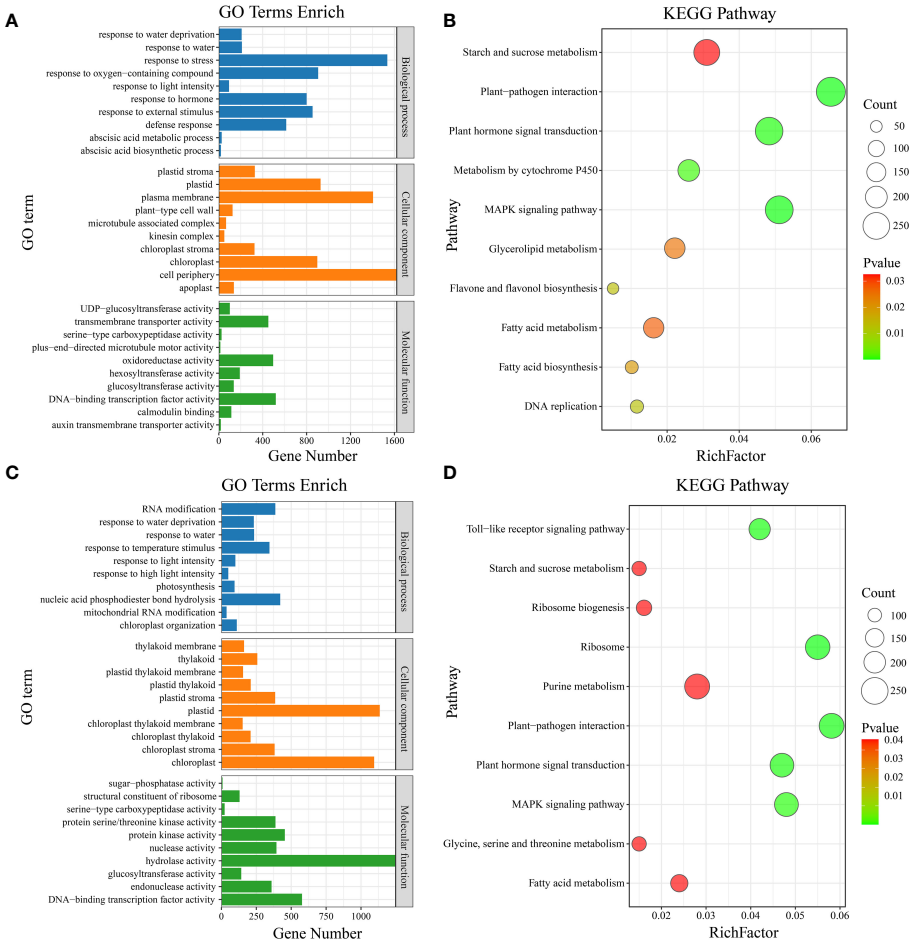


FIGURE 6
(A) Material-to-material DEG GO enrichment analysis. **(B)** Material-to-material DEG KEGG enrichment analysis. **(C)** GO enrichment analysis of DEGs at different temperatures. **(D)** KEGG enrichment analysis of DEGs at different temperatures.

conducted on the material-specific DEGs (Figure 5). The significantly enriched GO terms within each material included RNA modification, nucleic acid phosphodiester bond hydrolysis, mitochondrial RNA modification, response to light intensity, response to temperature stimulus, photosynthesis, response to water, response to water deprivation, chloroplast organization, and response to high light intensity (Figure 6C). KEGG pathway enrichment analysis of the material-specific DEGs showed enrichment in ribosome, plant-pathogen interaction, Toll-like receptor signaling pathway, plant hormone signal transduction, MAPK signaling pathway, ribosome biogenesis, purine metabolism, glycine, serine and threonine metabolism, fatty acid metabolism, and starch and sucrose metabolism (Figure 6D).

3.5 TF expression analysis

A total of 282 differentially expressed transcription factors (TFs) were identified between materials, and 136 differentially expressed TFs were identified under different temperature conditions, mainly including ERF, MYB, NAC, bHLH, and WRKY (Figures 7A, B). The expression patterns of differentially expressed TF genes were visualized using a heatmap, and most TFs showed the highest expression under T2 temperature conditions (Figure 7C). The expression levels of bHLH, C2H2, and HD-ZIP were predominantly higher in heat-tolerant materials, while ERF exhibited the highest expression at the T2 temperature in SM and JMG (Figure 7C). HSF showed high expression in both heat-tolerant materials and heat-sensitive materials, indicating the complex heat tolerance mechanism of HSF in grapes (Figure 7C). The expression patterns of MYB, NAC, and WRKY were also complex, similar to HSF, suggesting that more experiments are needed to validate and elucidate the roles and functions of these TFs in the heat tolerance process of grapes.

3.6 WGCNA exploration of heat-tolerant hub genes

Weighted gene coexpression network analysis (WGCNA) was performed on a total of 5,156 commonly differentially expressed genes (DEGs) between materials and within materials to construct a coexpression network associated with heat tolerance in grapes (β soft thresholding parameter set to 8, scale-free R^2 > 0.80), resulting

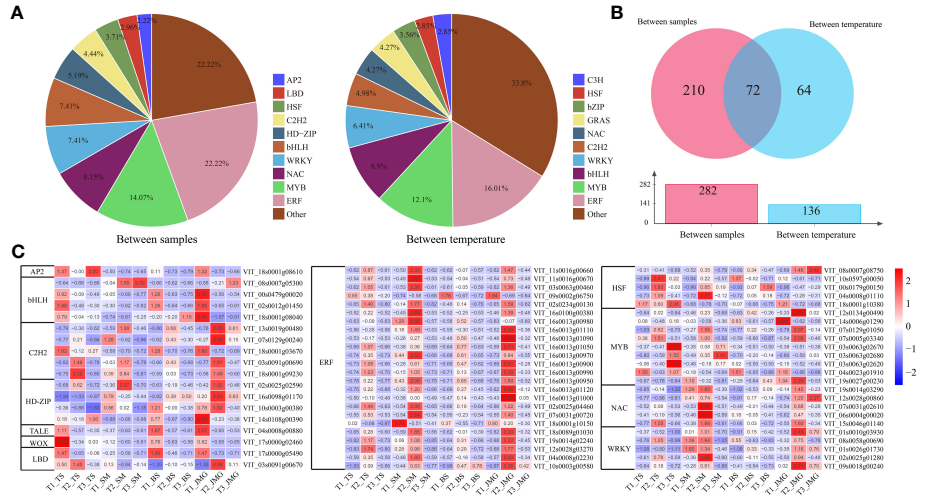


FIGURE 7
(A) TF pie chart of material and temperature differences. **(B)** TF Wayne diagram for differences between materials and different temperatures. **(C)** Heatmap of differential TF expression.

in 14 expression modules (Figure 8A). Based on the correlation results between modules and materials, the brown module showed a significant correlation with JMG at T2 temperature, the magenta module showed a significant correlation with SM at T2 temperature, and the pink and red modules showed a significant correlation with SM at T1 temperature (Figure 8B). The brown, magenta, pink, and red modules were selected to construct a gene interaction network and identify hub genes. Cytoscape was used for network visualization (Figure 8C). Five hub genes were determined for each module, resulting in a total of 20 hub genes (Figure 8C). The expression patterns of 20 hub genes at different temperatures were detected using qRT-PCR (Figure 8D; Figure S1). Among them, only *VIT_04s0044g01430*, *VIT_17s0000g09190*, *VIT_01s0011g01350*, *VIT_01s0011g03330*, *VIT_04s0008g05610* and *VIT_16s0022g00540* showed differential expression between heat-tolerant and heat-sensitive materials (Figure 8D). Among these, four genes (*VIT_04s0044g01430*, *VIT_17s0000g09190*, *VIT_01s0011g03330* and *VIT_04s0008g05610*) exhibited higher expression in heat-tolerant materials, suggesting their potential role as positive regulatory genes in heat tolerance, while two genes (*VIT_01s0011g01350* and *VIT_16s0022g00540*) showed higher expression in heat-sensitive materials, indicating their potential role as negative regulatory genes in heat tolerance (Figure 8D). In conclusion, we identified six candidate genes related to heat tolerance through qRT-PCR. These findings provide a theoretical basis for a deeper understanding of the molecular mechanisms underlying heat tolerance in grapes and offer new genetic resources for heat tolerance research in grapes.

4 Discussion

The impact of high-temperature stress on plant leaves directly reflects the plant's ability to cope with and resist heat stress (Zoong Lwe et al., 2021). Different varieties exhibit varying responses to high temperatures, with some maintaining better morphology and cellular structure under high-temperature conditions, demonstrating higher heat tolerance, while others are more susceptible to the effects of heat stress (Panigrahi et al., 1996; Li et al., 2022b). This variation may be attributed to genetic differences between varieties and variations in the plant's own adaptability to heat stress (Sadura and Janeczko, 2022; Kim et al., 2022). Leaf morphological changes and cellular structural damage induced by high-temperature stress may be associated with the interactions of multiple physiological and biochemical mechanisms (Hasanuzzaman et al., 2013; Sadok et al., 2021). For instance, leaf curling and yellowing may result from water imbalance within leaf cells and restricted photosynthesis (Kim et al., 2022; Li et al., 2022a). The reduced transpiration caused by high temperatures leads to decreased water evaporation and accumulation of moisture within the leaves, potentially causing leaf curling (Hasanuzzaman et al., 2013; Sadok et al., 2021). Simultaneously, the inhibition of photosynthesis by high temperatures can lower the chlorophyll content in leaves, resulting in leaf yellowing (Panigrahi et al., 1996; Li et al., 2022a). On the other hand, high-temperature stress may cause lipid peroxidation of cell membranes and ion imbalance, leading to cellular structural and functional damage (Hasanuzzaman et al., 2013; Sadok et al., 2021). Lipid peroxidation of the cell membrane disrupts its integrity, resulting in substance leakage and ion imbalance, which affects normal cell metabolism and function (Crockett, 2008; Thuwanut et al., 2009). Furthermore, high-temperature stress may also increase intracellular oxidative stress, further damaging cellular structure (Crockett, 2008; Thuwanut et al., 2009). Under different high-temperature stress conditions, we observed changes in the phenotypic characteristics of grape leaves and found that different grape varieties exhibited distinct morphological changes under high-temperature environments (Figure 2). This finding indicates variations in sensitivity and resistance to high temperatures among different varieties. Heat-tolerant materials may be able to regulate

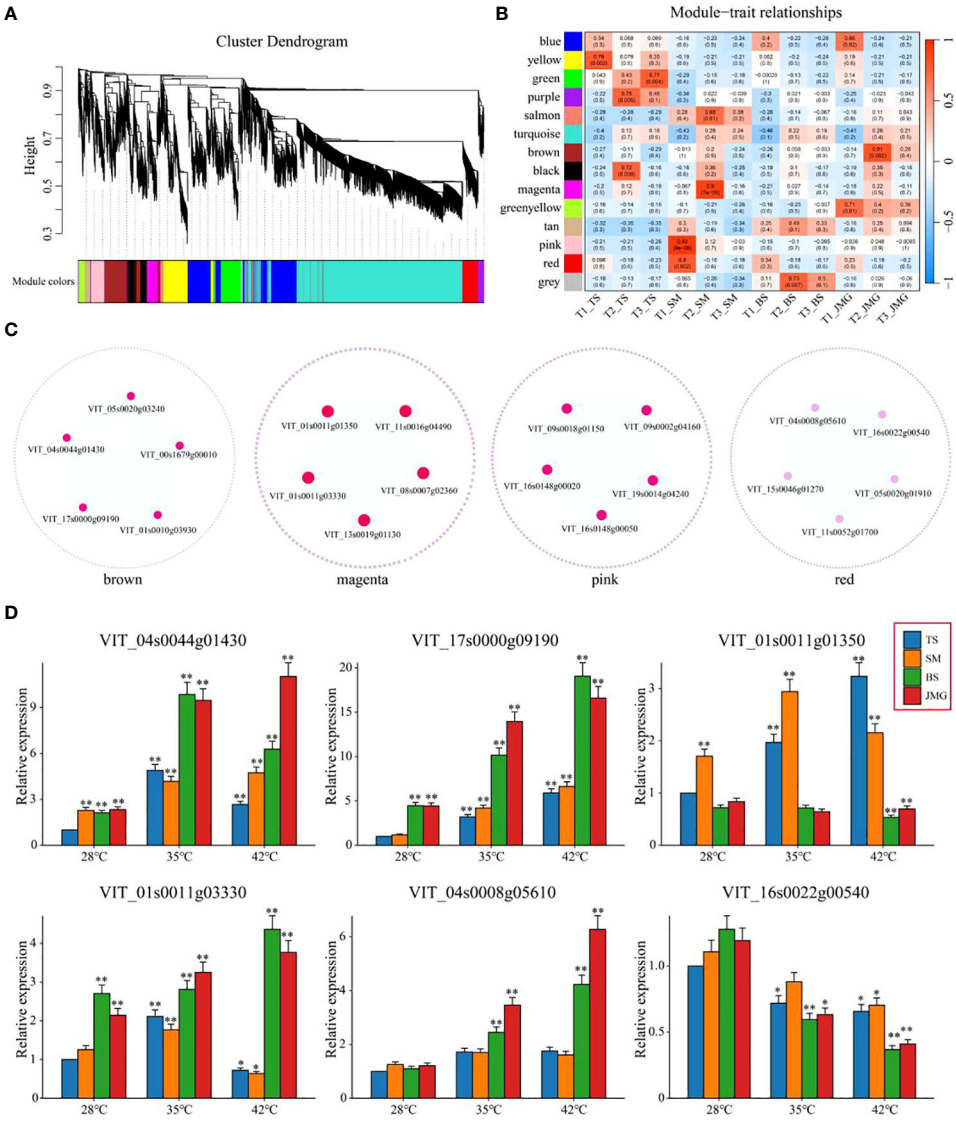


FIGURE 8
(A) Hierarchical clustering tree of genes based on coexpression network analysis. **(B)** Heatmap of correlation and significance between modules and materials at different temperatures. **(C)** Gene coexpression networks within specific modules. **(D)** qRT-PCR of grape heat-resistant hub genes. The results are presented as the means ± SDs (n = 3, $^{*}P < 0.05$,$^{**}P < 0.01$).

water balance and photosynthesis under high-temperature conditions, thereby reducing leaf dehydration and yellowing. Overall, the observed changes in the phenotypic characteristics of grape leaves and cellular structural damage under high-temperature stress reflect the plant's response and resistance level to heat stress.

Research on heat tolerance-related pathways in plants has always been a topic of intense interest in the fields of plant biology and agricultural science. With increasing attention being paid to global climate warming and heat stress, significant progress has been made in understanding the mechanisms of plant heat tolerance (Gouot et al., 2019a; Gouot et al., 2019b). Heat shock proteins (HSPs) are a class of proteins that are induced under high-temperature stress and help cells cope with protein instability and aggregation caused by heat (Li and Howell, 2021). The heat shock protein pathway includes processes such as synthesis, folding, and localization of HSPs, which protect cells from damage by maintaining protein stability and function under high-temperature stress (Khan et al., 2020). Moisture is one of the key factors in plant resistance to heat stress. Plants maintain water balance by regulating root water uptake and transpiration (Egawa et al., 2020). Studies have shown that some heat-tolerant plants have stronger root water uptake capacity and water retention ability, which can reduce water evaporation and dehydration under high-temperature environments, thereby lowering leaf temperature and damage. Plant hormones play an important role in regulating plant responses to high-temperature stress (Egawa et al., 2020). For example, hormones such as gibberellins, abscisic acid, and ethylene are involved in regulating plant growth, development, and stress tolerance. Studies have shown that the synthesis, signal transduction, and regulatory mechanisms of these hormones under

high-temperature stress have a significant impact on plant heat tolerance (Castroverde and Dina, 2021). Through enrichment analysis, we also identified the important roles of signaling pathways such as fatty acid metabolism, starch and sucrose metabolism, plant hormone signal transduction, the MAPK signaling pathway, and plant-pathogen interactions in grape heat tolerance processes.

In recent years, numerous heat-tolerant TFs have been identified in plants, and their regulation of target genes plays a crucial role in enhancing plant heat tolerance (Sahebi et al., 2018; Chauve et al., 2021). Heat shock transcription factors (HSFs) are considered to play a decisive role in this process. Studies in Arabidopsis have found that HSFA3, in addition to *HSFA2*, is another member of the HSF family that functions in heat stress memory (Friedrich et al., 2021). *HSFA3* can directly activate or maintain the hypermethylation of histone H3K4 to regulate the expression of genes related to heat stress memory (Friedrich et al., 2021). WRKY transcription factors play critical roles in plant responses to biotic and abiotic stresses. While there are reports demonstrating that overexpression of *OsWRKY11* under the promoter of the heat shock protein gene *HSP101* can enhance heat tolerance in rice, the *WRKY10* transcription factor negatively regulates rice heat tolerance through the regulation of ROS balance and hypersensitive responses, with its interacting protein *VQ8* playing an antagonistic role (Wu et al., 2009). In our study, we found that ERF, NAC, WRKY, MYB, and bHLH transcription factors in grapes may be associated with heat tolerance (Figure 7), providing reliable candidate genes for further investigation of the molecular mechanisms underlying heat tolerance in grapes.

Coexpression network analysis is a systems biology approach that involves analyzing the correlation of gene expression and constructing gene coexpression networks to discover functionally related gene modules. This analysis helps us identify candidate genes closely associated with our research (Langfelder and Horvath, 2008). Using WGCNA, 15 TFs related to poplar leaf blight, including *ATWRKY75*, *ANAC062*, *ATMYB23* and *ATEBP*, were identified, and these TFs exhibited high connectivity in the network (Wang et al., 2023). In potatoes, WGCNA identified 40 key candidate genes associated with development (Wei et al., 2023). In pepper, a key heat-tolerant gene, *CcBES1*, was discovered through WGCNA. *CcBES1* binds to the HSF promoter region in yeast, thereby regulating heat tolerance (Mumtaz et al., 2023). In our study, using transcriptomic data from leaf samples of grapes at different temperatures, we identified four highly significant gene modules through WGCNA, and a total of 20 heat stress-responsive genes were discovered. qRT-PCR analysis revealed that the expression of *VIT_04s0044g01430*, *VIT_17s0000g09190*, *VIT_01s0011g03330* and *VIT_04s0008g05610* was higher in heat-tolerant materials than in heat-sensitive materials. These four genes are potential positive regulators of heat tolerance in grapes. Among them *VIT_04s0044g01430* encode a Polyadenylate-binding protein (PABP) protein, and the functional annotation shows that it is mainly involved in response to light stimulus. *VIT_17s0000g09190* encodes a Phox and Bem1 (PB1) protein, and functional annotations show that it is primarily involved in the cellular lipid metabolic process. *VIT_01s0011g03330* encodes an Increased sodium tolerance protein 1 (IST1) protein, and the functional annotations show that it is primarily involved in Response to oxidative stress. *VIT_04s0008g05610* encodes a Coiled-Coil Domain-Containing (CCDC) protein, and the functional annotation shows that it is primarily involved in the Starch biosynthetic process. Conversely, the expression of *VIT_01s0011g01350* and *VIT_16s0022g00540* was higher in heat-sensitive materials than in heat-tolerant materials, suggesting that these two genes may be negative regulators of heat tolerance in grapes. *VIT_01s0011g01350* encodes a Valine-glutamine (VQ) protein, a gene with unknown function, *VIT_16s0022g00540* encodes a Glycerol-3-phosphate permease (G3Pp) protein, and functional annotations show that it is primarily involved in Transmembrane transport. Although the functions of these genes in grape heat tolerance require further validation, subsequent in-depth research can be conducted using biotechnological methods to elucidate their molecular mechanisms.

5 Conclusion

In summary, this study evaluated heat tolerance in four grape varieties and examined cellular structures in the naturally high-temperature environment of Turpan. RNA-seq analysis was performed, and the results showed that there were minimal differences in DEGs between the BS and JMG heat-tolerant materials, with a total of 1,890 DEGs identified. Additionally, the number of differentially expressed genes within materials was not significantly different, resulting in a total of 3,767 common DEGs. KEGG enrichment analysis revealed the enrichment of pathways such as fatty acid metabolism, starch and sucrose metabolism, plant hormone signal transduction, MAPK signaling pathway, and plant-pathogen interaction, both between and within the materials. By constructing a coexpression network, four specific modules highly associated with grape heat tolerance were identified, and six candidate genes related to heat tolerance were selected through qRT-PCR. These research findings provide a theoretical basis for a deeper understanding of the molecular mechanisms underlying grape heat tolerance and offer new genetic resources for studying grape heat tolerance.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://www.ncbi.nlm.nih.gov/sra/PRJNA914878.

Author contributions

JW: Data curation, Formal Analysis, Writing – original draft, Writing – review & editing. FZ: Methodology, Supervision, Validation, Writing – review & editing. GL: Methodology,

Supervision, Validation, Writing – review & editing. RA: Methodology, Supervision, Validation, Writing – review & editing. SB: Methodology, Supervision, Validation, Writing – review & editing. XYW: Methodology, Supervision, Validation, Writing – review & editing. CZ: Methodology, Supervision, Validation, Writing – review & editing. YM: Methodology, Supervision, Validation, Writing – review & editing. XPW: Methodology, Supervision, Validation, Writing – review & editing. QZ: Data curation, Formal Analysis, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing. HZ: Data curation, Formal Analysis, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

This research was funded by the Xinjiang Uygur Autonomous Region Tianshan Talents Training Program-Young top-notch scientific and technological talents (2022TSYCJC0036) ; by the Natural Science Foundation of Xinjiang Uygur Autonomous Region (2023D01A96); by the Xinjiang Uygur Autonomous Region Innovation Environment Construction Special Project (PT2314) ;by the Youth Science and Technology Backbone Innovation Ability Training Project of Xinjiang Academy of Agricultural Sciences (xjnkq-2021010), and by the Xinjiang Uygur Autonomous Region Tianchi Talent-Special Expert Project (Xiping Wang, 2022).

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.2023.1270933/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Correlation analysis of 36 RNA-seq samples.

SUPPLEMENTARY FIGURE 2
qRT-PCR of grape heat-resistant hub genes. The results are presented as the means ± SDs (n = 3).

References

Cao, L., Lu, X., Wang, G., Zhang, P., Fu, J., Wang, Z., et al. (2021). Transcriptional regulatory networks in response to drought stress and rewatering in maize (Zea mays L.). *Mol. Genet. Genomics* 296 (6), 1203–1219. doi: 10.1007/s00438-021-01820-y

Castroverde, C. D. M., and Dina, D. (2021). Temperature regulation of plant hormone signaling during stress and development. *J. Exp. Bot.* 72 (21), 7436–7458. doi: 10.1093/jxb/erab257

Chauve, L., Hodge, F., Murdoch, S., Masoudzadeh, F., Mann, H. J., Lopez-Clavijo, A. F., et al. (2021). Neuronal HSF-1 coordinates the propagation of fat desaturation across tissues to enable adaptation to high temperatures in C. Elegans. *PLoS Biol.* 19, e3001431. doi: 10.1371/journal.pbio.3001431

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34 (17), i884–i890. doi: 10.1093/bioinformatics/bty560

Correia, P. M. P., da Silva, A. B., Roitsch, T., Carmo-Silva, E., and Marques da Silva, J. (2021). Photoprotection and optimization of sucrose usage contribute to faster recovery of photosynthesis after water deficit at high temperatures in wheat. *Physiol. Plantarum* 172 (2), 615–628. doi: 10.1111/ppl.13227

Crockett, E. L. (2008). The cold but not hard fats in ectotherms: Consequences of lipid restructuring on susceptibility of biological membranes to peroxidation, a review. *J. Comp. Physiol. B.* 178 (7), 795–809. doi: 10.1007/s00360-008-0275-7

Egawa, M., Koizumi, K., and Hirao, T. (2020). Changes in facial moisture distribution and feelings of moisture/dryness among various environmental temperatures and humidities in summer and winter. *Skin Res. Technol.* 26 (6), 937–948. doi: 10.1111/srt.12898

Fleige, S., and Pfaffl, M. W. (2006). RNA integrity and the effect on the real-time qRT-PCR performance. *Mol. Aspects Med.* 27 (2-3), 126–139. doi: 10.1016/j.mam.2005.12.003

Friedrich, T., Oberkofler, V., Trindade, I., Altmann, S., Brzezinka, K., Lämke, J., et al. (2021). Heteromeric HSFA2/HSFA3 complexes drive transcriptional memory after heat stress in Arabidopsis. *Nat. Commun.* 12 (1), 3426. doi: 10.1038/s41467-021-23786-6

Gong, W., Proud, C., Fukai, S., and Mitchell, J. (2023). Low canopy temperature and high stomatal conductance contribute to high grain yield of contrasting japonica rice under aerobic conditions. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1176156

Gouot, J. C., Smith, J. P., Holzapfel, B. P., and Barril, C. (2019a). Grape berry flavonoid responses to high bunch temperatures post véraison: effect of intensity and duration of exposure. *Molecules* 24 (23), 4341. doi: 10.3390/molecules24234341

Gouot, J. C., Smith, J. P., Holzapfel, B. P., Walker, A. R., and Barril, C. (2019b). Grape berry flavonoids: a review of their biochemical responses to high and extreme high temperatures. *J. Exp. Bot.* 70 (2), 397–423. doi: 10.1093/jxb/ery392

Hasanuzzaman, M., Nahar, K., Alam, M. M., Roychowdhury, R., and Fujita, M. (2013). Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. *Int. J. Mol. Sci.* 14 (5), 9643–9684. doi: 10.3390/ijms14059643

IPCC (2021). *Climate Change 2021:The physical science basis//contribution of Working Group I contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge: Cambridge University Press).

Jin, J., Tian, F., Yang, D., Meng, Y., Kong, L., Luo, J., et al. (2017). PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45 (D1), D1040–D1045. doi: 10.1093/nar/gkw982

Ju, Y. L., Yue, X. F., Min, Z., Wang, X. H., Fang, Y. L., and Zhang, J. X. (2020). VvNAC17, a novel stress-responsive grapevine (*Vitis vinifera* L.) NAC transcription factor, increases sensitivity to abscisic acid and enhances salinity, freezing, and drought tolerance in transgenic Arabidopsis. *Plant Physiol. Biochem.* 146, 98–111. doi: 10.1016/j.plaphy.2019.11.002

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28 (1), 27–30. doi: 10.1093/nar/28.1.27

Khan, A., Ahmad, M., Ahmed, M., and Iftikhar Hussain, M. (2020). Rising atmospheric temperature impact on wheat and thermotolerance strategies. *Plants-Basel* 10 (1), 43. doi: 10.3390/plants10010043

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37 (8), 907–915. doi: 10.1038/s41587-019-0201-4

Kim, H. S., Shin, J. H., Lee, H. S., Kim, S., Jang, H. Y., Kim, E., et al. (2022). CsRCI2D enhances high-temperature stress tolerance in Camelina sativa L. through endo-membrane trafficking from the plasma membrane. *Plant Sci.* 320, 111294. doi: 10.1016/j.plantsci.2022.111294

Kusser, W., Javorschi, S., and Gleeson, M. A. (2006). Real-time RT-PCR: cDNA synthesis. *Cold Spring Harb. Protoc.* 2006 (1), pdb–prot4114. doi: 10.1101/pdb.prot4114

Landolt, L., Marti, H. P., Beisland, C., Flatberg, A., and Eikrem, O. S. (2016). RNA extraction for RNA sequencing of archival renal tissues. *Scand. J. Clin. Lab. Inv.* 76 (5), 426–434. doi: 10.1080/00365513.2016.1177660

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9, 559. doi: 10.1186/1471-2105-9-559

Li, Z., and Howell, S. H. (2021). Heat stress responses and thermotolerance in maize. *Int. J. Mol. Sci.* 22 (2), 948. doi: 10.3390/ijms22020948

Li, P., Lin, P., Zhao, Z., Li, Z., Liu, Y., Huang, C., et al. (2022a). Gene co-expression analysis reveals transcriptome divergence between wild and cultivated sugarcane under drought stress. *Int. J. Mol. Sci.* 23 (1), 569. doi: 10.3390/ijms23010569

Li, Y., Qi, X., Wang, K., Gu, J., Zhao, J., Hu, X., et al. (2022b). Response of the water footprint of maize production to high temperatures in the Huang-Huai-Hai region of China. *J. Sci. Food Agr.* 102 (14), 6539–6554. doi: 10.1002/jsfa.12020

Lian, P., Braber, S., Varasteh, S., Wichers, H. J., and Folkerts, G. (2021). Hypoxia and heat stress affect epithelial integrity in a Caco-2/HT-29 co-culture. *Sci. Rep.* 11 (1), 13186. doi: 10.1038/s41598-021-92574-5

Liu, S., Wang, Z., Zhu, R., Wang, F., Cheng, Y., and Liu, Y. (2021). Three differential expression analysis methods for RNA sequencing: limma, edgeR, DESeq2. *J. Vis. Exp.* 175, e62528. doi: 10.3791/62528

Liu, S., Ye, T., Li, Z., Li, J., Jamil, A. M., Zhou, Y., et al. (2019). Identifying hub genes for heat tolerance in water buffalo (Bubalus bubalis) using transcriptome data. *Front. Genet.* 10. doi: 10.3389/fgene.2019.00209

Meng, H. L., Sun, P. Y., Wang, J. R., Sun, X. Q., Zheng, C. Z., Fan, T., et al. (2022). Comparative physiological, transcriptomic, and WGCNA analyses reveal the key genes and regulatory pathways associated with drought tolerance in Tartary buckwheat. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.985088

Mori, K., Goto-Yamamoto, N., Kitayama, M., and Hashizume, K. (2007). oss of anthocyanins in red-wine grape under high temperature. *J. Exp. Bot.* 58 (8), 1935–1945. doi: 10.1093/jxb/erm055

Mumtaz, M. A., Zhou, Y., Gao, C., Kamran, H. M., Altaf, M. A., Hao, Y., et al. (2023). Interaction between transcriptional activator BRI1-EMS-SUPPRESSOR 1 and HSPs regulates heat stress tolerance in pepper. *Environ. Exp. Bot.* 211, 105341. doi: 10.1016/j.envexpbot.2023.105341

Panigrahi, S., Oguntona, E. B., and Roberts, B. R. (1996). Effects of oven-drying tubers of two high-protein sweet potato varieties at different temperatures on their feeding value in broilers. *Brit. Poultry Sci.* 37 (1), 173–188. doi: 10.1080/00071669608417846

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 (9), 1650–1667. doi: 10.1038/nprot.2016.095

Ren, C., Kuang, Y., Lin, Y., Guo, Y., Li, H., Fan, P., et al. (2022). Overexpression of grape ABA receptor gene VaPYL4 enhances tolerance to multiple abiotic stresses in Arabidopsis. *BMC Plant Biol.* 22 (1), 271. doi: 10.1186/s12870-022-03663-0

Sadok, W., Lopez, J. R., and Smith, K. P. (2021). Transpiration increases under high-temperature stress: Potential mechanisms, trade-offs and prospects for crop resilience in a warming world. *Plant Cell Environ.* 44 (7), 2102–2116. doi: 10.1111/pce.13970

Sadura, I., and Janeczko, A. (2022). Brassinosteroids and the tolerance of cereals to low and high temperature stress: photosynthesis and the physicochemical properties of cell membranes. *Int. J. Mol. Sci.* 23 (1), 342. doi: 10.3390/ijms23010342

Sahebi, M., Hanafi, M. M., Rafii, M. Y., Mahmud, T. M. M., Azizi, P., Osman, M., et al. (2018). Improvement of drought tolerance in rice (Oryza sativa L.): genetics, genomic tools, and the WRKY gene family. *BioMed. Res. Int.* 2018, 3158474. doi: 10.1155/2018/3158474

Shannon, P., Markiel, 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 (11), 2498–2504. doi: 10.1101/gr.1239303

Sun, Y., Gao, Y., Wang, H., Yang, X., Zhai, H., and Du, Y. (2018). Stimulation of cyclic electron flow around PSI as a response to the combined stress of high light and high temperature in grape leaves. *Funct. Plant Biol.* 45 (10), 1038–1045. doi: 10.1071/FP17269

Thuwanut, P., Axnér, E., Johanisson, A., and Chatdarong, K. (2009). Detection of lipid peroxidation reaction in frozen-thawed epididymal cat spermatozoa using BODIPY(581/591) C11. *Reprod. Domest. Anim.* 44, 373–376. doi: 10.1111/j.1439-0531.2009.01453.x

Tian, Z., He, W., Tang, J., Liao, X., Yang, Q., Wu, Y., et al. (2020). Identification of important modules and biomarkers in breast cancer based on WGCNA. *Oncotargets Ther.* 13, 6805–6817. doi: 10.2147/OTT.S258439

Wang, Y., Wang, Y., Liu, X., Zhou, J., Deng, H., Zhang, G., et al. (2022). WGCNA analysis identifies the hub genes related to heat stress in seedling of rice (Oryza sativa L.). *Genes* 13 (6), 1020. doi: 10.3390/genes13061020

Wang, R., Wang, Y., Yao, W., Ge, W., Jiang, T., and Zhou, B. (2023). Transcriptome sequencing and WGCNA reveal key genes in response to leaf blight in poplar. *Int. J. Mol. Sci.* 24 (12), 10047. doi: 10.3390/ijms241210047

Wei, H., Xu, X., Feng, G., Shao, S., Chen, X., and Yang, Z. (2023). Candidate genes potentially involved in molting and body size reduction in the male of the horned gall aphid, Schlechtendalia chinensis. *Front. Physiol.* 14. doi: 10.3389/fphys.2023.1097317

Wu, X., Shiroto, Y., Kishitani, S., Ito, Y., and Toriyama, K. (2009). Enhanced heat and drought tolerance in transgenic rice seedlings overexpressing OsWRKY11 under the control of HSP101 promoter. *Plant Cell Rep.* 28, 21–30. doi: 10.1007/s00299-008-0614-x

Zhao, D. Q., Li, T. T., Hao, Z. J., Cheng, M. L., and Tao, J. (2019). Exogenous trehalose confers high temperature stress tolerance to herbaceous peony by enhancing antioxidant systems, activating photosynthesis, and protecting cell structure. *Cell Stress Chaperon.* 24, 247–257. doi: 10.1007/s12192-018-00961-1

Zoong Lwe, Z., Sah, S., Persaud, L., Li, J., Gao, W., Raja Reddy, K., et al. (2021). Alterations in the leaf lipidome of Brassica carinata under high-temperature stress. *BMC Plant Biol.* 21, 404. doi: 10.1186/s12870-021-03189-x

frontiers | Frontiers in Plant Science

TYPE Opinion
PUBLISHED 27 October 2023
DOI 10.3389/fpls.2023.1295528

OPEN ACCESS

EDITED BY
Yi Han,
Anhui Agricultural University, China

REVIEWED BY
Yongping Li,
Hainan University, China
Tao Zhao,
Northwest A&F University, China

*CORRESPONDENCE
Zhao Zhang
zhangzhao@cau.edu.cn

RECEIVED 16 September 2023
ACCEPTED 18 October 2023
PUBLISHED 27 October 2023

CITATION
Ding C and Zhang Z (2023) Effective omics tools are still lacking for improvement of stress tolerance in polyploid crops. *Front. Plant Sci.* 14:1295528. doi: 10.3389/fpls.2023.1295528

Effective omics tools are still lacking for improvement of stress tolerance in polyploid crops

Chao Ding[1] and Zhao Zhang[2*]

[1]Shanxi Center for Testing of Functional Agro-Products, Shanxi Agricultural University, Taiyuan, China, [2]Beijing Key Laboratory of Development and Quality Control of Ornamental Crops, Department of Ornamental Horticulture, China Agricultural University, Beijing, China

KEYWORDS

abiotic stress, disease resistance, epigenetic regulation, post-transcriptional regulation, phenomics

Introduction

After decades of development, today's Omics technologies have advanced tremendously. It is important to note, however, that the vast majority of methodologies in bioinformatics were originally developed for the study of human genetics. Unlike diploid humans, many crop genomes are polyploid, such as tetraploid potato (Sun et al., 2022), hexaploid wheat (Ling et al., 2018; Walkowiak et al., 2020), tetraploid sugarcane (Zhang et al., 2018), octoploid strawberry (Edger et al., 2019), and so on. Existing Omics methods and tools, including genome assembly strategies, algorithm, statistical methods, and efficient data processing and analysis, are far from adequate for the complex genomes of polyploid crops.

Challenges of genome assembly and annotation

Genome assembly and annotation of polyploid plants is a challenging task due to the complexity of their genomes. Although some new technological tools, such as the highly accurate long-read sequencing technologies, e.g. PacBio HiFi (Hon et al., 2020), have made it possible to accurately obtain genomic fragments that distinguish haplotypes. However, the presence of highly similar sequences tends to make assembly of repetitive regions difficult. Therefore, obtaining haplotype-resolved polyploid genomes of high quality is still a difficult task. It is worth noting that compared with allopolyploid plants, this issue is more prominent for autopolyploid plants. People have attempted to assemble the chromosome-level genome of polyploid plants using High-throughput chromosome conformation capture (Hi-C) technology (Dekker, 2006) and ALLHiC software (Zhang et al., 2019),

but this approach fails in autopolyploid plants in many cases (Bao et al., 2022; Sun et al., 2022). Due to the lack of reliable bioinformatic methods, obtaining high-quality autopolyploid plant genomes can only be achieved through experimental methods. For example, the genome of autotetraploid potato was obtained by sequencing a total of 1034 individuals from a selfing population (F2 population; Bao et al., 2022), which obviously involves significant time and cost.

Furthermore, this challenge becomes even more pronounced for crops with highly heterozygous genomes. The presence of multiple alleles and variations within the genome adds an extra layer of complexity to the assembly and annotation processes. Resolving the allelic diversity and accurately assigning variations to specific loci can be particularly challenging in highly heterozygous crops. Advanced computational algorithms and innovative sequencing technologies are continually being developed to address these complexities and improve the accuracy of genome assembly and annotation in highly heterozygous crops.

Complex genome structure

Polyploid plants typically possess large genomes, e.g. approximately 16 Gb for wheat (Walkowiak et al., 2020) and 7.68 Gb for tetraploid rhubarb (*Rheum officinale*, a perennial medicinal herb) (Zhang et al., 2023a). For the crops that rely mainly on asexual propagation (many of which are horticultural), e.g. the tetraploid modern rose (*Rosa hybrida*), their genomes maintain a high level of heterozygosity due to a lack of sexual reproduction (Zhang et al., 2023b). The presence of multiple homologous copies (alleles) in the genome increases complexity to the analysis and interpretation of genomic information. Distinguishing different alleles and resolving dosage effect of gene can be challenging. Different studies have found that the expression between alleles in polyploid plants can vary significantly or may differ only slightly (Zhang et al., 2023b).

Complexity of epigenetic regulation and post-transcriptional regulation

In addition to genomic and transcriptomic complexity, polyploid plants have complex epigenetic mechanisms. Epigenetic regulation involves dynamic modifications of nucleic acids and histones, encompassing both genomic and transcriptional levels (Shen et al., 2017). Additionally, non-coding RNAs, such as miRNA, siRNA, piRNA and long non-coding RNAs, are involved in the epigenetic regulation of gene expression by targeting mRNA stability and translation (Shi et al., 2022). Epigenetic modifications may have important effects on gene expression in plants, but the complexity of the epigenetic regulatory network makes it difficult to accurately understand and interpret (Ma et al., 2015). The complexity of epigenetic regulation lies in its multilevel nature, involving interactions among DNA methylation (and methylome), histone modifications, chromatin remodeling, noncoding RNAs, and various regulatory proteins. Moreover, the three-dimensional organization of chromatin within the nucleus, refers as 3D genome, have significant implications for gene regulation and other cellular processes. Understanding the 3D genome and its dynamics provides insights into how genetic information is regulated spatially and temporally, impacting processes like plant growth, development, stress responses and so on. The interactions between different subgenomes in polyploid plants undoubtedly add another layer of complexity of their epigenetic regulation.

Reliable phenomics is necessary for effective interpretation of complex traits

In most cases, stress tolerance in plants is a complex trait. This applies not only to abiotic stress tolerance but also to resistance against biotic stress. Plant complex traits are typically determined by multiple genes and their interactions. These genes may be located on different chromosomes, making direct linkage analysis challenging, and polyploidy further exacerbates this complexity. Moreover, the effects of these genes are often subtle, resulting in minimal phenotypic differences from changes in a single gene. Therefore, reliable and high-throughput phenomics is required. Accurate assessment of phenotypes is essential for effective genetic analysis.

Data analysis require appropriate statistical models and efficient computation

The extensive data generated from polyploid plant genomics research requires complex analysis and interpretation. Challenges include standardized data processing, establishing effective analysis pipelines, and developing suitable statistical models. Additionally, integrating genomic data with phenotype information poses a further challenge in gaining deeper insights into the characteristics of polyploid plants. These tasks often rely on powerful parallel computational capabilities, and thus, developing efficient computations based on Graphics Processing Units (GPUs) can significantly enhance performance.

Prospects

In the past decade, many software tools applicable to polyploids have been developed, such as fitTetra (Voorrips et al., 2011; Zych et al., 2019) and polymapR (Bourke et al., 2018). However, genomic research on polyploid plants still faces challenges regarding complex genome structure, genome assembly and annotation, the complexity of epigenetic regulation, as well as the complexity of data analysis and interpretation (Figure 1). Overcoming these

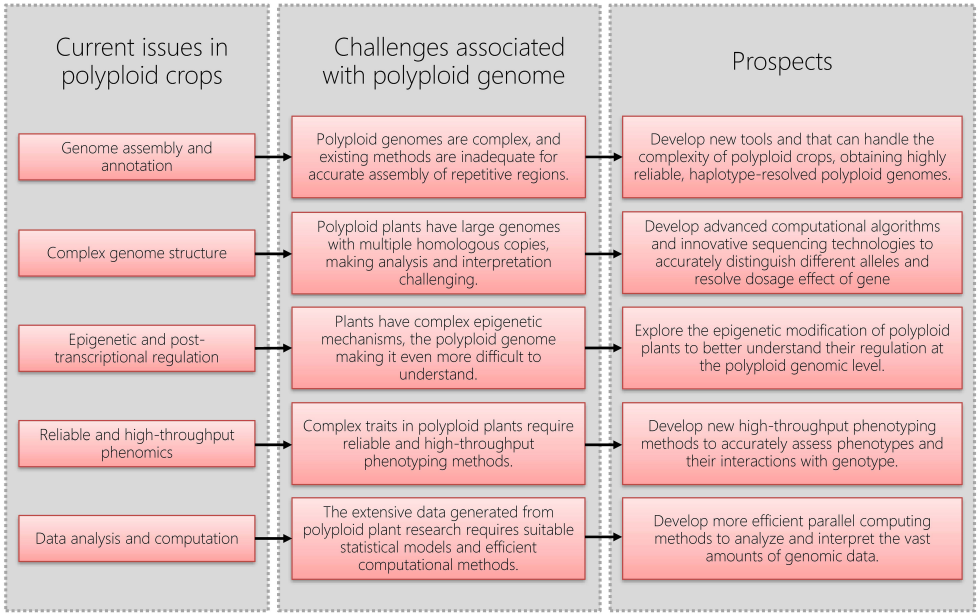


FIGURE 1
Challenges and prospects in polyploid crop genome research.

challenges requires continuous technological development and innovation, as well as enhanced collaboration and communication across different disciplines, to further advance genomic research on polyploid plants.

Author contributions

CD: Writing – original draft, Writing – review & editing. ZZ: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Special Project for Science and Technology Cooperation and Exchange of Shanxi province (Grant No.202204041101017) to Chao Ding and Zhao Zhang.

Conflict of interest

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

Publisher's note

References

Bao, Z., Li, C., Li, G., Wang, P., Peng, Z., Cheng, L., et al. (2022). Genome architecture and tetrasomic inheritance of autotetraploid potato. *Mol. Plant* 15 (7), 1211–1226. doi: 10.1016/j.molp.2022.06.009

Bourke, P. M., van Geest, G., Voorrips, R. E., Jansen, J., Kranenburg, T., Shahin, A., et al. (2018). polymapR-linkage analysis and genetic map construction from F1 populations of outcrossing polyploids. *Bioinformatics* 34 (20), 3496–3502. doi: 10.1093/bioinformatics/bty371

Dekker, J. (2006). The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nat. Methods* 3 (1), 17–21. doi: 10.1038/nmeth823

Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., et al. (2019). Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* 51 (3), 541–547. doi: 10.1038/s41588-019-0356-4

Hon, T., Mars, K., Young, G., Tsai, Y.-C., Karalius, J. W., Landolin, J. M., et al. (2020). Highly accurate long-read HiFi sequencing data for five complex genomes. *Sci. Data* 7 (1), 399. doi: 10.1038/s41597-020-00743-4

Ling, H. Q., Ma, B., Shi, X., Liu, H., Dong, L., Sun, H., et al. (2018). Genome sequence of the progenitor of wheat A subgenome Triticum urartu. *Nature* 557 (7705), 424–428. doi: 10.1038/s41586-018-0108-0

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

Shen, Y., Sun, S., Hua, S., Shen, E., Ye, C. Y., Cai, D., et al. (2017). Analysis of transcriptional and epigenetic changes in hybrid vigor of allopolyploid Brassica

napus uncovers key roles for small RNAs. *Plant J.* 91 (5), 874–893. doi: 10.1111/tpj.13605

Shi, S., Zhang, S., Wu, J., Liu, X., and Zhang, Z. (2022). Identification of long non-coding RNAs involved in floral scent of Rosa hybrida. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.996474

Sun, H., Jiao, W. B., Krause, K., Campoy, J. A., Goel, M., Folz-Donahue, K., et al. (2022). Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. *Nat. Genet.* 54 (3), 342–348. doi: 10.1038/s41588-022-01015-0

Voorrips, R. E., Gort, G., and Vosman, B. (2011). Genotype calling in tetraploid species from bi-allelic marker data using mixture models. *BMC Bioinf.* 12, 172. doi: 10.1186/1471-2105-12-172

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M. T., Brinton, J., et al. (2020). Multiple wheat genomes reveal global variation in modern breeding. *Nature* 588 (7837), 277–283. doi: 10.1038/s41586-020-2961-x

Zhang, H., He, Q., Xing, L., Wang, R., Wang, Y., Liu, Y., et al. (2023a). The haplotype-resolved genome assembly of autotetraploid rhubarb Rheum officinale provides insights into the genome evolution and massive accumulation of anthraquinones. *Plant Commun.* 100677. doi: 10.1016/j.xplc.2023.100677

Zhang, Z., Liu, Y., Yang, T., Wu, S., Sun, H., Wu, J., et al. (2023b). Haplotype-resolve genome assembly and resequencing provide insights into the origin and domestication of modern rose. *bioRxiv* 2006. doi: 10.1101/2023.06.02.543351

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., et al. (2018). Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum L. *Nat. Genet.* 50 (11), 1565–1573. doi: 10.1038/s41588-018-0237-2

Zhang, X., Zhang, S., Zhao, Q., Ming, R., and Tang, H. (2019). Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nat. Plants* 5 (8), 833–845. doi: 10.1038/s41477-019-0487-8

Zych, K., Gort, G., Maliepaard, C. A., Jansen, R. C., and Voorrips, R. E. (2019). FitTetra 2.0 – improved genotype calling for tetraploids with multiple population and parental data support. *BMC Bioinf.* 20 (1), 148. doi: 10.1186/s12859-019-2703-y

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 19 December 2023
DOI 10.3389/fpls.2023.1246677

OPEN ACCESS

EDITED BY
Jian Chen,
Jiangsu University, China

REVIEWED BY
Muhammad Waheed Riaz,
Zhejiang Agriculture and Forestry University, China
Abdul Rehman,
Zhengzhou University, China

*CORRESPONDENCE
Wuwei Ye
yew158@163.com
Quanjia Chen
chqjia@126.com

†These authors have contributed equally to this work

RECEIVED 04 August 2023
ACCEPTED 27 November 2023
PUBLISHED 19 December 2023

CITATION
Chen W, Cui Y, He Y, Zhao L, Cui R, Liu X, Huang H, Zhang Y, Fan Y, Feng X, Ni K, Jiang T, Han M, Lei Y, Liu M, Meng Y, Chen X, Lu X, Wang D, Wang J, Wang S, Guo L, Chen Q and Ye W (2023) Raffinose degradation-related gene *GhAGAL3* was screened out responding to salinity stress through expression patterns of *GhAGALs* family genes.
Front. Plant Sci. 14:1246677.
doi: 10.3389/fpls.2023.1246677

Raffinose degradation-related gene *GhAGAL3* was screened out responding to salinity stress through expression patterns of *GhAGALs* family genes

Wenhua Chen[1,2†], Yupeng Cui[1†], Yunxin He[3†], Lanjie Zhao[1†], Ruifeng Cui[1†], Xiaoyu Liu[1], Hui Huang[1], Yuexin Zhang[1], Yapeng Fan[1], Xixian Feng[1], Kesong Ni[1], Tiantian Jiang[1], Mingge Han[1], Yuqian Lei[1], Mengyue Liu[1], Yuan Meng[1], Xiugui Chen[1], Xuke Lu[1], Delong Wang[1], Junjuan Wang[1], Shuai Wang[1], Lixue Guo[1], Quanjia Chen[2*] and Wuwei Ye[1,2*]

[1]Institute of Cotton Research of Chinese Academy of Agricultural Sciences/Research Base, Anyang Institute of Technology, National Key Laboratory of Cotton Bio-breeding and Integrated Utilization, Anyang, Henan, China, [2]Engineering Research Centre of Cotton, Ministry of Education/College of Agriculture, Xinjiang Agricultural University, Urumqi, China, [3]Hunan Institute of Cotton Science, Changde, Hunan, China

A-galactosidases (AGALs), the oligosaccharide (RFO) catabolic genes of the raffinose family, play crucial roles in plant growth and development and in adversity stress. They can break down the non-reducing terminal galactose residues of glycolipids and sugar chains. In this study, the whole genome of AGALs was analyzed. Bioinformatics analysis was conducted to analyze members of the AGAL family in *Gossypium hirsutum*, *Gossypium arboreum*, *Gossypium barbadense*, and *Gossypium raimondii.* Meanwhile, RT-qPCR was carried out to analyze the expression patterns of AGAL family members in different tissues of terrestrial cotton. It was found that a series of environmental factors stimulated the expression of the *GhAGAL3* gene. The function of *GhAGAL3* was verified through virus-induced gene silencing (VIGS). As a result, *GhAGAL3* gene silencing resulted in milder wilting of seedlings than the controls, and a significant increase in the raffinose content in cotton, indicating that *GhAGAL3* responded to NaCl stress. The increase in raffinose content improved the tolerance of cotton. Findings in this study lay an important foundation for further research on the role of the *GhAGAL3* gene family in the molecular mechanism of abiotic stress resistance in cotton.

KEYWORDS

α-galactosidase (AGALs), *Gossypium hirsutum*, raffinose family oligosaccharides (RFOs), abiotic stresses, functional verification

1 Introduction

Plants are subjected to a variety of abiotic stresses in their growth process, including drought, salt, high temperature, and cold stresses. To enable the normal growth of plants under various stress conditions, plants have developed multiple signaling and regulatory pathways to withstand stress and thrive. Raffinose family oligosaccharides (RFOs), the essential substances mediating stress response in plants, play an important role in enhancing plant resistance to abiotic stresses (Li et al., 2011). Studies have reported that the expression of genes related to RFO metabolites is significantly upregulated under stress conditions such as low temperature, drought, and high salt. These findings suggest that RFOs may have a crucial effect on enhancing resistance to abiotic stresses (Joshi et al., 2021; Lee et al., 2021). In addition, studies have reported that RFOs can facilitate the elimination of reactive oxygen species (ROS) in plants experiencing stress (Ma et al., 2021). Raffinose can be transported to the chloroplast, protects the thylakoid, and maintains the stability of the photosynthetic system, PSII (Schneider and Keller, 2009). As ROS scavengers, both raffinose and galactose can trap hydroxyl radicals and reduce oxidative damage in plants under stressful conditions (Nishizawa-Yokoi et al., 2008). AGAL, the gene that breaks down RFOs in plants, has received little attention, but it is just as important as the gene that synthesizes RFOs.

AGAL is a ubiquitous enzyme in the plant kingdom (Wang et al., 2022). AGALs of Family 26 are of eukaryotic origin, whereas the AGALs of Family 37 is mainly of prokaryotic origin (Fialho Lda et al., 2008). A notable function of this enzyme is in the germination of seeds and tubers (Hughes et al., 2016). AGAL exhibits a high activity during seed maturation, during germination, and in the seedling stage, which can be attributed to hydration factors that result in the complete degradation of soluble sugars in the embryo axis and primarily in the cotyledons (Mittal and Sharma, 1991). AGAL is involved in numerous aspects of plant metabolism, including the hydrolysis of α-1,6 chains of cotton seed oligosaccharides during deforming. Downregulation of *α-Gal* gene expression in Petunia spp. results in increased whole plant frost resistance in non-domesticated and cold-domesticated plants. In contrast, overexpression of the *α-Gal* gene leads to the reduced endogenous raffinose and decreased frost resistance (Pennycooke et al., 2003). *Osh69*, a rice alkaline galactosidase, was found to be located in chloroplasts and to play a role in leaf senescence. It is upregulated in darkness and in response to injury (Lee et al., 2004). Meanwhile, the *AGA3* protein was identified in cucumber chloroplasts and its expression was downregulated under cold stress. However, it was upregulated after temperature recovery. This strongly suggests that *AGA3* plays an important role in the catabolism of chloroplast RFOs under cold stress. Notably, the alkaline environment of the chloroplast stroma is suitable for *AGA3* to exert its catalytic action. *AGA2*, another alkaline galactosidase, is responsible for the catabolism of RFOs in the cytoplasm once the temperature returns to the normal level. RFOs accumulate in different subcellular compartments of cucumber leaves under cold stress, whereas upon the removal of the stress, they can be broken down *in situ* by various galactosidases (Sui et al., 2012; Gu et al., 2018). In the research on spinach in New Zealand, the expression of the *TtAGAl1* gene was affected by various abiotic stresses. Among them, drought was found to be a particularly strong promoter of *TtAGAl* expression (Hara et al., 2008). Moreover, the expression of related genes in *Vitis vinifera* was studied under salt and drought stresses as well as and during seed development. It was found that the *Vv-a-gal/SIP* gene is differentially expressed under different osmotic stress conditions. Additionally, *Vv-α-gal/SIP*-specific transcripts were preferentially accumulated in salt-tolerant *Vitis vinifera* varieties (Daldoul et al., 2012). *ZmAGA1* expression was enhanced in maize seedlings under cold and drought stress conditions, but not upon sodium chloride stress (Zhao et al., 2006). *AGAL* is an essential enzyme for plant growth and development, which plays a crucial role in plant development Although characterized *AGAL* has been studied in several species, it has not been reported in cotton. It is crucial to delve into the mechanisms underlying the salinity stress resistance in cotton.

Cotton is an economically important crop that is grown worldwide for the production of fiber and cottonseed oil. However, stress conditions often affect the growth and development of cotton, thereby reducing its yield. Different abiotic stresses may have diverse effects on cotton (Bawa et al., 2022). With the rapid development of sequencing technology, the genome sequences of *Gossypium hirsutum*, *Gossypium arboreum*, *Gossypium barbadense*, and *Gossypium raimondii* have been sequenced and resolved. In this study, the *AGAL* family of cotton was identified and characterized by bioinformatics analyses (Huang et al., 2022).The expression pattern of the *GhAGALs* gene was also analyzed, and the results showed that *GhAGALs* expression varied in different tissues. In addition, the *AGAL* gene plays a role in the function of *Gossypium hirsutum* and its specific site of action. Findings in this study provide a theoretical foundation for future studies on the utilization of the *AGAL* family in *Gossypium hirsutum.*

2 Experimental materials and methods

2.1 Identification of AGAL family members and construction of evolutionary tree

The protein sequences of four major cotton species *Gossypium hirsutum*, *Gossypium arboreum*, *Gossypium barbadense*, and *Gossypium raimondii* were downloaded from the cotton database Cotton FGD (https://cottonfgd.org), *Gossypium hirsutum* (ZJU), *Gossypium arboreum* (CRI), *Gossypium barbadense* (ZJU), *Gossypium raimondii* (JGI), the genome sequence, and the CDS sequence (Zhu et al., 2017). Additionally, protein sequences, genomes, and other biological information of five diploid species were downloaded using the online database JGI Phytozome v12.1. *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, *Vitis vinifera*, and *Zea mays* (Wang et al., 2021). Furthermore, the hidden Markov model (HMM) file for the conserved domain

(PF16499) was downloaded from the Pfam database (https://pfam.xfam.org/). Hidden Markov models were used with (PF16499) as the query file to search for candidate *AGAL* family genes in the genomes of cotton and other species. This search was conducted using HMMER (version 3.3.1) (http://www.hmmer.org/) and BLASTP. Manual removal of redundant sequences from HMMER (version 3.3.1) and BLASTP results for incomplete genes (Henderson et al., 1997). These genes were identified as members of the *AGAL* gene family. The identified *AGAL* family members were renamed *GhAGAL1-GhAGAL15* based on their chromosomal positions (Zhang et al., 2021).

To clearly understand the evolutionary relationship of the *AGAL* gene, those identified *AGAL* gene family members were used to search for protein sequences of cotton and five diploid species: *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, *Vitis vinifera*, and *Zea mays*. We employed using HMMER (version 3.3.1) and BLASTP for this purpose. Multiple sequence alignment was performed using MEGA 7.0 software (Cui et al., 2021b). Construction of intraspecific and interspecific evolutionary trees for *AGAL* using the online software ChiPlot (Xie et al., 2023) online tools to enhance the visualization of evolutionary trees (Kumar et al., 2016).

2.2 Chromosome localization in the *AGAL* family

Biological information concerning the location and structure of *AGAL* family members was extracted from the genome gff3 annotation files of four cotton species (Chen et al., 2018). Afterward, locations of the cotton *AGAL* family genes on the chromosome were analyzed and mapped with TBtools software (Chen et al., 2020; Cai et al., 2021), so as to reveal the coevolutionary relationships between the *AGAL* families of cotton species.

2.3 Gene structure and conserved motif analysis

Phylogenetic tree, gene structure, and conserved protein motif were triply mapped with MAST files, GhAGALs evolutionary tree NWK files, and gff3 files by adopting TBtools software.

2.4 Analysis of *GhAGALs* gene promoter elements and expression level analysis

To investigate the relationships of the *AGAL* family with hormones and stress, the upstream 2,000-bp sequence (Malik et al., 2020) of the start codon in the *AGAL* family was selected from the *Gossypium hirsutum* genome sequence by using the PlantCARE data database query system (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al., 2002). Afterward, *cis*-acting elements in the upstream regions of genes were identified, analyzed and later visualized with TBtools after screening (Chen et al., 2020).

2.5 Analysis of the tissue expression pattern of *GhAGALs* family genes

To analyze the expression of *GhAGALs* in various tissues, *GhAGALs* data were obtained from the online database of the Cotton Research Institute (http://grand.cricaas.com.cn/page/tools/expressionVisualization). In addition, representative organization values for root, stem, and leaf tissues were selected based on the characteristics of the *GhAGALs* family. The original data were analyzed to obtain the gene expression levels, which were subsequently visualized using a histogram.

2.6 Gene interaction network

The GhAGAL protein interaction network was analyzed based on the STRING database (https://string-db.org/) after taking *Arabidopsis thaliana* orthologs into consideration, so as to predict the interactions between GhAGAL family genes and other genes in cotton.

2.7 RNA isolation and quantitative reverse transcription-polymerase chain reaction of *GhAGALs* family genes

Gossypium hirsutum plants, provided by the Institute of Cotton Research of the Chinese Academy of Agricultural Sciences (Zhong 9807), were grown in an incubator at 25°C.When the cotton grew to three leaves and one heart stage, the plants were subjected to the stress of 100 mM NaCl, and the leaves were collected at 0 h, 6 h, 12 h, and 24 h. RNA was then extracted using the Aidlab kit. cDNA was synthesized through reverse transcription using TransGen Biotech. Clear water-treated plants were used as the control. The procedure was as follows: 94°C for 30 s; 94°C for 5 s, 55°C for 15 s, and 72°C for 10 s for 45 cycles; followed by storage at 4°C (Supplementary Table S1). The relative expression of genes was calculated using the $2^{-\Delta\Delta Ct}$ method, with actin being the internal reference (Chen et al., 2020).

2.8 Construction of the *GhAGAL3* recombinant vector and transformation by *Agrobacterium tumefaciens*

Firstly, we downloaded the CDS sequence of *GhAGAL3* from the Cotton FGD database. Then, a 300-bp region was selected from the silent site to design VIGS primers for amplifying the target fragment. The pYL156 silencing vector was constructed by double digesting the pYL156 vector with two restriction enzymes, *BamHI* and *SacI*, using the In-Fusion technique. Next, the recombinant vector product was transfected into *E. coli*, PCR was performed using VIGS primers, and the resulting bands were sent to the testing company for sequencing. After obtaining the correct sequencing results, re-transformation with *Agrobacterium tumefaciens* was

conducted. In brief, the LBA4404 bacteriophage carrying the following constructs, including control pYL156 (empty vector), pYL156: *GhAGAL3*, pYL156: PDS (positive control), and pYL192 (help vector), was inserted into the cotyledons of Zhong 9807. After 24 h of dark treatment, cotton was grown in an incubator at 25°C/16-h light and 23°C/8-h dark cycle conditions (Fan et al., 2022). When cotton grew to three leaves and one heart stage, lines injected with pYL156 and successfully silenced were immersed in the 100-mM NaCl solution, whereas control lines were immersed in ddH_2O. After 36 h, phenotypic differences were observed among the various treatments and individual plants were sampled. Each sample weighed 0.1 g.

2.9 Determination of raffinose, D-galactose, and D-glucose contents

The leaves of control and NaCl-treated cotton seedlings were collected at each time point, immediately frozen in liquid nitrogen, and stored at −80°C to determine the raffinose content. The content of raffinose was detected using a plant raffinose ELISA kit, that of D-galactose content using the D-galactose content kit (Article number: ADS-W-TDX046). D-Glucose content test was performed using the D-glucose content (GOPOD oxidase method) kit (Article number: ADS-W-TDX002).

2.10 Determination of Pro and MDA contents

For determination of proline (Pro) and malondialdehyde (MDA) contents, the leaves of control and NaCl-treated cotton seedlings were collected at each time point and immediately frozen in liquid nitrogen and stored at −80°C to determine the content of Pro and MDA. The content of Pro was determined using the test kit from Nanjing Jiancheng Bioengineering Institute, the content of MDA was determined using the test kit from Beijing Solarbio Science & Technology Co., Ltd. (Liu et al., 2022; Meng et al., 2023).

2.11 Determination of chlorophyll content

Chlorophyll was extracted from 0.1 g of fresh cotyledon leaves by overnight immersion in an 85% (v/v) acetone solution, the supernatant was centrifuged at 4,000 rpm for 10 min and diluted with 85% acetone to the appropriate concentration, and absorbance values were measured by absorbance at 424.5, nm 644 nm, and 663 nm, with 85% acetone serving as a blank (Farooq et al., 2013).

2.12 Leaf DAB staining

First, the DAB working solution was made in accordance with the DAB stain kit's instructions and kept chilled at 4°C. The leaves were removed from the staining solution and soaked in anhydrous ethanol. The ethanol can be changed several times during the decolorization process to ensure that the green color of the leaves disappears. Cotton samples were submerged in the staining solution and stained overnight in the dark. When the green tint had finally faded completely, ROS distribution and accumulation was seen.

2.13 Statistical analysis

A minimum of three biological replicates was deemed necessary for each dataset. Spass19.0 was used for histogram drawing, and GraphPad Prism 9.0 was used for data significance analysis.

3 Results

3.1 Identification of *AGAL* family members and construction of the evolutionary trees

To further understand the evolutionary relationships of the *AGAL* family in cotton, nine different species were identified in this study, namely, *Gossypium hirsutum*, *Gossypium arboreum*, *Gossypium barbadense*, *Gossypium raimondii*, *Oryza sativa*, *Arabidopsis thaliana*, *Zea mays*, *Populus trichocarpa*, and *Vitis vinifera*. According to our identification results, 15 genes were in *Gossypium hirsutum*, 15 in *Gossypium barbadense*, 7 in *Gossypium raimondii*, 7 in *Gossypium arboreum*, 9 in *Oryza sativa*, 7 in *Vitis vinifera*, 18 in *Zea mays*, 22 in *Populus trichocarpa*, and 4 in *Arabidopsis thaliana*. Using MEGA 7.0 software, the amino acid sequences were compared and analyzed by the maximum likelihood method, and a phylogenetic tree was constructed (Figure 1A). The classification of the genes within the species' family can be clearly observed through the phylogenetic tree. The genes are divided into three subclasses: I, II, and III. Subclass I was further divided into two subfamilies, and subclass II and subclass III were also divided into two subfamilies. The (Figure 1B) shows that the genes of subclass I and subclass III families tend to converge, whereas subclass II had the lowest number of family genes. The closest relationship was seen between cotton and *Zea mays*. Thereafter, the evolutionary trees were compared, which showed that the number of genes in *Gossypium hirsutum* and *Gossypium barbadense* was higher than that in *Vitis vinifera*, *Oryza sativa*, and *Arabidopsis thaliana* (Rui et al., 2022). This suggested that cotton underwent massive amplification during the evolutionary process (Figure 1B).

To further analyze the evolutionary relationships among the common ancestors of the four cotton species, their amino acid sequences were compared and analyzed with MEGA 7.0 software. In addition, the neighbor-joining method was applied in constructing the phylogenetic tree of the *AGAL* family (Figure 1B). As clearly observed from the phylogenetic tree, the species' family genes were classified into three subclasses I, II, and III. Among the homologous genes of *Gossypium hirsutum*, the *AGAL* gene was subjected to stable selection across different cotton species. Meanwhile, there were only a few genes undergoing natural selection, as indicated by the conserved amino acid phenotype of the protein.

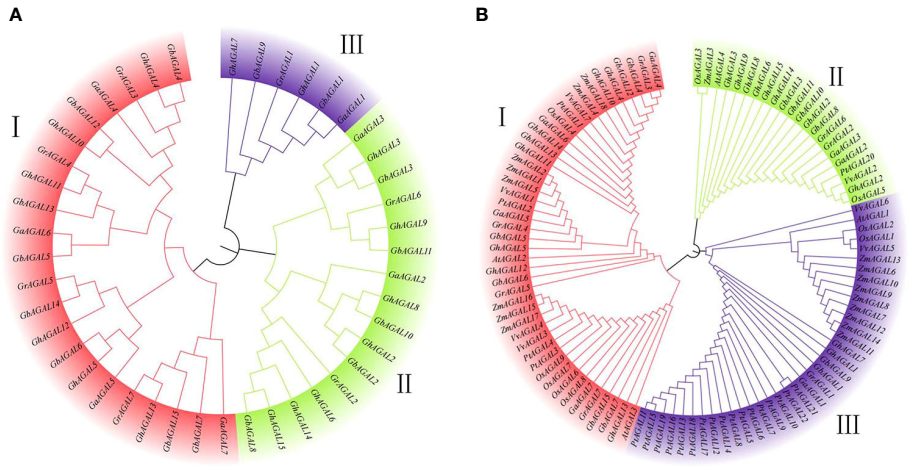


FIGURE 1
Phylogenetic tree constructed with MEGA7 by the neighbor-joining (NJ) method. **(A)** Evolutionary analysis of the *AGAL* family members of four cotton species, **(B)** Phylogenetic relationships of *AGAL* genes from four cotton species and five other plant species.

3.2 Chromosome distribution of the *AGAL* family genes in cotton

We further investigated the distribution of *AGAL* family genes on chromosomes. It was found that the genes were unevenly distributed among certain chromosomes or closely aligned regions (Figure 2). Among them, 14 *GhAGALs* family genes of cotton were located on 11 chromosomes and distributed on each of these chromosomes. However, *GhAGAL15* did not appear in chromosome localization, possibly indicating that it was not located on any specific chromosome. Moreover, the 15 *GbAGALs* family genes in *Gossypium barbadense* were also localized on 11 chromosomes, with distribution on each of these chromosomes. The localization of the *AGAL* family genes on chromosomes was generally similar in both *Gossypium barbadense* and *Gossypium hirsutum*, suggesting a certain degree of similarity between these different species. Additionally, all the seven genes of evolutionary *Gossypium arboreum* and *Gossypium raimondii* were evenly distributed on each chromosome. Upon further analysis, most of the genes were highly conserved in their location on the chromosomes. For example, the positions of *GhAGAL1-GbAGAL1-GaAGAL1-GrAGAL1* and *GhAGAL2-GbAGAL2-GaAGAL2-GrAGAL2* on the chromosomes all show similar locations of gene distribution (Supplementary Table S2).

3.3 Conserved protein motifs and gene structure analysis

Triadic joint analysis of four major cotton species was conducted by phylogenetic tree, gene structure, and conserved motifs (Figure 3). From the figure, it was seen that family members on the same evolutionary branch displayed a similar distribution of conserved motif, which suggested that the family members were functionally similar and structurally evolutionarily conserved (Bailey and Elkan, 1994).

3.4 Analysis of cis-acting elements and expression levels in *Gossypium hirsutum*

To further investigate the mechanisms of the response of *GhAGAL* gene family members to abiotic stresses, relevant data were downloaded to analyze the differential expression of genes under cold, heat, salt, and PEG stresses. As observed from Figure 3, the expression of different genes changed to varying degrees under different stresses. Therefore, it was speculated that the *GhAGAL* gene family was involved in the regulation of abiotic stresses (Figure 3). In addition, members of the *GhAGAL* gene family also played a crucial role in the physiological and biochemical processes of plants. According to Figure 4, the *GhAGAL* gene family is involved in various environmental stimuli, including light response, methyl jasmone, low-temperature stress, abscisic acid, gibberellin, defense, and stress response (Hu et al., 2019) (Supplementary Table S3). Therefore, it inferred that the *Gossypium hirsutum GhAGALs* family primarily consists of phytohormones and *cis*-acting elements related to adversity. Afterward, the expression of 15 genes in the *AGAL* family of *Gossypium hirsutum* was analyzed under different abiotic stresses and the results were visualized in the form of heat maps (Chen et al., 2020). According to the results, under cold stress, *GhAGAL4* and *GhAGAL11*, *GhAGAL2*, *GhAGAL15*, *GhAGAL6*, *GhAGAL3*, *GhAGAL9*, *GhAGAL13*, *GhAGAL5*, *GhAGAL12*, and *GhAGAL11* were upregulated whereas *GhAGAL1*, *GhAGAL2*, *GhAGAL3*, *GhAGAL8*, and *GhAGAL11* were downregulated. Under heat stress, *GhAGAL2*, *GhAGAL3*, *GhAGAL5*, *GhAGAL8*, *GhAGAL9*, *GhAGAL12*, *GhAGAL11*, and *GhAGAL13* were upregulated. *GhAGAL1*, *GhAGAL5*, *GhAGAL11*, and *GhAGAL13* were downregulated. Under PEG stress, the upregulated expressions of *GhAGAL8*, *GhAGAL2*, *GhAGAL3*, *GhAGAL9*, *GhAGAL11*, and *GhAGAL1* were downregulated. Under NaCl stress, *GhAGAL2*, *GhAGAL3*, *GhAGAL5*, *GhAGAL6*, *GhAGAL9*, *GhAGAL12*, *GhAGAL11*, and *GhAGAL1* were upregulated, whereas *GhAGAL1*, *GhAGAL3*, *GhAGAL11*, *GhAGAL12*, and *GhAGAL13* were downregulated.

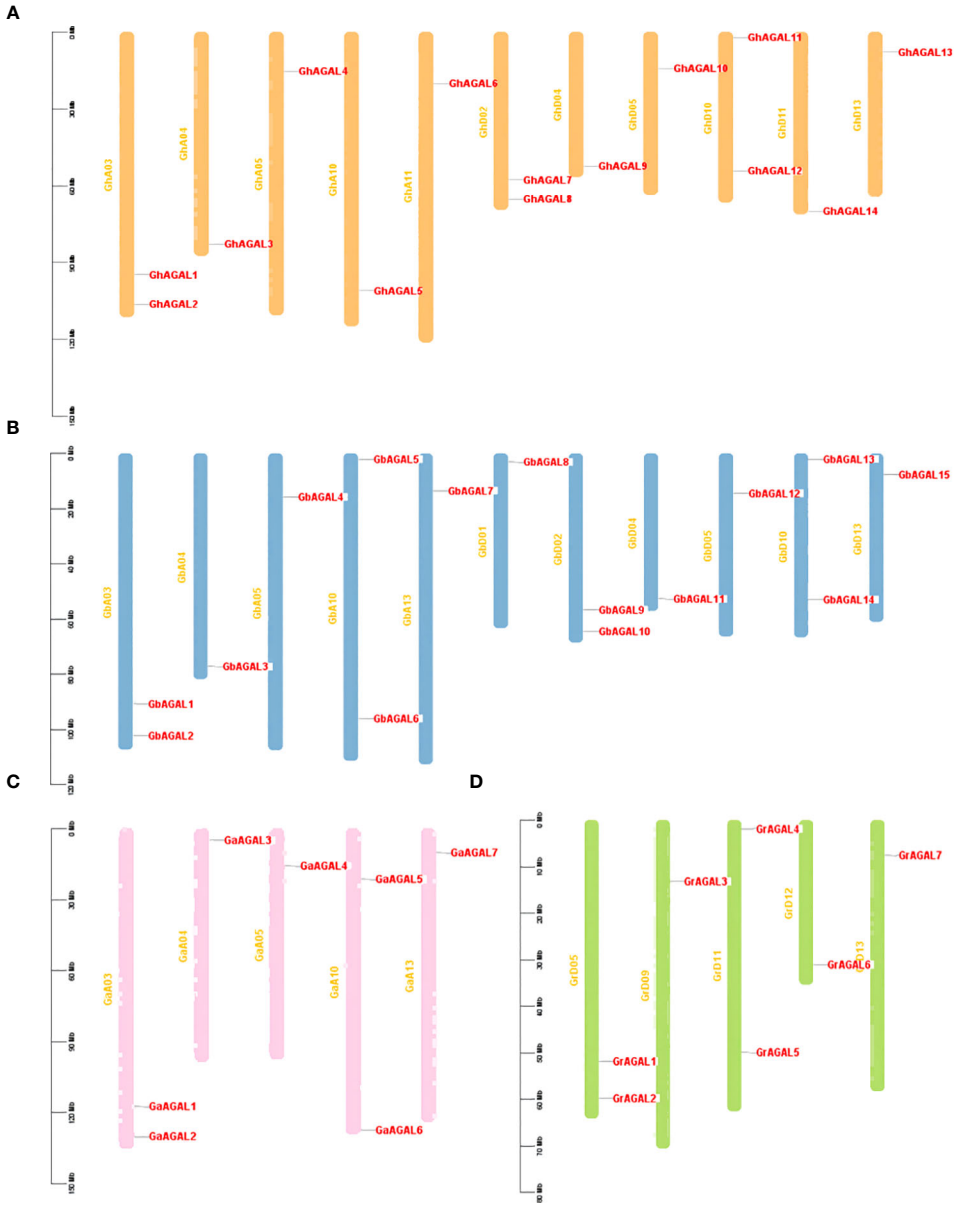


FIGURE 2
Chromosomal localization of *AGAL* Chromosome distribution of *AGAL* genes in four cotton species. **(A)** The chromosomes of *Gossypium hirsutum*, **(B)** The chromosomes of *Gossypium barbadense*. **(C)** The chromosomes of *Gossypium arboreum*. **(D)** The chromosomes of *Gossypium raimondii*.

3.5 Interaction network of GhAGAL proteins

A protein interaction network map of GhAGAL was constructed based on the protein sequence of GhAGAL by adopting the online STRING database (https://string-db.org/) (Yu et al., 2009). In the predicted network, 10 proteins were predicted to interact with GhAGAL, including DIN10, STS, SIP2, RFS5, BGAL17, HEXO1, HEXO3, HEXO2, and BGAL2, (Figure 5). Afterward, the CottonFGD database was searched to identify the related genes that interact with GhAGAL. Among them, RFS5 interacted with GhAGAL and played a crucial role in mitigating salt stress in cotton. RFS5 not only helps scavenge free radicals but also enhances salt tolerance in cotton (Cui et al., 2021a). In this study, GhAGAL was identified as the decomposition gene of RFS5. Consequently, it was speculated that there might be a relationship between GhAGAL and RFS5.

3.6 Tissue-specific expression pattern of the *GhAGALs* gene and RT-qPCR analysis

To understand the expression of *AGAL* family genes in *Gossypium hirsutum*, the expression levels of 15 *AGAL* genes were analyzed in different tissues (roots, leaves, and stems) of *Gossypium hirsutum* (Figure 6). As a result, various genes exhibited significant differential expression within the same tissue, meanwhile the same also gene showed significant differential expression across different tissues. The expression pattern maps were created through clustering and analyzing

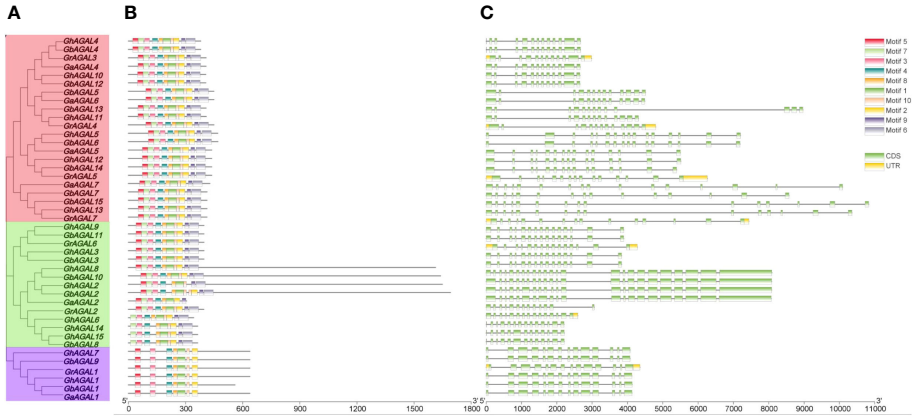


FIGURE 3
Conserved motifs and gene structure analysis of the GhAGAL family. **(A)** Phylogenetic tree of the GhAGAL family. **(B)** Conserved motifs of the GhAGAL family. **(C)** Gene structure of the GhAGAL family gene structure.

differences in gene expression from a holistic perspective. These maps were divided into three categories, namely, high expression, no expression, and low expression. It was obviously observed from the graph that the relative expressions of *GhAGAL3* and *GhAGAL9* were the highest in leaves, whereas those of *GhAGAL5*, *GhAGAL11*, and *GhAGAL12* were the highest. The relative expression levels of *GhAGAL1*, *GhAGAL6*, and *GhAGAL8* were the highest in roots, where the relative expressions of *GhAGAL2*, *GhAGAL4*, *GhAGAL14*, and *GhAGAL15* were expressed only in roots and not in stems and leaves. *GhAGAL7* and *GhAGAL13* are expressed in stems, but not in roots and leaves.

To explore the effects of *AGAL* family genes on the plant tissue of *Gossypium hirsutum*. Quantitative reverse transcription-polymerase chain reaction (RT-qPCR) analysis was conducted, and a histogram was created to visualize the expression of 15 *AGAL* family genes in *Gossypium hirsutum* leaves under NaCl stress at 0 h (CK), 6 h, 12 h, and 24 h. As clearly observed from Figure 7, the expression level of *GhAGAL1* increased at 6 h and 12 h but decreased at 24 h compared with the control (CK). *GhAGAL2* expression decreased at 6 h, 12 h, and 24 h relative to the control group (CK), whereas *GhAGAL3* expression continued to increase at 0 h, 6 h, 12 h, and 24 h, and *GhAGAL4* expression decreased at 6 h, 12 h, and 24 h. The expression of *GhAGAL5* increased at 6 h and 12 h but decreased significantly at 24 h. *GhAGAL6* expression decreased at 6 h and then increased rapidly at 12 h compared with CK but decreased again at 24 h. *GhAGAL7* expression increased at 6 h, 12 h, and 24 h but decreased again at 24 h after increasing at 12 h. *GhAGAL8* expression decreased at 24 h compared with CK and increased at 0 h, 6 h, 12 h, and 24 h relative to CK. In addition, the expression of *GhAGAL8* decreased compared with the control (CK). *GhAGAL12* expression initially increased at 6 h

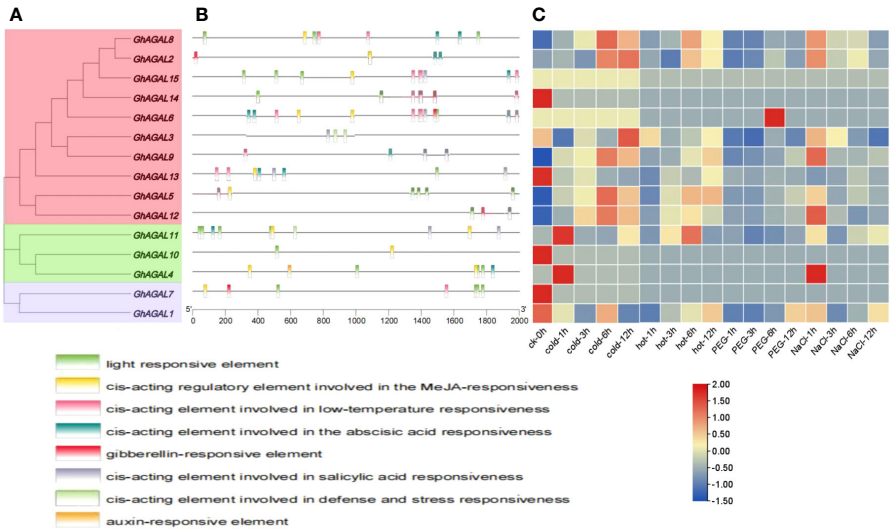


FIGURE 4
Analysis of *cis*-acting elements and differential expression patterns of the *GhAGAL* gene family members. **(A)** Phylogenetic tree of the *GhAGAL* gene family. **(B)** *cis*-Acting elements in the promoters of the *GhAGAL* gene family members. **(C)** Differential expression levels of *GhAGAL* gene family members in response to cold, heat, salt, and PEG stresses.

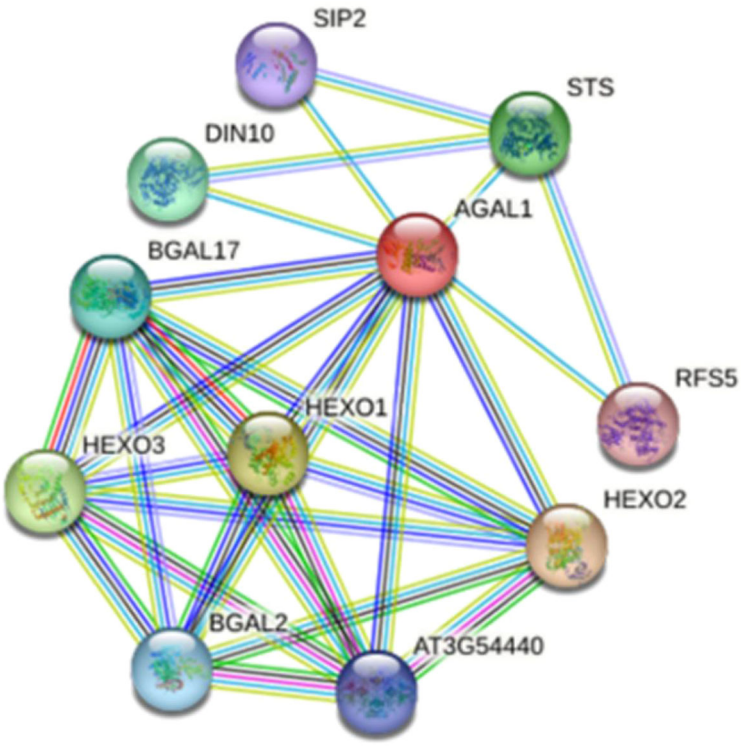


FIGURE 5
Interaction network of *AGAL* protein.

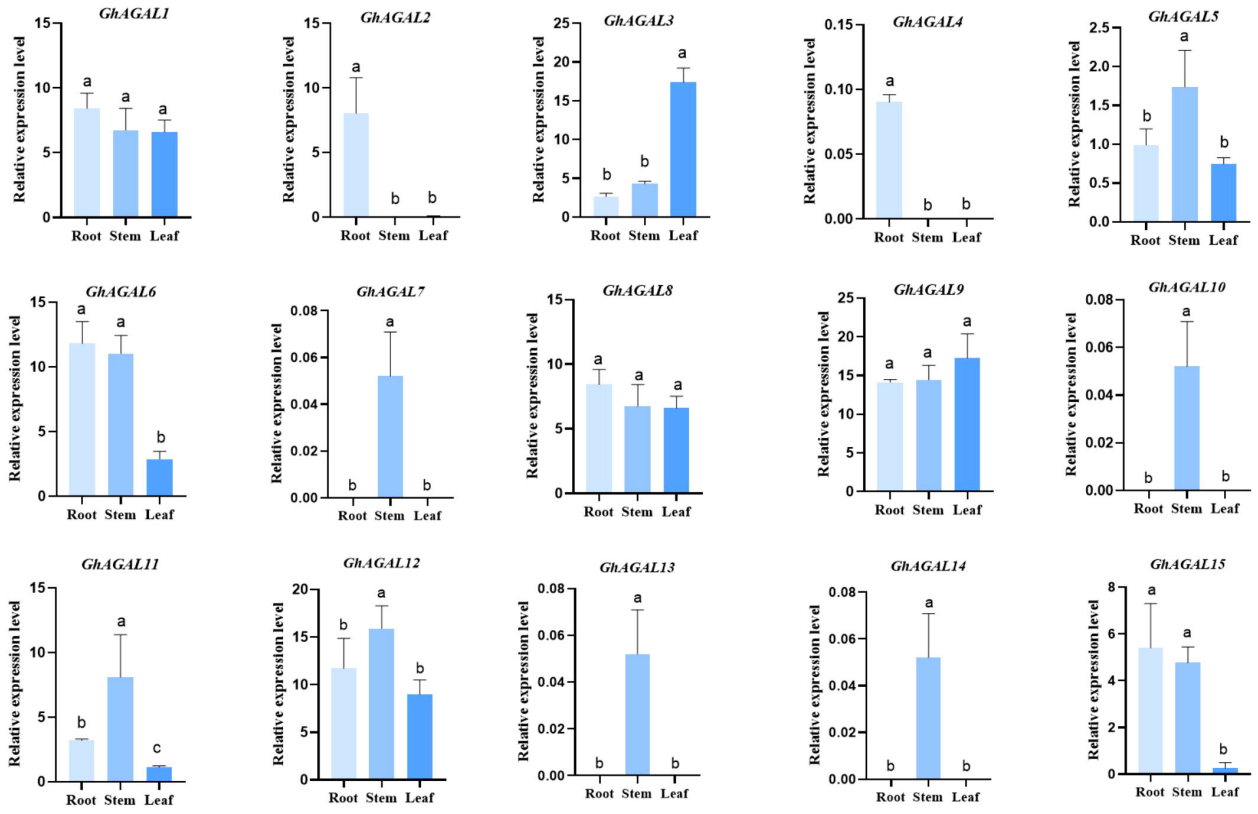


FIGURE 6
Specific expression of *AGAL* family genes in roots, stems, and leaves. Signifcance level α_i=0.05: The resulting values are expressed in relative units. The error bar in the figure is the standard deviation (SD) of the three biological replicates in each treatment group.

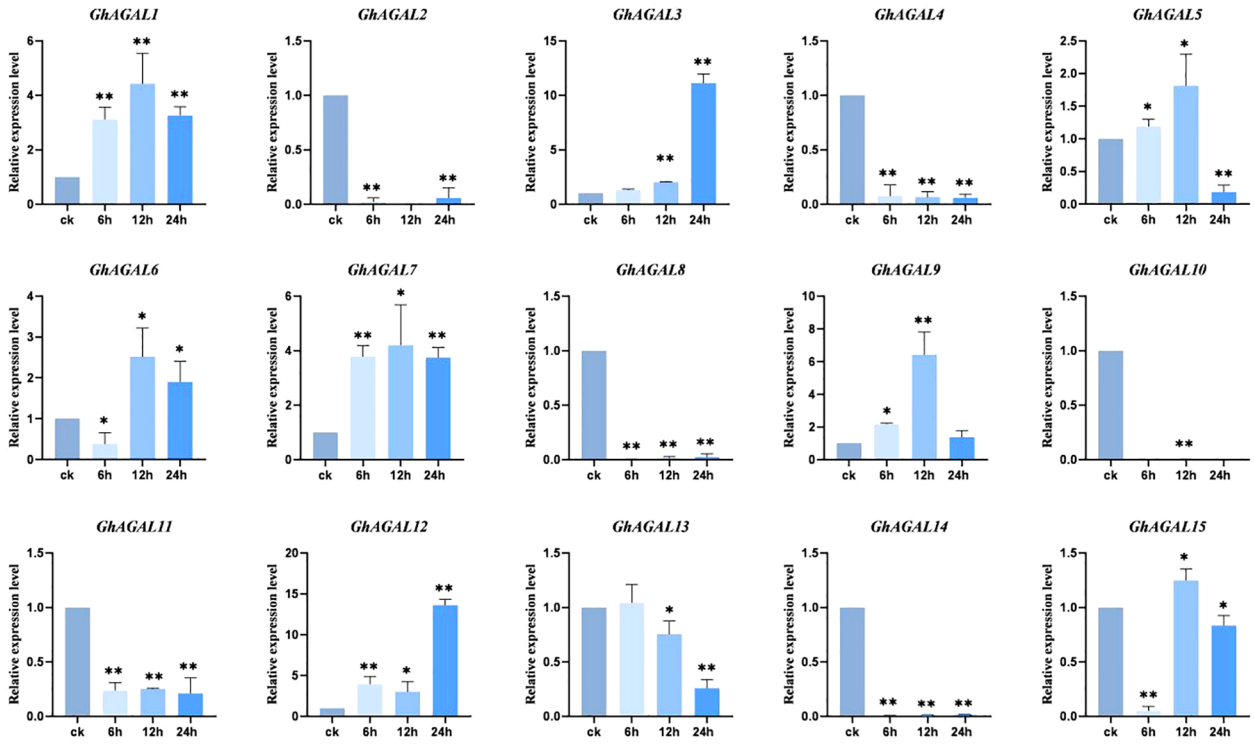


FIGURE 7
Expression of *AGAL* family genes in leaves at 0 h, 6 h, 12 h, and 24 h under NaCl stress. *0.01 < p < 0.05, **p < 0.01; the resulting mean values are presented as relative units.

and then slightly decreased at 12 h and later rapidly increased at 24 h compared with the control. *GhAGAL13* expression did not show any significant change at 6 h compared with the control but then continued to decrease at 12 h and 24 h. *GhAGAL14* expression decreased at 6 h, 12 h, and 24 h relative to the control, whereas *GhAGAL15* expression was also decreased at 6 h, 12 h, and 24 h compared with the control. The expression of *GhAGAL15* decreased significantly at 6 h, 12 h, and 24 h. It decreased significantly at 6 h compared with CK and then increased significantly at 12 h but decreased again at 24 h. *GhAGAL3* expression continued to rise with time under NaCl stress (Figure 7). Therefore, it was screened for functional verification by combining the tissue specificity of the *GhAGAL* family and conducting RT-qPCR analysis in leaves.

3.7 Phenotype of cotton plants with *GhAGAL3* gene silencing by VIGS under NaCl stress

To validate the above result, the genetic virus-mediated gene silencing technique was performed on the *GhAGAL3* gene (Figure 8). When the cotton reached the three-leaf stage, PDS plants exhibited albinism. Then, plants injected with the pYL156 carrier and pYL156:*GhAGAL3* were transferred to a tripod and further treated with 100 mM NaCl. As a result, the plant phenotype of pYL156 was more wilted than that of pYL156: *GhAGAL3* at 36 h. The expression level of the plants was analyzed using RT-qPCR. The results showed that the expression level of *GhAGAL3* plants was lower than that of pYL156 carriers, which further indicated the successful gene silencing. At the same time, the MDA content of *GhAGAL3* gene-silenced plants decreased slightly after salt stress, whereas the Pro content increased significantly. In contrast, the MDA content of pYL156: *GhAGAL3* was higher than that of pYL156 in the case of CK treatment. However, under NaCl stress, there was a slight downward trend in the MDA content of pYL156: *GhAGAL3* compared with that of pYL156. This suggests that the degree of membrane lipolysis damage was also reduced, resulting in lesser damage to the plants. As revealed by DAB staining results, the leaves of pYL156 plants under NaCl stress were darker than those of *GhAGAL3* plants. Based on the above results, after *GhAGAL3* silencing, cotton maintained normal growth by regulating the contents of osmotic substances such as proline. Additionally, cotton scavenged ROS by improving the activities of antioxidant enzymes, thus enabling cotton to resist NaCl stress.

3.8 Changes of raffinose content under NaCl stress

To further investigate the potential relationship between raffinose and salt tolerance in cotton, the raffinose content was measured in both control and salt-treated plants before and after treatment (Figure 9). According to our results, the content of raffinose increased after silencing *GhAGAL3*. Meanwhile, the levels of D-glucose and D-galactose, which are the downstream metabolic substances controlled by *GhAGAL3*, were determined, respectively. As a result, the levels of

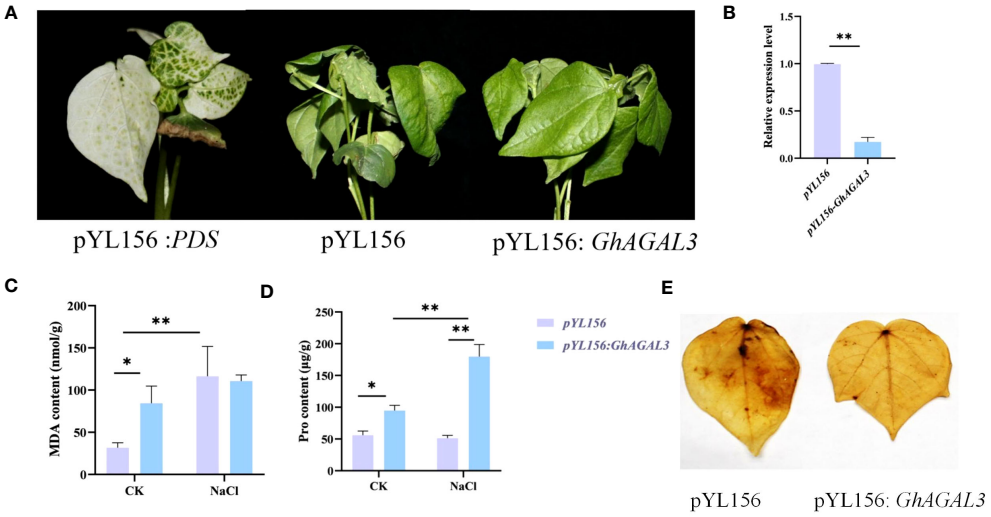


FIGURE 8
Phenotype of cotton leaves after virus infection and expression of *GhAGAL3* under NaCl stress. **(A)** Phenotype of cotton after *GhAGAL3* gene silencing under NaCl stress. pYL 156: PDS as a positive control, pYL156 was an empty vector as control, and pYL156 *GhAGAL3* was the *GhAGAL3*-silenced lines. **(B)** Relative expression level of *GhAGAL3* under NaCl stress. **(C)** MDA content of empty control and VIGS plants under normal growth and NaCl stress. **(D)** Pro content of empty control and VIGS plants under normal growth and NaCl stress. **(E)** DAB staining. *0.01 < p < 0.05, **p < 0.01; the resulting mean values are presented as relative units.

D-glucose and D-galactose decreased under salt stress. the chlorophyll content of *GhAGAL3* plant increased significantly compared with that of the control plant pYL156. These results indicated that salt stress was alleviated by increasing the raffinose content.

4 Discussion

Due to global climate change, the abiotic environment has become one of the most crucial factors influencing plant growth and development. Salt stress, in particular, has a direct impact on crop growth, development, and yield (Vaughan et al., 2018). AGAL is a hydrolyzing enzyme related to raffinose, which has an irreplaceable role in plant growth and development, as well as in response to adversity (Tsaniklidis et al., 2016). Studies on the *AGAL* gene family have revealed that AGAL exists not only in cotton but also in other evolutionary branches of species, including *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, *Vitis vinifera*, and *Zea mays*. To be specific, 15 AGAL genes were identified in *Gossypium hirsutum*, 15 in *Gossypium barbadense*, 7

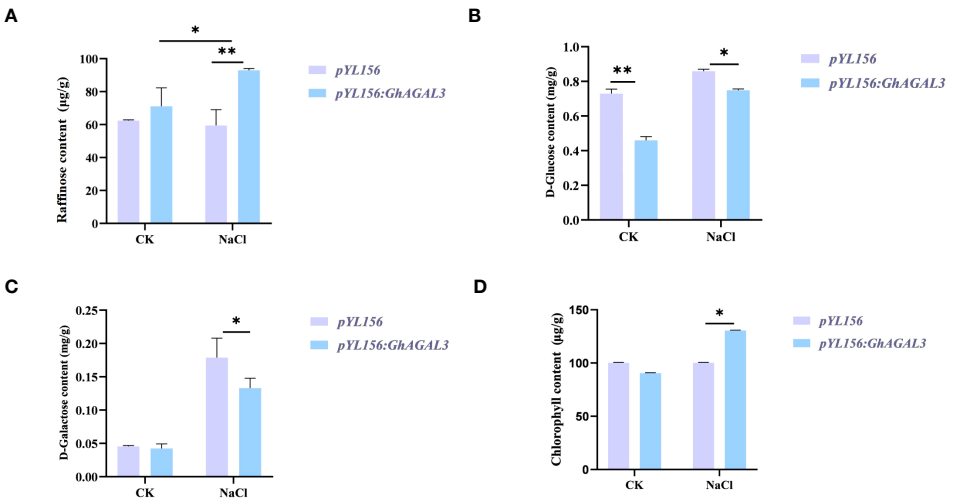


FIGURE 9
Determination of raffinose, D-glucose, D-galactose, and chlorophyll content after *GhAGAL3* silencing. **(A)** Raffinose content of empty control and VIGS plants under normal growth and NaCl stress. **(B)** D-glucose content of empty control and VIGS plants under normal growth and NaCl stress. **(C)** D-galactose content of empty control and VIGS plants under normal growth and NaCl stress. **(D)** content of empty control and VIGS plants under normal growth and NaCl stress. *0.01 < p < 0.05, **p < 0.01; the resulting mean values were presented as relative units.

in *Gossypium raimondii*, 7 in *Gossypium arboreum*, 9 in *Oryza sativa*, 7 in *Vitis vinifera*, 18 in *Zea mays*, 22 in *Populus trichocarpa*, and 4 in *Arabidopsis thaliana* (Chrost et al., 2007). There were 44 *AGAL* genes identified in the four major cotton species, with the same number being identified in *Vitis vinifera*, *Gossypium arboreum*, and *Gossypium raimondii*. Therefore, it was speculated that amplification might occur during the evolutionary process. Although less studied in plant *AGAL*, four members have been identified in *Arabidopsis thaliana*, but none of them has been reported in cotton as of now. At the same time, the constructed phylogenetic tree showed that cotton was closely related to *Zea mays* and distantly related to *Vitis vinifera*. In addition, *AGAL* genes tended to be conserved throughout the cotton genome. *AGAL* plays an important role in plant growth, development, and response to adversity stress (Zhou et al., 2012; Han et al.,2015).

From the perspectives of chromosome localization and motif structure analysis, it was found that the gene structure of the *AGAL* family was highly conserved. In addition, the structure of the *GhAGAL* gene family members was highly consistent with that of the *GbAGAL* gene family members both of which evolve from diploid to tetraploid.

The expression of *AGAL* genes is regulated by various environmental factors, including light, phytohormones, and adversity stress. The present study revealed the patterns of response to salt, PEG, cold, and heat stresses, as well as the tissue-specific expression of the *AGAL* gene family. Notably, the *cis*-acting element has an important function when plants are subjected to abiotic stresses(Yadav et al., 2011). According to our results, *cis*-acting elements responded and generated excitons to regulate gene expression. A large number of hormone response elements, such as salicylic acid, jasmonic acid, and abscisic acid, are present in the promoter of *GhAGALs*. In this study, corresponding *cis*-acting elements of salicylic acid, jasmonic acid, abscisic acid, and other plant hormones exerted a crucial effect plant adaptation to abiotic stresses. When cotton is exposed to abiotic stress, it undergoes a variety of physiological and biochemical responses to mitigate the adverse effects of the stress (Zhang et al., 2022). Abiotic stresses including drought, cold, heat, and salt can reduce cotton yield. This study provided the first comprehensive analysis of the *AGAL* gene family in four cotton species. Finally, to verify the function of the *GhAGAL3* gene, we silenced it through the VIGS experiment. The results showed that silencing *GhAGAL3* reduced the salt tolerance of cotton.

Studies have reported that the metabolism of RFOs is a complex regulatory network in plants, and numerous associated enzymes are involved in the accumulation of RFOs (Sui et al., 2012). The first step in the catabolism of raffinose is the hydrolysis mediated by α-Gal, which produces sucrose and galactose (Peng et al., 2014). RFOs are distributed in plants and exert a protective effect in response to a wide range of abiotic stresses (Sonali et al., 2015). Overexpression of the *α-Gal* gene decreases the level of raffinose, which in turn reduces the cold tolerance of plants (Salvi et al., 2022). In the meantime, low temperature induced an increase in fructose content and a decrease in sucrose content within cucumber stems (Dos Santos et al., 2011). Moreover, the alkaline *α-Gal* activity of stem samples was lower than that of control samples after low temperature stress. Thus, *α-Gal* plays an important role in inhibiting glucose metabolism.

According to the expression patterns of *AGAL* gene in different tissues under NaCl stress and under different abiotic stresses, *GhAGAL3* was selected for a functional verification study. After silencing *GhAGAL3* gene, the wilting degree of seedlings decreased relative to the negative control. It was inferred that *GhAGAL3* might play an important role in responding to NaCl stress (Figure 10). Under NaCl stress, the expression of *GhAGAL3* decreased and the raffinose content increased. Additionally, the proline content of *GhAGAL3* significantly increased. Typically, the proline content in plants reflects the extent of their resistance to stress. Plants accumulate proline under stressful conditions. The higher the proline content, the stronger the resistance (Xie et al., 2011). When plants accumulate an excessive amount of reactive oxygen species, lipid peroxidation occurs, leading to the production of free radicals. Under such circumstances, plants are affected by oxidative stress, and as a result, cells cannot function normally. Lipid peroxidation is the most serious injury process in an organic body. As revealed by the DAB staining in this study, the leaves of pYL156 plants treated with NaCl were darker than those of *GhAGAL3*, which might be ascribed oxidative stress. Brown spots were clearly visible on cotton leaves under stress, and the color of veins was darker. This suggested the significant buildup deepening, indicating the accumulation of ROS in cotton leaves after the stress treatment (Montillet et al., 2005). Therefore, it was concluded that increasing the raffinose content in cotton seedlings effectively eliminated hydroxyl free radicals and reduced the salt stress-induced damage to cotton seedlings. To enhance the salt stress tolerance of cotton to salt stress, raffinose is used as an osmotic regulator in previous studies to alleviate salt stress and eliminate hydroxyl radicals from the body. ROS production in plants is reduced under stress. (Nishizawa et al., 2019). In some studies, the mechanism by which *CsGolS4* enhances drought resistance is illustrated, and whether *CsGolS4* improves oxidative damage caused by low temperature and drought stresses is examined by measuring the activities of reactive oxygen enzymes (Ma et al., 2021). Various environmental stresses induce the accumulation of ROS whereas an excessive amount of ROS can cause cell damage, including lipid peroxidation damage (Van den Ende and Valluru, 2008). Collectively, these studies suggest that RFOs play an important role in enhancing stress resistance in plants.

5 Conclusion

According to the findings in this study, raffinose played a vital role in salt stress. The phenotype and raffinose content of cotton leaves under salt stress were investigated in this study. *AGAL* genes were discovered in cotton, and 7, 15, 15, and 7 *AGAL* genes were found in *Gossypium arboreum*, *Gossypium barbadense*, *Gossypium hirsutum*, and *Gossypium raimondii*, respectively. The AGAL genes were divided into three branches based on the phylogenetic tree, gene structure, and motifs. The *AGAL* family was engaged in a variety of abiotic stressors. Raffinose played a crucial role in salt

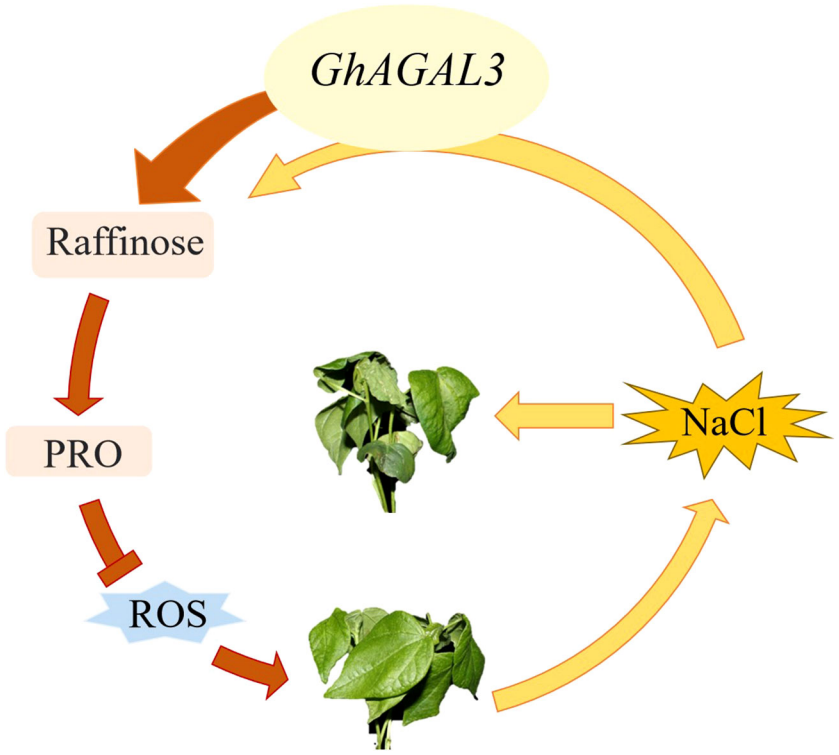


FIGURE 10
Mechanism model of *GhAGAL3* in regulating cotton response to NaCl stress.

stress. Therefore, silencing *GhAGAL3* resulted in a more favorable phenotype under NaCl stress. Results in this study can lay a theoretical foundation for further research on the link between *GhAGAL3* and NaCl stress.

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

WC, YC, YH, LZ, and RC: Designed the experiments, methodology, experiment, analysis of data, writing-original draft preparation, writing-review and editing. XYL: Methodology, experiment. HH: Methodology. YF and YZ: Experiment. XF, KN, and TJ: Experiment. YM and ML: Experiment. MH and YL: Experiment. XKL, XC, and DW: Methodology. LZ, LG, JW, and SW: Methodology. QC: Writing-review and editing. WY: Conceived and designed the experiments, supervision. All authors contributed to the article and approved the submitted version.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences, China Agriculture Research System of MOF and MARA.

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.2023.1246677/full#supplementary-material

References

Bailey, T. L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in bipolymers. *Intenational Conf. On Intelligent Syst. Mol. Biol.* 2 (1), 28–36.

Bawa, G., Liu, Z., Zhou, Y., Fan, S., Ma, Q., Tissue, D. T., et al. (2022). Cotton proteomics: Dissecting the stress response mechanisms in cotton. *Front. Plant Sci.* 13, 1035801. doi: 10.1186/s12864-021-07622-1

Cai, K., Liu, H., Chen, S., Liu, Y., Zhao, X., and Chen, S. (2021). Genome-wide identification and analysis of class III peroxidases in Betula pendula. *BMC Genomics* 22 (1), 314. doi: 10.1186/s12864-021-07622-1

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 (8), 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chen, W., Si, G. Y., Zhao, G., Abdullah, M., Guo, N., Li, D. H., et al. (2018). Genomic comparison of the P-ATPase gene family in four cotton species and their expression patterns in *Gossypium hirsutum*. *Molecules* 23 (5), 1092. doi: 10.3390/molecules23051092

Chrost, B., Kolukisaoglu, U., Schulz, B., and Krupinska, K. (2007). An α-galactosidase with an essential function during leaf development. *Planta* 225, 311–320. doi: 10.1007/s00425-006-0350-9

Cui, R., Lu, X., Chen, X., Malik, W. A., Wang, D., Wang, J., et al. (2021a). A novel raffinose biological pathway is observed by symbionts of cotton≡Verticillium dahliae to improve salt tolerance genetically on cotton. *J. Agron. Crop Sci.* 207 (6), 956–969. doi: 10.1111/jac.12556

Cui, R., Wang, X., Malik, W. A., Lu, X., Chen, X., Wang, D., et al. (2021b). Genome-wide identification and expression analysis of Raffinose synthetase family in cotton. *BMC Bioinf.* 22 (1), 356. doi: 10.1186/s12859-021-04276-4

Daldoul, S., Toumi, I., Reustle, G. M., Krczal, G., Ghorbel, A., Mliki, A., et al. (2012). Molecular cloning and characterisation of a cDNA encoding a putative alkaline alpha-galactosidase from grapevine (*Vitis vinifera* L.) that is differentially expressed under osmotic stress. *Acta physiologiae plantarum* 34, 891–903. doi: 10.1007/s11738-011-0887-5

Dos Santos, T. B., Budzinski, I. G., Marur, C. J., Petkowicz, C. L., Pereira, L. F., and Vieira, L. G. (2011). Expression of three galactinol synthase isoforms in *Coffea arabica* L. and accumulation of raffinose and stachyose in response to abiotic stresses. *Plant Physiol. Biochem.* 49 (4), 441–448. doi: 10.1016/j.plaphy.2011.01.023

Fan, Y., Zhang, Y., Rui, C., Zhang, H., Xu, N., Wang, J., et al. (2022). Molecular structures and functional exploration of NDA family genes respond tolerant to alkaline stress in *Gossypium hirsutum L. Biol. Res.* 55 (1), 4. doi: 10.1186/s40659-022-00372-8

Farooq, M. A., Ali, S., Hameed, A., Ishaque, W., Mahmood, K., and Iqbal, Z. (2013). Alleviation of cadmium toxicity by silicon is related to elevated photosynthesis, antioxidant enzymes; suppressed cadmium uptake and oxidative stress in cotton. *Ecotoxicology Environ. Saf.* 96, 242–249. doi: 10.1016/j.ecoenv.2013.07.006

Fialho Lda, S., Guimaraes, V. M., Callegari, C. M., Reis, A. P., Barbosa, D. S., Borges, E. E., et al. (2008). Characterization and biotechnological application of an acid alpha-galactosidase from *Tachigali multijuga* Benth. seeds. *Phytochemistry* 69 (14), 2579–2585. doi: 10.1016/j.phytochem.2008.08.017

Gu, H., Lu, M., Zhang, Z., Xu, J., Cao, W., and Miao, M. (2018). Metabolic process of raffinose family oligosaccharides during cold stress and recovery in cucumber leaves. *J. Plant Physiol.* 224-225, 112–120. doi: 10.1016/j.jplph.2018.03.012

Han, Q., Li, T., Zhang, L., Yan, J., Dirk, L.M., Downie, B., et al. (2015). Functional analysis of the 5′-regulatory region of the maize alkaline alpha-galactosidase1 gene. *Plant Mol. Biol. Reporter* 33, 1361–1370.

Hara, M., Tokunaga, K., and Kuboi, T. (2008). Isolation of a drought-responsive alkaline α-galactosidase gene from New Zealand spinach. *Plant Biotechnol.* 25 (5), 497–501. doi: 10.5511/plantbiotechnology.25.497

Henderson, J., Salzberg, S., and Fasman, K. H. (1997). Finding genes in DNA with a hidden markov model. *J. Comput. Biol.* 4 (2), 127–141. doi: 10.1089/cmb.1997.4.127

Hu, Y., Chen, J., Fang, L., Zhang, Z., Ma, W., Niu, Y., et al (2019). Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat. Genet.* 51 (4), 739–748.

Huang, H., He, Y., Cui, A., Sun, L., Han, M., Wang, J., et al. (2022). Genome-wide identification of GAD family genes suggests *GhGAD6* functionally respond to Cd^{2+} stress in cotton. *Front. Genet.* 13. doi: 10.3389/fgene.2022.965058

Hughes, D., Boyd, S., Giraldo, P., Gonzalez, D., Holida, M., Goker-Alpan, O., et al. (2016). Novel treatment for Fabry disease: IV administration of plant derived alpha-gal-A enzyme safety and efficacy interim report. *Mol. Genet. Metab.* 2 (117), S59. doi: 10.1016/j.ymgme.2015.12.296

Joshi, J., Hasnain, G., Logue, T., Lynch, M., Wu, S., Guan, J. C., et al. (2021). A core metabolome response of maize leaves subjected to long-duration abiotic stresses. *Metabolites* 11 (11), 797. doi: 10.3390/metabo11110797

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33 (7), 1870–1874. doi: 10.1093/molbev/msw054

Lee, C., Chung, C. T., Hong, W. J., Lee, Y. S., Lee, J. H., Koh, H. J., et al. (2021). Transcriptional changes in the developing rice seeds under salt stress suggest targets for manipulating seed quality. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.748273

Lee, R.-H., Lin, M.-C., and Chen, S.-C. (2004). A novel alkaline α-galactosidase gene is involved in rice leaf senescence. *Plant Mol. Biol.* 55, 281–295. doi: 10.1007/s11103-004-0641-0

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

Li, H. W., Zang, B. S., Deng, X. W., and Wang, X. P. (2011). Overexpression of the trehalose-6-phosphate synthase gene *OsTPS1* enhances abiotic stress tolerance in rice. *Planta* 234 (5), 1007–1018. doi: 10.1007/s00425-011-1458-0

Liu, X., Cui, Y., Kang, R., Zhang, H., Huang, H., Lei, Y., et al. (2022). GhAAO2 was observed responding to NaHCO(3) stress in cotton compared to AAO family genes. *BMC Plant Biol.* 22 (1), 603. doi: 10.1186/s12870-022-03999-7

Ma, S., Lv, J., Li, X., Ji, T., and Gao, L. (2021). Galactinol synthase gene 4 (*CsGolS4*) increases cold and drought tolerance in *Cucumis sativus* L. by inducing RFO accumulation and ROS scavenging. *Environ. Exp. Bot.* 185, 104406. doi: 10.1016/j.envexpbot.2021.104406

Malik, W. A., Wang, X., Wang, X., Shu, N., Cui, R., Chen, X., et al. (2020). Genome-wide expression analysis suggests glutaredoxin genes response to various stresses in cotton. *Int. J. Biol. Macromolecules* 153 (prepublish), 492. doi: 10.1016/j.ijbiomac.2020.03.021

Meng, Y., Cui, Y., Peng, F., Guo, L., Cui, R., Xu, N., et al. (2023). GhCYS2 governs the tolerance against cadmium stress by regulating cell viability and photosynthesis in cotton. *Ecotoxicology Environ. Saf.* 263, 115386. doi: 10.1016/j.ecoenv.2023.115386

Mittal, Y., and Sharma, C. B. (1991). Development of α-galactosidase isoenzymes in chickpea seeds. *Plant Sci.* 77 (2), 185–190. doi: 10.1016/0168-9452(91)90087-O

Montillet, J.-L., Chamnongpol, S., Rusterucci, C., Dat, J., van de Cotte, B., Agnel, J.-P., et al. (2005). Fatty acid hydroperoxides and H_2O_2 in the execution of hypersensitive cell death in tobacco leaves. *Plant Physiol.* 138 (3), 1516–1526. doi: 10.1104/pp.105.059907

Nishizawa, A., Yabuta, Y., and Shigeoka, S. (2019). Galactinol and raffinose constitute a novel function to protect plants from oxidative damage. *Plant Physiol.* 147 (3), 1251–1263. doi: 10.1104/pp.108.122465

Nishizawa-Yokoi, A., Yabuta, Y., and Shigeoka, S. (2008). The contribution of carbohydrates including raffinose family oligosaccharides and sugar alcohols to protection of plant cells from oxidative damage. *Plant Signaling Behav.* 3 (11), 1016–1018. doi: 10.4161/psb.6738

Peng, Z., He, S., Gong, W., Sun, J., Pan, Z., Xu, F., et al (2014). Comprehensive analysis of differentially expressed genes and transcriptional regulation induced by salt stress in two contrasting cotton genotypes. *BMC Genomics* 15 (1), 1–28.

Pennycooke, J. C., Jones, M. L., and Stushnoff, C. (2003). Down-regulating alpha-galactosidase enhances freezing tolerance in *transgenic petunia*. *Plant Physiol.* 133 (2), 901–909. doi: 10.1104/pp.103.024554

Rui, C., Chen, X., Xu, N., Wang, J., Zhang, H., Li, S., et al. (2022). Identification and structure analysis of KCS family genes suggest their reponding to regulate fiber development in long-staple cotton under salt-alkaline stress. *Front. Genet.* 13, 812449. doi: 10.3389/fgene.2022.812449

Salvi, P., Varshney, V., and Majee, M. (2022). Raffinose family oligosaccharides (RFOs): role in seed vigor and longevity. *Bioscience Rep.* 42 (10), BSR20220198. doi: 10.1042/bsr20220198

Schneider, T., and Keller, F. (2009). Raffinose in chloroplasts is synthesized in the cytosol and transported across the chloroplast envelope. *Plant Cell Physiol.* 50 (12), 2174–2182. doi: 10.1093/pcp/pcp151

Sonali, S., Sritama, M., Papri, B., and Majumder, A. L. (2015). Significance of galactinol and raffinose family oligosaccharide synthesis in plants. *Front. Plant Sci.* 6 (656), 656. doi: 10.3389/fpls.2015.00656

Sui, X. L., Meng, F. Z., Wang, H. Y., Wei, Y. X., Li, R. F., Wang, Z. Y., et al. (2012). Molecular cloning, characteristics and low temperature response of raffinose synthase gene in *Cucumis sativus* L. *J. Plant Physiol.* 169 (18), 1883–1891. doi: 10.1016/j.jplph.2012.07.019

Tsaniklidis, G., Benovias, A., Delis, C., and Aivalakis, G. (2016). Acidic alpha galactosidase during the maturation and cold storage of cherry tomatoes. *Acta Physiologiae Plantarum* 38 (2), 57. doi: 10.1007/s11738-016-2075-0

Van den Ende, W., and Valluru, R. (2008). Sucrose, sucrosyl oligosaccharides, and oxidative stress: scavenging and salvaging? *J. Exp. Bot.* 60 (1), 9–18. doi: 10.1093/jxb/ern297

Vaughan, M. M., Block, A., Christensen, S. A., Allen, L. H., and Schmelz, E. A. (2018). The effects of climate change associated abiotic stresses on maize phytochemical defenses. *Phytochem. Rev.* 17 (1), 37–49. doi: 10.1007/s11101-017-9508-2

Wang, J., Zhang, Y., Xu, N., Zhang, H., Fan, Y., Rui, C., et al. (2021). Genome-wide identification of CK gene family suggests functional expression pattern against Cd^{2+}

stress in *Gossypium hirsutum* L. *Int. J. Biol. Macromolecules* 188, 272–282. doi: 10.1016/j.ijbiomac.2021.07.190

Wang, Y., Wang, C., Chen, Y., Cui, M., Wang, Q., and Guo, P. (2022). Heterologous expression of a thermostable α-galactosidase from parageobacillus thermoglucosidasius isolated from the lignocellulolytic microbial consortium TMC7. *J. Microbiol. Biotechnol.* 32 (6), 749–760. doi: 10.4014/jmb.2201.01022

Xie, H., Yang, L., and Li, Z. G. (2011). The roles of proline in the formation of plant tolerance to abiotic stress. *Biotechnol. Bull.* 51 (2), 23–25.

Xie, J., Chen, Y., Cai, G., Cai, R., Hu, Z., and Wang, H. (2023). Tree Visualization By One Table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees. *Nucleic Acids Res.* 51 (W1), W587–w592. doi: 10.1093/nar/gkad359

Yadav, N. R., Taunk, J., Rani, A., Aneja, B., and Yadav, R. C. (2011). Role of transcription factors in abiotic stress tolerance in crop plants. *Climate Change Plant Abiotic Stress Tolerance* 25 (3), 2433–2442. doi: 10.5504/BBEQ.2011.0072

Yu, M., Lin, S., Bei, F., Zhao, D., Zheng, Y., and Sheng, J. (2009). The effect of MeJA on ethylene biosynthesis and induced disease resistance to *Botrytis cinerea* in tomato. *Postharvest Biol. Technol.* 54 (3), 153–158. doi: 10.1016/j.postharvbio.2009.07.001

Zhang, H., Mao, L., Xin, M., Xing, H., Zhang, Y., Wu, J., et al. (2022). Overexpression of *GhABF3* increases cotton(*Gossypium hirsutum* L.) tolerance to salt and drought. *BMC Plant Biol.* 22 (1), 1–15. doi: 10.1186/s12870-022-03705-7

Zhang, H., Zhang, Y., Xu, N., Rui, C., Fan, Y., Wang, J., et al. (2021). Genome-wide expression analysis of phospholipase A1 (*PLA1*) gene family suggests phospholipase A1-32 gene responding to abiotic stresses in cotton. *Int. J. Biol. Macromolecules* 192, 1058–1074. doi: 10.1016/j.ijbiomac.2021.10.038

Zhao, T.-Y., Corum Iii, J. W., Mullen, J., Meeley, R. B., Helentjaris, T., Martin, D., et al. (2006). An alkaline α-galactosidase transcript is present in maize seeds and cultured embryo cells, and accumulates during stress. *Seed Sci. Res.* 16 (2), 107–121. doi: 10.1093/glycob/cws121

Zhou, M. L., Zhang, Q., Zhou, M., Sun, Z. M., Zhu, X. M., Shao, J. R., et al. (2012). Genome-wide identification of genes involved in raffinose metabolism in maize. *Glycobiology* 22 (12), 1775–1785. doi: 10.1093/glycob/cws121

Zhu, T., Liang, C., Meng, Z., Sun, G., Meng, Z., Guo, S., et al. (2017). CottonFGD: an integrated functional genomics database for cotton. *BMC Plant Biol.* 17 (1), 101. doi: 10.1186/s12870-017-1039-x

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 08 January 2024
DOI 10.3389/fpls.2023.1308085

OPEN ACCESS

EDITED BY
Weicong Qi,
Jiangsu Academy of Agricultural Sciences (JAAS), China

REVIEWED BY
Jiangxin Wang,
Shenzhen University, China
Zhi-Kai Yang,
Guangzhou Medical University, China

*CORRESPONDENCE
Si Li
jackeikee@aliyun.com

†These authors have contributed equally to this work

RECEIVED 05 October 2023
ACCEPTED 18 December 2023
PUBLISHED 08 January 2024

CITATION
Liu C, Li L, Yang S, Wang M, Zhang H and Li S (2024) Multi-omic insights into the cellular response of *Phaeodactylum tricornutum* (Bacillariophyta) strains under grazing pressure.
Front. Plant Sci. 14:1308085.
doi: 10.3389/fpls.2023.1308085

Multi-omic insights into the cellular response of *Phaeodactylum tricornutum* (Bacillariophyta) strains under grazing pressure

Chenqi Liu†, Liang Li†, Shuo Yang, Mingye Wang, Hang Zhang and Si Li*

School of Chemical Engineering, Hebei University of Technology, Tianjin, China

Background/Aims: *Phaeodactylum tricornutum*, a model organism of diatoms, plays a crucial role in Earth's primary productivity. Investigating its cellular response to grazing pressure is highly significant for the marine ecological environment. Furthermore, the integration of multi-omics approaches has enhanced the understanding of its response mechanism.

Methods: To assess the molecular and cellular responses of *P.tricornutum* to grazer presence, we conducted transcriptomic, proteomic, and metabolomic analyses, combined with phenotypic data from previous studies. Sequencing data were obtained by Illumina RNA sequencing, TMT Labeled Quantitative Proteomics and Non-targeted Metabolomics, and WGCNA analysis and statistical analysis were performed.

Results: Among the differentially expressed genes, we observed complex expression patterns of the core genes involved in the phenotypic changes of *P.tricornutum* under grazing pressure across different strains and multi-omics datasets. These core genes primarily regulate the levels of various proteins and fatty acids, as well as the cellular response to diverse signals.

Conclusion: Our research reveals the association of multi-omics in four strains responses to grazing effects in *P.tricornutum*. Grazing pressure significantly impacted cell growth, fatty acid composition, stress response, and the core genes involved in phenotype transformation.

KEYWORDS

***Phaeodactylum tricornutum*, grazing stress, transcriptome, proteome, metabolome, phenotype**

1 Introduction

Phytoplankton, particularly marine diatoms, are key contributors to Earth's primary productivity (Shukla et al., 2023), responsible for about one-fifth of the total. The impact of global climate change intensifies the importance of understanding their response to grazing pressure, crucial for marine ecosystem conservation. *Phaeodactylum Tricornutum* (Bowler et al., 2008), a model organism for diatoms, showcases a diverse range of responses to various environmental stressors, including nitrogen and phosphorus deprivation and heavy metal ion exposure (Thamatrakoln, 2021). Moreover, it affects the behavior of its predators (Kim et al., 2022), such as copepods, altering their grazing behavior and consequently, the marine food web. The advent of high-throughput sequencing technologies has allowed for a more in-depth exploration of these responses, extending to many omics levels.

P.tricornutum, a pennate diatom with an evolutionary history spanning 90 million years (Sims et al., 2006), displays remarkable plasticity under environmental stressors. In response to nitrogen deprivation, it reorganizes its proteome to facilitate nitrogen removal and minimize lipid degradation, thereby limiting nitrogen use to high-demand pathways (Kang et al., 2023). Phosphorus scarcity triggers cellular changes, increasing polyphosphate production, phosphorus transport, and metalloenzyme production to dissolve organophosphates (Helliwell et al., 2021). The diatom has a defined response to heavy metal ions, where aluminum toxicity targets photosynthesis, inducing increased reactive oxygen species (ROS) and enhancing lipid peroxidation. This effect is mitigated by up-regulated glycolysis and the pentose phosphate pathway, providing the necessary cellular energy and carbon skeletons for growth (Xie et al., 2015).

In the intricate marine food web, the impact of the diatom *P.tricornutum* on its predators, particularly copepods, stands as a significant ecological dynamic. Grazing-induced signals, stemming from mechanical damage or specific compounds, can alter the morphology, physiology, and life history of *P.tricornutum* (Zhang et al., 2021). Toxins produced can impact diatoms' nutritional quality, thus modulating copepods' grazing behavior (Park et al., 2023). Furthermore, *P.tricornutum*'s synthesis of aldehydes strategically curtails zooplankton reproduction, effectively regulating predator populations and modulating grazing pressures (Lauritano et al., 2016; Øie et al., 2017; Martino et al., 2019).

High-throughput sequencing technologies have propelled *P.tricornutum* studies to new depths. Detailed transcriptomic and proteomic analyses under low CO_2 conditions have unearthed profound alterations in metabolic pathways, underscoring changes in carbon acquisition, signaling, and nitrogen metabolism (Clement et al., 2017). Particularly under nitrogen limitation, there is an observable upregulation in nitrogen fixation, central carbon metabolism, and the tricarboxylic acid (TCA) cycle, paralleled by lipid rearrangements favoring triglyceride distribution (Remmers et al., 2018). The advent of high-throughput sequencing technologies has provided an unprecedented opportunity to delve deeper into the intricate biological processes at multiple "omics" levels - namely transcriptomics, proteomics, metabolomics, and non-coding RNAs (Clement et al., 2017; Li and Ismar, 2018; Remmers et al., 2018).

Recent investigations have highlighted substantial variations in growth rate, biovolume, and nutritional composition among individual *P.tricornutum* strains (Li and Ismar, 2018). The sequence replicates of 1055/1 strain corroborates the consistency of these samples. Quantitative simulations and observed shifts in growth rate, biovolume, and nutritional composition have illuminated the effects of copepod grazing pressure on *P.tricornutum*. However, given the genetic diversity and unique response profiles among different strains, a comprehensive understanding of the differential gene expression mechanisms is imperative. Such insights are critical to grasp the species' adaptive responses to grazing pressure. Therefore, there is a pressing need for extensive, multi-strain, and multi-omics studies to refine our understanding and unravel the gene regulatory networks that drive growth discrepancies and morphological adaptations.

In this study, our objective is to elucidate the genetic underpinnings that govern the diverse strain-specific responses of *P.tricornutum* to grazing pressure. By integrating phenotype-associated gene modules with comprehensive transcriptomic, proteomic, and metabolomic datasets, we aspire to decode the intricate molecular mechanisms driving these responses. This endeavor is not merely an exploration of *P.tricornutum*'s adaptive strategies; it represents a broader initiative to enhance our comprehension of marine ecological interactions. Furthermore, it aims to contribute strategically to the development of informed approaches for mitigating the effects of global environmental changes on these vital ecosystems.

2 Materials and methods

2.1 Cultivation of *Phaedactylum tricornutum* strains and experimental conditions

Four strains of *P.tricornutum* (CCAP 1052/1A, CCAP 1052/1B, CCAP 1055/3, CCAP 1055/7) were provided by Génomique, Environnementale et Evolutive Section 3 CNRS UMR8197, Institut de Biologie de l'ENS (IBENS). The setting of culture conditions comes from the paper of Li (Li and Ismar, 2018). The culture medium was prepared using sterile-filtered natural seawater (Minisart High-Flow 0.1µm syringe filter; Sartorius Stedim Biotech GmbH, Goettingen, Germany), fortified with macronutrients and micronutrients in accordance with a modified Provasoli medium. *P.tricornutum* cultures were incubated at 18°C under constant illumination of 100 µmol photons·m^{-2}·s^{-1}, following a 12:12 h light-dark cycle with 2 hours of simulated sunrise and sunset periods.

2.2 Experimental design

The design of grazing experiments followed the protocol presented in Li (Li and Ismar, 2018). The grazing pressure was

applied using over 2000 *Acartia tonsa* individuals, fed with 1×10^7*P.tricornutum* cells daily in 25 L seawater. *Acartia* were maintained in a double-bucket system, with a 100 µm mesh sieve replacing the bottom of the inner bucket (Supplementary Figure 1). Daily, seawater in the outer bucket was replaced and *Acartia* faeces and eggs were removed. Grazing treatment replicates were generated by filtering *Acartia*'s culture medium using a 0.1 µm syringe filter and adding it to *P.tricornutum* cultures. Control groups were provides with the same volume of sterile-filtered seawater, maintaining the nutrient ratio of the original medium used for *P.tricornutum* cultivation.

Batch culture experiments were conducted for each *P.tricornutum* strain under two treatments (control and grazing). The 4 strains, under distinct conditions, were grouped into 8 sample sets: pt52_A_c (CCAP 1052/1A strain control group), pt52_A_g (CCAP 1052/1A strain grazing group), pt52_B_c (CCAP 1052/1B strain control group), pt52_B_g (CCAP 1052/1B strain grazing group), pt55_3_c (CCAP 1055/3 strain control group), pt55_3_g (CCAP 1055/3 strain grazing group), pt55_7_c (CCAP 1055/7 strain control group), pt55_7_g (CCAP 1055/7 strain grazing group). Each strain was cultured in 150 ml of medium and subjected to Reference Genome, Transcriptome Sequencing on an Illumina sequencing platform, TMT (Tandem Mass Tag) Labeled Quantitative Proteomics Standard Analysis, and non-targeted metabolomics Bioinformatics Standard Analysis. The grazing treatment groups only received *Acartia* water, ensuring that the responses to the chemical cues from grazing pressure was obtained without within-strain selection due to grazing selectivity.

In a preceding study (Li and Ismar, 2018), three independent experiments were conducted on multiple strains of *P.tricornutum*, and the data demonstrated high repeatability. Rigorous quality control measures were employed to ensure the removal of outliers or any disproportionately impactful data points. Transcriptomic, proteomic, and metabolomic approaches were employed to improve the accuracy of the results. A principal component analysis (PCA) was conducted on the proportion of omics data with K-means clusters among the eight sample groups to examine correlations between different omics within each strain, ensuring mutual validation of data from diverse omics layers (Supplementary Figure 2). Employing a multi-level data approach significantly clarifies our analytical results. Despite these constraints, our findings provide valuable preliminary insights and establish a foundational basis for future, more extensive research.

2.3 Illumina RNA sequencing

Total RNA was extracted from 4 strains under two conditions (control vs grazing). The quality of the extracted RNA, including its concentration and purity, was confirmed by RNA-specific agarose electrophoresis and Agilent 2100 Bioanalyzer. After the extraction, purification, and library construction processes, Next-Generation Sequencing (NGS) was conducted on the libraries using the Illumina HiSeq sequencing platform, aligning with the Phaeodactylum_tricornutum.ASM15095v2.dna.toplevel.fa reference genome (accessible at https://protists.ensembl.org/Phaeodactylum_tricornutum/Info/Index). The raw sequencing data have been deposited to the NCBI with the dataset identifier PRJNA1008380 (https://www.ncbi.nlm.nih.gov/sra/PRJNA1008380).

Complete transcriptome analysis involved the filtration of raw data and comparison of the filtered high-quality sequence (Clean Data) with the species' reference genome. Based on the comparison results, the expression level of each gene was determined. After splicing the resulting reads to restore the transcript sequence, they were compared with known mRNA and LncRNA transcripts in order to identify new LncRNA. An analysis of expression differences, target gene enrichment, and cluster analysis was conducted between known LncRNA and newly identified LncRNA. For unmatched sequences, 20bp were intercepted from both ends to obtain Anchor Reads. The Anchor Reads were then re-matched with the genome, and the CircRNA was identified using find_circ based on the comparison results. Then, basic statistical analysis, quantitative and differential expression analysis, functional and pathway enrichment analysis, and miRNA targeting relationship prediction were conducted for CircRNA. The raw miRNA data was filtered for quality, compared with the Rfan and miRBase databases, and various sRNA annotation information was obtained. The characteristics and expression of miRNA were analyzed, and miRNAs with significant differential expression were analyzed using clustering techniques. The predicted target genes were also subjected to enrichment analysis. Differential expression analysis involved standardized mRNA, LncRNA, CircRNA, and miRNA for significant differential expression screening, employing Log2Foldchange treatment and p-value screening. KEGG and GO functional enrichment analysis were employed to ascertain the primary biological functions of the differentially expressed genes.

2.4 TMT labeled quantitative proteomics

TMT™ (Tandem Mass Tag™), an *in vitro* labeling technology by Thermo Scientific, was employed to assess relative protein content across our samples. Using labels of 2, 6, 10 or 16 isotopes, peptides' amino groups were specifically labeled, followed by tandem mass spectrometry to analyze protein content variation in 2, 6, 10, or 16 groups simultaneously.

Our eight samples (four strains) from the two treatments (control and grazing) underwent protein extraction through the TCA acetone precipitation method, and SDT (4%(w/v) SDS, 100mM Tris/HCl pH7.6, 0.1M DTT) cleavage method (Wisniewski et al., 2009). Then the protein was quantified by BCA method. 200µg of protein from each sample was used for tryptic enzymatic hydrolysis via the Filter aided proteome preparation (FASP) method (Wisniewski et al., 2009), followed by enzymatic enzymatic peptide desalination using C_{18} Cartridge. The desalted peptide was then lyophilized and dissolved with a 40µL Dissolution buffer (OD280). Subsequent to this, 100µg of the peptide segment from each sample was labeled using the TMT labeling kit (Thermo Fisher), and the proteins were identified and quantitatively analyzed after SCX chromatography and LC-MS/MS data collection. The mass spectrometry proteomics data have been

deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the iProX partner repository (Ma et al., 2019; Chen et al., 2021) with the dataset identifier PXD044954.

2.5 Non-targeted metabolomics

HILIC UHPLC-Q-EXACTIVE MS technology coupled with a data-dependent acquisition method was adopted for full-spectrum analysis of the samples. Compound Discoverer 3.0 (Thermo Fisher Scientific) facilitated peak extraction and metabolite identification (Benton et al., 2015). The chromatographic separation of the extracted metabolites from the eight samples was performed using an ACQUITY UPLC BEH C18 column (100 mm*2.1 mm, 1.7μm, Waters, USA) with a column temperature set at 40°C and a flow rate of 0.3 ml/min. Both positive and negative ion modes of electrospray ionization (ESI) were utilized for detection. The samples post- UHPLC separation were analyzed with a Q-Exactive four-pole and a Thermo Fisher Scientific mass spectrometer. The resulting raw mass spectrometry data were processed by Compound Discoverer 3.0 (Thermo Fisher Scientific) software for peak extraction, peak alignment, peak correction, and standardization, resulting in a 3D data matrix composed of the sample name, spectral peak information (including retention time and molecular weight), and peak area. The metabolite structure was identified by precise mass number matching (<25 ppm) and secondary spectrogram matching method to search the laboratory database and various other databases including Bio cyc, HMDB, Metlin, HFMDB, Lipidmaps. The metabolomics data have been deposited to the Metabolights (Haug et al., 2020) with the dataset identifier MTBLS8485 (www.ebi.ac.uk/metabolights/MTBLS8485).

2.6 Weighted gene co-expression network analysis

Co-presentation network construction employed the WGCNA (Version 1.71) (Langfelder and Horvath, 2008). Hierarchical clustering of the sample data was first conducted to calculate the test of delocalization value. During co-expression network construction, we screened soft threshold values. Hierarchical clustering and dynamic tree cutting algorithms (Langfelder et al., 2008) were used to group genes with similar expression patterns and categorize them into different gene co-expression modules. We used the Pearson correlation coefficient to assess the relationship between the characteristic gene (Pei et al., 2017) of the gene co-expression module and the phenotype. Hierarchical clustering of the sample data was carried out and the Pearson correlation coefficient was calculated to link the sample to the phenotype.

In WGCNA, the core genes within each co-expression module can be evaluated by Gene Significance and Module Membership index (Pei et al., 2017). In this study, the genes with |KME|≥0.8 and |GS|≥0.8 in each module were identified as the core genes in the module. We then used the Gene Significance between genes and phenotypes within each module to map the Connectivity of genes within the module to verify the correlation between modules and phenotypes. The essential code for WGCNA analysis is available in the appendix for reference.

2.7 Softwares

Statistical analysis were performed using RStudio (R Core Team, 2022).

3 Results

3.1 Transcriptome profiling of *P.tricornutum*

The genomes of 8 samples of *P.tricornutum* were sequenced using the Illumina HiSeq sequencing platform. Following the removal of adapter and the filteration of low-quality sequences, the resultant clean reads counts were as follows in Table 1. For accurate genome alignment, clean reads were mapped to the reference genome ASM15095v2 using the BWT algorithm

TABLE 1 Basic information about the sequencing data after data filtering.

Sample	Reads No.	Clean Read No.	Bases (bp)	Clean Data (bp)	Clean Reads %	Clean Data %
pt52_A_c	107600456	98856954	16140068400	14828543100	91.87	91.87
pt52_A_g	118828944	109112418	17824341600	16366862700	91.82	91.82
pt52_B_c	109030966	99613970	16354644900	14942095500	91.36	91.36
pt52_B_g	110369976	100799184	16555496400	15119877600	91.32	91.32
pt55_3_c	106200518	96171702	15930077700	14425755300	90.55	90.55
pt55_3_g	104258976	94442576	15638846400	14166386400	90.58	90.58
pt55_7_c	121164860	107697388	18174729000	16154608200	88.88	88.88
pt55_7_g	114128898	103032006	17119334700	15454800900	90.27	90.27

Reads No.: Total Reads; Clean Read No.: High quality sequence read number; Bases (bp): Total number of bases; Clean Data (bp): High quality sequence base number; Clean Reads %: High quality sequence reads accounted for the percentage of sequencing reads; Clean Data %: High quality sequence bases accounted for the percentage of sequencing bases.

implemented in the HISAT2 software. The mapping efficiencies for each sample are detailed in Supplementary Table 2.

The mRNA expression of *P.tricornutum* was normalized using the Fragments Per Kilo base of transcript per Million mapped reads (FPKM) metric across various genes and samples. This study identified12392 mRNA genes. Gene prevalence in each sample was enumerated based on mRNA expression profiles, with the statistical interplay of unique and common genes across all samples illustrated in Figure 1. The inter-sample correlation, based on FPKM values, revealed varying degrees of differential expression amongst the four strains under grazing stress (Figure 2). Notably, the correlation coefficient between the control and grazing groups for all four strains exceeded 0.75, with strain pt52_B exhibiting the highest correlation (0.93), indicating relatively lesser differential expression under grazing pressure. Differential gene expression analysis, conducted using DESeq, adhered to criteria of |log2FoldChange| >1 and P-value < 0.05. This approach identified 1235 mRNA genes as significantly differentially expressed. Comparative analysis between control and grazing pressure groups uncovered distinct expression patterns for the four strains under grazing stress, with additional expression data from the CCAP1055_1 strain represented in Figure 3.

For LncRNA gene expression, FPKM normalization was similarly differences applied. Our analysis identified a total of 505 LncRNAs, with correlation analysis indicating varied degrees of differential expression among the strains under grazing pressure. Notably, between pt52_B_c and pt52_B_g were statistically insignificant, while pronounced differences were observed in the pt55_3_c vs pt55_3_g and pt55_7_c vs pt55_7_g pairings, as evidenced in Figure 2. Differential analysis of LncRNA expression set the criteria for differentially expressed genes at |log2FoldChange| >1 and P-value < 0.05. After this filtration, 65 LncRNA transcriptional genes showed significant differential expression, and cis-acting target gene prediction was performed for these genes, with corresponding target mRNAs identified (the relationships are delineated in the Supplementary Table 3).

For CircRNA expression, normalization employed the Transcripts per Million of a gene (TPM) metric, identifying a total of 536 CircRNAs. Inter-sample expression correlations, based on TPM, indicated varied differential expressions across strains under grazing pressure, with no significant variations among control groups (Figure 2). Despite using consistent differential expression criteria of |log2FoldChange| > 1and P-value < 0.05, no CircRNA showcased significant differential expression upon further analysis.

miRNA expression was normalized using the Counts per Million (CPM) metric, identifying 297 known miRNAs. Correlation analyses of sample expressions, based on CPM, showed notably low relationships among the four strains under grazing stress (Figure 2).

3.2 Proteomic profiling of *P.tricornutum*

In our comprehensive proteomic analysis of eight *P.tricornutum* samples, 4984 proteins and 22985 peptides were identified. Proteins exhibiting differential expression, defined by an expression alternation exceeding 1.5-fold (upregulated beyond 1.5 times or downregulated below 0.67-fold), are detailed in Table 2.

To glean deeper biological insights, significant differentially expressed proteins were subjected to KEGG pathway enrichment analysis, the details of which are presented in Supplementary

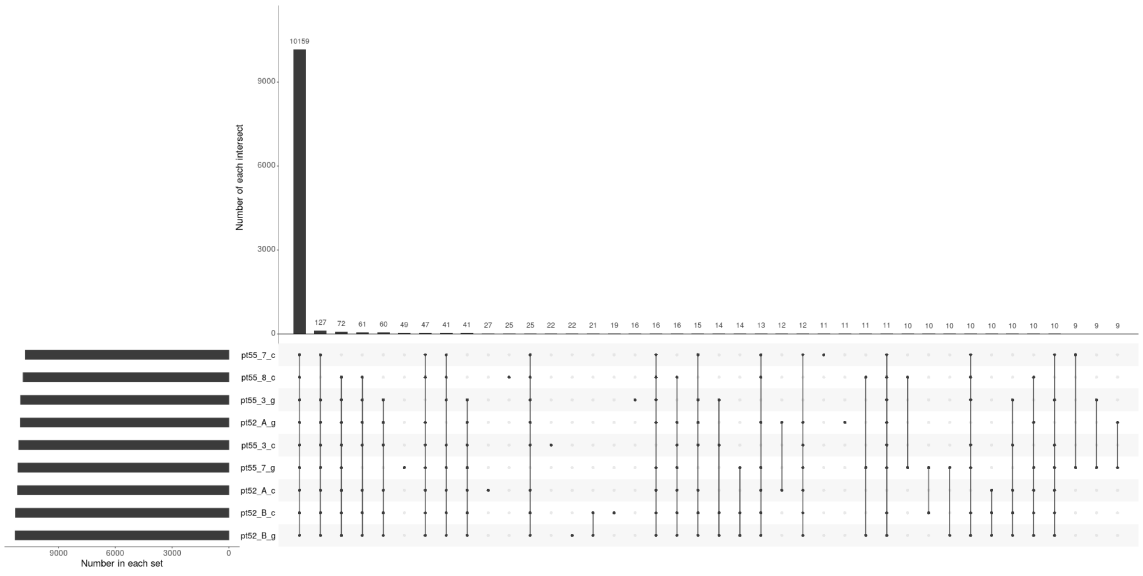


FIGURE 1
Upset plot representing gene identification across samples. Each set's cardinality represents the total count of genes identified within that particular sample. Intersection cardinalities denote the number of genes detected in multiple samples. The horizontal line connecting all points signifies the universally identified genes across all samples, while the remaining lines, connecting single or multiple points, represent genes exclusive to their respective samples.

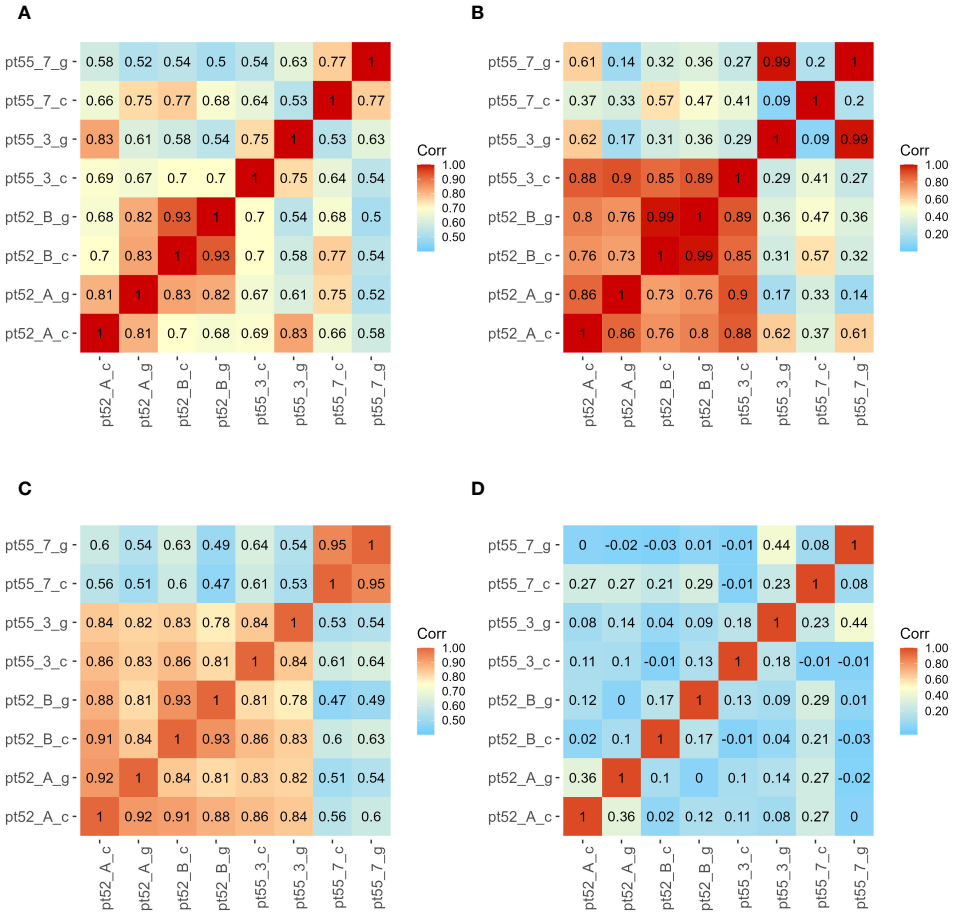


FIGURE 2
Sample correlation Analysis. Sample identifiers are presented on the left and bottom. The color gradient within each square quantifies the correlation strength between corresponding sample pairs. Subfigure (A) corresponds to mRNA. Subfigure (B) denotes LncRNA. Subfigure (C) represents CircRNA. Subfigure (D) highlights miRNA.

Table 5. This proteomic exploration was further enriched by integrating mRNA, LncRNA, and metabolomic data for an encompassing metabolic pathway enrichment and analysis.

3.3 Metabolomic profiling of *P.tricornutum*

Metabolomic analysis of eight *P.tricornutum* samples, using Compound Discoverer 3.0 software, yielded 5967 positive ion peaks and 5814 negative ion peaks. Noteworthy Total Ion Chromatogram (TIC) patterns emerged from the Mass Spectrometry (MS) evaluation. The significant differential metabolites were clustered and enriched by KEGG pathway (Supplementary Table 6). This analysis was further integrated with mRNA, LncRNA, and proteomic data to perform comprehensive metabolic pathway enrichment and analysis.

3.4 Pathway enrichment and analysis

Significantly differentially expressed genes within *P.tricornutum* mRNA under grazing pressure were aligned with both Gene Ontology (GO) and KEGG metabolic pathways (Figure 4). This revealed prominent involvement in pathways such as Glycolysis and Gluconeogenesis, Fatty Acid Biosynthesis, the Fatty Acid Elongation, Fatty Acid Degradation and Pyruvate Metabolism, Protein Processing in Endoplasmic Reticulum, Carbon Fixation in Photosynthetic Organisms, Nitrogen Metabolism. Furthermore, pathways linked with stress responses, including Calcium Signaling, Oxidative Stress, Nitrosative Stress, and Antioxidant Activity, were also noteworthy. The specific data of KEGG Pathway Enrichment in Transcriptomic (Supplementary Table 7), Proteomic (Supplementary Table 5) and Metabolomic (Supplementary Table 6) are in the Appendix.

Concurrently, genes markedly differentially expressed across the transcriptome, proteome and metabolome of *P.tricornutum* under grazing conditions were compared against the KEGG metabolic pathway (Figure 5).

3.5 Weighted gene co-expression network analysis

Utilizing the expression data of 12802 selected genes across eight samples, a weighted gene co-expression network was constructed.

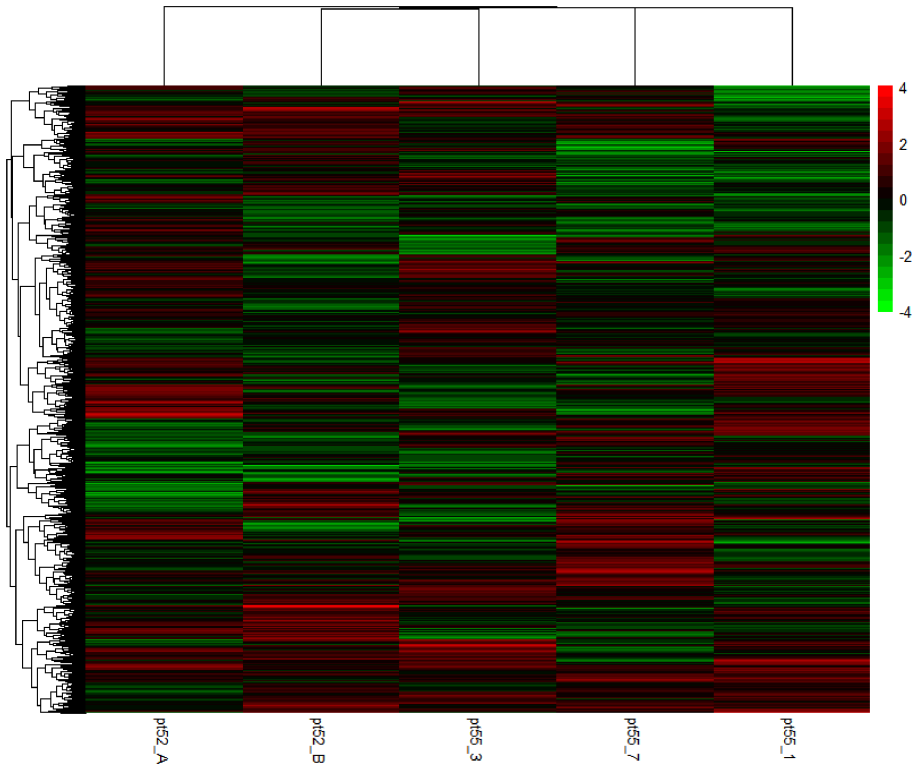


FIGURE 3
Cluster analysis of heat map of DEGs. Green bars represent the down-regulated genes. Red bars represent the up-regulated genes. The color depth represents the log2 ratio.

Network Construction: A cluster analysis was performed on the eight samples to identify outlying values. A soft threshold was determined to achieve a scale-free topological fit coefficient (R^2) of 0.8 (Supplementary Figure 8). Hierarchical clustering was employed to segregate genes into modules based on their correlations, followed by merging modules exhibiting similar expression patterns. Ultimately, 13 distinct modules were identified, ranging in size from 11 to 2066 genes, each represented by different colors (Supplementary Table 8, Figure 6).

Phenotypic Association and Interaction Analysis: Phenotypic data were associated with gene co-expression modules following outlier detection. The characteristic genes within each module were considered as representatives for module-specific expression profiles. Correlations between module characteristic genes and various phenotypes were computed, providing insights into the relationships within gene co-expression modules (Figures 7, 8).

Identification of Core Genes: Within the network, Gene Significance (|GS|) and Module Membership (|KME|) indices of ≥0.8 were used to ascertain core genes within modules. A comprehensive analysis was conducted to identify the module with the highest phenotype correlation. Correlations between phenotypes and co-expression modules were further validated by mapping gene connectivity within each module (Figure 9).

KEGG Pathway Annotation: The R KEGGREST Package (Version 1.36.0) was employed for annotation, screening with a significance level of $p < 0.05$ (Supplementary Table 9). A total of 29 phenotypic core genes were annotated across 38 pathways, with varying correlations to distinct phenotypic modules.

Integration of Multi-Omics Data: The core genes were compared and associated with the transcriptome, proteome, and metabolome. The resulting multi-omics co-expression network was visualized via Cytoscape (Version 3.9.1), elucidated in Figure 10.

TABLE 2 Protein quantification and differences.

Comparisons	Up	Down	All
1052_1A_c VS 1052_1A_g	134	128	262
1052_1B_c VS 1052_1B_g	116	202	318
1055_3_c VS 1055_3_g	227	389	616
1055_7_c VS 1055_7_g	434	326	760

4 Discussion

4.1 Transcriptomic insights and omics convergence in *P.tricornutum* Strains under grazing pressure

In our exploration of the transcriptome across four *P.tricornutum* strainsunder grazing pressure, a discernible convergence emerged: genes consistently exhibited up-regulation

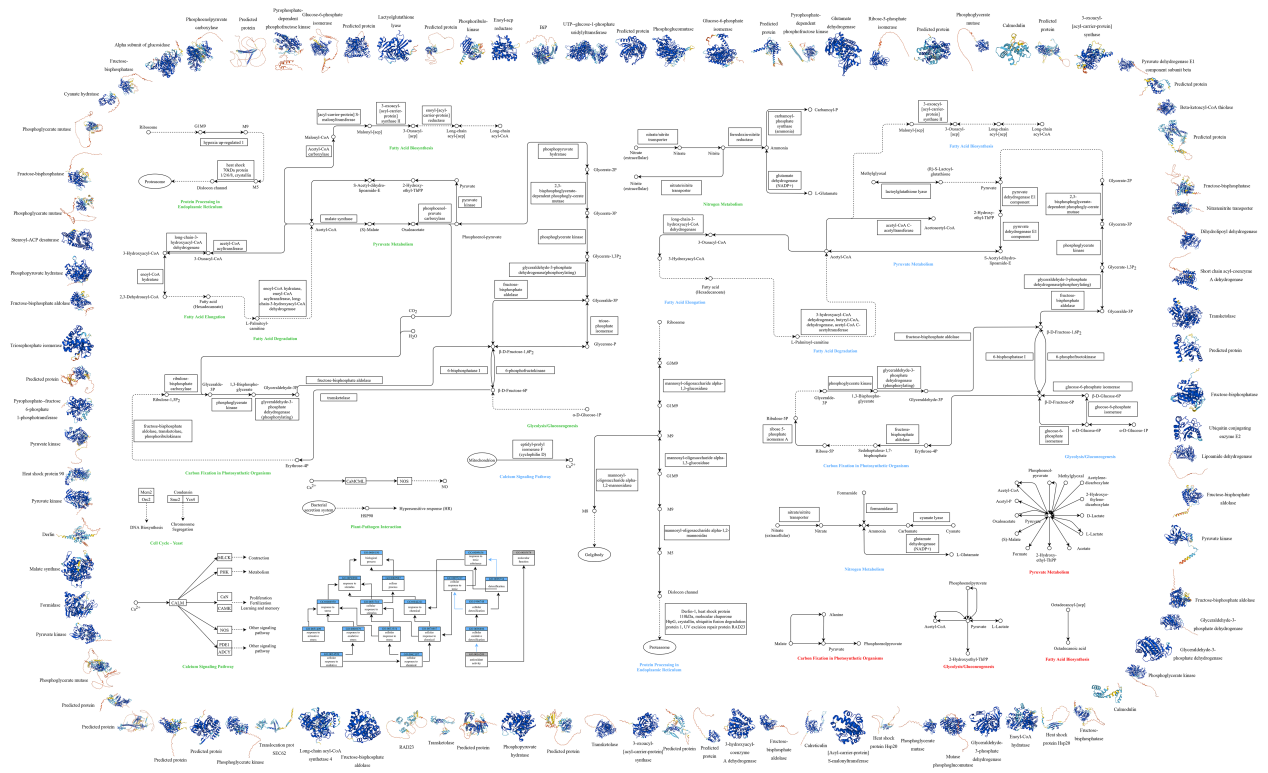


FIGURE 4
Multi-omics Cell Function Map of P.tricornutum. This figure illustrates the integrated analysis of transcriptomics, proteomics, and metabolomics, along with GO and KEGG pathway analysis, highlighting the significant variations under grazing pressure. Outer Ring: The periphery of the diagram represents the protein structure involved, color-coded by Model Confidence according to the AlphaFold's per-residue confidence score (pLDDT): Blue: pLDDT > 90, Cyan: 70 ≤ pLDDT ≤ 90, Yellow: 50 ≤ pLDDT ≤ 70,Orange: pLDDT < 50. Regions with low pLDDT may signify isolated or unstructured segments. Middle Section: Encompasses GO and KEGG pathway analysis for P.tricornutum under grazing pressure, integrating transcriptomics, proteomics, and metabolomics: GO Analysis (Lower Left Corner): Blue rectangles indicate Process IDs and names; gray rectangles indicate Function IDs and names; black arrows represent "is" relationships; blue arrows denote "part of" relationships. KEGG Metabolic Pathway Analysis (Remaining Middle Section): Circles: Various compounds, DNA, or other molecules; Solid Arrows: Direct interactions or relationships; Dashed Arrows: Indirect connections or unidentified reactions; Double Arrows: Reversible conversions; Rectangles: Required genes or enzymes; Ellipses: Specific organelles; Font Colors: Represent the metabolic pathways obtained through different analyses - green for transcriptomic, blue for proteomic, and red for metabolomic.

or down-regulation, albeit with varying intensities. While there are shared defensive responses among strains when exposed to herbivorous stressors, distinct transcriptomic regulatory mechanisms also exist across different strains. Notably, the pt55_1 strain from the prior studies (Li and Ismar, 2018) exhibited transcriptomic patterns similar to those of the pt55_3 strain in our analysis.

The advancement in omics methodologies, notably RNA, protein, and metabolic level analyses (Blum et al., 2022; Cheng et al., 2023), has allowed for a more integrated and comprehensive examination of complex networks, from genes to phenotypes. The omic patterns observed in *P.tricornutum* strains underpin a nuanced understanding of organismal responses to external pressures.

Notably:

Cellular Processes: A uniform down-regulation of genes was observed at the transcriptomic level, while the proteome exhibited an equitable distribution of up- and down-regulated genes, demonstrating overall subdued differential expression.

Organic Systems and Environmental Information Processing: Consistent up-regulation of genes was noted at both transcriptomic and proteomic levels, with a more tempered up-regulation in the proteome compared to the transcriptome.

Genetic Information Processing: A stark contrast between the transcriptome and proteome emerged, with most transcriptomic alterations showing down-regulation, whereas the proteome predominantly exhibited up-regulation. The metabolome added another dimension, displaying varied degrees of up-regulation in significantly altered genes.

Metabolism Pathways: Under grazing stress, *P.tricornutum* employed diverse gene regulatory strategies across different metabolic pathways as a defense mechanism. Noteworthy trends include: down-regulation across both transcriptomic and proteomic dimensions in the Cell Cycle pathway, with a more subdued effect at the protein level; down-regulation in pathways related to Plant-pathogen interaction and Calcium signaling, in response to grazing pressures; contrasting trends in Aminoacyl-tRNA biosynthesis with

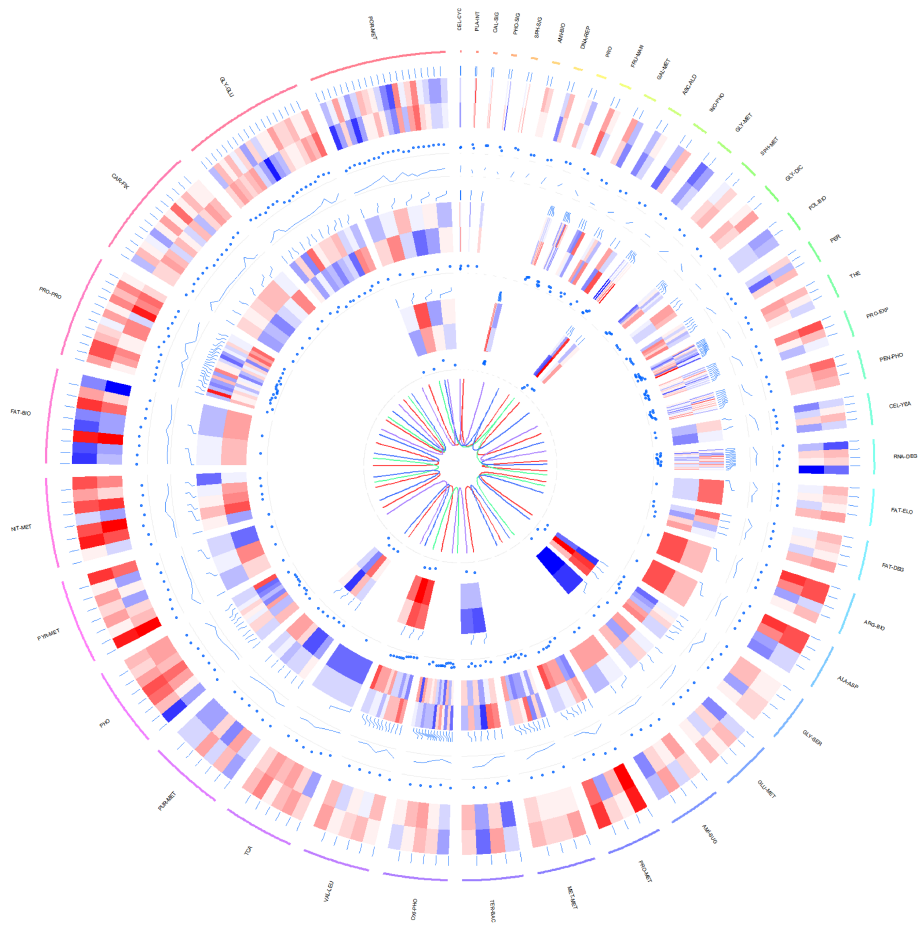


FIGURE 5
Circos plot illustrating differential expression across transcriptomic, proteomic, and metabolomic data in P.tricornutum's KEGG pathway under grazing stress. Transcriptomic Data: Represents the log10-transformed gene transcription levels for each KO. Proteomic Data: Demonstrates the log2-transformed expression levels of proteins, comparing the grazing pressure group with the control group. Metabolomic Data: Features the log10-transformed expression profiles of metabolites, contrasting the grazing pressure and control groups. From the outside in: The first circle is the metabolic path corresponding to each gene and its text annotation, the bar color is meaningless; The second circle is a heat map of the data expression at the transcriptomic level in the metabolic pathway (red to blue with red representing up-regulation and blue representing down-regulation); The third circle is the dot plot of differential expression values of each gene corresponding to the grazing pressure group and the control group in transcriptomic level. The fourth circle is the line chart of the P-Value of each gene in the transcriptomic level; The fifth circle is a heat map of data expression at the proteomic level of metabolic pathways (color meaning is the same as transcriptomic data); The sixth circle is the dot plot of the differential expression values of each gene corresponding to the grazing pressure group and the control group in the proteomic level. The seventh circle is a heat map of data expression at the metabolomic level in the metabolic pathway (color meaning is the same as the transcriptomic data); The eighth circle is the dot plot of the differential expression values of each gene corresponding to the grazing pressure group and the control group in the data of the metabolomic level. The ninth circle is the relationship diagram of the four strains distinguished by color, where brown represents pt52_A, seagreen represents pt52_B, royalblue represents pt55_3, and mediumpurple represents pt55_7. The outermost metabolic path annotation is: CEL-CYC (Cell cycle); PLA-INT (Plant-pathogen interaction); CAL-SIG (Calcium signaling pathway); PHO-SIG (Phosphatidylinositol signaling system); SPH-SIG (Sphingolipid signaling pathway); AMI-BIO (Aminoacyl-tRNA biosynthesis); DNA-REP (DNA replication); PRO (Proteasome); FRU-MAN (Fructose and mannose metabolism); GAL-MET (Galactose metabolism); ASC-ALD (Ascorbate and aldarate metabolism); INO-PHO (Inositol phosphate metabolism); GLY-MET (Glycerophospholipid metabolism); SPH-MET (Sphingolipid metabolism); GLY-DIC (Glyoxylate and dicarboxylate metabolism); FOL-BIO (Folate biosynthesis); PER (Peroxisome); THE (Thermogenesis); PRO-EXP (Protein export); PEN-PHO (Pentose phosphate pathway); RNA-DEG (RNA degradation); FAT-ELO (Fatty acid elongation); FAT-DEG (Fatty acid degradation); ARG-BIO (Arginine biosynthesis); ALA-ASP (Alanine, aspartate and glutamate metabolism); GLY-SER (Glycine, serine and threonine metabolism); GLU-MET (Glutathione metabolism); AMI-SUG (Amino sugar and nucleotide sugar metabolism); PRO-MET (Propanoate metabolism); MET-MET (Methane metabolism); TER-BAC (Terpenoid backbone biosynthesis); OXI-PHO (Oxidative phosphorylation); VAL-LEU (Valine, leucine and isoleucine degradation); TCA (Citrate cycle (TCA cycle)); PUR-MET (Purine metabolism); PHO (Photosynthesis); PYR-MET (Pyruvate metabolism); NIT-MET (Nitrogen metabolism); FAT-BIO (Fatty acid biosynthesis); PRO-PRO (Protein processing in endoplasmic reticulum); CAR-FIX (Carbon fixation in photosynthetic organisms); GLY-GLU (Glycolysis/Gluconeogenesis); POR-MET (Porphyrin metabolism).

down-regulation at the transcriptome but pronounced up-regulation at the metabolome; varied gene expression in Terpenoid backbone biosynthesis across transcriptome and proteome; yet a unified up-regulation in the metabolome. In sum, the contrasting gene expressions across omics, especially in pathways like Valine, leucine, and isoleucine degradation, reaffirm the intricate and multifaceted nature of gene regulation in *P.tricornutum* under grazing pressure.

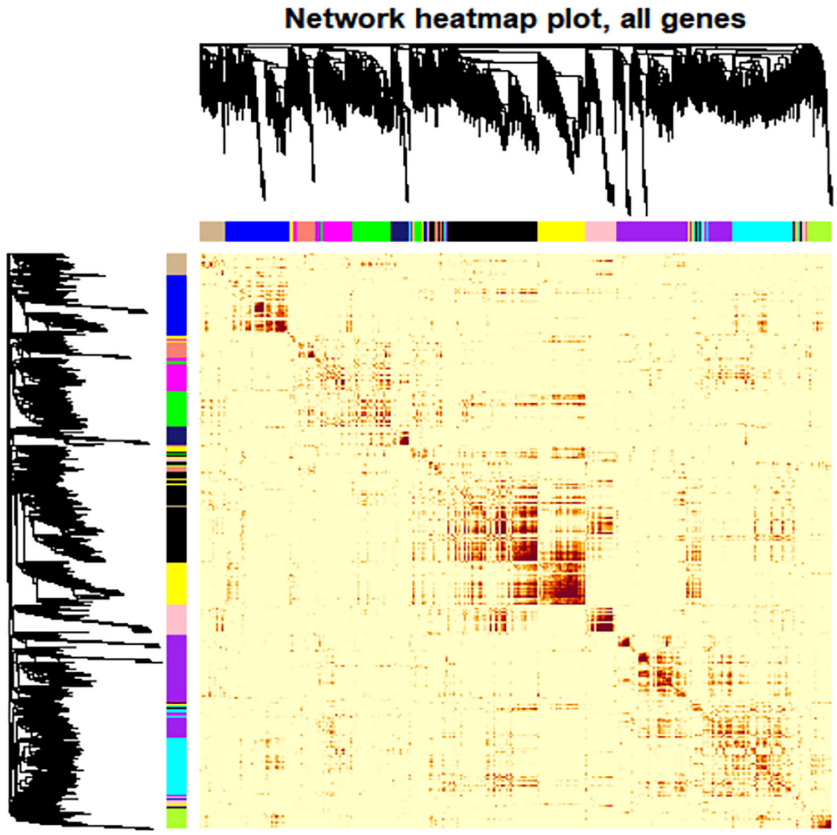


FIGURE 6
Heat map depiction of the topological overlapping matrix (TOM). This visualization represents the gene values within the TOM, providing insights into gene interactions within and across modules. Hierarchical Clustering and Modules: The upper and left sides of the graph depict a hierarchical clustering tree and the corresponding core presentation modules identified by a dynamic tree cutting algorithm. TOM Values: The color gradient within the figure illustrates the TOM values, while progressively darker shades indicate higher value. Interpretation: The most profound darkness near the diagonal line signifies the strongest gene interaction within individual modules. In contrast, The darker regions further from the diagonal reveal interactions between corresponding modules.

4.2 Mechanisms of cellular response under grazing pressure in *P.tricornutum* Strains

4.2.1 Cellular processes

Cellular Processes primarily reflect alterations in the abundance, size, and cell cycle of *P.tricornutum* cells in response to grazing pressure. A consistent up-regulation in genes associated with protein synthesis was observed across different strains at both the transcriptomic and proteomic levels. In particular, triradiate strains demonstrated an enhanced capacity to modulate the cell cycle under grazing pressure compared to fusiform strains, adapting more effectively to planktonic lifestyles under environmental stress (Song et al., 2020).

The endoplasmic reticulum (ER) serves as a crucial hub for protein synthesis and maturation, facilitating the post-translational modification, folding, and oligomerization of newly synthesized proteins (Supplementary Figure 14). Under grazing pressures, transcriptomic analysis of pt55_7 strains highlighted the role of CALR (calreticulin) in recognizing G1M9 glycoproteins in the ER. This process involves the inhibition of activity through hydrophobic entrapment and the promotion of folding in newly synthesized glycoproteins (Hirano et al., 2015). Concurrently, an increase in ER chaperone BiP, integral for luminal chaperones recognition, was noted, bolstering ER protein homeostasis (Kyeong and Lee, 2022). BiP, acting as a sentinel of ER integrity, targets aberrant proteins for proteasomal degradation and curtails aggregation, as anchoring the protein quality control system (Pobre et al., 2019). Chaperones such as DNAJA1 and CRYAA play a crucial role in marking aberrant proteins for ER-associated degradation (ERAD) (Shen et al., 2002). Proteomic scrutiny of pt55_7 strains under grazing pressures, accentuated the up-regulation of GANAB, facilitating deglycosylation processes (Gallo et al., 2018), and ERManI, responsible for mannose glycol-groups cleavage (Maki et al., 2022). The increased of HSP110 activity further substantiates the role of heat proteins in ER-associated degradation (Hrizo et al., 2007), while a decline in SKP1 activity implies attenuated proteasomal ubiquitination (Yoshida and Tanaka, 2010). This integrative analysis illuminates a nuanced regulation of protein synthesis and quality control in the pt55_7 strain's ER, characterized by enhanced protein synthesis alongside proteasomal ubiquitination. This nuanced protein regulation under grazing pressures, varies across strains, highlighting the heterogeneity and complexity of their cellular responses.

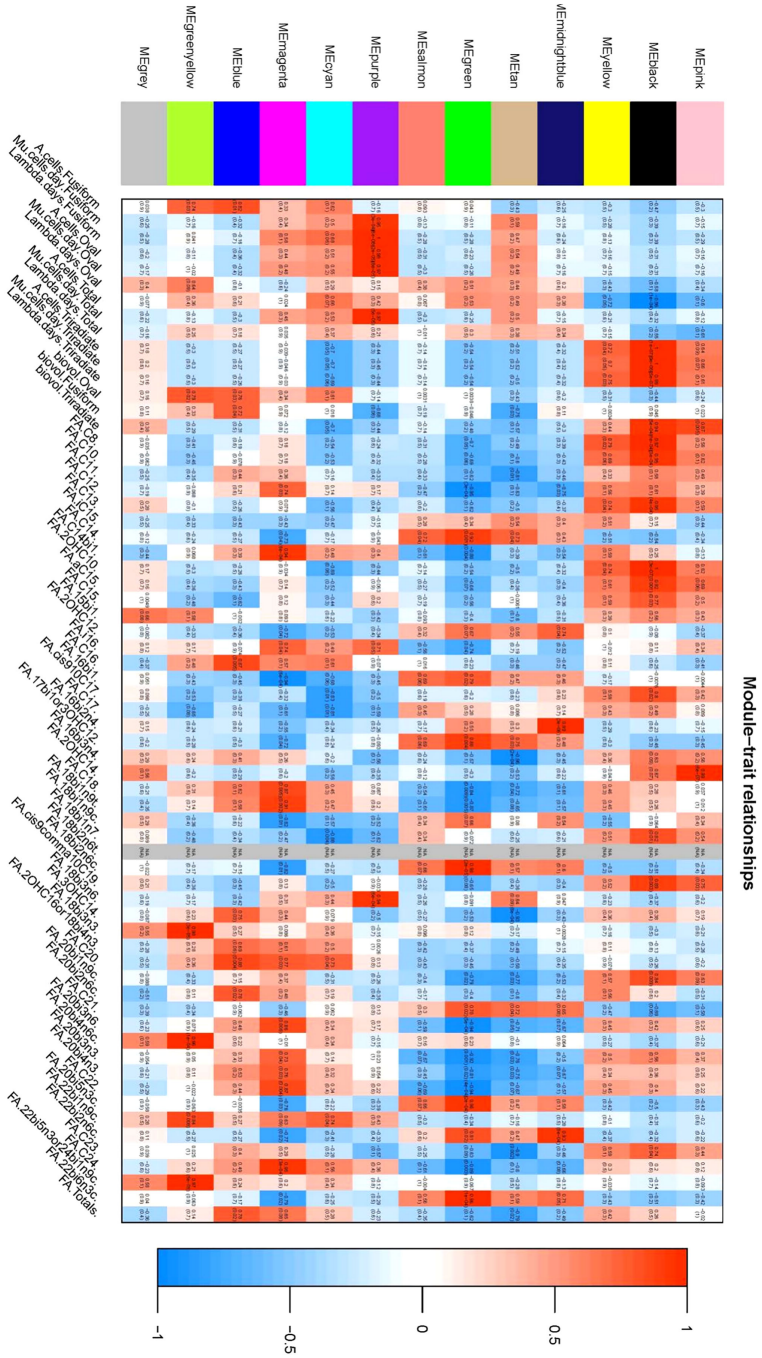


FIGURE 7
Heat map of correlation illustrating the correlation between coexpression modules and phenotypes: This figure delineates the relationship between individual coexpression modules and specific phenotypes. The X-axis represents individual co-expression modules, depicted in the left color block, while the Y-axis represents all phenotypes, shown on the lower side of the figure. within the graph, color coding is used to represent the correlation values. Red indicates a positive correlation, blue denotes a negative correlation, and white signifies no correlation. Each cell within the figure contains two values; The upper value is the Pearson correlation coefficient, and the lower value represents the P-value. the figure elucidates that the two co-expression modules most positively correlated with each phenotype are identified as the black module and the purple module.

In the Cell Cycle-yeast metabolic pathway (Supplementary Figure 15), specific genes (MCM2, ORC2, SMC2, YCS4) in the pt55_7 strain are down-regulated in response to grazing pressure, suggesting potential impediments in DNA replication and metaphase chromosome compaction. These intricate interactions among McM2-7 proteins, ORC components, and other essential factors like CDC6 and CDT1 are fundamental to DNA replication processes (Remus et al., 2009; Shibata and Dutta, 2020). Additionally, the collaboration between SMC2 and YCS4 establish a protein condensin complex pivotal for DNA morphology (Stray et al., 2005; Hassler et al., 2019). Among the strains studied, only pt55_7 exhibited significant alterations in the cell cycle, suggesting

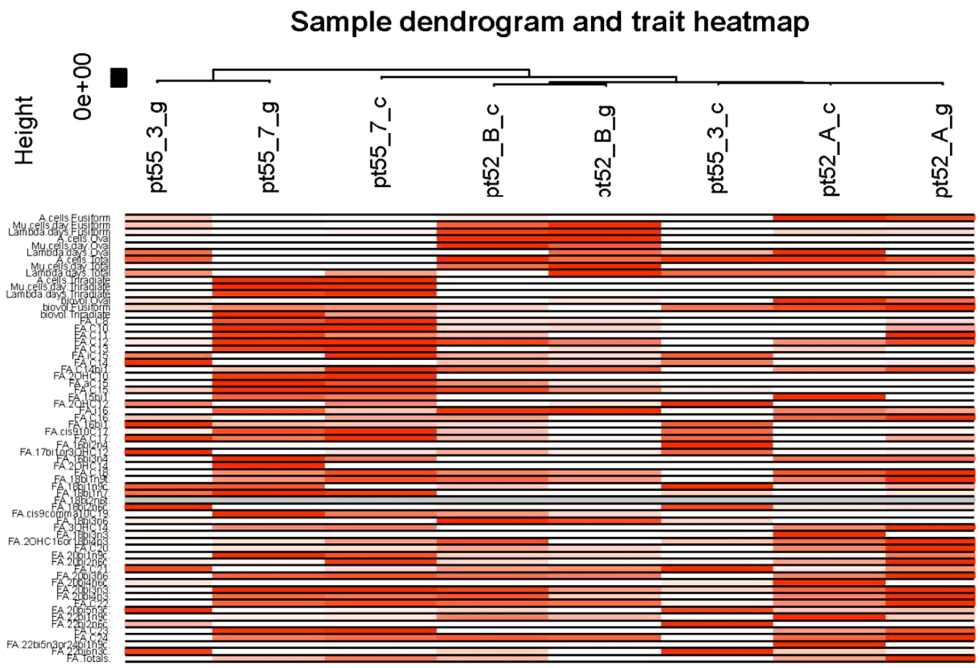


FIGURE 8
Sample and phenotypic correlation heat map. The X-axis designates the samples while the Y-axis illustrates various phenotypes. Progressive shades of red reflect the strength of the correlation between samples and phenotypes. White signifies the absence of correlation, while gray denotes missing data points.

that the triradiate pt55_7 strain of *P.tricornutum* is more likely to modulate its cell cycle in response to grazing pressures than its fusiform strains.

4.2.2 Fatty acid metabolism

Fatty acids play vital physiological roles in organisms, and the intricate alterations of fatty acid composition in *P.tricornutum* in response to grazing pressure are crucial in multi-omics analysis. The synthesis efficiency of long-chain fatty acids decreased at the transcriptome level and exhibited a slight increase at the proteome level, alongside an overall reduction in fatty acid abundance. This aligns with previous findings that environmental changes impact lipid profiles (Martino et al., 2011), with the triradiate strain showing greater resilience to grazing stress compared to the fusiform strain.

The study of fatty acid metabolism, especially in pt55_3, clarified the influence of grazing pressure on the biosynthesis, elongation, and degradation pathways. Within the fatty acid biosynthesis pathway (Supplementary Figure 16), a significant down-regulation of ACACA (Acetyl-CoA carboxylase) implies a reduction in ATP-dependent carboxylation of Acetyl-CoA to Malonyl-CoA under grazing pressure (Pereira et al., 2022). Enoyl ACP reductases, including FabI and its isomers (FabL, FabK, FabV, and InhA), play a crucial role in many microorganisms (Rana et al., 2020). A decline in the expression of FabI leads to a reduced processing of trans-2-Enoyl-[ACP] and Long-chain acyl-[ACP]. Prior to utilization within organisms, fatty acids require CoA conjugation (Ma et al., 2022). The down-regulation of FadD results in a slower conversion of Long-chain fatty acid to Long-chain acyl-CoA. Collectively, these findings highlight a decreased efficiency in fatty acid biosynthesis at the transcriptomic level due to grazing pressure. Within the fatty acid elongation metabolic pathway (Supplementary Figure 17), the diminished activity of HADH (3-hydroxyacyl-CoA dehydrogenase) during the processing of fatty acids containing between 4 and 16 carbon atoms suggests a suppressed oxidation reaction, consequently hindering fatty acid elongation at the proteomic level. Within the degradation pathway (Supplementary Figure 18), a decrease in the activities of HADH and ACADS results in a reduction of fatty acid dehydrogenation during the process of carbon chain reduction. Notably, the expression of ACAT, which is crucial in the terminal cleavage reactions of fatty acid degradation, is increased. Simultaneously, a down-regulated lncRNA, MSTRG.9570, has been found to be associated with ACAT modulation. Taken together, these findings suggest that grazing pressure hampers fatty acid oxidation while enhancing terminal cleavage reactions in degradation.

Collectively, omics analyses reveal that grazing pressures on *P.tricornutum* impede its fatty acid synthesis efficiency, favoring degradation processes, leading to a significant reduction in fatty acid quantity, which is in line with previous findings (Li and Ismar, 2018). At the transcriptomic level, the strains pt52_A and pt52_B exhibit enhanced fatty acid synthesis efficiencies, with variable reductions in elongation and degradation processes,not mirrored at the proteome and metabolome levels. Consequently, under grazing stress, both strains appear to favor the synthesis of short-chain fatty acids. In contrast, the pt55_7 strain exhibits only a modest proteomic increase in fatty acid synthesis, indicating

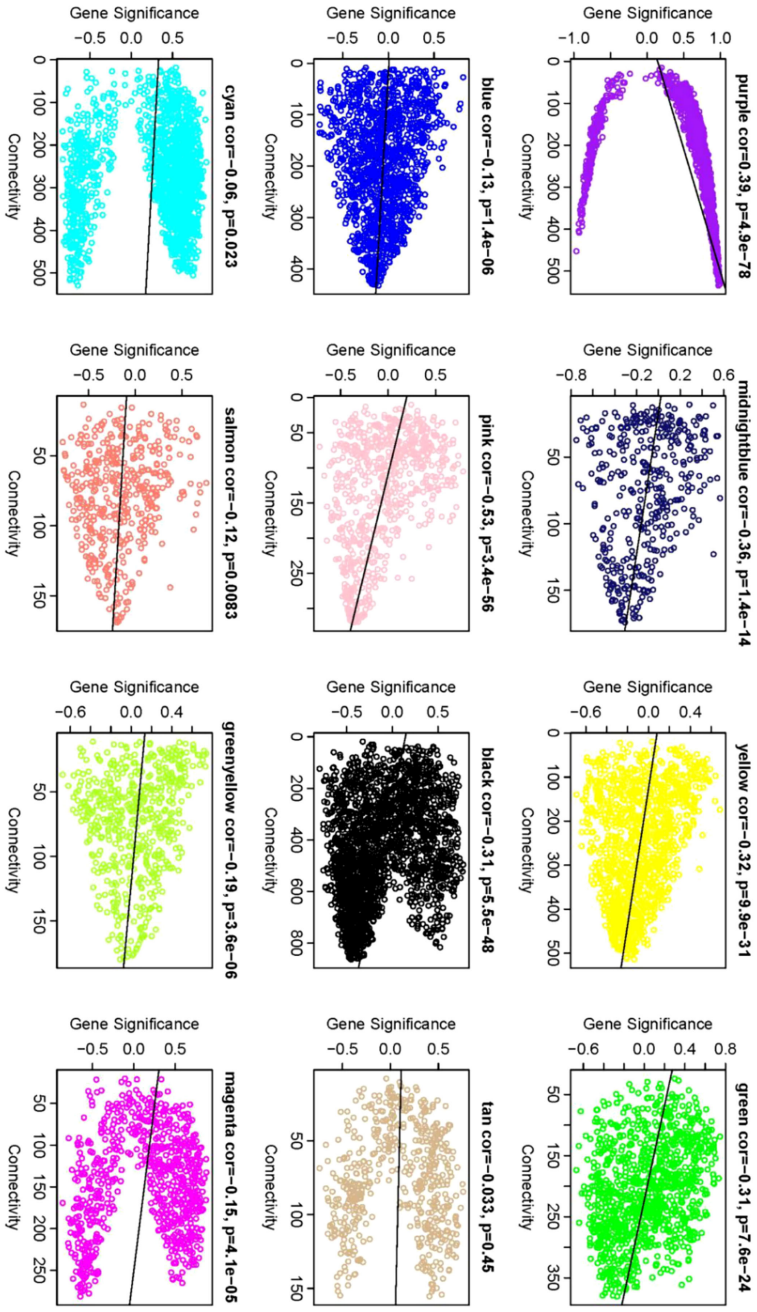


FIGURE 9
Gene significance vs. connectivity scatter plot. The scatter plot illustrates the relationship between a gene's significance relative to phenotype and its connectivity within modules. Each data point corresponds to an individual gene. Notably, core genes display heightened |GS| values and increased connectivity. Therefore, the genes located in the upper and lower right corners are classified as core genes. This visual allows for the discernment of modules most closely correlated with phenotype, providing complementary information to that depicted in Figure 7.

distinct response patterns between fusiform (pt52_A, pt52_B, pt55_3) and triradiate (pt55_7) strains under grazing pressures.

4.2.3 Signal response

Cells respond to environmental stress by undergoing complex signal transduction and regulatory actions. In the presence of grazing stress, *P.tricornutum* regulates Ca^{2+} levels through a variety of sensor proteins, resulting in changes at both the transcriptome and proteome levels.

Within the metabolic pathway of plant-pathogen interaction (Supplementary Figure 19), under grazing pressure, an up-regulation of CaMCML (Ca^{2+} calmodulin-like protein) in pt52_A strains suggests the translation of intracellular Ca^{2+} fluctuations into downstream signals through numerous sensor proteins, facilitating a defense against diverse stressors (Melo-Braga et al., 2012). This regulation could potentially lead to stomatal closure and increase the production of nitric oxide (NO), a molecule that is considered to be a sentinel against external threats and is thus crucial for plant

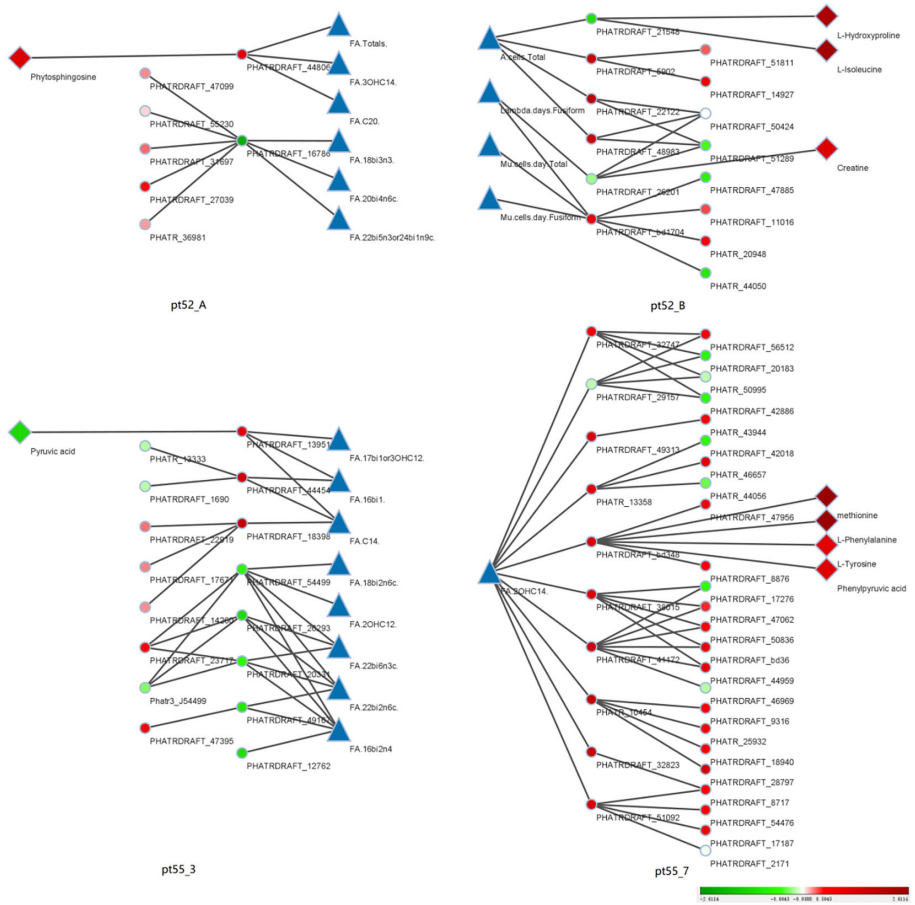


FIGURE 10
Coexpression relationships across phenotypic data: This figure portrays the coexpression interactions among core genes, transcriptional genes, proteins, and metabolites across various phenotypes, as discerned through WGCNA. Triangles represent distinct phenotypic data. Dots represent individual transcriptional gene IDs or protein gene IDs. Diamond symbolize specific metabolite names. Lines joining the icons denote co-expression relationships. A gradient from red to green depicts the degree of differential expression of transcriptional genes, protein genes or metabolites. Intense shades of red indicate significant up-regulation, while deeper greens indicate pronounced down-regulation. White suggests negligible differential expression. Blue is not representative of any specific data. The right periphery of each icon is labeled with its respective name. Additionally, each of the four segments is labeled with a specific strain name. The sequence from left to right consists of metabolites, protein genes, transcription genes, phenotypic data, and is mirrored for the right half of the figure.

defense (Robbins and Cowen, 2023). HSP90B, a universally conserved molecular chaperone, regulates the structure and function of numerous client proteins, many of which act as crucial signal transduction nodes (Kumar and Ohri, 2023). Intriguingly, under grazing stress, pt52_A strains manifest a down-regulation of HSP90B, potentially in response to stressors like the Hypersensitive Reaction (HR). Conversely, in the pt55_3 strain, an opposite trend is observed, where up-regulation of HSP90B may enhance signal transduction pathways to counteract stress.

In the Ca^{2+} signaling pathway (Supplementary Figure 20), the transcriptomic up-regulation of SPHK (sphingosine kinase) in pt52_A strains promotes sphingosine phosphorylation, which results in 1-phosphate sphingosine. Its equilibrium with ceramide plays a significant role in influencing sphingolipid dynamics (Adams et al., 2020; Smith et al., 2023). Moreover, with the up-regulation of CALM (calmodulin), Ca^{2+} modulates the MAPK signaling cascade, thus governing cellular processes including contraction, metabolism, and proliferation. The proteome of pt55_3 exhibits a remarkable up-regulation of PPIF (peptidyl-prolyl isomerase F (cyclophilin D)), which leads to compromised mitochondrial stability and functionality under Ca^{2+} stress (Gainutdinov et al., 2015). Consequently, it is apparent that under the influence of Ca^{2+}, both pt52_A and pt55_3 strains promote cell apoptosis at the transcriptomic level; proteomic evidence simultaneously suggests impaired mitochondrial function. This observation provides a coherent explanation for the diminished cellular efficacy observed in the Ca^{2+} signaling pathway under grazing pressure in these strains, while other strains appear to remain relatively unaffected. In response to grazing pressure, intricate regulatory modifications across signaling pathways are discernible, emphasizing the dynamic predator-prey relationship between the grazers and *P.tricornutum*.

4.3 Weighted gene co-expression network analysis

Multi-Omics Insights into *P.tricornutum*'s Adaptation Under Grazing Stress Utilizing Weighted Gene Co-expression Network

Analysis (WGCNA), core genes were correlated with transcriptomic, proteomic, and metabolomic datasets, constructing a multi-omics co-expression network pertinent to the observed phenotype (Li and Ismar, 2018). This analysis underscored the morphological modifications in the pt52_B and pt55_7 strains of *P.tricornutum* under grazing stress.

In pt52_B strain, phenotypic traits like total cell count, growth rate, and specific morphological variants of *P.tricornutum* were linked to metabolic routes such as Glycolysis/Gluconeogenesis (Supplementary Figure 21) and Carbon fixation in photosynthetic organisms (Supplementary Figure 22) at both the transcriptomic and proteomic levels. Notably, down-regulation of PGAM (2,3-bisphosphoglycerate-dependent phosphoglycerate mutase) inhibits the interconversion between 3-phosphoglyceric and 2-phosphoglyceric acid, highlighting PGAM's pivotal role in metabolic flux and redox balance, as supported by its modulation of mitochondrial ROS and activation of the pentose phosphate pathway (Mikawa et al., 2020). Concurrently, up-regulation of GAPDH (aldehyde 3-phosphate dehydrogenase (phosphorylating)) and PGK (phosphoglycerate) underscores their importance in the carbon flux during Glycolysis/Gluconeogenesis, influencing cellular growth dynamics under grazing stress. Earlier work has similarly indicated such morphological adaptations in Phaeocystis globose in response to grazing, spotlighting the universality of this response (Yang et al., 2023).

For pt55_7, a correlation was found between phenotypic measures and the down-regulation of POLA1 (DNA polymerase alpha subunit A) at the transcriptomic level. Given that DNA polymerase alpha orchestrates replication initiation, the diminished activity of POLA1 hints at perturbed DNA replication dynamics (Begemann et al., 2022). In light of our multi-omics findings, this suggests that the triradiate *P.tricornutum*'s pt55_7 strain employs cellular replication regulation as an adaptive mechanism under grazing stress.

In strains pt52_A, pt55_3 and pt55_7, phenotypic alterations, particularly in fatty acid profiles, showed correlations across all omics layers. A collective downtrend in fatty acid synthesis emerged, intimating a reduction in the overall fatty acid pool. This hints at a strategic recalibration by *P.tricornutum* under grazing stress, aligning with the notion that modulation of fatty acid profiles is a crucial defense mechanism for plankton against predators. Such adjustments, mirroring changes in environmental pressures, are evident in other marine phytoplankton, reinforcing the idea that shifts in fatty acid content are universal stress markers in phytoplankton under grazer-induced stress (Antacli et al., 2021; Yang et al., 2023).

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA1008380 http://www.proteomexchange.org/, PXD044954 https://www.ebi.ac.uk/metabolights/, MTBLS8485.

Author contributions

CL: Writing – original draft, Writing – review & editing, Data curation, Formal analysis. LL: Writing – review & editing, Writing – original draft. SY: Writing – review & editing, Data curation, Formal analysis. MW: Writing – review & editing, Investigation. HZ: Writing – review & editing, Investigation. SL: Writing – original draft, Writing – review & editing, Data curation, Formal analysis, Investigation, Funding acquisition.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The study was supported by Natural Science Foundation of Hebei Province, China (Grant No. D2020202004, Grant No. C2021202005).

Acknowledgments

The study was supported by Natural Science Foundation of Hebei Province, China (Grant No. D2020202004, Grant No. C2021202005).

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.2023.1308085/full#supplementary-material

References

Adams, D. R., Pyne, S., and Pyne, N. J. (2020). Structure-function analysis of lipid substrates and inhibitors of sphingosine kinases. *Cell Signal* 76, 109806. doi: 10.1016/j.cellsig.2020.109806

Antacli, J. C., Hernando, M. P., Troch, M. D., Malanga, G., Mendiolar, M., Hernández, D. R., et al. (2021). Ocean warming and freshening effects on lipid metabolism in coastal Antarctic phytoplankton assemblages dominated by sub-Antarctic species. *Sci. Total Environ.* 790, 147879. doi: 10.1016/j.scitotenv.2021.147879

Begemann, A., Oneda, B., Baumer, A., Guldimann, M., Tutschek, B., and Rauch, A. (2022). A Xp22.11-p21.3 microdeletion in a three-generation family supports male lethality of POLA1 nullisomy resulting in reduced fertility of female carriers. *Eur. J. Med. Genet.* 65 (12), 104628. doi: 10.1016/j.ejmg.2022.104628

Benton, H. P., Ivanisevic, J., Mahieu, N. G., Kurczy, M. E., Johnson, C. H., Franco, L., et al. (2015). Autonomous metabolomics for rapid metabolite identification in global profiling. *Anal. Chem.* 87 (2), 884–891. doi: 10.1021/ac5025649

Blum, B. C., Lin, W., Lawton, M. L., Liu, Q., Kwan, J., Turcinovic, I., et al. (2022). Multiomic metabolic enrichment network analysis reveals metabolite-protein physical interaction subnetworks altered in cancer. *Mol. Cell Proteomics* 21 (1), 100189. doi: 10.1016/j.mcpro.2021.100189

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008). The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* 456 (7219), 239–244. doi: 10.1038/nature07410

Chen, T., Ma, J., Liu, Y., Chen, Z. G., Xiao, N., Lu, Y. T., et al. (2021). iProX in 2021: connecting proteomics data sharing with big data. *Nucleic Acids Res.* 50 (D1), D1522–D1527. doi: 10.1093/nar/gkab1081

Cheng, Q., Zhang, J., Ding, H., Wang, Z. Y., Fang, J. Y., Fang, X., et al. (2023). Integrated multiomics analysis reveals changes in liver physiological function in Aqp9 gene knockout mice. *Int. J. Biol. Macromol* 245, 125459. doi: 10.1016/j.ijbiomac.2023.125459

Clement, R., Lignon, S., Mansuelle, P., Jensen, E., Pophillat, M., Lebrun, R., et al. (2017). Responses of the marine diatom Thalassiosira pseudonana to changes in CO2 concentration: a proteomic approach. *Sci. Rep-UK* 7, 42333. doi: 10.1038/srep42333

Gainutdinov, T., Molkentin, J. D., Siemen, D., Ziemer, M., Debska-Vielhaber, G., Vielhaber, S., et al. (2015). Knockout of cyclophilin D in Ppif(-)/(-) mice increases stability of brain mitochondria against Ca(2)(+) stress. *Arch. Biochem. Biophys.* 579, 40–46. doi: 10.1016/j.abb.2015.05.009

Gallo, G. L., Valko, A., Aramburu, S. I., Etchegaray, E., Völker, C., Parodi, A. J., et al. (2018). Abrogation of glucosidase I-mediated glycoprotein deglucosylation results in a sick phenotype in fission yeasts: Model for the human MOGS-CDG disorder. *J. Biol. Chem.* 293 (52), 19957–19973. doi: 10.1074/jbc.RA118.004844

Hassler, M., Shaltiel, I. A., Kschonsak, M., Simon, B., Merkel, F., Thärichen, L., et al. (2019). Structural basis of an asymmetric condensin ATPase cycle. *Mol. Cell* 74 (6), 1175–1188. doi: 10.1016/j.molcel.2019.03.037

Haug, K., Cochrane, K., Nainala, V. C., Williams, M., Chang, J. K., Jayaseelan, K. V., et al. (2020). MetaboLights: a resource evolving in response to the needs of its scientific community. *Nucleic Acids Res.* 48 (D1), D440–D444. doi: 10.1093/nar/gkz1019

Helliwell, K. E., Harrison, E. L., Christie-Oleza, J. A., Rees, A. P., Kleiner, F. H., Gaikwad, T., et al. (2021). A novel Ca2+ Signaling pathway coordinates environmental phosphorus sensing and nitrogen metabolism in marine diatoms. *Curr. Biol.* 31 (5), 978–989.e4. doi: 10.1016/j.cub.2020.11.073

Hirano, M., Adachi, Y., Ito, Y., and Totani, K. (2015). Calreticulin discriminates the proximal region at the N-glycosylation site of Glc1Man9GlcNAc2 ligand. *Biochem. Biophys. Res. Commun.* 466 (3), 350–355. doi: 10.1016/j.bbrc.2015.09.026

Hrizo, S. L., Gusarova, V., Habiel, D. M., Goeckeler, J. L., Fisher, E. A., and Brodsky, J. L. (2007). The Hsp110 molecular chaperone stabilizes apolipoprotein B from endoplasmic reticulum-associated degradation (ERAD). *J. Biol. Chem.* 282 (45), 32665–322675. doi: 10.1074/jbc.M705216200

Kang, W. L., Sun, S., and Hu, X. G. (2023). Microplastics trigger the Matthew effect on nitrogen assimilation in marine diatoms at an environmentally relevant concentration. *Water Res.* 223, 119762. doi: 10.1016/j.watres.2023.119762

Kim, S. Y., Hedberg, P., Winder, M., and Rydberg, S. (2022). Evidence of 2,4-diaminobutyric acid (DAB) production as a defense mechanism in diatom Thalassiosira pseudonana. *Aquat Toxicol.* 249, 106210. doi: 10.1016/j.aquatox.2022.106210

Kumar, D., and Ohri, P. (2023). Say "NO" to plant stresses: Unravelling the role of nitric oxide under abiotic and biotic stress. *Nitric. Oxide* 130, 36–57. doi: 10.1016/j.niox.2022.11.004

Kyeong, M., and Lee, J. S. (2022). Endogenous BiP reporter system for simultaneous identification of ER stress and antibody production in Chinese hamster ovary cells. *Metab. Eng.* 72, 35–45. doi: 10.1016/j.ymben.2022.02.002

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 29 (9), 559. doi: 10.1186/1471-2105-9-559

Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* 24 (5), 719–720. doi: 10.1093/bioinformatics/btm563

Lauritano, C., Romano, G., Roncalli, V., Amoresano, A., Fontanarosa, C., Bastianini, M., et al. (2016). New oxylipins produced at the end of a diatom bloom and their effects on copepod reproductive success and gene expression levels. *Harmful Algae* 55, 221–229. doi: 10.1016/j.hal.2016.03.015

Li, S., and Ismar, S. M. H. (2018). Transcriptome, biochemical and growth responses of the marine phytoplankter phaeodactylum tricornutum bohlin (bacillariophyta) to copepod grazer presence. *Cell Physiol. Biochem.* 46, 1091–1111. doi: 10.1159/000488839

Ma, J., Chen, T., Wu, S. F., Yang, C. Y., Bai, M. Z., Shu, K. X., et al. (2019). iProX: an integrated proteome resource. *Nucleic Acids Res.* 47 (D1), D1211–D1217. doi: 10.1093/nar/gky869

Ma, F., Zou, Y., Ma, L., Ma, R., and Chen, X. (2022). Evolution, characterization, and immune response function of long-chain acyl-CoA synthetase genes in rainbow trout (Oncorhynchus mykiss) under hypoxic stress. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 260, 110737. doi: 10.1016/j.cbpb.2022.110737

Maki, Y., Otani, Y., Okamoto, R., Izumi, M., and Kajihara, Y. (2022). Isolation and characterization of high-mannose type glycans containing five or six mannose residues from hen egg yolk. *Carbohydr Res.* 521, 108680. doi: 10.1016/j.carres.2022.108680

Martino, A. D., Amato, A., and Bowler, C. (2019). Mitosis in diatoms: rediscovering an old model for cell division. *Bioessays* 31 (8), 874–884. doi: 10.1002/bies.200900007

Martino, A. D., Bartual, A., Willis, A., Meichenin, A., Villazán, B., Maheswari, U., et al. (2011). Physiological and molecular evidence that environmental changes elicit morphological interconversion in the model diatom *Phaeodactylum tricornutum*. *Protist* 162 (3), 462–481. doi: 10.1016/j.protis.2011.02.002

Melo-Braga, M. N., Verano-Braga, T., Leon, I. R., Antonacci, D., Nogueira, F. C. S., Thelen, J. J., et al. (2012). Modulation of protein phosphorylation, N-glycosylation and Lys-acetylation in grape (Vitis vinifera) mesocarp and exocarp owing to Lobesia botrana infection. *Mol. Cell Proteomics* 11 (10), 945–956. doi: 10.1074/mcp.M112.020214

Mikawa, T., Shibata, E., Shimada, M., Ito, K., Ito, T., Kanda, H., et al. (2020). Phosphoglycerate mutase cooperates with chk1 kinase to regulate glycolysis. *iScience* 23 (7), 101306. doi: 10.1016/j.isci.2020.101306

Øie, G., Galloway, T., Sørøy, M., Hansen, M. H., Norheim, I. A., Halseth, C. K., et al. (2017). Effect of cultivated copepods (acartia tonsa) in first-feeding of atlantic cod (gadus morhua) and ballan wrasse (labrus bergylta) larvae. *Aquacult Nutr.* 23 (1), n/a–n/a. doi: 10.1111/anu.12352

Park, G., Norton, L., Avery, D., and Dam, H. G. (2023). Grazers modify the dinoflagellate relationship between toxin production and cell growth. *Harmful Algae* 126, 102439. doi: 10.1016/j.hal.2023.102439

Pei, G., Chen, L., and Zhang, W. (2017). WGCNA application to proteomic and metabolomic data analysis. *Method Enzymol.* 585, 135–158. doi: 10.1016/bs.mie.2016.09.016

Pereira, H., Azevedo, F., Domingues, L., and Johansson, B. (2022). Expression of Yarrowia lipolytica acetyl-CoA carboxylase in Saccharomyces cerevisiae and its effect on *in-vivo* accumulation of Malonyl-CoA. *Comput. Struct. Biotechnol. J.* 20, 779–787. doi: 10.1016/j.csbj.2022.01.020

Pobre, K. F. R., Poet, G. J., and Hendershot, L. M. (2019). The endoplasmic reticulum (ER) chaperone BiP is a master regulator of ER functions: Getting by with a little help from ERdj friends. *J. Biol. Chem.* 294 (6), 2098–2108. doi: 10.1074/jbc.REV118.002804

Rana, P., Ghouse, S. M., Akunuri, R., Madhavi, Y. V., Chopra, S., and Nanduri, S. (2020). FabI (enoyl acyl carrier protein reductase) - A potential broad spectrum therapeutic target and its inhibitors. *Eur. J. Med. Chem.* 208, 112757. doi: 10.1016/j.ejmech.2020.112757

R Core Team (2022). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: https://www.R-project.org/.

Remmers, I. M., Sarah, D., Martens, D. E., de Vos, R. C. H., Mumm, R., America, A. H. P., et al. (2018). Orchestration of transcriptome, proteome and metabolome in the diatom Phaeodactylum tricornutum during nitrogen limitation. *Algal Res.* 35, 33–49. doi: 10.1016/j.algal.2018.08.012

Remus, D., Beuron, F., Tolun, G., Griffith, J. D., Morris, E. P., and Diffley, J. F. X. (2009). Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* 139 (4), 719–730. doi: 10.1016/j.cell.2009.10.015

Robbins, N., and Cowen, L. E. (2023). Roles of Hsp90 in Candida albicans morphogenesis and virulence. *Curr. Opin. Microbiol.* 75, 102351. doi: 10.1016/j.mib.2023.102351

Shen, Y., Meunier, L., and Hendershot, L. M. (2002). Identification and characterization of a novel endoplasmic reticulum (ER) DnaJ homologue, which stimulates ATPase activity of BiP in *vitro* and is induced by ER stress. *J. Biol. Chem.* 277 (18), 15947–15956. doi: 10.1074/jbc.M112214200

Shibata, E., and Dutta, A. (2020). A human cancer cell line initiates DNA replication normally in the absence of ORC5 and ORC2 proteins. *J. Biol. Chem.* 295 (50), 16949–16959. doi: 10.1074/jbc.RA120.015450

Shukla, S. K., Crosta, X., and Ikehara, M. (2023). Synergic role of frontal migration and silicic acid concentration in driving diatom productivity in the Indian sector of the Southern Ocean over the past 350 ka. *Mar. Micropaleontol* 181, 102245. doi: 10.1016/j.marmicro.2023.102245

Sims, P. A., Mann, D. G., and Linda, K. (2006). Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia* 45 (4), 361–402. doi: 10.2216/05-22.1

Smith, C. J., Williams, J. L., Hall, C., Casas, J., Caley, M. P., O'Toole, E. A., et al. (2023). Ichthyosis linked to sphingosine 1-phosphate lyase insufficiency is due to aberrant sphingolipid and calcium regulation. *J. Lipid Res.* 64 (4), 100351. doi: 10.1016/j.jlr.2023.100351

Song, Z. D., Lye, G. J., and Parker, B. M. (2020). Morphological and biochemical changes in *Phaeodactylum tricornutum* triggered by culture media: Implications for industrial exploitation. *Algal Res.* 47, 101822. doi: 10.1016/j.algal.2020.101822

Stray, J. E., Crisona, N. J., Belotserkovskii, B. P., Lindsley, J. E., and Cozzarelli, N. R. (2005). The Saccharomyces cerevisiae Smc2/4 condensin compacts DNA into (+) chiral structures without net supercoiling. *J. Biol. Chem.* 280 (41), 34723–34734. doi: 10.1074/jbc.M506589200

Thamatrakoln, K. (2021). Diatom ecophysiology: crossing signals on the road to recovery from nutrient deprivation. *Curr. Biol.* 31 (5), 253–254. doi: 10.1016/j.cub.2021.01.016

Wisniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6 (5), 359–362. doi: 10.1038/nmeth.1322

Xie, J., Bai, X. C., Lavoie, M., Lu, H. P., Fan, X. J., Pan, X. L., et al. (2015). Analysis of the Proteome of the Marine Diatom Phaeodactylum tricornutum Exposed to Aluminum Providing Insights into Aluminum Toxicity Mechanisms. *Environ. Sci. Technol.* 49 (18), 11182–11190. doi: 10.1021/acs.est.5b03272

Yang, X., Yan, Z., Li, X. D., Li, Y. X., and Li, K. (2023). Chemical cues in the interaction of herbivory-prey induce consumer-specific morphological and chemical defenses in Phaeocystis globosa. *Harmful Algae* 126, 102450. doi: 10.1016/j.hal.2023.102450

Yoshida, Y., and Tanaka, K. (2010). Lectin-like ERAD players in ER and cytosol. *BBA-Biomembranes* 1800 (2), 172–180. doi: 10.1016/j.bbagen.2009.07.029

Zhang, S. W., Zheng, T. T., Lundholm, N., Huang, X. F., Jiang, X. H., Li, A. F., et al. (2021). Chemical and morphological defenses of Pseudo-nitzschia multiseries in response to zooplankton grazing. *Harmful Algae* 104, 102033. doi: 10.1016/j.hal.2021.102033

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 January 2024
DOI 10.3389/fpls.2023.1309678

OPEN ACCESS

EDITED BY
Weicong Qi,
Jiangsu Academy of Agricultural Sciences (JAAS), China

REVIEWED BY
Qijie Guan,
University of Mississippi, United States
Wenqin Wang,
Shanghai Normal University, China

*CORRESPONDENCE
Yueqiang Zhang
zhangyqyhm@163.com
Zheru Fan
fzr640814@qq.com

RECEIVED 08 October 2023
ACCEPTED 22 December 2023
PUBLISHED 18 January 2024

CITATION
Shi J, Wang L, Wang Z, Li J, Zhang H, Gao X, Wang C, Xia J, Zhao Z, Wang Z, Yang Z, Xu Z, Zhang Y and Fan Z (2024) Comparison of the transcriptome and metabolome of wheat (*Triticum aestivum* L.) proteins content during grain formation provides insight. *Front. Plant Sci.* 14:1309678. doi: 10.3389/fpls.2023.1309678

Comparison of the transcriptome and metabolome of wheat (*Triticum aestivum* L.) proteins content during grain formation provides insight

Jia Shi, Lihong Wang, Zhong Wang, Jianfeng Li, Hongzhi Zhang, Xin Gao, Chunsheng Wang, Jianqiang Xia, Zhun Zhao, Zhenlong Wang, Zhenyu Yang, Zihan Xu, Yueqiang Zhang* and Zheru Fan*

Institute of Nuclear and Biological Technologies, Xinjiang Academy of Agricultural Sciences/Xinjiang Key Laboratory of Crop Biotechnology/Crop Chemical Regulation Engineering Technology Research Center in Xinjiang, Urumqi, China

Introduction: Wheat is a food crop with a large global cultivation area, and the content and quality of wheat glutenin accumulation are important indicators of the quality of wheat flour.

Methods: To elucidate the gene expression regulation and metabolic characteristics related to the gluten content during wheat grain formation, transcriptomic and metabolomic analyses were performed for the high gluten content of the Xinchun 26 cultivar and the low proteins content of the Xinchun 34 cultivar at three periods (7 d, 14 d and 21 d) after flowering.

Results: Transcriptomic analysis revealed that 5573 unique differentially expressed genes (DEGs) were divided into two categories according to their expression patterns during the three periods. The metabolites detected were mainly divided into 12 classes. Lipid and lipid-like molecule levels and phenylpropanoid and polyketide levels were the highest, and the difference analysis revealed a total of 10 differentially regulated metabolites (DRMs) over the three periods. Joint analysis revealed that the DEGs and DRMs were significantly enriched in starch and sucrose metabolism; the citrate cycle; carbon fixation in photosynthetic organisms; and alanine, aspartate and glutamate metabolism pathways. The genes and contents of the sucrose and gluten synthesis pathways were analysed, and the correlation between gluten content and its related genes was calculated. Based on weighted correlation network analysis (WGCNA), by constructing a coexpression network, a total of 5 specific modules and 8 candidate genes that were strongly correlated with the three developmental stages of wheat grain were identified.

Discussion: This study provides new insights into the role of glutenin content in wheat grain formation and reveals potential regulatory pathways and candidate genes involved in this developmental process.

KEYWORDS

wheat, transcriptome, metabolome, glutenin content, WGCNA

1 Introduction

Wheat (*Triticum aestivum* L.) is the largest crop in the world, providing approximately 20% of the food available for humankind and one of the most important food crops (Chawade et al., 2018). With the improvements in living standards, the quality of wheat has received increasing attention (Chawade et al., 2018). Wheat grain protein is divided into albumin, globulin, gliadin and gluten according to solubility; gliadin and gluten are the main storage proteins and are the main components of gluten (Zheng et al., 2018; Hackenberg et al., 2019). Because of the presence of gliadin and gluten components, wheat flour can be kneaded with water to form a dough, which can be fermented, steamed or baked to obtain a variety of foods (Zheng et al., 2018; Hackenberg et al., 2019). The content and quality of gluten are important indicators for determining the quality of wheat flour, determining the process performance of dough and determining the quality of steamed and baked goods (Zheng et al., 2018; Hackenberg et al., 2019). Gluten imparts characteristics of water retention, cohesion, viscoelasticity, etc., that play a decisive role in the rheological properties and baking quality of dough (Zheng et al., 2018; Hackenberg et al., 2019). Therefore, exploring candidate genes for the study of wheat gluten content, analysing the underlying molecular mechanism, and improving the quality of wheat cultivars through breeding pathways are important tasks for modern wheat breeding.

Transcriptomics is a discipline that studies gene expression and transcriptional regulation in cells as a whole, and transcriptome analysis is necessary for exploring genome function and differential expression and plays an important role in studying plant growth and development (Chen et al., 2014; Pankievicz et al., 2016; Stelpflug et al., 2016; Hsu and Tung, 2017; Ji et al., 2022). Metabolomics refers to the inheritance and development of genomics, transcriptomics and proteomics and can directly reflect the biochemical pathways and potential molecular mechanisms in organisms by elucidating the metabolites downstream of the genome as a whole and subsequently revealing the relevant metabolic pathways and metabolic networks (Li et al., 2022; Prakash et al., 2023). In recent years, transcriptomic and metabolomic-based techniques have provided powerful tools and methods for revealing molecular characteristics and identifying candidate genes related to plant growth and development and fruit quality (Jiang et al., 2022; Wan et al., 2022; Zhang and Fernie, 2023). Coexpression network analysis is a systems biology method in which gene coexpression networks are constructed by analysing the correlation of gene expression to discover functionally relevant gene modules (Ma et al., 2021). Transcriptome and metabolome techniques have been used to detect differences in fruit flavour and carotenoid content in the early ripening (MG) and postripening (TR) stages of mango fruits (Peng et al., 2022). Transcriptome and metabolome data were used to study the accumulation of metabolites and transcriptional changes in the late-maturing cultivar Kate Mango at different stages of fruit development, and a regulatory network related to mango fruit ripening was constructed (Wu et al., 2022). By analysing clusters of metabolites and genes with the same tendencies to change in expression in cashews, 17 genes involved in phosphatidylinositol (PI) synthesis were found, and the transcription factor *WRKY11*, which can potentially regulate PI synthesis, was also identified (Zhao et al., 2022). Through poplar transcriptomics and metabolomics, the effects of miR156 on other microRNAs and their targets associated with anthocyanin biosynthesis were revealed (Wang et al., 2020).

The gluten content during the process of wheat grain formation is an important index for quality evaluation, and studying the gluten content is highly important for improving the quality of wheat (Zheng et al., 2018; Hackenberg et al., 2019). However, the gluten content during wheat grain formation involves complex polygenetic mechanisms, multisignalling pathways and metabolic processes. Therefore, in this study, the high-gluten-content Xinchun 26 cultivar and the low-gluten-content Xinchun 34 cultivar were selected for application of transcriptomic and metabolomic methods to conduct a cluster analysis of differentially expressed genes (DEGs) and differentially regulated metabolites (DRMs), KEGG enrichment analysis, and transcription factor (TF) expression analysis and to determine the key genes related to gluten content in wheat grain formation through coexpression analysis and qRT–PCR. This study provides new insights into the gluten content during wheat grain formation and reveals potential regulatory pathways and candidate genes involved in this developmental process.

2 Materials and methods

2.1 Plant material

The high gluten spring wheat cultivar Xinchun 26 and the low gluten spring wheat cultivar Xinchun 34 were chosen for the study. The above two cultivars were sown according to the designated community area of 4.8 m^2 in the military household experimental base of Changji city, Xinjiang, and the management method was the same as that used for conventional fields. Fertilization and watering were applied at the same time to ensure that the growth environment of the two cultivars was the same. Mid-spike grains exhibiting consistent growth were collected at 7 d, 14 d, and 21 d after flowering, and 14 replicates were collected for each variety (3 for RNA-seq sequencing, 5 for metabolome sequencing, 3 for physiological index determination, and 3 for qRT–PCR). Immediately after collection, the samples were flash frozen with liquid nitrogen, brought back to the laboratory and stored in a -80° C freezer.

2.2 RNA-seq sequencing and analysis

After DNase I (Illumina, USA) digestion of the sample total RNA, the mRNA was purified from 1 μg of total RNA using oligo (dT) magnetic beads, followed by mRNA fragmentation in ABclonal First Strand Synthesis Reaction Buffer. Subsequently, the first strand of cDNA was synthesized with random primers

and reverse transcriptase (RNase H) using fragmented mRNA as a template, and the second strand of cDNA was subsequently synthesized with dNTPs, RNAseH, DNA polymerase I and buffer and ligated to perform PCR amplification. The PCR products were purified, and the library quality was evaluated using an Agilent Bioanalyzer 4150 (Kusser et al., 2006). The constructed library was sequenced on the Illumina HiSeq 2500 sequencing platform. Sequencing was performed by Nanjing Jisi Huiyuan Biotechnology Co., Ltd. (Nanjing, China). After the original sequence was obtained, Fastp software (version 0.23.4) was used to remove the barcode sequence and filter out the N sequences with low masses and ratios greater than 5%, etc., to obtain clean reads that could be used for subsequent analysis (Chen et al., 2018). HISAT2 was used to align the clean reads with the wheat reference genome (https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Assemblies/v2.1/, version iwgsc_refseqv2.1) (Pertea et al., 2016). The number of transcripts per thousand bases per million mapped fragments (FPKM) was used for the characterization of expression. The read counts (raw counts) of the genes were calculated, and the p values and fold changes were calculated with DESeq2 software. A P value ≤ 0.05 and |log2fold change|>1 were used as the screening criteria for identifying DEGs (Liu et al., 2021). The DEGs were annotated based on the KEGG database (http://www.genome.jp/kegg/) (Kanehisa and Goto, 2000).

2.3 Metabolite extraction

One hundred milligrams of the sample was measured, and 800 μL of the extraction solution (methanol–acetonitrile–water volume ratio = 2:2:1, internal standard concentration = 20 mg/L) was added to the internal standard and added to each sample. Two small steel balls were added, and the samples were placed into a tissue grinder for grinding (at 50 Hz for 5 min; special samples that are difficult to break can be appropriately extended). After ultrasonication in a 4°C water bath for 10 min, the samples were allowed to rest at -20°C for 1 h. The samples were subsequently centrifuged at 4°C at 25,000 rpm for 15 min. After centrifugation, 600 μL of the supernatant was added to a 96-well plate. Using a 96-well filter plate for filtration, 200 μl of 70% methanol was first added to rinse the filter plate, and then 500 μl of supernatant was added for filtration. The filtered samples were collected, and each sample was transferred to a 96-well plate with 100 μL of supernatant. The plates were divided into positive and negative ions and were spared for a total of 3 plates (Jones and Kinghorn, 2012).

2.4 UPLC–MS analysis

In this experiment, a Waters UPLC I-Class Plus (Waters, USA) tandem Q Autonomous high-resolution mass spectrometer (Thermo Fisher Scientific, USA) was used for the separation and detection of metabolites. The column used was a Hypersil GOLD aQ Dim column (1.9 μm 2.1*100 mm, Thermo Fisher Scientific, USA). The mobile phases were 0.1% formic acid in water (liquid A) and acetonitrile (liquid B) containing 0.1% formic acid. The flow rate was 0.3 mL/min, the column temperature was 40°C, and the injection volume was 5 μL. A Q Autonomous mass spectrometer (Thermo Fisher Scientific, USA) was used for primary and secondary mass spectrometry data acquisition. The mass spectrometry scanning mass-core ratio range was 125~1500 positive ions, 100~1500 negative ions, 70,000 first-order resolution, 1e6 AGC, and 100 ms injection time (IT, injection time). The MS data were imported into Compound Discoverer 3.2 (Thermo Fisher Scientific, USA) software, combined with the mzCloud database and the ChemSpider online database for MS data analysis, and a data matrix containing the metabolite peak area and identification results was obtained (Jones and Kinghorn, 2012; Wen et al., 2017; Huang et al., 2019; Yu et al., 2021).

2.5 Metabolomic analysis

Based on the metabolite content data matrix, principal component analysis (PCA) was performed on each sample using R. The first principal component was first modelled and analysed by OPLS-DA, and the quality of the model was tested by 7-fold cross-validation. The validity of the model was judged by the R2Y (interpretability of the model to the categorical variable Y) and Q2 (predictability of the model) obtained by cross-validation. Finally, by permutation test, the order of the categorical variable Y was randomly adjusted several times to obtain different random Q2 values, and further tests of the effectiveness of the model were performed. Using the Human Metabolome Database (HMDB) and KEGG database, the classification of metabolites and the functional annotation of the pathway were carried out, and the main biochemical metabolic pathways and signal transduction pathways associated with the metabolites were determined (Kanehisa and Goto, 2000). Partial least squares regression was used to establish a model of the relationship between metabolite expression and sample class to model and predict sample class (Barker and Rayens, 2003; Westerhuis et al., 2008). The fold change in the expression of each metabolite in each comparison group was calculated. Student's t test was used to test the significance of the expression of each metabolite in each comparison group, and a fold change ≥1.2 or ≤0.83 and a q value < 0.05 were used as the standards for screening for differentially abundant metabolites (Dunn et al., 2011).

2.6 WGCNA

To ensure the distribution of scale-free networks, the weighting coefficient β should meet the correlation coefficient close to 0.8 and have a certain degree of gene connectivity. In this study, β=7 was selected as the weighting coefficient. The automatic network building function of blockwise modules was used to construct the network, and multiple valid modules were obtained. The number of genes contained in each module was different. MinModuleSize = 30

and Merge Cut Height = 0.25 were used as the standards, and modules with a combined similarity of 0.75 were obtained. The correlation coefficients between the module's characteristic vector ME (module eigengene) and different durations of hormone content and treatment were calculated. R>0.80 and P<0.05 were used as criteria for screening the specificity modules. Cytoscape (version 3.10.0) software was used for visualization of coexpression networks (Shannon et al., 2003).

2.7 qRT–PCR

Total RNA was extracted using an EZNA. Plant RNA Kit (Omega Bio-Tek, Doraville, GA, USA). The concentration of each RNA sample was determined using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), followed by the use of 1 μg of isolated RNA to obtain first-strand cDNA via a PrimeScript reverse transcription RT kit with gDNA^{TM} erasure (Takara Bio, Inc., Shiga, Japan). qRT–PCR analysis was performed using Roche LC480 equipment (Roche Diagnostics GmbH, Mannheim, Germany) and SYBR Green (Takara Bio, Inc.). Using a two-step PCR amplification procedure, predenaturation was carried out at 95°C for 30 sec, followed by 40 cycles of denaturation at 95°C for 5 sec and annealing at 60°C for 34 sec. The relative expression levels of the target genes were calculated using geNorm software, with the reference gene Actin and three biological replicates for each gene. All primers used in this study are shown in Supplementary Table S1.

3 Results

3.1 Determination of Xinchun 26 and Xinchun 34 protein content

The content and quality of gluten are important indicators for determining the quality of wheat flour and determine the process performance of the dough and the quality of steamed and baked goods (Zheng et al., 2018; Hackenberg et al., 2019). To do this, we first determined the levels of four proteins (albumin, globulin, gliadin and glutenin) in Xinchun 26 and Xinchun 34 seeds 7 d, 14 d and 21 d after flowering (Figure 1). Compared with those at 7 d after flowering, the expression of the four proteins at 14 d and 21 d increased significantly in both materials. The serum ALB concentration significantly differed among the three treatment groups, and the globulin concentration significantly differed between the two treatment groups at 21 d. Gliadin and glutamine levels were significantly greater in Xinchun 26 than in Xinchun 34 at 7 d, 14 d and 21 d. These results showed that the gluten content in Xinchun 26 was significantly greater than that in Xinchun 34. To further explore key genes and key metabolites related to gluten content during wheat grain formation, RNA-seq and metabolome sequencing were performed on Xinchun 26 and Xinchun 34 grains at 7 d, 14 d and 21 d after flowering.

3.2 RNA-seq analysis

A total of 18 RNA-seq samples from 2 materials and 3 periods produced a total of 130.37 Gb of data, and the amount of clean data from each sample reached 5.45 Gb or more. The Q30 percentage was 89.52%, the percentage of sequences that were shared with the reference genome was between 86.08% and 94.31%, and the average alignment rate was 89.62% (Supplementary Table S2). The correlation between the same biological replicates was good, the correlation coefficient range was 0.84~1.00, the PCA and correlation analysis results were consistent, and the replicates were clustered together (Supplementary Figure S1). Ten genes were randomly selected for 3 independent replicates of qRT–PCR analysis, and the transcriptome data were significantly correlated with the qRT–PCR data (R^2 = 0.9207; Supplementary Figure S2). The results showed that the test sampling was reasonable and that the RNA-seq data quality was reliable.

3.3 RNA-seq differential analysis

Differential analysis was performed over 3 developmental periods for Xinchun 26, which revealed 32,315 DEGs between 7 d and 14 d, 34,399 DEGs between 7 d and 21 d, and 10,313 DEGs between 14 d and 21 d, for a total of 2,697 DEGs over three periods (Figure 2A). In Xinchun 34, there were 10,077 DEGs between 7 d and 14 d, 22,123 DEGs between 7 d and 21 d, and 13,933 DEGs between 14 d and 21 d, for a total of 1,872 DEGs (Figure 2B) over the three periods. Among the two materials, 7 d had 33036 DEGs, 14 d had 15249 DEGs, 21 d had 14925 DEGs, and 5573 DEGs were detected over the three periods (Figure 2C). A total of 5573 DEGs in the three periods were divided into two categories according to their expression patterns; the expression of Cluster 1 in Xinchun 26 was greater than that in Xinchun 34, and the expression gradually decreased with the development of grains (Figures 2D, E). Similarly, the expression of Cluster2 in Xinchun 34 was greater than that in Xinchun 26, but the expression did not change with grain development (Figures 2D, E).

3.4 Metabolomic analysis

UPLC–MS identified a total of 863 metabolites, and PCA revealed that the first principal component could explain 71.23% of the total variance, the second principal component could explain 8.51% of the total variance, and the first principal component could distinguish different materials and periods (Figure 3A). To understand the classification and functional characteristics of the different metabolites, we classified and annotated the identified metabolites, which were divided into 12 main categories. The contents of lipids and lipid-like molecules accounted for 22.72%, the contents of phenylpropanoids and polyketides accounted for 18.25%, the contents of organoheterocyclic compounds accounted for 16.39%, and the contents of organic acids and derivatives

10.3389/fpls.2023.1309678

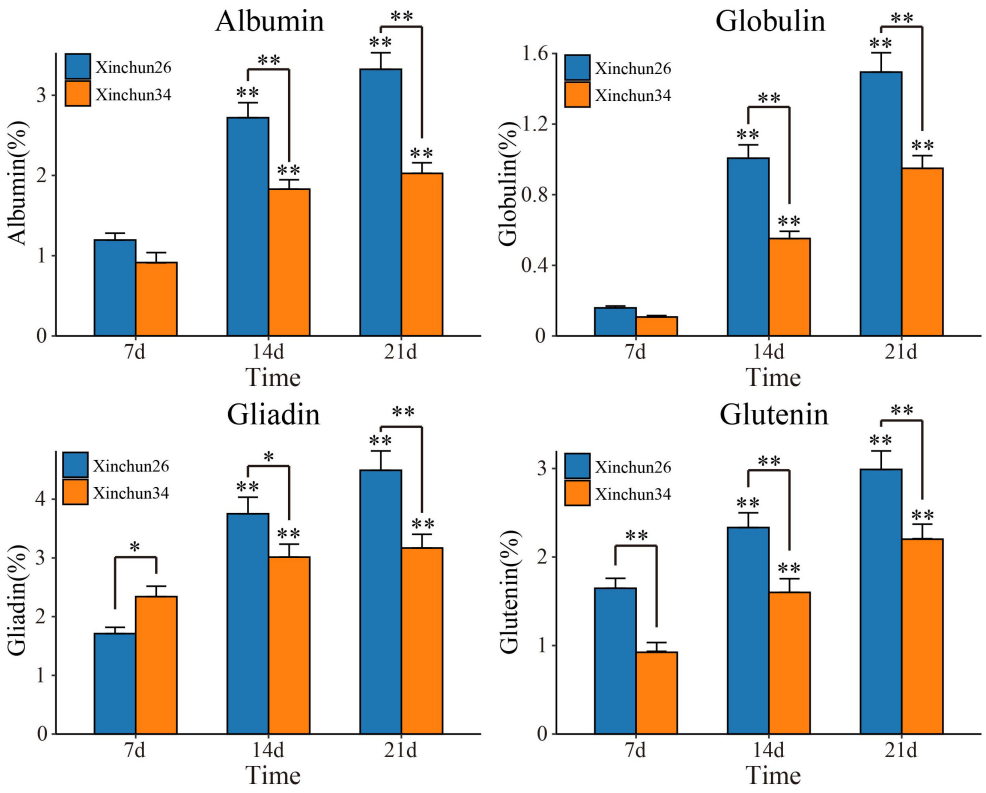


FIGURE 1
Albumin, globulin, gliadin and glutenin contents of Xinchun 26 and Xinchun 34 seeds 7 d, 14 d and 21 d after flowering. The results are presented as the means ± SDs (n = 3, **P < 0.01).

accounted for 14.89%. Benzenoids accounted for 11.55%; organic oxygen compounds, 6.15%; nucleosides, nucleotides, and analogues, 4.66%; alkaloids and derivatives, 1.86%; and organic nitrogen compounds, 1.31%. The percentages of organooxygen compounds, lignans, neolignans, related compounds and organic compounds were 0.74% (Figure 3B).

3.5 Metabolomic difference analysis

Differences were observed over 3 developmental periods in Xinchun 26, with 147 DRMs occurring between the 7 d mark and 14 d mark, 205 DRMs occurring between the 7 d mark and 21 d mark, 138 DRMs occurring between the 14 d mark and 21 d mark, and a total of 35 DRMs occurring over the three periods (Figure 4A). In Xinchun 34, there were 166 DRMs between 7 d and 14 d, 115 DRMs between 7 d and 21 d, 48 DRMs between 14 d and 21 d, and 17 DRMs in the three periods (Figure 4B). Among the two materials, 7 d had 137 DRMs, 14 d had 138 DRMs, 21 d had 34 DRMs, and 10 DRMs occurred during the three periods (Figure 4C). The four DRMs (3-methyl-L-histidine, L-arginine, L-citrulline, L-citrulline and L-asparagine) had the highest 7d content in Xinchun26, which decreased with development (Figure 4D). The six DRMs (L-alanine, beta-alanine, 4-aminobutyric acid, xanthosine, N1-methyl-2-pyridone-5-carboxamide and N1-methyl-4-pyridone-3-carboxamide) had the highest 14d content in Xinchun 26 (Figure 4D).

3.6 RNA-seq and metabolome combined analysis

KEGG enrichment analysis was performed on the DRMs and DEGs, and the DEGs were enriched mainly in starch and sucrose metabolism; photosynthesis-antenna proteins; glycolysis/gluconeogenesis; carbon fixation in photosynthetic organisms; pyruvate metabolism; fructose and mannose metabolism; the pentose phosphate pathway; alanine, aspartate and glutamate metabolism; glyoxylate and dicarboxylate metabolism; fatty acid degradation; and the citrate cycle (Figure 5A). The DRMs were mainly enriched in aminobenzoate degradation; starch and sucrose metabolism; ABC transporters; protein digestion and absorption; biosynthesis of amino acids; biosynthesis of various secondary metabolites; the citrate cycle; carbon fixation in photosynthetic organisms; alanine, aspartate and glutamate metabolism; mineral absorption; and flavone and flavonol biosynthesis (Figure 5B). The common enrichment pathways for DRMs and DEGs were involved in starch and sucrose metabolism; the citrate cycle; carbon fixation in photosynthetic organisms; and alanine, aspartate and glutamate metabolism (Figures 5A, B). The sucrose synthesis pathway genes

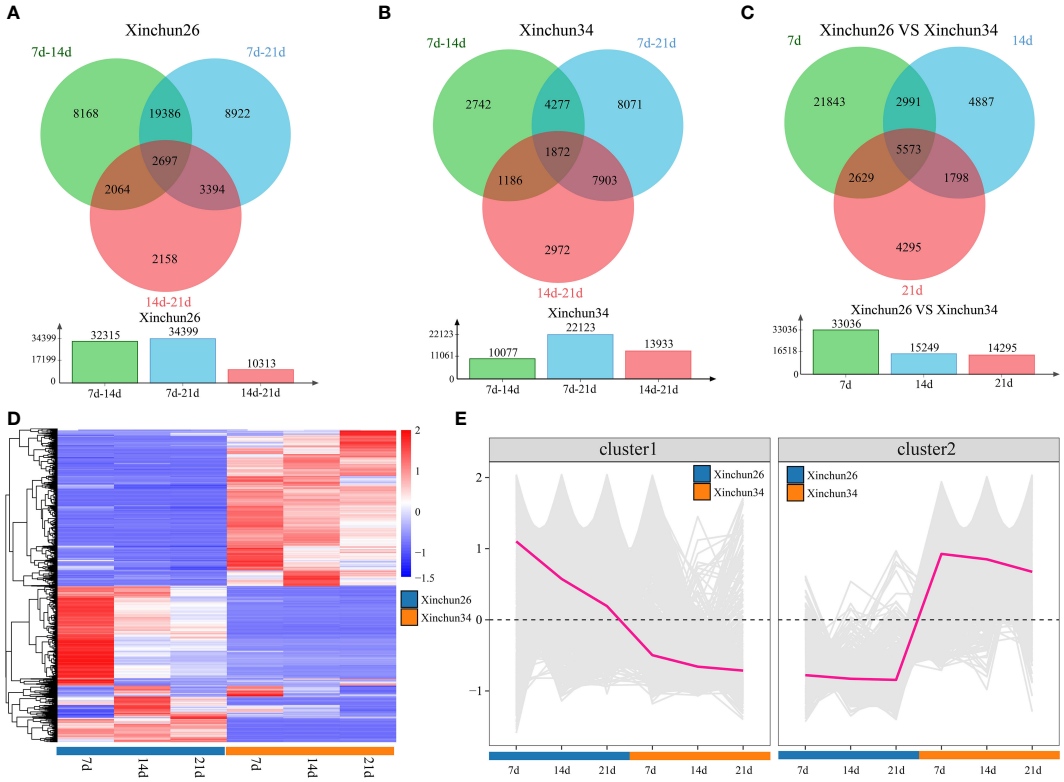


FIGURE 2
(A) There were Xinchun (26) DEGs at 7 d, 14 d and 21 d after flowering, **(B)** There were Xinchun (34) DEGs and corresponding quantities at 7 d, 14 d and 21 d after flowering, **(C)** The Venn diagrams and quantities of DEGs in Xinchun (26) and Xinchun (34) flowers after 7, 14 and 21 d, **(D)** shows an expression heatmap of DEGs between 7 d, 14 d and 21 d after Xinchun 26 and Xinchun 34 treatment, **(E)** Line chart of the expression trends of DEGs among the 7 d, 14 d and 21 d periods after Xinchun 26 and Xinchun 34 flowered.

and sucrose content were analysed, and the sucrose synthesis pathway genes were sucrose phosphate synthase (SPS), sucrose-phosphate phosphatase (SPP), sucrose synthase (SUS) and 1,4-alpha-glucan branching enzyme (GEB) (Figure 5C, Supplementary Table S3). The sucrose content in Xinchun 34 was significantly greater than that in Xinchun 26, and the sucrose content in both materials at 14 d and 21 d was significantly greater than that in the seeds of rose flowers at 7 d (Figure 5D).

3.7 Gluten-related gene expression analysis

Among the wheat storage proteins, gluten plays a key role in the processing quality of wheat. For this purpose, glutenin-related genes were analysed among the DEGs, and a total of 59 genes were identified. The expression patterns of glutenin-related genes were visualized using a heatmap, which was divided into 7 main clusters (Figure 6A, Supplementary Table S3). Cluster 1 included 4 genes; the

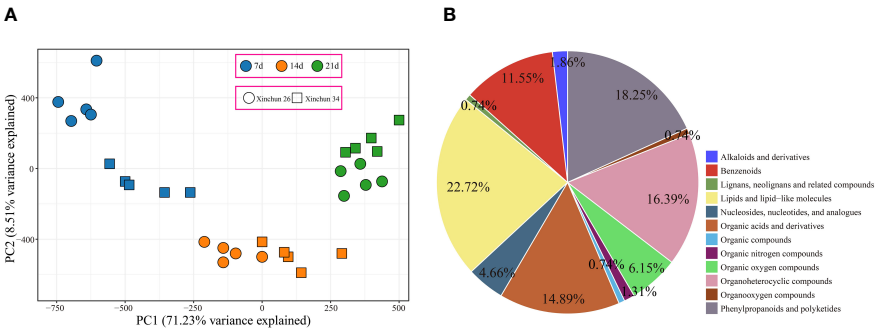


FIGURE 3
(A) Metabolome PCA. **(B)** Metabolome classification pie chart.

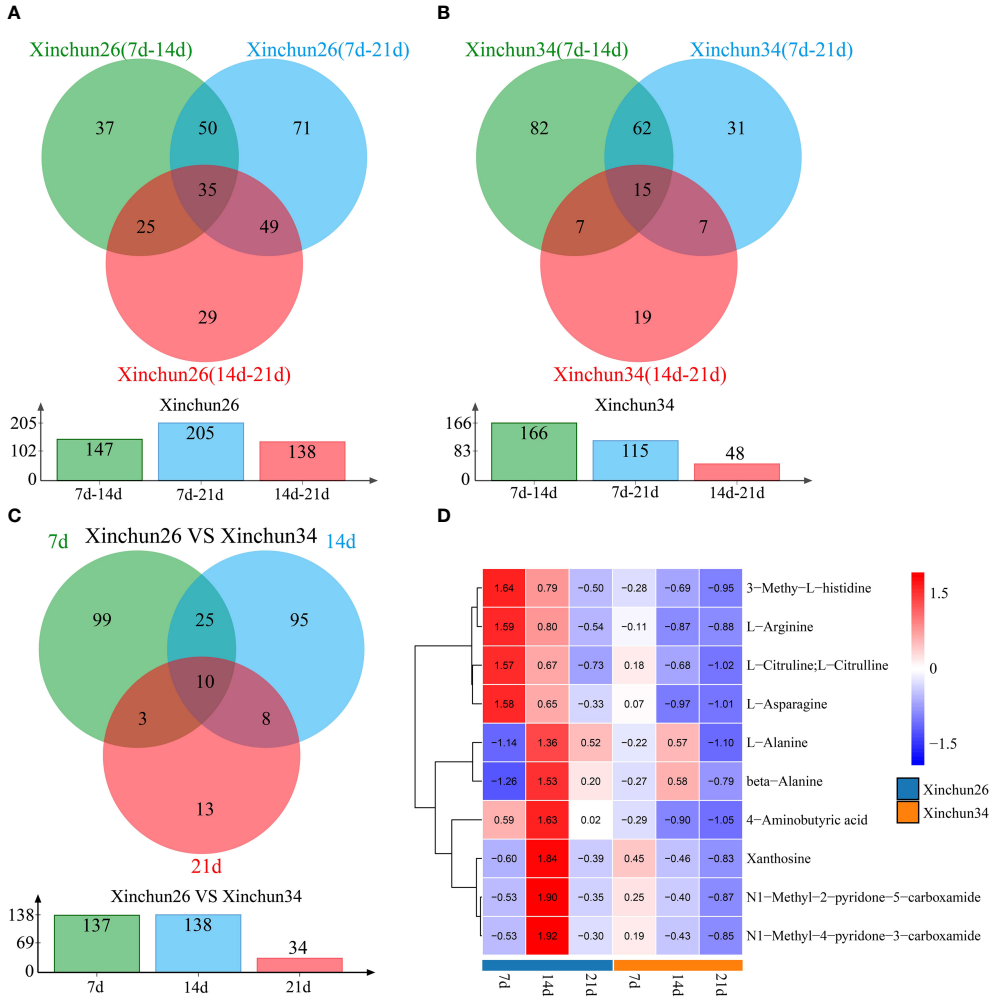


FIGURE 4
(A) Xinchun 26, 7, 14 and 21 d after flowering, a Venn diagram and quantity of DRMs were generated, **(B)** Xinchun 34, 7, 14 and 21 d after flowering, a Venn diagram and number of DRMs were generated, **(C)** The Venn diagrams and quantities of the Xinchun 26 and Xinchun 34 flowers after 7, 14 and 21 d of treatment, **(D)**. Heatmap of DRM levels 7, 14 and 21 d after Xinchun 26 and Xinchun 34 flowering.

expression of 4 genes in Xinchun 34 was greater than that in Xinchun 26, and the expression level gradually increased with the development of grains. Cluster 2 included 8 genes, the expression level of Cluster 2 gradually increased with the development of grains in Xinchun 26, and the expression trend in Xinchun 34 remained basically unchanged. Cluster 3 included 11 genes whose expression levels increased with grain development, and the expression patterns of the two materials were basically the same. Cluster 4 included 6 genes, and with increasing grain development, the expression patterns of the two materials were basically the same, while the expression levels at 14 d and 21 d were basically unchanged. Cluster 5 included 14 genes, the expression level of Cluster 5 gradually decreased with the development of grains, and the expression level in Xinchun 26 was greater than that in Xinchun 34. Cluster 6 included 5 genes, whose expression gradually decreased with grain development, and the expression level in Xinchun 34 was greater than that in Xinchun 26. Cluster 7 included 12 genes, and the expression levels of Cluster 7 gradually decreased with the development of grains on Lunar New Year 34; the expression levels were basically the same at 7 d and 14 d in New Spring 26, and the lowest expression was observed at 21 d. To further explore the relationship between these genes and gluten content, we calculated the correlation between gene expression and wheat glutenin and screened for absolute correlation coefficients greater than 0.5 for visualization (Figure 6B, Supplementary Table S3). A total of 25 genes were positively correlated with glutenin, with correlation coefficients ranging from 0.53 to 0.95, and 20 genes were negatively correlated with glutenin content, with correlation coefficients ranging from 0.50 to 0.95.

3.8 TF expression analysis

We analysed all the DEGs, which included B3 (4.43%), C2H2 (4.43%), AP2/ERF (6.33%), HSF (4.43%), NAC (6.96%), MYB (10.13%) and FAR1 (20.25%) (Figure 7A). The expression patterns of the differentially expressed TF genes are shown using a heatmap (Figures 7B-D). B3 exhibited completely opposite modes of expression in both materials (Figure 7B). C2H2 expression was downregulated in

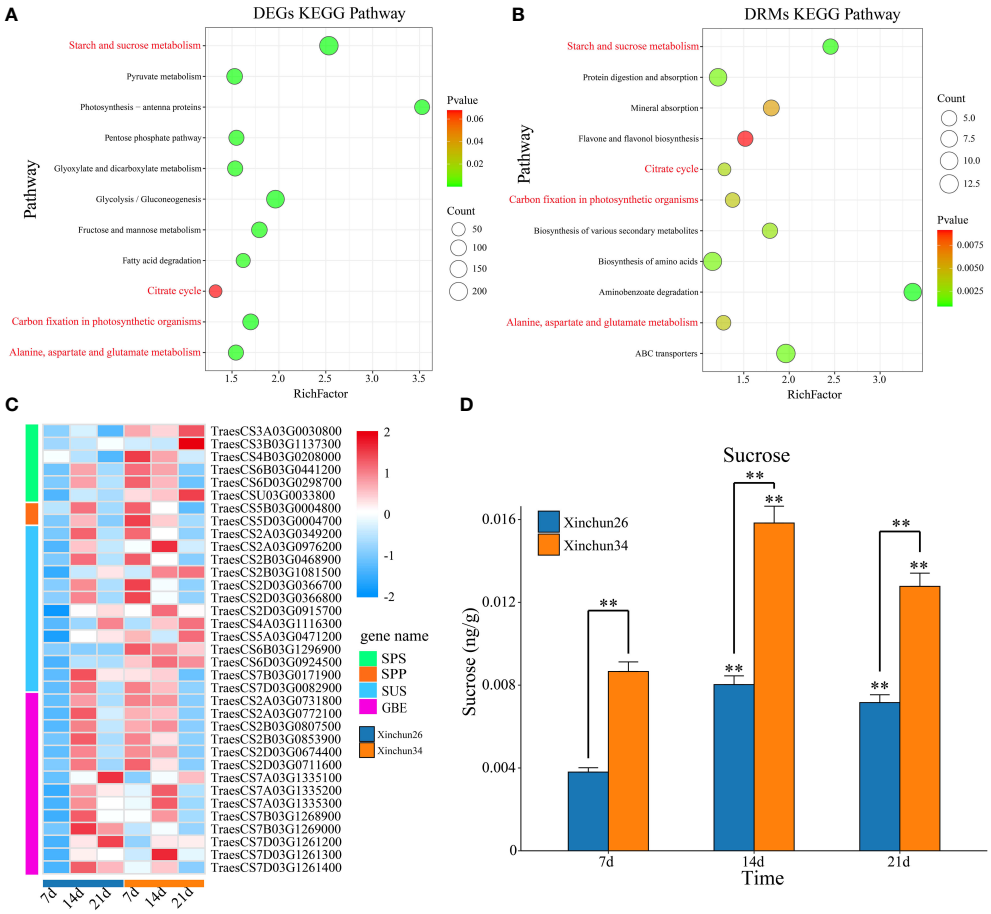


FIGURE 5
(A) Bubble map of DEG KEGG enrichment between materials. **(B)** Bubble map of the DRM KEGG enrichment between materials. **(C)** Gene expression calorimetry of the sucrose synthesis pathway, **(D)** The results are presented as the means ± SDs (n = 3, **P < 0.01).

Xinchun 26 and upregulated or unchanged in Xinchun 34 (Figure 6B, Supplementary Table S3). The expression of most of the AP2/ERF genes gradually decreased with development (Figure 6B, Supplementary Table S3). The expression of all the genes in the HSF family except TraesCS4B03G0978300 was downregulated (Figure 6C, Supplementary Table S3). The expression pattern of NAC is complex, with expression downregulated in Xinchun 26, upregulated or unchanged in Xinchun 34, and upregulated in 26 (Figure 6C). MYB was expressed mainly in Yanghua plants after 14 and 21 d (Figure 6C). FAR1 was expressed mainly on Xinchun 34 (Figure 6C).

3.9 WGCNA

Based on the FPKM values of the genes, according to the soft threshold calculation results, β=7 was selected for network construction; a total of 12 coexpression modules were identified by combining and expressing similar modules via the dynamic shearing tree method, and each module is represented by a different colour (Figure 7A). Five of the 12 modules were strongly correlated with 7 d, 14 d and 21 d (Figure 7B), and four candidate genes were identified (*TraesCS7B03G1102000* (ATPase), *TraesCS1A03G0797600* (SpoU), *TraesCS2B03G0927000* (G6PD4) and *TraesCS4D03G0099800* (ADA1E)) (Figure 7C). Overall, 2 candidate genes were identified (*TraesCS5B03G1060800* (PUP4) and *TraesCS1B03G1120700* (PHO2)) (Figure 7D). Brown identified four candidate genes (*TraesCS1B03G0703000* (ERF), *TraesCS3D03G0849200* (PICALM4A), *TraesCS5B03G0681600* (O-glycosyl hydrolases) and *TraesCS3B03G0727200* (xanthine/uracil permease)) (Figure 7E). Four candidate genes were identified (*TraesCS6D03G0025600* (SWEET12), *TraesCS7B03G0941900* (DIR1-like), *TraesCS1A03G0577300* (CHX) and *TraesCS1A03G1007800* (DALL)) (Figure 7F). Four candidate genes were identified (*TraesCS1A03G0201700* (PIF3), *TraesCS2A03G1077700* (PPR), *TraesCS5B03G0905200* (COG) and *TraesCS3D03G0349800* (GRF)) (Figure 7G).

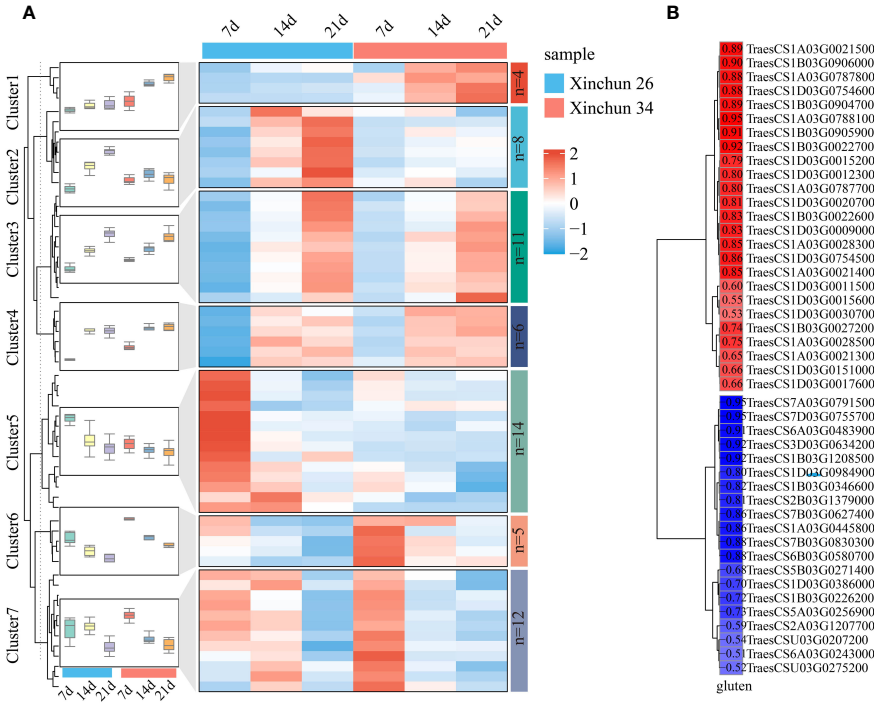


FIGURE 6
(A) A heatmap of the expression patterns of gluten-related genes in wheat was constructed, and the box plot represents the overall expression trend of genes in each cluster. **(B)** For the correlation between gluten-related gene expression and gluten content, red represents a positive correlation, blue represents a negative correlation, and colour depth represents the size of the image relationship number.

3.10 qRT–PCR of candidate genes

Based on the 18 candidate genes associated with the gluten content of wheat grains mined by WGCNA, we performed qRT–PCR on Xinchun 26 and Xinchun 34, and 8 genes (*TraesCS7B03G1102000* (ATPase), *TraesCS1A03G0797600* (SpoU), *TraesCS2B03G0927000* (G6PD4), *TraesCS5B03G1060800* (PUP4), *TraesCS1B03G0703000* (ERF), *TraesCS6D03G0025600* (SWEET12), *TraesCS1A03G0201700* (PIF3) and *TraesCS1B03G1120700* (PHO2)) were significantly differentially expressed between the two materials (Figure 8, Supplementary Figure S3). The expression of *TraesCS7B03G1102000* (ATPase) and *TraesCS1B03G0703000* (ERF) decreased in both materials, and the expression in Xinchun 26 was significantly greater than that in Xinchun 34. Similarly, the expression of *TraesCS1A03G0797600* (SpoU) decreased in Xinchun 26 but did not change significantly in Xinchun 34. In Xinchun 26, TraesCS2B03G0927000 (G6PD4) expression decreased on 14 d and was 3 fold change than on 7 d. In Xinchun 34, TraesCS2B03G0927000 expression increased on 14d and was 4 fold change than that on 7d. In Xinchun 34, TraesCS5B03G1060800 (PUP4) expression increased on 21d and was 1.5 fold change than that on 7d. In Xinchun 26, TraesCS5B03G1060800 expression decreased on 14d and was 2 fold change than that on 7d. In both materials, *TraesCS6D03G0025600* (SWEET12), *TraesCS1A03G0201700* (PIF3) and *TraesCS1B03G1120700* (PHO2) increased significantly (Figure 8).

4 Discussion

Wheat is an important food crop that provides energy and a variety of nutrients, such as protein and dietary fibre, for humans (Chawade et al., 2018). In China, wheat is the main food for northerners, and more than 85% of wheat is used to make bread, biscuits, noodles and other flour products (Chawade et al., 2018; Zheng et al., 2018; Hackenberg et al., 2019)[1-3]. High-gluten wheat is suitable for making bread, medium-gluten wheat is suitable for making steamed buns and noodles, and low-gluten wheat is suitable for making biscuits (Chawade et al., 2018)[1]. As people pursue a higher quality of life, the demand for better quality specialty wheat is continually increasing (Zheng et al., 2018; Hackenberg et al., 2019). Therefore, an in-depth study of the quality formation mechanism of different types of wheat is highly important for the selection and breeding of wheat cultivars with high gluten levels. Wheat grain protein can be divided into nongluten protein (approximately 15%~20%) and gluten protein (approximately 80%~85%) (Zheng et al., 2018). The solubility of different reagents can be divided into four categories (i.e., albumin, globulin, gliadin and gluten) (Delcour et al., 2012; Zheng et al., 2018). Among these proteins, albumin and globulin are metabolic proteins that play a role in plant growth and seed development; moreover, both are structural proteins, the former giving rise to dough ductility and stickiness, and the latter giving rise to dough elasticity and strength (Delcour et al., 2012). Compared with those

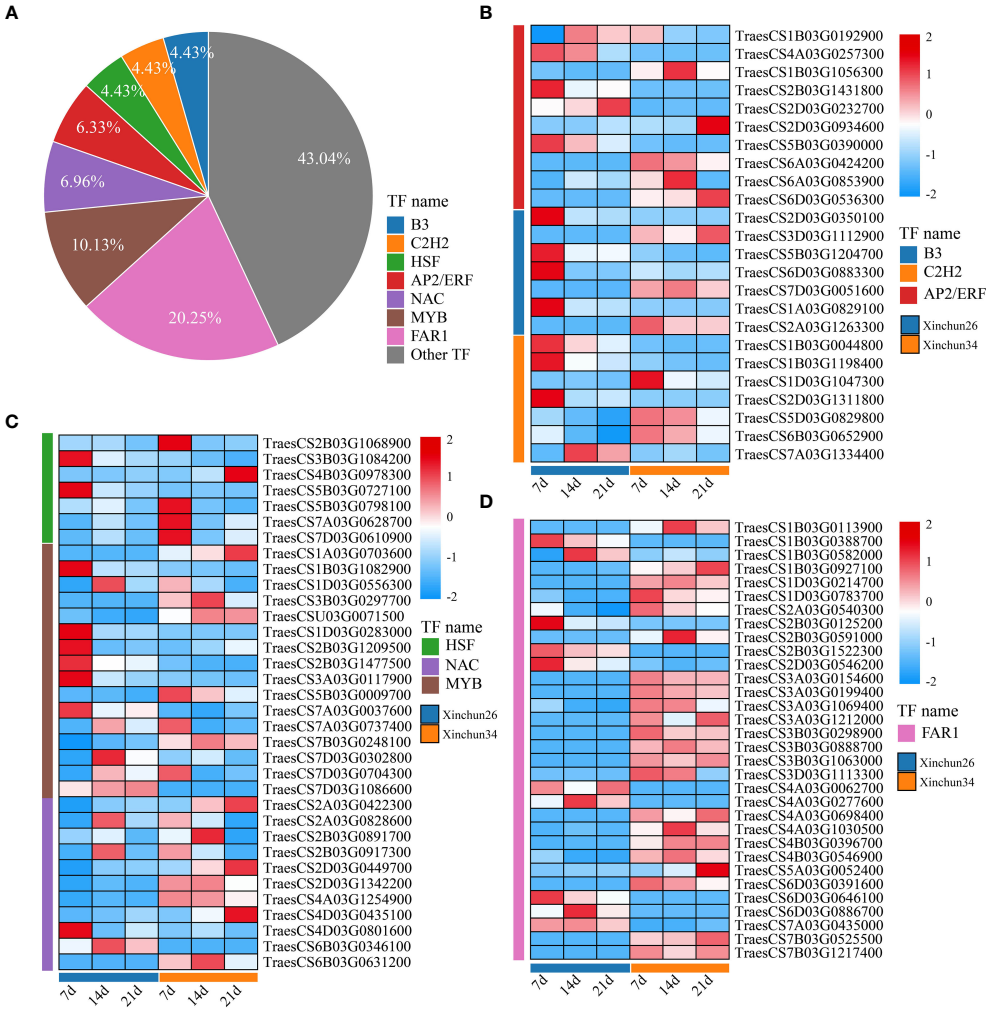


FIGURE 7
(A) The Xinchun 26 and Xinchun 34 difference TF pie charts are shown in Figure 5B. Figure 5C shows the expression heatmaps of B3, C2H2 and AP2/ERF between Xinchun 26 and Xinchun 34. Heatmap of HSF, NAC and MYB expression between Xinchun 26 and Xinchun 34, Figure 5D. FAR1 expression calorimetry between Xinchun 26 and Xinchun 34.

in the 7-d seed period after flowering, the expression of the four proteins at 14 d and 21 d increased significantly in both materials. The serum ALB concentration significantly differed among the three treatment groups, and the globulin concentration significantly differed between the two treatment groups at 21 d. The levels of gliadin and gluten were significantly greater in Xinchun 26 than in Xinchun 34 (Figure 1). These results showed that the gluten content of Xinchun 26 was significantly greater than that of Xinchun 34, and the contents of these four proteins also increased significantly with grain development.

To investigate the potential link between DEGs and metabolites during grain formation and gluten protein components in wheat, transcriptome and metabolome analyses were performed. PCA of the transcriptome and metabolome revealed that the differences between the periods were greater than the differences between materials, and the first principal component could distinguish between different materials and periods (Supplementary Figure S1, Figure 3A). There were 2697 DEGs in Xinchun 26 for the three periods, 1872 DEGs in Xinchun 34 for the three periods, and 5573 DEGs in the two materials for the three periods (Figure 2). According to the metabolomic differential analysis, 4 DRMs (3-methyl-L-histidine, L-arginine, L-citrulline, L-citrulline and L-asparagine) had the highest 7d content in Xinchun 26, which decreased with development (Figure 4D). The six DRMs (L-alanine, beta-alanine, 4-aminobutyric acid, xanthosine, N1-methyl-2-pyridone-5-carboxamide and N1-methyl-4-pyridone-3-carboxamide) had the highest 14d content in Xinchun 26 (Figure 4D). These genes and metabolites can provide a reference for the elucidation of gene and molecular mechanisms related to the content of gluten in wheat.

The accumulation of protein in wheat grains mainly depends on nitrogen metabolism and amino acid synthesis (Gao et al., 2021). In wheat, gluten is the main storage protein that determines gluten content (Yu et al., 2018). Gluten is first synthesized by the mRNA of the gluten gene family in the rough endoplasmic reticulum to form a 57 kDa precursor and then transported to the protein reservoir vacuole by Golgi modification, and the hydrolytic endonuclease in the vacuole results in the formation of an acidic subunit of 37-39 kDa and a basic subunit of 22-23 kDa (Yu et al., 2018; Gao et al., 2021). To this end, we identified 59 differentially expressed gluten synthesis-related genes, including 24

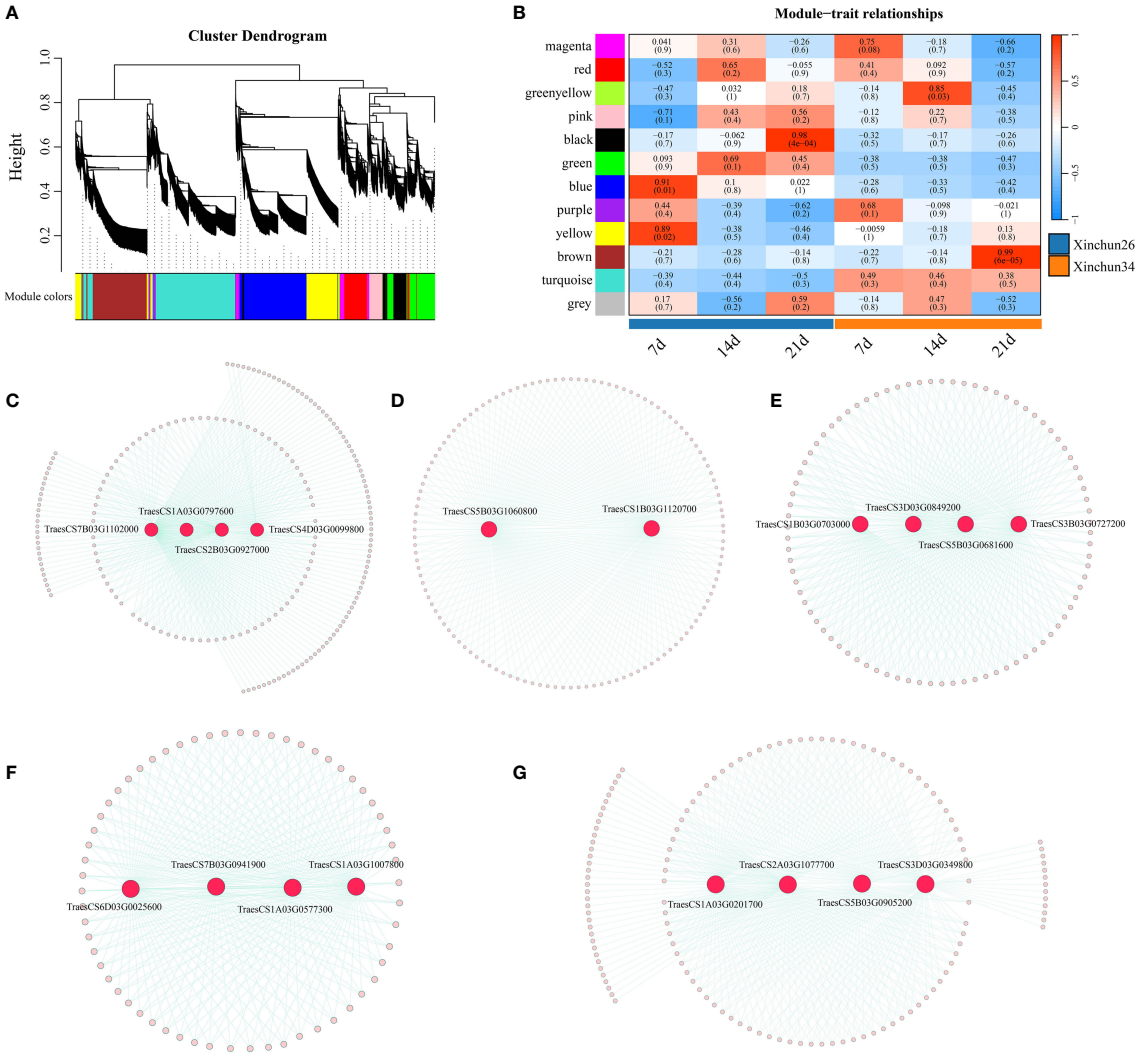


FIGURE 8
(A) A hierarchical clustering tree of genes was constructed based on the coexpression network analysis, **(B)** A heatmap of the correlations and significance between different developmental time points and materials, **(C)** The gene coexpression networks within the black modules, **(D)** The gene coexpression networks are shown within blue modules, **(E)** Gene coexpression networks within brown modules. **(F)** The gene coexpression networks within the green–yellow modules, **(G)** Gene coexpression networks within the green–yellow module.

glutamine synthetases (GS), in wheat. Correlation analysis revealed that a total of 25 genes were positively correlated with glutenin, with correlation coefficients ranging from 0.53 to 0.95, and that 20 genes were negatively correlated with glutenin content, with correlation coefficients ranging from 0.50 to 0.95. GS is closely related to the high grain protein content and nitrogen efficiency of wheat, and its activity significantly affects protein and amino acid contents (Yin et al., 2022). The nitrogen absorbed by wheat is catalysed by GS to produce glutamine (Gln) by the ATP-dependent condensation reaction of ammonium with glutamic acid (Glu), which then provides the N group directly through Glu for the biosynthesis of proteins, amino acids, and other nitrogen-containing compounds (Nigro et al., 2016). We found that these DEGs related to gluten synthesis, especially GS, can be used as important candidate genes for improving wheat quality in the future.

Sugars are not only the energy source of plants but also important structural material components. Many kinds of sugars can also bind proteins to complex compounds (such as glycoproteins) and participate in cell recognition, intercellular material transport and other life activities, regulating plant growth and development (Kanwar and Jha, 2019; Nägele et al., 2022). The common enrichment pathways for DRMs and DEGs were starch and sucrose metabolism; the citrate cycle; carbon fixation in photosynthetic organisms; and alanine, aspartate and glutamate metabolism (Figures 5A, B). The sucrose content in Xinchun 34 was significantly greater than that in Xinchun 26, and the sucrose content in both materials at 14 d and 21 d was significantly greater than that in the seeds of Yanghua 7 d (Figure 5D). This means that the higher the sugar content of the wheat kernel is, the lower the gluten content. The sugar content of wheat not only reduces the yield of gluten but also may affect the composition and quality of gluten. Changes in light and dark times can significantly affect the synthesis of plant sugars because changes in light time can cause changes in the efficiency of plant cells to use light energy, which in turn affects cell division (Julius et al., 2017; Yoon et al., 2022). In addition, the photocycle can not only change the synthesis ability of photosynthetic pigments in plant cells but also affect

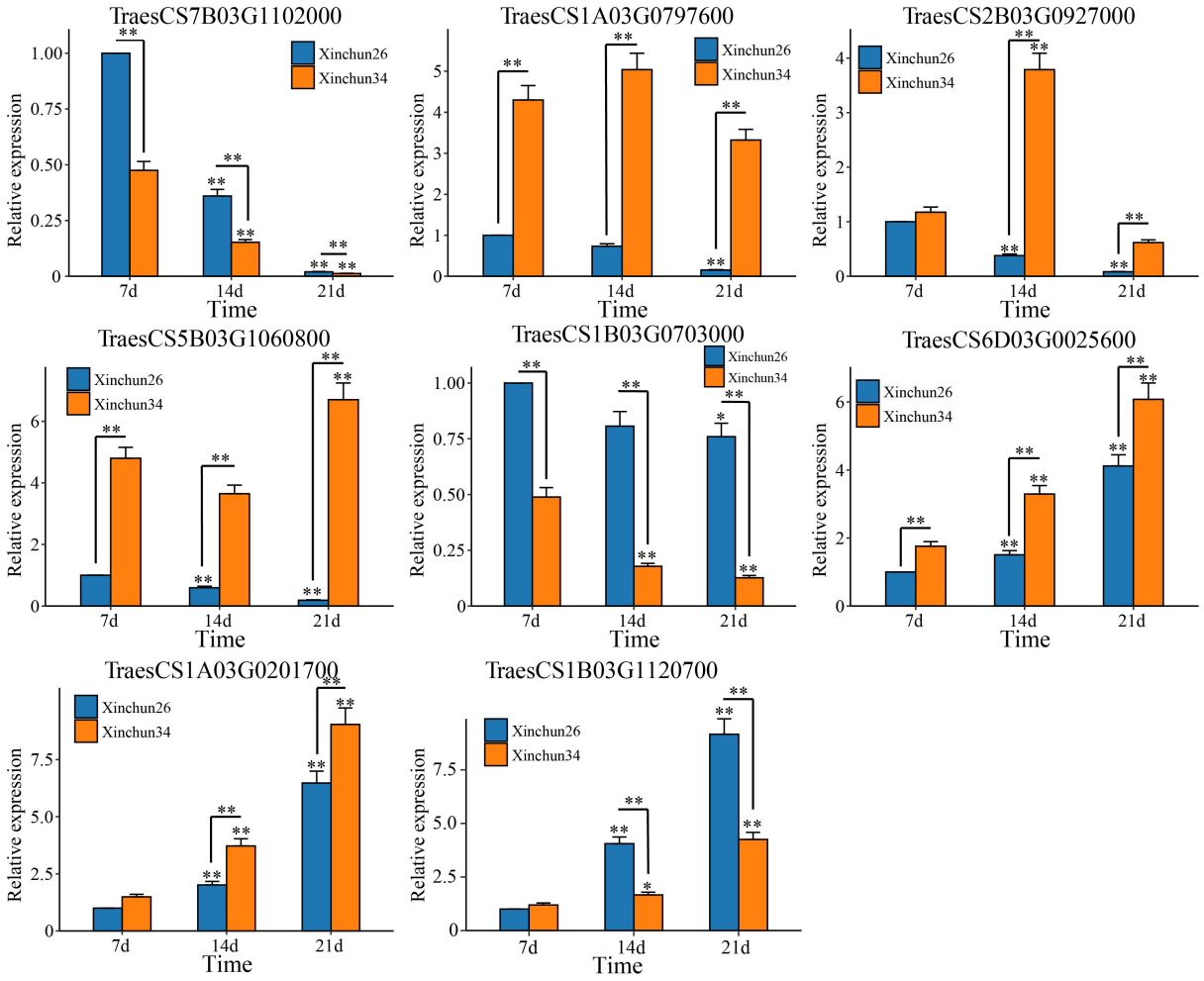


FIGURE 9
qRT–PCR of the wheat grain gluten content hub genes. The results are presented as the means ± SDs (n = 3, **P < 0.01, *P < 0.05).

the absorption and utilization of nutrients and other substances by cells and ultimately affect the synthesis and metabolism of the main active substances in cells (Ahmad, 2016; Hart et al., 2019). Our joint analysis also revealed that carbon fixation in photosynthetic organisms is an important pathway affecting the gluten content of wheat grains. The citrate cycle is a common metabolic pathway for the complete oxidation of the three main types of organic matter in the body: sugar, fat and protein (Fernie et al., 2004). The citrate cycle is a catabolic pathway that provides precursor molecules for the biosynthesis of several substances (Zhang and Fernie, 2018). For example, sugar and glycerol are metabolized in the body to produce α-ketoglutaric acid, oxaloacetic acid and other intermediate products of the tricarboxylic acid cycle; oxaloacetic acid is a precursor for the synthesis of aspartic acid; and α-ketoglutaric acid is a precursor for the synthesis of glutamic acid (Zhang and Fernie, 2023). Some amino acids can also produce sugars through different pathways through gluconeogenesis (Zhang and Fernie, 2023). These findings suggested that changes in the sugar and gluten contents of wheat grains may have been transformed by this process.

The method of combining transcriptome data with the WGCNA algorithm to study core genes related to plant growth and development and resistance has been widely used in the study of morphological formation and developmental regulatory mechanisms of plants, flowers, leaves and fruits, as well as the prediction of unknown gene functions (Dai et al., 2021; Wang et al., 2022; Yao et al., 2023). In this study, RNA-seq data from 18 Xinchun 26 and Xinchun 34 samples were used to screen 8 candidate genes related to the gluten content of wheat grains by combining wheat grains with the WGCNA algorithm combined with qRT–PCR. Among them, the expression of *TraesCS2B03G0927000* and *TraesCS5B03G1060800* in the low-gluten-content Xinchun 34 variety increased with grain development, while the expression in the high-gluten-content Xinchun 26 gradually decreased. These two genes may be the most important candidates for wheat gluten content. Based on the annotation of homologous Arabidopsis genes, we found that *TraesCS2B03G0927000* encodes a BTB/POZ domain-containing protein that mediates nutrient uptake and growth and development in Arabidopsis (Mao et al., 2017). *TraesCS5B03G1060800* encodes a purine permease that is involved in ATP-dependent cytokinin translocations and controls the spatiotemporal pattern of cytokinin signalling (Qi and Xiong, 2013). The depletion of ligands in the ectoplast leads to the inhibition of the cytokinin response (Qi and Xiong, 2013). Cytokinins can stimulate plant growth and flowering, control growth and differentiation, and delay ageing and are related to increased yield

(Gupta and Rashotte, 2012). These two genes may regulate wheat grain gluten content through different mechanisms; however, the biological function of these genes (specifically, *TraesCS2B03G0927000* and *TraesCS5B03G1060800*) in regulating wheat grain gluten content requires further study. In conclusion, these findings provide new insights and ideas for the study of wheat grain gluten content and lay a foundation for in-depth analysis of the molecular mechanism of wheat grain gluten content.

5 Conclusion

In this study, the RNA-seq and metabolome of Xinchun 26 plants with high gluten content and low gluten content Xinchun 34 were analysed at 7, 14 and 21 d after flowering. Transcriptomic analysis revealed 5573 DEGs between the materials in the three periods, which were divided into two categories according to their expression patterns. Metabolomic analysis revealed that lipids, lipid-like molecules, phenylpropanoids and polyketides were the two most abundant metabolites, and the difference analysis revealed a total of 10 DRMs over the three periods. Combined RNA-seq and metabolome analysis revealed that starch and sucrose metabolism; the citrate cycle; carbon fixation in photosynthetic organisms; and alanine, aspartate and glutamate metabolism pathways were more important for determining the gluten content of wheat grains. By constructing a coexpression network, five specific modules that were strongly correlated with wheat grain development were identified, and eight candidate genes were screened via qRT–PCR. These findings provide new insights into the gluten content during wheat grain formation and reveal potential regulatory pathways and candidate genes involved in this developmental process.

Data availability statement

The generated raw reads have been uploaded to NCBI's SRA database and are available under the accession number PRJNA1021180. 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

JS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. LW: Formal analysis, Investigation, Methodology, Writing – review & editing. ZhoW: Data curation, Formal analysis, Methodology, Writing – review & editing. JL: Formal analysis, Methodology, Writing – review & editing. HZ: Formal analysis, Methodology, Writing – review & editing. XG: Formal analysis, Methodology, Writing – review & editing. CW: Formal analysis, Methodology, Writing – review & editing. JX: Formal analysis, Methodology, Writing – review & editing. ZZ: Formal analysis, Methodology, Writing – review & editing. ZheW: Formal analysis, Methodology, Writing – review & editing. ZY: Formal analysis, Methodology, Writing – review & editing. ZX: Formal analysis, Methodology, Writing – review & editing. YZ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – review & editing. ZF: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Xinjiang Academy of Agricultural Sciences Young Science and Technology Backbone Innovation Ability Training Project (xjnkq-2021005); National Natural Science Foundation International (regional) Cooperation and Exchange Project (32061143040); and Silk Road Economic Belt Innovation-driven Development Pilot Zone, Wuchangshi National Independent Innovation Demonstration Zone Science and Technology Development Plan (2022LQ03017).

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.2023.1309678/full#supplementary-material

SUPPLEMENTARY FIGURE 1
RNA-seq sample correlation and PCA.

SUPPLEMENTARY FIGURE 2
Scatter plot of the correlation between the gene expression levels from the transcriptome and qRT–PCR data.

SUPPLEMENTARY FIGURE 3
qRT–PCR analysis of the wheat grain gluten content hub genes; the results are presented as the means ± SDs (n = 3, **P < 0.01, *P < 0.05).

References

Ahmad, M. (2016). Photocycle and signaling mechanisms of plant cryptochromes. *Curr. Opin. Plant Biol.* 33, 108–115. doi: 10.1016/j.pbi.2016.06.013

Barker, M., and Rayens, W. (2003). Partial least squares for discrimination. *J. Chemometr* 17, 166–173. doi: 10.1002/cem.785

Chawade, A., Armoniené, R., Berg, G., Brazauskas, G., Frostgård, G., Geleta, M., et al. (2018). A transnational and holistic breeding approach is needed for sustainable wheat production in the Baltic Sea region. *Physiol. Plant* 164, 442–451. doi: 10.1111/ppl.12726

Chen, Q. F., Ya, H. Y., Wang, W. D., and Jiao, Z. (2014). RNA-seq reveals the downregulated proteins related to photosynthesis in growth-inhibited rice seedlings induced by low-energy N+ beam implantation. *Genet. Mol. Res.* 13, 7029–7036. doi: 10.4238/2014.March.26.9

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Dai, Y., Sun, X., Wang, C., Li, F., Zhang, S., Zhang, H., et al. (2021). Gene co-expression network analysis reveals key pathways and hub genes in Chinese cabbage (*Brassica rapa* L.) during vernalization. *BMC Genomics* 22, 236. doi: 10.1186/s12864-021-07510-8

Delcour, J. A., Joye, I. J., Pareyt, B., Wilderjans, E., Brijs, K., and Lagrain, B. (2012). Wheat gluten functionality as a quality determinant in cereal-based food products. *Annu. Rev. Food Sci. Technol.* 3, 469–492. doi: 10.1146/annurev-food-022811-101303

Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., et al. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* 6, 1060–1083. doi: 10.1038/nprot.2011.335

Fernie, A. R., Carrari, F., and Sweetlove, L. J. (2004). Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. *Curr. Opin. Plant Biol.* 7, 254–261. doi: 10.1016/j.pbi.2004.03.007

Gao, Y., An, K., Guo, W., Chen, Y., Zhang, R., Zhang, X., et al. (2021). The endosperm-specific transcription factor TaNAC019 regulates glutenin and starch accumulation and its elite allele improves wheat grain quality. *Plant Cell.* 33, 603–622. doi: 10.1093/plcell/koaa040

Gupta, S., and Rashotte, A. M. (2012). Down-stream components of cytokinin signaling and the role of cytokinin throughout the plant. *Plant Cell Rep.* 31, 801–812. doi: 10.1007/s00299-012-1233-0

Hackenberg, S., Vogel, C., Scherf, K. A., Jekle, M., and Becker, T. (2019). Impact of altered starch functionality on wheat dough microstructure and its elongation behaviour. *Food Chem.* 290, 64–71. doi: 10.1016/j.foodchem.2019.03.016

Hart, J. E., Sullivan, S., Hermanowicz, P., Petersen, J., Diaz-Ramos, L. A., Hoey, D. J., et al. (2019). Engineering the phototropin photocycle improves photoreceptor performance and plant biomass production. *Proc. Natl. Acad. Sci. U S A.* 116, 12550–12557. doi: 10.1073/pnas.1902915116

Hsu, S. K., and Tung, C. W. (2017). RNA-seq analysis of diverse rice genotypes to identify the genes controlling coleoptile growth during submerged germination. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00762

Huang, Z. W., Yang, Y. X., Huang, L. H., and Zhang, S. Q. (2019). Pharmacokinetics and metabolism of icaritin in rats by UPLC-MS/MS. *Food Sci. Nutr.* 7, 4001–4006. doi: 10.1002/fsn3.1263

Ji, X., Jin, B., Zhuang, Z., Chang, F., Wang, F., and Peng, Y. (2022). Study on *zmRPN10* regulating leaf angle in maize by RNA-seq. *Int. J. Mol. Sci.* 24, 189. doi: 10.3390/ijms24010189

Jiang, Z., Zhang, H., Jiao, P., Wei, X., Liu, S., Guan, S., et al. (2022). The integration of metabolomics and transcriptomics provides new insights for the identification of genes key to auxin synthesis at different growth stages of maize. *Int. J. Mol. Sci.* 23, 13195. doi: 10.3390/ijms232113195

Jones, W. P., and Kinghorn, A. D. (2012). Extraction of plant secondary metabolites. *Methods Mol. Biol.* 864, 341–366. doi: 10.1007/978-1-61779-624-1_13

Julius, B. T., Leach, K. A., Tran, T. M., Mertz, R. A., and Braun, D. M. (2017). Sugar transporters in plants: new insights and discoveries. *Plant Cell Physiol.* 58, 1442–1460. doi: 10.1093/pcp/pcx090

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Kanwar, P., and Jha, G. (2019). Alterations in plant sugar metabolism: signatory of pathogen attack. *Planta.* 249, 305–318. doi: 10.1007/s00425-018-3018-3

Kusser, W., Javorschi, S., and Gleeson, M. A. (2006). Real-time RT-PCR: cDNA synthesis. *Cold Spring Harbor Protoc* 2006. pdb-prot4114. doi: 10.1101/pdb.prot4114

Li, H., Li, Y., Song, L., Cheng, J., Ge, J., Yu, X., et al. (2022). Effects of tebuconazole application at different growth stages on rice grain quality of rice-based untargeted metabolomics. *Chemosphere.* 303, 134920. doi: 10.1016/j.chemosphere.2022.134920

Liu, S., Wang, Z., Zhu, R., Wang, F., Cheng, Y., and Liu, Y. (2021). Three differential expression analysis methods for RNA sequencing: limma, edgeR, DESeq2. *J. Vis. Exp.* 18, 175. doi: 10.3791/62528

Ma, L., Zhang, M., Chen, J., Qing, C., He, S., Zou, C., et al. (2021). GWAS and WGCNA uncover hub genes controlling salt tolerance in maize (*Zea mays* L.) seedlings. *Theor. Appl. Genet.* 134, 3305–3318. doi: 10.1007/s00122-021-03897-w

Mao, H., Aryal, B., Langenecker, T., Hagmann, J., Geisler, M., and Grebe, M. (2017). Arabidopsis BTB/POZ protein-dependent PENETRATION3 trafficking and disease susceptibility. *Nat. Plants.* 3, 854–858. doi: 10.1038/s41477-017-0039-z

Nägele, T., Gibon, Y., and Hir., R. (2022). Plant sugar metabolism, transport and signalling in challenging environments. *Physiol. Plant* 174, e13768. doi: 10.1111/ppl.13768

Nigro, D., Fortunato, S., Giove, S. L., Paradiso, A., Gu, Y. Q., Blanco, A., et al. (2016). Glutamine synthetase in durum wheat: genotypic variation and relationship with grain protein content. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00971

Pankievicz, V. C., Camilios-Neto, D., Bonato, P., Balsanelli, E., Tadra-Sfeir, M. Z., Faoro, H., et al. (2016). RNA-seq transcriptional profiling of Herbaspirillum seropedicae colonizing wheat (*Triticum aestivum*) roots. *Plant Mol. Biol.* 90, 589–603. doi: 10.1007/s11103-016-0430-6

Peng, L., Gao, W., Song, M., Li, M., He, D., and Wang, Z. (2022). Integrated Metabolome and Transcriptome Analysis of Fruit Flavor and Carotenoids Biosynthesis Differences Between Mature-Green and Tree-Ripe of cv. "Golden Phoenix" Mangoes (*Mangifera indica* L.). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.816492

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

Prakash, S., Kumar, M., Radha., , Kumar, S., Jaconis, S., Parameswari, E., et al. (2023) The resilient cotton plant: uncovering the effects of stresses on secondary metabolomics and its underlying molecular mechanisms. *Funct Integr Genomics*. 23(2):183. doi: 10.1007/s10142-023-01118-9

Qi, Z., and Xiong, L. (2013). Characterization of a purine permease family gene OsPUP7 involved in growth and development control in rice. *J. Integr. Plant Biol.* 55, 1119–1135. doi: 10.1111/jipb.12101

Shannon, P., Markiel, 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

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 (1). doi: 10.3835/plantgenome2015.04.0025

Wan, H., Qian, J., Zhang, H., Lu, H., Li, O., Li, R., et al. (2022). Combined transcriptomics and metabolomics analysis reveals the molecular mechanism of salt tolerance of huayouza 62, an elite cultivar in rapeseed (*Brassica napus* L.). *Int. J. Mol. Sci.* 23, 1279. doi: 10.3390/ijms23031279

Wang, Y., Liu, W., Wang, X., Yang, R., Wu, Z., Wang, H., et al. (2020). MiR156 regulates anthocyanin biosynthesis through *SPL* targets and other microRNAs in poplar. *Hortic. Res.* 7, 118. doi: 10.1038/s41438-020-00341-w

Wang, Z., Yang, H., Ma, Y., Jiang, G., Mei, X., Li, X., et al. (2022). WGCNA analysis revealing molecular mechanism that bio-organic fertilizer improves pear fruit quality by increasing sucrose accumulation and reducing citric acid metabolism. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1039671

Wen, B., Mei, Z., Zeng, C., and Liu, S. (2017). metaX: a flexible and comprehensive software for processing metabolomics data. *BMC Bioinf.* 18, 183. doi: 10.1186/s12859-017-1579-y

Westerhuis, J. A., Hoefsloot, H. C., Smit, S., Vis, D. J., Smilde, A. K., van Velzen, E. J., et al. (2008). Assessment of PLSDA cross validation. *Metabolomics* 4, 81–89. doi: 10.1007/s11306-007-0099-6

Wu, S., Wu, D., Song, J., Zhang, Y., Tan, Q., Yang, T., et al. (2022). Metabolomic and transcriptomic analyses reveal new insights into the role of abscisic acid in modulating mango fruit ripening. *Hortic. Res.* 9, uhac102. doi: 10.1093/hr/uhac102

Yao, Y., Xiong, E., Qu, X., Li, J., Liu, H., Quan, L., et al. (2023). WGCNA and transcriptome profiling reveal hub genes for key development stage seed size/oil content between wild and cultivated soybean. *BMC Genomics* 24, 494. doi: 10.1186/s12864-023-09617-6

Yin, H., Sun, Q., Lu, X., Zhang, L., Yuan, Y., Gong, C., et al. (2022). Identification of the glutamine synthetase (GS) gene family in four wheat species and functional analysis of Ta4D.GSe in Arabidopsis thaliana. *Plant Mol. Biol.* 110, 93–106. doi: 10.1007/s11103-022-01287-4

Yoon, J., Cho, L. H., Tun, W., Jeon, J. S., and An, G. (2012). Sucrose signaling in higher plants. *Plant Sci.* 302, 110703. doi: 10.1016/j.plantsci.2020.110703

Yu, Z., Islam, S., She, M., Diepeveen, D., Zhang, Y., Tang, G., et al. (2018). Wheat grain protein accumulation and polymerization mechanisms driven by nitrogen fertilization. *Plant J.* 96, 1160–1177. doi: 10.1111/tpj.14096

Yu, X., Liu, H., Xu, X., Hu, Y., Wang, X., and Wen, C. (2021). Pharmacokinetics of yunaconitine and indaconitine in mouse blood by UPLC-MS/MS. *J. Chromatogr B Analyt Technol. BioMed. Life Sci.* 1179, 122840. doi: 10.1016/j.jchromb.2021.122840

Zhang, Y., and Fernie, A. R. (2018). On the role of the tricarboxylic acid cycle in plant productivity. *J. Integr. Plant Biol.* 60, 1199–1216. doi: 10.1111/jipb.12690

Zhang, Y., and Fernie, A. R. (2023). The role of TCA cycle enzymes in plants. *Adv. Biol. (Weinh).* 7, e2200238. doi: 10.1002/adbi.202200238

Zhao, L., Zhang, B., Huang, H., Huang, W., Zhang, Z., Wang, Q., et al. (2022). Metabolomic and transcriptomic analyses provide insights into metabolic networks during cashew fruit development and ripening. *Food Chem.* 404, 134765. doi: 10.1016/j.foodchem.2022.134765

Zheng, T., Qi, P. F., Cao, Y. L., Han, Y. N., Ma, H. L., Guo, Z. R., et al. (2018). Mechanisms of wheat (*Triticum aestivum*) grain storage proteins in response to nitrogen application and its impacts on processing quality. *Sci. Rep.* 8, 11928. doi: 10.1038/s41598-018-30451-4

frontiers | Frontiers in Plant Science

TYPE Mini Review
PUBLISHED 02 February 2024
DOI 10.3389/fpls.2024.1342639

OPEN ACCESS

EDITED BY
Jian Chen,
Jiangsu University, China

REVIEWED BY
Zhao Zhang,
China Agricultural University, China
Bo Wei,
Hefei University, China

*CORRESPONDENCE
Bing He
hebing1980@aliyun.com
Fushan Sun
sunfushan@caas.cn

†These authors have contributed equally to this work

RECEIVED 22 November 2023
ACCEPTED 16 January 2024
PUBLISHED 02 February 2024

CITATION
Wang C, Meng L, Zhang G, Yang X, Pang B, Cheng J, He B and Sun F (2024) Unraveling crop enzymatic browning through integrated omics.
Front. Plant Sci. 15:1342639.
doi: 10.3389/fpls.2024.1342639

Unraveling crop enzymatic browning through integrated omics

Chunkai Wang[1†], Lin Meng[1†], Guochao Zhang[1], Xiujun Yang[1], Bingwen Pang[2], Junjie Cheng[2], Bing He[2*] and Fushan Sun[1*]

[1]Key Laboratory of Tobacco Biology and Processing, Ministry of Agriculture, Tobacco Research Institute, Chinese Academy of Agricultural Sciences (CAAS), Qingdao, China, [2]Institute of Germplasm Resources and Biotechnology, Jiangsu Academy of Agricultural Sciences, Nanjing, China

Enzymatic browning reactions, triggered by oxidative stress, significantly compromise the quality of harvested crops during postharvest handling. This has profound implications for the agricultural industry. Recent advances have employed a systematic, multi-omics approach to developing anti-browning treatments, thereby enhancing our understanding of the resistance mechanisms in harvested crops. This review illuminates the current multi-omics strategies, including transcriptomic, proteomic, and metabolomic methods, to elucidate the molecular mechanisms underlying browning. These strategies are pivotal for identifying potential metabolic markers or pathways that could mitigate browning in postharvest systems.

KEYWORDS
enzymatic browning, ROS, PPO activity, multi-omics, postharvest, oxidative stress

1 Introduction

Postharvest crops remain biologically active to sustain physiological functions and metabolic processes. However, this continued vitality renders them vulnerable to abiotic stresses during storage and processing phases, which often leads to enzymatic browning. Such browning not only diminishes the nutritional value, sensory appeal, and safety of the produce but also poses a significant challenge within the postharvest processing sector (Tinello and Lante, 2018). The escalating prevalence and severity of browning during storage and processing underscore the critical need to identify the mechanisms that bolster crop resilience to browning hazards.

During postharvest handling, crops undergo significant biochemical and metabolic transformations that lead to the accumulation of reactive oxygen species (ROS) and subsequent membrane disruption. The biochemical underpinnings of enzymatic browning involve the oxidation of phenolic substrates by oxidases to form ortho-quinone compounds. These compounds then polymerize with various substrates, culminating in the formation of brown pigments (Paudel et al., 2020). Prior studies have established that the incidence of enzymatic browning in harvested crops is associated with multiple stress

factors, including senescence, desiccation, chilling injury, pathogen infection, mechanical damage, heat stress, and other processes (Yi et al., 2009). Commercially, a range of postharvest strategies is implemented to mitigate these effects, including maintaining low temperatures, modifying atmospheric conditions, and applying chemical treatments. These strategies are all aimed at reducing respiration rates, delaying senescence and browning, inhibiting pathogen proliferation, and preserving crop quality.

Identifying new compounds relevant to browning and elucidating their biosynthetic pathways in postharvest crops present significant challenges. Over the last two decades, omics research has made substantial strides across various biological and chemical fields, mainly owing to rapid advancements in high-throughput nucleic acid sequencing (e.g., next-generation sequencing, NGS) and the identification of proteins and metabolites through diverse mass spectrometry techniques (He et al., 2022). More recently, a paradigm shift has occurred in omics research based on the increasing application of these methods for evaluations of harvested crops processing and storage, with studies transitioning from single-omics to integrated multi-omics approaches (Yun et al., 2019; Sirangelo et al., 2022). This transition has inspired many integrated multi-omics studies, resulting in the exponential growth of omics-based research on the browning reaction (Liu et al., 2022b). Therefore, this review aims to consolidate current knowledge on integrated omics strategies, including transcriptomic, proteomic, and metabolomic analyses, and explore how these omics techniques contribute to a deeper understanding of the molecular mechanisms underlying the browning of harvested crops during the postharvest period.

2 Enzymatic browning during postharvest crop processing

Several studies have established that is primarily accountable for the discoloration observed in crops during storage and processing (Figure 1). This browning occurs when crop polyphenols, or phenolic molecules, are oxidized by specific enzymes in the presence of oxygen, leading to the formation of quinones, which subsequently undergo chemical polymerization, resulting in a brown coloration (Parveen et al., 2010; Sui et al., 2023). Crops present a diverse array of phenolic compounds and oxidizing enzymes and thus are predisposed to rapid browning upon slicing, crushing, or any form of processing (Hamdan et al., 2022). The enzymatic browning mechanism is a process in which enzymes in the cytoplasm act upon substrates, mainly polyphenols located in plastids. This reaction does not occur in fresh fruit and vegetable due to separation of their enzymes from its substrate by cell compartments. During the handling of crops, tissue damage may cause plastids to rupture, facilitating the interaction between oxidase enzymes and their polyphenolic substrates, and thereby triggering the browning reaction (Sui et al., 2023).

2.1 Substrates of enzymatic browning

Phenolic compounds, widely recognized as polyphenols, are prevalent chemical constituents in a variety of fruits and vegetables and play a pivotal role in enzymatic browning by acting as substrates for the enzymes that catalyze this reaction (Singh et al., 2016).

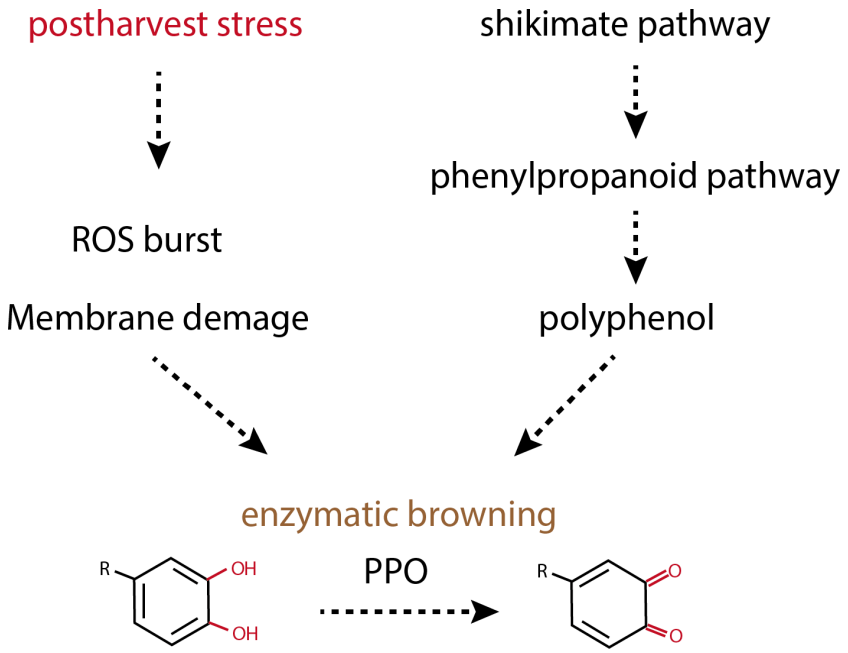


FIGURE 1
Biochemical mechanism of enzymatic browning during postharvest.

These compounds belong to a class of secondary metabolites synthesized via complex and irreversible pathways, notably the shikimate and phenylpropanoid pathways, which are exclusive to plants. Such pathways contribute to plant defense mechanisms against herbivores, microbial pathogens, invertebrate pests, and environmental stressors (Sun et al., 2022).

The structure of phenolic compounds is characterized by an aromatic ring bearing one or more hydroxyl groups, and their biosynthesis is primarily routed through the shikimate pathway, which leverages intermediates from carbohydrate metabolism (Singh et al., 2017). The oxidation potential of monocyclic phenolics is greatest for compounds with 2,4,5-trihydroxy substituents and lowest in monophenols. Typically, a wide range of 1,2-dihydroxyarenes, also referred to as ortho-dihydroxyphenols, are particularly prone to oxidation by polyphenol oxidase (PPO) due to the ortho positioning of the hydroxyl groups, which facilitates oxidation (Parveen et al., 2010).

The oxidation process mediated by PPOs is significantly influenced by the characteristics of the phenolic side chain, including the nature and number of hydroxyl groups and their placement on the aromatic ring (Winters et al., 2008). These structural attributes also determine the extent to which these compounds can undergo indirect oxidation through interactions with PPO reaction products. The substrate specificity of PPOs varies depending on the plant species and isoform, with a generally higher affinity for diphenols over monophenols. Consequently, polyphenols are recognized as the principal contributors to enzymatic browning in postharvest storage and processing of crops.

2.2 Enzymes of enzymatic browning

Enzymatic browning is a biochemical process in which specific enzymes oxidize phenolic compounds to form quinones, which then non-enzymatically polymerize to produce brown pigments (Zhu et al., 2023). The chief enzyme implicated in this browning reaction is polyphenol oxidase (PPO), which is categorized into two main types: EC 1.10.3.1 (including o-diphenol oxygen oxidoreductase, diphenol oxidase, and catechol oxidase) and EC 1.14.18.1 (encompassing tyrosinase, cresolase, and monophenol monooxygenase) (Moon et al., 2020). PPO enzymes are located within the cytoplasm, while their phenolic substrates are typically housed within plastids. Tissue damage in plants triggers the migration of PPO enzymes from the cytoplasm to the plastids, facilitating their interaction with the substrates.

Ortho-quinones, the products of PPO activity, display electrophilic characteristics, making them prone to nucleophilic attacks from a variety of biomolecules, such as proteins, peptides, amino acids, water, and other polyphenols. These interactions lead to the formation of Michael-type adducts, contributing to the complexity of the browning pigments (Loizzo et al., 2012; Schieber, 2018). Additionally, peroxidase (POD, EC 1.11.1.7) is another thermostable enzyme that significantly contributes to oxidative browning reactions (Shrestha et al., 2020). Utilizing hydrogen peroxide (H2O2) as a cofactor, POD catalyzes the single-electron oxidation of a wide array of substrates, further promoting enzymatic browning. It simultaneously reduces hydrogen peroxide to water, thus participating actively in the browning process (Nokthai et al., 2010).

Therefore, both PPO and POD are instrumental in the enzymatic browning process. They not only share common substrates but also their concerted action on diphenolic substrates can lead to melanin formation, which is a key component of the browning phenotype in plants.

2.3 The influence of reactive oxygen species on enzymatic browning

Reactive oxygen species (ROS) are byproducts of cellular processes such as photosynthesis, respiration, and other metabolic activities. These include singlet oxygen ($1O_2$), hydrogen peroxide (H_2O_2), superoxide anions (O_2–), hydroxyl radicals (OH·), lipid peroxide (ROOR′), and alkyl radicals (R) (Li et al., 2019a). At low concentrations, ROS act as signaling molecules within plant cells, playing a pivotal role in various physiological functions. It can induce both reversible and irreversible oxidative post-translational modifications to proteins involved in many signaling cascades (Waszczak et al., 2015). However, when produced in excess, ROS can cause oxidative stress, which is detrimental to the organism.

High levels of ROS can induce senescence, compromising the integrity and functionality of cellular membranes. This interaction between ROS and cellular biomolecules can lead to their modification or inactivation, resulting in organelle dysfunction and structural cellular changes (Rezayian et al., 2019). During postharvest storage and processing, the accumulation of ROS and free radicals can trigger browning reactions. Specifically, peel browning is often attributed to the release of cellular contents following the loss of membrane permeability caused by ROS surges (Lin et al., 2020).

Controlling the ROS pathway is essential for reducing browning and maintaining the quality of fruits and vegetables after harvest. Plants can mitigate browning by enhancing the activity of antioxidative enzymes and increasing the antioxidant content within the fruit, thus counteracting the accumulation of ROS (Ali et al., 2019; Li et al., 2019a; Liu et al., 2021b). The ROS metabolic pathway is regulated by three primary antioxidant enzymes: superoxide dismutase (SOD), ascorbate peroxidase (APX), and catalase (CAT). These enzymes play a critical role in maintaining ROS balance and protecting plants against oxidative stress (Li et al., 2019b). SOD catalyzes the dismutation of O_2^- into H_2O_2 and O_2, while APX and CAT efficiently detoxify H_2O_2. By reducing ROS levels, these enzymatic ROS scavengers not only prevent oxidative damage but also inhibit enzymatic browning processes (Li et al., 2019c).

2.4 The role of intracellular membrane integrity in enzymatic browning

The integrity of the plasma membrane plays a pivotal role in preventing browning in postharvest produce. Compromised cell

membrane integrity facilitates the contact between phenolic compounds and oxidative enzymes, catalyzing the enzymatic browning process (Ma et al., 2022). This degradation is often indicated by a shift from unsaturated to saturated fatty acids in the membrane, coupled with an increase in lipid peroxidation products and enhanced membrane permeability (Lin et al., 2016).

Phospholipids are essential components of cellular membranes, and enzymes such as Phospholipase D (PLD) and lipoxygenase (LOX) are critical in the breakdown of membrane phospholipids, an early sign of senescence and browning in harvested tissues (Li et al., 2009; Sun et al., 2020). PLD catalyzes the hydrolysis of phosphatidylcholine (PC) or phosphatidylethanolamine (PE) to produce phosphatidic acid (PA), whereas LOX initiates the peroxidation of membrane lipids. This peroxidation can lead to the deterioration of the cell membrane's structural integrity, resulting in the disruption of the phospholipid bilayer and loss of cellular compartmentalization, which are critical events preceding browning (Trabelsi et al., 2012).

3 Omics technologies and postharvest crop browning studies

The browning reaction significantly impacts marketability during the storage and processing of postharvest crops. Unraveling the molecular basis of the browning reaction is key to understanding and developing effective strategies to inhibit it. Omics technologies, comprising transcriptomics, proteomics, and metabolomics, provide comprehensive insights when applied either individually or synergistically. These methodologies have been instrumental in assessing the effects of various postharvest treatments aimed at mitigating browning during storage and processing (Table 1).

3.1 The application of transcriptome in postharvest browning

Transcriptomics is the study of RNA expression profiles, including both coding and regulatory non-coding RNA sequences, within a given temporal context (Pandit et al., 2018; Luo et al., 2022). This analysis can be conducted through hybridization-based methods, such as microarrays, or sequence-based approaches, like direct cDNA sequencing. The browning process in postharvest produce involves significant changes in gene expression.

Early genetic studies focused on identifying key factors in the browning of postharvest fruits and vegetables, providing foundational insights into the underlying regulatory mechanisms (Coetzer et al., 2001). However, as whole-genome transcriptomic studies advanced, genes related to the browning process, including those involved in downstream signaling and anti-browning responses, have been somewhat neglected. Our current study employs transcriptome analysis of fresh-cut potato tubers to pinpoint critical genes implicated in browning, highlighting the

TABLE 1 Recent articles on integrated omics analysis of crops, type of omics integration and major output.

Crop	Integrated omics	Major Output	Reference
Longan	Transcriptome Metabolome	Polymerization reaction of PAs and lignin monomers mediated by LACs/PRXs is induced by water-loss leading to the browning of longan pericarp.	(Liu et al., 2024)
Lilium bulb	Transcriptome	Phenol and fatty biosynthesis are responsible for browning and a complex hormone signaling network and most genes responsive to injury transcription factors significantly change.	(Wang et al., 2023a)
Fresh-cut apples	Transcriptome Metabolome	Selenium inhibits browning of fresh-cut apples by reducing membrane lipid degradation and increase gene expression of the antioxidant.	(Wang et al., 2023b)
Apple	Transcriptome Proteome Methylome	Methylated-*NCA1* and O-methyltransferase 1 (OMT1) significantly increased in apple browning	(Wang et al., 2022)
Flesh-cut eggplant	Transcriptome Metabolome	Chlorogenic acid act as the main browning substrate in fresh-cut eggplant	(Liu et al., 2022a)
Fresh-cut lettuce	Transcriptome Metabolome	6-Benzylaminopurine reduce browning by inhibiting phenolic-related metabolite biosynthesis, especially scopoletin.	(Liu et al., 2022b)
Morel	Metabolome	Tyrosine metabolism is involved in browning of morels during storage	(Gao et al., 2022)
Fresh-Cut Sand Pear	Transcriptome	Fresh-cut sand pear fruit enzymatic browning is due to the expression of *PbrPPO4* that was probably regulated by lncRNA *PB.156.1*.	(Fan et al., 2021)
Grape	Proteome	Browning is primary involved in phenylpropanoid biosynthesis, tyrosine metabolism, phenylalanine metabolism, oxidative phosphorylation metabolism, glutathione metabolism, peroxisome pathway, and fatty acid degradation	(Liu et al., 2021a)
Fresh-cut apples	Transcriptome	γ-aminobutyric acid reduce browning by regulating the genes expression related to the synthesis of browning enzymes and phenolic substances	(Zhao et al., 2021)
Grape	Proteome	The browning-related proteins are primarily involved in the	(Liu et al., 2021a)

(Continued)

TABLE 1 Continued

Crop	Integrated omics	Major Output	Reference
		phenylpropanoid biosynthesis, oxidative phosphorylation metabolism, peroxisome pathway and fatty acid degradation.	
Litchi	Transcriptome Proteome	Anthocyanin metabolism is involved in litchi pericarp browning	(Qu et al., 2021)
Flesh-cut eggplant	Transcriptome	Browning is involved in expression regulatory networks was set up based on tyrosine metabolism and phenylpropanoid biosynthesis	(Liu et al., 2021c)
Litchi	Long-read sequencing transcriptome	During the 'browning' stage, the expression of isoforms related to cell wall degradation, oxidation, and disease response was significantly up-regulated.	(Zhou et al., 2020b)
Pear	Transcriptome	DEGs indicates redox reaction, membrane lipid metabolism account for the browning disorder.	(Zhou et al., 2020a)
'Nanguo' pear	Proteome	Peel browning is primarily involved in the phenylpropanoid pathway, linoleic acid pathways, fatty acid biosynthesis pathway, glutathione metabolism pathway, photosynthesis pathway, oxidative phosphorylation pathway, and glycolysis pathway.	(Wang et al., 2017)
Fresh-Cut lettuce	Metabolome	Browning process kinetics is associated with a higher level of constitutive lysophospholipids and constitutive levels of caffeoylquinic derivatives,	(García et al., 2017)

roles of plant hormone biosynthesis, signaling molecules, and respiratory burst oxidase in this process (Wang et al., 2023a).

Senescence, a crucial stage in plant development, is intimately connected with browning. Zhou et al. (2020b) utilized long-read sequencing technology to identify a two-phase browning pattern in postharvest litchi, distinguishing the onset of senescence from the browning stage. Their findings underscored significant stage-specific biological pathway activations, such as cell wall degradation, oxidative processes, and disease responses during browning. Comparative transcriptomic analyses between browning-resistant and susceptible cultivars have revealed candidate genes and mechanisms involved in postharvest browning. A recent time-course study on fresh-cut eggplant employed transcriptome analysis to unravel the transcriptional regulation of browning, proposing two key regulatory networks related to tyrosine metabolism and phenylpropanoid biosynthesis (Liu et al., 2021c).

Transcriptomics has also been applied to understand how postharvest crops respond to anti-browning treatments. Zhao et al. (2021) explored how γ-aminobutyric acid (GABA) prevents browning in fresh-cut apples, finding that GABA modified the expression of genes related to the synthesis of browning enzymes and phenolic compounds. Similarly, the application of melatonin, an effective anti-browning agent, was shown to enhance the antioxidant system by regulating ROS-metabolism-related genes in a comparative transcriptome study (Min et al., 2023). Plant hormones, which play a critical role in regulating postharvest browning, have been studied as well; for instance, Liu et al. (2022b) demonstrated that 6-Benzylaminopurine (6-BA) effectively delayed browning in fresh-cut lettuce, with transcriptome analysis indicating a significant impact on phenolic-related metabolic pathways.

3.2 Proteome and its impact on postharvest browning studies

Proteomics investigates the abundance, expression, and interactions of proteins within organisms or cells. Unlike transcriptomic or genomic data, proteomic information provides a more direct understanding of the actual functional molecules in biological processes, as protein levels and activities cannot be fully predicted from RNA or DNA data alone (Mathabe et al., 2020). Proteins are central to the regulation and execution of nearly all biological functions, making proteomics essential for uncovering the molecular mechanisms that govern the development and browning of postharvest crops.

Recent proteomic studies have focused on gene expression and the accumulation of proteins during the storage and processing of crops (Feng et al., 2016). These studies have the potential to reduce postharvest losses due to browning by identifying and selecting crop varieties with enhanced tolerance to browning and other desirable quality traits. For instance, Qu et al. (2022) analyzed the proteome of postharvest mushroom fruiting bodies during storage, identifying 168 significantly regulated proteins involved in processes such as translation, carbohydrate metabolism, signal transduction, and amino acid metabolism. Their research also highlighted the role of AMPK and FOXO signaling pathways in the browning of mushrooms during storage. Ban et al. (2018) applied proteomics to study the effect of chitosan and carboxymethyl cellulose coatings on strawberries packaged in polyethylene terephthalate containers and stored at low temperatures. Their findings included the identification of a set of proteins related to browning, primarily associated with primary and secondary metabolism (Wang et al., 2017).

Beyond the linear sequence and three-dimensional structure of proteins, post-translational modifications (PTMs) greatly influence protein function and activity. For example, quantitative phosphoproteome analysis has shown that SN2 can reduce browning in fresh-cut potatoes by altering the phosphorylation levels of kinases. A network involving serine/arginine-rich proteins and mitogen-activated protein kinases has been proposed as a

potential kinase-substrate interaction system that influences browning (Li et al., 2023).

3.3 Metabolome and postharvest browning studies

Metabolomics is the holistic study of metabolites within a biological system under defined conditions, providing a snapshot of the physiological state of an organism. These metabolites reflect the final products of gene expression and regulatory interactions, often showing a more direct correlation with observed phenotypes than mRNA or protein levels. As a branch of omics science, metabolomics is particularly effective for its comprehensive coverage of biological processes and its ability to link genotype to phenotype. Current metabolomics research employs targeted, widely targeted, and untargeted strategies, requiring the use of sensitive and precise techniques like mass spectrometry (MS) coupled with gas chromatography (GC), liquid chromatography (LC), or nuclear magnetic resonance (NMR) spectroscopy (Meng et al., 2022).

The role of metabolomics in understanding the browning of postharvest crops has gained significant momentum. It is instrumental in identifying and characterizing bioactive compounds and chemical markers that may influence the quality and shelf-life of produce (Qiu et al., 2021; Gao et al., 2022). For instance, a study on pomegranate storage identified key secondary metabolites involved in aril browning, particularly those related to the biosynthesis pathways of flavonoids, flavonols, and isoflavonoids, with a focus on phenylpropanoid biosynthesis (Shi et al., 2022). Metabolomics has also revealed that the metabolic profile of morel mushrooms changes substantially during storage, with increases in amino acids and fatty acids and decreases in soluble sugars, organic acids, and certain phenolic compounds (Gao et al., 2022). Comparative metabolomics between lettuce cultivars with varying browning rates has pinpointed metabolites implicated in the browning process, such as differences in lysophospholipid levels, phospholipase and lipoxygenase activity, and the presence of caffeoylquinic acid derivatives (García et al., 2017).

3.4 Integrated omics approaches applied in understanding postharvest browning

The integration of various omics datasets is pivotal for unraveling the complex molecular mechanisms of enzymatic browning in postharvest crops. By combining multi-omics data, such as genomics, transcriptomics, proteomics, and metabolomics, researchers can gain a comprehensive understanding of the molecular events that affect postharvest crop quality (Aiese Cigliano et al., 2022; Xie et al., 2023). Data integration is a sophisticated process that requires advanced computational methods to merge and analyze diverse omics datasets effectively. Current research often utilizes functional and statistical networks to facilitate the visualization and interpretation of these complex data relationships, aiming to identify key metabolic hubs associated with the browning process.

Recent studies have demonstrated the power of multi-omics approaches in pinpointing the primary factors contributing to browning during postharvest treatments and storage (Qiao et al., 2022; Romero et al., 2022). For example, a combined transcriptomic and metabolomic investigation into the browning of fresh-cut eggplant identified fluctuations in membrane phospholipid and unsaturated fatty acid metabolites, as well as changes in the expression of genes related to membrane lipid metabolism (Liu et al., 2022a). Similarly, the analysis of harvested litchi using both transcriptomic and proteomic techniques has provided insights into the regulation of the anthocyanin biosynthesis pathway during the browning process (Qu et al., 2021).

4 Prospects

This review has highlighted the potential applications of an integrated omics approach to study the browning reactions that occur during the storage and processing of postharvest crops. It offers an in-depth look at how omics integration can enhance the development of anti-browning treatments in postharvest management. Omics platforms, including those for transcriptomics, proteomics, and metabolomics, have proven to be more effective than traditional methods, providing a powerful suite of tools for elucidating molecular markers, regulatory networks, and genes involved in the browning reaction.

Despite the advantages, the application of diverse omics techniques comes with the substantial challenge of managing and making sense of the massive amounts of data they generate. Integrating this high-throughput data from various sources remains a daunting task, necessitating sophisticated analytical strategies. The successful integration of omics data will depend on the development of user-friendly analytical tools that are tightly linked to biological processes. Advanced data integration tools, such as machine learning should be considered to take advantage of omics datasets for constructing more biologically realistic network models across different biological layers from gene expression to postharvest browning reaction.

Metabolomics, in particular, faces hurdles such as the sheer complexity of the metabolome, gaps in our understanding of metabolic pathways, and difficulties in identifying molecules by their structural detector signals, compounded by the lack of extensive, metabolite-specific libraries. Moreover, due to significant advancements in redox proteomics, the iodoacetyl tandem mass tag (iodoTMT)-based redox proteomic approach has been successfully utilized for detecting redox-sensitive proteins during tomato fruit ripening (Wang et al., 2021). This approach can now be applied to gain a deeper understanding of the mechanism by which ROS-mediated oxidative post-translational modifications occur during the browning reaction in postharvest crops. There is a pressing need for further research in multi-omics to directly link biochemical activities to biomarkers and to advance our understanding of the postharvest browning process. Future studies should aim to fill these knowledge gaps and leverage the power of integrated omics for practical applications in crop postharvest biology.

Author contributions

CW: Writing – original draft. LM: Writing – original draft. GZ: Data curation, Investigation, Writing – original draft. XY: Data curation, Investigation, Writing – original draft. BP: Data curation, Investigation, Writing – original draft. JC: Data curation, Investigation, Methodology, Writing – review & editing. BH: Conceptualization, Writing – original draft, Writing – review & editing. FS: Conceptualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was financially supported by the Agricultural Science and Technology Innovation Program (ASTIP-TRIC03).

Conflict of interest

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

Publisher's note

References

Aiese Cigliano, R., Aversano, R., Di Matteo, A., Palombieri, S., Termolino, P., Angelini, C., et al. (2022). Multi-omics data integration provides insights into the post-harvest biology of a long shelf-life tomato landrace. *Horticulture Res.* 9, uhab042. doi: 10.1093/hr/uhab042

Ali, S., Sattar Khan, A., Ullah Malik, A., Anjum, M. A., Nawaz, A., and Shoaib Shah, H. M. (2019). Modified atmosphere packaging delays enzymatic browning and maintains quality of harvested litchi fruit during low temperature storage. *Scientia Hortic.* 254, 14–20. doi: 10.1016/j.scienta.2019.04.065

Ban, Z., Yan, J., Wang, J., Zhang, J., Yuan, Q., and Li, L. (2018). Effects of postharvest application of chitosan-based layer-by-layer assemblies on regulation of ribosomal and defense proteins in strawberry fruit (fragaria x ananassa). *Scientia horticulturae,* 240. doi: 10.1016/j.scienta.2018.06.035

Coetzer, C., Corsini, D., Love, S., Pavek, J., and Tumer, N. (2001). Control of enzymatic browning in potato (Solanum tuberosum L.) by sense and antisense RNA from tomato polyphenol oxidase. *J. Agric. Food Chem.* 49 (2), 652–657. doi: 10.1021/jf001217f

Fan, J., Du, W., Chen, Q.-L., Zhang, J.-G., Yang, X.-P., Hussain, S. B., et al. (2021). Comparative transcriptomic analyses provide insights into the enzymatic browning mechanism of fresh-cut sand pear fruit. *Horticulturae* 7, (11). doi: 10.3390/horticulturae7110502

Feng, X., An, Y., Zheng, J., Sun, M., and Wang, L. (2016). Proteomics and SSH analyses of ALA-promoted fruit coloration and evidence for the involvement of a MADS-box gene, mdMADS1. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01615

Gao, F., Xie, W., Zhang, H., Li, Z., Li, S., and Li, T. (2022). Metabolomic analysis of browning mechanisms of morels (Morchella sextelata) during storage. *Postharvest Biol. Technol.* 185, 111801. doi: 10.1016/j.postharvbio.2021.111801

García, C. J., García-Villalba, R., Gil, M. I., and Tomas-Barberan, F. A. (2017). LC-MS untargeted metabolomics to explain the signal metabolites inducing browning in fresh-cut lettuce. *J. Agric. Food Chem.* 65 (22), 4526–4535. doi: 10.1021/acs.jafc.7b01667

Hamdan, N., Lee, C. H., Wong, S. L., Fauzi, C. E., Zamri, N. M., and Lee, T. H. (2022). Prevention of enzymatic browning by natural extracts and genome-editing: A review on recent progress. *Molecules* 27, (3). doi: 10.3390/molecules27031101

He, B., Hu, F. Q., Du, H. Y., Cheng, J. J., Pang, B. W., Chen, X., et al. (2022). Omicsdriven crop potassium use efficiency breeding. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1076193

Li, M., Hong, Y., and Wang, X. (2009). Phospholipase D- and phosphatidic acid-mediated signaling in plants. *Biochim. Biophys. Acta (BBA) - Mol. Cell Biol. Lipids* 1791 (9), 927–935. doi: 10.1016/j.bbalip.2009.02.017

Li, M., Li, X., Han, C., Ji, N., Jin, P., and Zheng, Y. (2019b). Physiological and metabolomic analysis of cold plasma treated fresh-cut strawberries. *J. Agric. Food Chem.* 67 (14), 4043–4053. doi: 10.1021/acs.jafc.9b00656

Li, M., Li, X., Han, C., Ji, N., Jin, P., and Zheng, Y. (2019c). UV-C treatment maintains quality and enhances antioxidant capacity of fresh-cut strawberries. *Postharvest Biol. Technol.* 156, 110945. doi: 10.1016/j.postharvbio.2019.110945

Li, L.-Q., Mu, Y.-L., Chen, J., Wang, Q., Lu, Y.-F., Xin, S., et al. (2023). Molecular mechanism by which StSN2 overexpression inhibits the enzymatic browning of potato. *Postharvest Biol. Technol.* 203, 112416. doi: 10.1016/j.postharvbio.2023.112416

Li, J., Zhou, X., Wei, B., Cheng, S., Zhou, Q., and Ji, S. (2019a). GABA application improves the mitochondrial antioxidant system and reduces peel browning in 'Nanguo' pears after removal from cold storage. *Food Chem.* 297, 124903. doi: 10.1016/j.foodchem.2019.05.177

Lin, Y., Chen, G., Lin, H., Lin, M., Wang, H., and Lin, Y. (2020). Chitosan postharvest treatment suppresses the pulp breakdown development of longan fruit through regulating ROS metabolism. *Int. J. Biol. Macromolecules* 165, 601–608. doi: 10.1016/j.ijbiomac.2020.09.194

Lin, Y., Lin, H., Lin, Y., Zhang, S., Chen, Y., and Jiang, X. (2016). The roles of metabolism of membrane lipids and phenolics in hydrogen peroxide-induced pericarp browning of harvested longan fruit. *Postharvest Biol. Technol.* 111, 53–61. doi: 10.1016/j.postharvbio.2015.07.030

Liu, B., Fang, F., Guan, H., Zhang, J., Luo, H., Zhong, R., et al. (2024). Integrated function of proanthocyanidin and lignin polymerization mediated by LAC/PRXs in pericarp browning of longan fruit. *Postharvest Biol. Technol.* 207, 112618. doi: 10.1016/j.postharvbio.2023.112618

Liu, F., Huang, W., Feng, Z., Tao, Y., Fan, Y., He, W., et al. (2021a). Proteomic analyses on the browning of shade-dried Thompson seedless grape. *Appl. Biol. Chem.* 64 (1), 41. doi: 10.1186/s13765-021-00612-7

Liu, Y., Liao, L., Yin, F., Song, M., Shang, F., Shuai, L., et al. (2022b). Integration of metabolome and transcriptome profiling reveals the effect of 6-Benzylaminopurine on the browning of fresh-cut lettuce during storage. *Postharvest Biol. Technol.* 192, 112015. doi: 10.1016/j.postharvbio.2022.112015

Liu, X., Xiao, K., Zhang, A., Zhu, W., Zhang, H., Tan, F., et al. (2022a). Metabolomic analysis, combined with enzymatic and transcriptome assays, to reveal the browning resistance mechanism of fresh-cut eggplant. *Foods* 11, (8). doi: 10.3390/foods11081174

Liu, X., Zhang, A., Shang, J., Zhu, Z., Li, Y., Wu, X., et al. (2021b). Study on browning mechanism of fresh-cut eggplant (Solanum melongena L.) based on metabolomics, enzymatic assays and gene expression. *Sci. Rep.* 11 (1), 6937. doi: 10.1038/s41598-021-86311-1

Liu, X., Zhang, A., Zhao, J., Shang, J., Zhu, Z., Wu, X., et al. (2021c). Transcriptome profiling reveals potential genes involved in browning of fresh-cut eggplant (Solanum melongena L.). *Sci. Rep.* 11 (1), 16081. doi: 10.1038/s41598-021-94831-z

Loizzo, M. R., Tundis, R., and Menichini, F. (2012). Natural and synthetic tyrosinase inhibitors as antibrowning agents: an update. *Compr. Rev. Food Sci. Food Saf.* 11 (4), 378–398. doi: 10.1111/j.1541-4337.2012.00191.x

Lou, C. P., He, B., Shi, P. B., Xi, J. L., Gui, H. B., Pang, B. W., et al. (2022). Transcriptome dynamics uncovers long non-coding RNAs response to salinity stress in *Chenopodium quinoa*. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.988845

Ma, W., Li, J., Murtaza, A., Iqbal, A., Zhang, J., Zhu, L., et al. (2022). High-pressure carbon dioxide treatment alleviates browning development by regulating membrane lipid metabolism in fresh-cut lettuce. *Food Control* 134, 108749. doi: 10.1016/j.foodcont.2021.108749

Mathabe, P. M. K., Belay, Z. A., Ndlovu, T., and Caleb, O. J. (2020). Progress in proteomic profiling of horticultural commodities during postharvest handling and storage: A review. *Scientia Hortic.* 261, 108996. doi: 10.1016/j.scienta.2019.108996

Meng, L., Song,W. J., Chen,S. W., Hu,F. Q., Pang,B. W., Cheng,J. J., et al. (2022). Widely targeted metabolomics analysis reveals the mechanism of quality improvement of flue-cured tobacco. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1074029

Min, T., Lu, K., Chen, J., Niu, L., Lin, Q., Yi, Y., et al. (2023). Biochemical mechanism of fresh-cut lotus (Nelumbo nucifera gaertn.) root with exogenous melatonin treatment by multiomics analysis. *Foods* 12, (1). doi: 10.3390/foods12010044

Moon, K. M., Kwon, E.-B., Lee, B., and Kim, C. Y. (2020). Recent trends in controlling the enzymatic browning of fruit and vegetable products. *Molecules* 25, (12). doi: 10.3390/molecules25122754

Nokthai, P., Lee, V. S., and Shank, L. (2010). Molecular modeling of peroxidase and polyphenol oxidase: substrate specificity and active site comparison. *Int. J. Mol. Sci.* 11 (9), 3266–3276. doi: 10.3390/ijms11093266

Pandit, A. A., Shah, R. A., and Husaini, A. M. Transcriptomics: A time-efficient tool with wide applications in crop and animal biotechnology. *J. Pharm. Phyt.* 2018;7 (2):1701–1704.

Parveen, I., Threadgill, M. D., Moorby, J. M., and Winters, A. (2010). Oxidative phenols in forage crops containing polyphenol oxidase enzymes. *J. Agric. Food Chem.* 58 (3), 1371–1382. doi: 10.1021/jf9024294

Paudel, P., Seong, S. H., Wagle, A., Min, B. S., Jung, H. A., and Choi, J. S. (2020). Antioxidant and anti-browning property of 2-arylbenzofuran derivatives from Morus alba Linn root bark. *Food Chem.* 309, 125739. doi: 10.1016/j.foodchem.2019.125739

Qiao, L., Gao, M., Wang, Y., Tian, X., Lu, L., and Liu, X. (2022). Integrated transcriptomic and metabolomic analysis of cultivar differences provides insights into the browning mechanism of fresh-cut potato tubers. *Postharvest Biol. Technol.* 188, 111905. doi: 10.1016/j.postharvbio.2022.111905

Qiu, S., Tu, Y., Huang, D., Dong, Z., Huang, M., Cheng, J., et al. (2021). Selection of appropriate post-harvest processing methods based on the metabolomics analysis of Salvia miltiorrhiza Bunge. *Food Res. Int.* 144, 110366. doi: 10.1016/j.foodres.2021.110366

Qu, S., Li, M., Wang, G., Yu, W., and Zhu, S. (2021). Transcriptomic, proteomic and LC-MS analyses reveal anthocyanin biosynthesis during litchi pericarp browning. *Scientia Hortic.* 289, 110443. doi: 10.1016/j.scienta.2021.110443

Qu, H., Zhou, H., Ma, T., Zheng, Z., Zheng, E., Yang, H., et al. (2022). TMT-based quantitative proteomic analysis of postharvest Coprinus comatus fruiting body during storage. *Postharvest Biol. Technol.* 185, 111786. doi: 10.1016/j.postharvbio.2021.111786

Rezayian, M., Niknam, V., and Ebrahimzadeh, H. (2019). Oxidative damage and antioxidative system in algae. *Toxicol. Rep.* 6, 1309–1313. doi: 10.1016/j.toxrep.2019.10.001

Romero, I., Escribano, M. I., Merodio, C., and Sanchez-Ballesta, M. T. (2022). Postharvest high-CO2 treatments on the quality of soft fruit berries: an integrated transcriptomic, proteomic, and metabolomic approach. *J. Agric. Food Chem.* 70 (28), 8593–8597. doi: 10.1021/acs.jafc.2c01305

Schieber, A. (2018). Reactions of quinones—Mechanisms, structures, and prospects for food research. *J. Agric. Food Chem.* 66 (50), 13051–13055. doi: 10.1021/acs.jafc.8b05215

Shi, J., Wang, S., Tong, R., Wang, S., Chen, Y., Wu, W., et al. (2022). Widely targeted secondary metabolomics explored pomegranate aril browning during cold storage. *Postharvest Biol. Technol.* 186, 111839. doi: 10.1016/j.postharvbio.2022.111839

Shrestha, L., Kulig, B., Moscetti, R., Massantini, R., Pawelzik, E., Hensel, O., et al. (2020). Optimisation of physical and chemical treatments to control browning development and enzymatic activity on fresh-cut apple slices. *Foods* 9, (1). doi: 10.3390/foods9010076

Singh, B., Singh, J. P., Kaur, A., and Singh, N. (2016). Bioactive compounds in banana and their associated health benefits - A review. *Food Chem.* 206, 1–11. doi: 10.1016/j.foodchem.2016.03.033

Singh, B., Singh, J. P., Kaur, A., and Singh, N. (2017). Phenolic composition and antioxidant potential of grain legume seeds: A review. *Food Res. Int.* 101, 1–16. doi: 10.1016/j.foodres.2017.09.026

Sirangelo, T. M., Rogers, H. J., and Spadafora, N. D. (2022). Multi-omic approaches to investigate molecular mechanisms in peach post-harvest ripening. *Agriculture* 12, (4). doi: 10.3390/agriculture12040553

Sui, X., Meng, Z., Dong, T., Fan, X., and Wang, Q. (2023). Enzymatic browning and polyphenol oxidase control strategies. *Curr. Opin. Biotechnol.* 81, 102921. doi: 10.1016/j.copbio.2023.102921

Sun, Y., Luo, M., Ge, W., Zhou, X., Zhou, Q., Wei, B., et al. (2022). Phenylpropanoid metabolism in relation to peel browning development of cold-stored 'Nanguo' pears. *Plant Sci.* 322, 111363. doi: 10.1016/j.plantsci.2022.111363

Sun, Y., Sun, H., Luo, M., Zhou, X., Zhou, Q., Wei, B., et al. (2020). Membrane lipid metabolism in relation to core browning during ambient storage of 'Nanguo' pears. *Postharvest Biol. Technol.* 169, 111288. doi: 10.1016/j.postharvbio.2020.111288

Tinello, F., and Lante, A. (2018). Recent advances in controlling polyphenol oxidase activity of fruit and vegetable products. *Innovative Food Sci. Emerging Technol.* 50, 73–83. doi: 10.1016/j.ifset.2018.10.008

Trabelsi, H., Cherif, O. A., Sakouhi, F., Villeneuve, P., Renaud, J., Barouh, N., et al. (2012). Total lipid content, fatty acids and 4-desmethylsterols accumulation in developing fruit of Pistacia lentiscus L. *growing Wild Tunisia. Food Chem.* 131 (2), 434–440. doi: 10.1016/j.foodchem.2011.08.083

Wang, P., Li, X., Wang, Y., Wang, W., Tian, S., and Qin, G. (2021). Redox proteomic analysis reveals the involvement of oxidative post-translational modification in tomato fruit ripening. *Postharvest Biol. Technol.* 178, 111556. doi: 10.1016/j.postharvbio.2021.111556

Wang, L., Tang, T., Wang, W., Zhang, J., Wang, Z., and Wang, F. (2022). Multi-omics landscape of DNA methylation regulates browning in "Fuji" Apple. *Front. Nutr.* 8. doi: 10.3389/fnut.2021.800489

Wang, L., Wang, W., Shan, J., Li, C., Suo, H., Liu, J., et al. (2023a). A genome-wide view of the transcriptome dynamics of fresh-cut potato tubers. *Genes* 14, (1). doi: 10.3390/genes14010181

Wang, X., Zhang, X., Jia, P., Luan, H., Qi, G., Li, H., et al. (2023b). Transcriptomics and metabolomics provide insight into the anti-browning mechanism of selenium in freshly cut apples. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1176936

Wang, J.-W., Zhou, X., Zhou, Q., Liu, Z.-Y., Sheng, L., Wang, L., et al. (2017). Proteomic analysis of peel browning of 'Nanguo' pears after low-temperature storage. *J. Sci. Food Agric.* 97 (8), 2460–2467. doi: 10.1002/jsfa.8060

Waszczak, C., Akter, S., Jacques, S., Huang, J., Messens, J., and Van Breusegem, F. (2015). Oxidative post-translational modifications of cysteine residues in plant signal transduction. *J. Exp. Bot.* 66 (10), 2923–2934. doi: 10.1093/jxb/erv084

Winters, A. L., Minchin, F. R., Michaelson-Yeates, T. P. T., Lee, M. R. F., and Morris, P. (2008). Latent and active polyphenol oxidase (PPO) in red clover (Trifolium pratense) and use of a low PPO mutant to study the role of PPO in proteolysis reduction. *J. Agric. Food Chem.* 56 (8), 2817–2824. doi: 10.1021/jf0726177

Xie, Y., Lv, Y., Jia, L., Zheng, L., Li, Y., Zhu, M., et al. (2023). Plastid-localized amino acid metabolism coordinates rice ammonium tolerance and nitrogen use efficiency. *Nat. Plants.* 9 (9), 1514–1529. doi: 10.1038/s41477-023-01494-x

Yi, C., Jiang, Y., Shi, J., Qu, H., Duan, X., Yang, B., et al. (2009). Effect of adenosine triphosphate on changes of fatty acids in harvested litchi fruit infected by Peronophythora litchii. *Postharvest Biol. Technol.* 54 (3), 159–164. doi: 10.1016/j.postharvbio.2009.06.008

Yun, Z., Li, T., Gao, H., Zhu, H., Gupta, V. K., Jiang, Y., et al. (2019). Integrated transcriptomic, proteomic, and metabolomics analysis reveals peel ripening of harvested banana under natural condition. *Biomolecules* 9, (5). doi: 10.3390/biom9050167

Zhao, P., Li, W., Zhen, C., Wang, K., Qin, Z., and Gao, H. (2021). Transcriptomic analysis of the effects of γ-aminobutyric acid treatment on browning and induced disease resistance in fresh-cut apples. *Postharvest Biol. Technol.* 181, 111686. doi: 10.1016/j.postharvbio.2021.111686

Zhou, Y., Chen, Z., He, M., Gao, H., Zhu, H., Yun, Z., et al. (2020b). Unveiling the complexity of the litchi transcriptome and pericarp browning by single-molecule long-read sequencing. *Postharvest Biol. Technol.* 168, 111252. doi: 10.1016/j.postharvbio.2020.111252

Zhou, H., Tian, M., Huang, W., Luo, S., Hu, H., Zhang, Y., et al. (2020a). Physiological and transcriptomic analysis of 'Whangkeumbae' pear core browning during low-temperature storage. *Gene Expression Patterns* 36, 119113. doi: 10.1016/j.gep.2020.119113

Zhu, L., Hu, W., Murtaza, A., Iqbal, A., Kong, M., Zhang, J., et al. (2023). Browning inhibition in fresh-cut Chinese water chestnut under high pressure CO2 treatment: Regulation of reactive oxygen species and membrane lipid metabolism. *Food Chem.* 427, 136586. doi: 10.1016/j.foodchem.2023.136586

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 05 March 2024
DOI 10.3389/fpls.2024.1336513

OPEN ACCESS

EDITED BY
Weicong Qi,
Jiangsu Academy of Agricultural Sciences (JAAS), China

REVIEWED BY
Jim Tokuhisa,
Virginia Tech, United States
Xiaofeng Su,
Chinese Academy of Agricultural Sciences, China

*CORRESPONDENCE
Janak R. Joshi
janak.joshi@montana.edu
Adam L. Heuberger
adam.heuberger@colostate.edu

RECEIVED 10 November 2023
ACCEPTED 12 February 2024
PUBLISHED 05 March 2024

CITATION
Joshi JR, Paudel D, Eddy E, Charkowski AO and Heuberger AL (2024) Plant necrotrophic bacterial disease resistance phenotypes, QTL, and metabolites identified through integrated genetic mapping and metabolomics in *Solanum* species.
Front. Plant Sci. 15:1336513.
doi: 10.3389/fpls.2024.1336513

Plant necrotrophic bacterial disease resistance phenotypes, QTL, and metabolites identified through integrated genetic mapping and metabolomics in *Solanum* species

Janak R. Joshi[1,2*], Dev Paudel[3], Ethan Eddy[2], Amy O. Charkowski[4] and Adam L. Heuberger[2*]

[1]Department of Plant Sciences & Plant Pathology, Montana State University, Bozeman, MT, United States, [2]Department of Horticulture & Landscape Architecture, Colorado State University, Fort Collins, CO, United States, [3]Department of Environmental Horticulture, University of Florida Gulf Coast Research and Education Center, Wimauma, FL, United States, [4]Department of Agricultural Biology, Colorado State University, Fort Collins, CO, United States

Most food crops are susceptible to necrotrophic bacteria that cause rotting and wilting diseases in fleshy organs and foods. All varieties of cultivated potato (*Solanum tuberosum* L.) are susceptible to diseases caused by *Pectobacterium* species, but resistance has been demonstrated in wild potato relatives including *S. chacoense*. Previous studies demonstrated that resistance is in part mediated by antivirulence activity of phytochemicals in stems and tubers. Little is known about the genetic basis of antivirulence traits, and the potential for inheritance and introgression into cultivated potato is unclear. Here, the metabolites and genetic loci associated with antivirulence traits in *S. chacoense* were elucidated by screening a sequenced *S. tuberosum* x *S. chacoense* recombinant inbred line (RIL) population for antivirulence traits of its metabolite extracts. Metabolite extracts from the RILs exhibited a quantitative distribution for two antivirulence traits that were positively correlated: quorum sensing inhibition and exo-protease inhibition, with some evidence of transgressive segregation, supporting the role of multiple loci and metabolites regulating these resistance-associated systems. Metabolomics was performed on the highly resistant and susceptible RILs that revealed 30 metabolites associated with resistance, including several alkaloids and terpenes. Specifically, several prenylated metabolites were more abundant in resistant RILs. We constructed a high-density linkage map with 795 SNPs mapped to 12 linkage groups, spanning a length of 1,507 cM and a density of 1 marker per 1.89 cM. Genetic mapping of the antivirulence and metabolite data identified five quantitative trait loci (QTLs) related to quorum sensing inhibition that explained 8-28% of the phenotypic variation and two QTLs for protease activity inhibition that explained 14-19% of the phenotypic variation. Several candidate genes including alkaloid, and secondary metabolite biosynthesis that are related to disease resistance were identified within these QTLs. Taken together, these data support that

quorum sensing inhibition and exo-protease inhibition assays may serve as breeding targets to improve resistance to nectrotrophic bacterial pathogens in potato and other plants. The identified candidate genes and metabolites can be utilized in marker assisted selection and genomic selection to improve soft- rot and blackleg disease resistance.

KEYWORDS

potato, soft rot and blackleg resistance, virulence screening, QTL, metabolomics, genetic mapping, pectobacterium

1 Introduction

Many cultivated plants are susceptible to necrotrophic bacterial pathogens that infect stems and nutrient-rich organs. These bacteria can survive in the natural environment in soil, alternative hosts, irrigation water, and plant debris. Bacteria including *Pectobacterium* and *Dickeya* spp., enter plants through wounds or natural openings and these pathogens may stay latent until favorable conditions occur, at which time the bacteria shift from latent to virulent states. This shift is mediated by acyl homoserine lactone mediated quorum sensing (Põllumaa et al., 2012). During pathogenesis, these bacteria colonize interior parts of the host plant and synthesize exo-enzymes, including pectate lyases, polygalacturonases, exo-proteases, and cellulases, that break down plant cell walls to obtain nutrients from the plant cell walls and from within plant cells (Davidsson et al., 2013). This results in a total collapse of the host tissue leading to fleshy organ decay and stem necrosis that cause major pre- and post-harvest losses. These necrotrophic pathogens are especially destructive due to their ability to remain latent, and infect a wide range of hosts (Ma et al., 2007). Notable, effective management strategies for these diseases are still lacking. Sanitation, quarantine, and exclusion practices are often ineffective as *Pectobacterium* and *Dickeya* spp. are prevalent in the environment. Moreover chemical management options can be phytotoxic, are not systemic, further limiting their efficacy. Consequently, the development of resistant cultivars is critical for improved disease management (Vyska et al., 2016).

Potato (*Solanum tuberosum* L.) is an integral part of the world's agriculture ranking as third most important food crop. Potato has been cultivated for more than 8,000 years and during this time the crop has been continuously selected for many agronomic and nutritional traits through breeding. Despite these efforts, including disease-resistance breeding cultivated, potatoes remain vulnerable to necrotrophic bacterial pathogens like *Pectobacterium* and *Dickeya* spp. (known to cause soft rot, aerial stem rot, and blackleg in the potatoes).

In contrast, several wild *Solanum* relatives of cultivated potato are resistant to these diseases. Recent phenotyping studies support several independent mechanisms of resistance in these related species. For example, some wild potatoes have a rapid form of wound healing correlated with soft rot resistance (Chung et al., 2017). We recently characterized a different resistance mechanism that is driven by antimicrobial metabolites and proteins extracted from wild potato stems and tubers. These molecules, found in the wild potato *S. chacoense*, may inhibit virulence of *Pectobacterium* by reducing exo-enzyme activity, swimming motility, and quorum sensing (Joshi et al., 2021b, 2022). In these studies, chemical extracts were isolated from a single *S. chacoense* line M6. However, the genetics of *S. chacoense* that control these resistance phenotypes remain uncharacterized, as well as the plant metabolic processes that enable this type of biochemical resistance is understudied.

While cultivated potato shares many morphological and physiological traits with its wild diploid relatives, the tetraploid genetics and excessive heterozygosity of *S. tuberosum* makes it extremely difficult to map loci associated with disease resistance and to introgress traits from wild species. For necrotrophic bacteria, wild species do exhibit some variation in quantitative measures of resistance and the germplasm tends to be highly heterozygous (McCauley, 2021; Lebecka et al., 2021b). Subsequent genetic mapping within wild potato species shows that many necrotrophic bacterial resistance traits are quantitative (Jansky et al., 2014; Chung et al., 2017; Lebecka et al., 2021a; Ma et al., 2022), showing that bacterial disease resistance can be improved via introgression breeding. Currently, the potato industry is experiencing a revolutionary shift in breeding towards diploid genetics, and double monoploid lines (DM) are being used for research (Xu et al., 2011; Jansky et al., 2016; Bethke et al., 2022). These DM lines are still susceptible to pathogens, but they can be readily crossed with diploid wild potato species as an intermediate step in introgression breeding. While these genetic mapping populations tend to be small due to poor tuberization and fruit formation attributed to the mixing of very diverse genetics (*i.e.*, wild x domesticated species crosses), they can still be studied to understand the distribution of resistance traits, their qualitative or quantitative nature, and to begin to understand the integrative relationship of *Solanum* genes, metabolites, and resistance traits. Some studies have performed quantitative trait loci (QTL) analysis of DM x wild diploid *Solanum* populations and identified soft rot resistance across multiple chromosomes (Zimnoch-Guzowska et al.,

2000; Lebecka et al., 2021a), although the links between these QTL and biochemical resistance traits such as with *S. chacoense* are unknown.

Here, we utilized a recombinant inbred line (RIL) population derived from a cross of *S. chacoense* M6 x *S. tuberosum* DM1 and evaluated the disease resistance performance of the population. This population is segregated for tuberization, tuber morphology, and stem thickness leading to no tubers, degenerative tubers, and thin stem, making typical virulence assays unreliable or impossible. Previously, we identified metabolite-based resistance in *S. chacoense* M6 that directly affects virulence of bacteria i.e quorum sensing and exo-protease (Joshi et al., 2021a). Therefore traits of interest in this study were the ability of metabolites from RIL population to inhibit quorum sensing (QS-I) and exo-protease activity (EP-I). Our investigation focused on determining whether QS-I and EP-I are quantitative or qualitative traits. Additionally, we explored the inheritance pattern of these traits in a biparental mapping population. Finally, we performed a non-targeted metabolomic analyses on the parents alongside transgressive segregants for the QS-I and EP-I traits to identify components of the potato metabolome that co-vary with the QS-I and EP-I resistance traits.

2 Materials and methods

2.1 Plant materials, bacterial strains, and chemicals

We generated a recombinant inbred line (RIL) population of potato by crossing *S. tuberosum* DM1-3 (as a female) with *S. chacoense* M6 (male) (McCauley, 2021; Jansky et al., 2024). For this population, a single F_1 plant was self-pollinated to generate a large F_2 population. Then the fertile individuals in this population were self-pollinated for 5 or 6 (F_5/F_6) generations to develop a potato RIL population consisting of approximately 100 inbred lines. Plants were obtained from the Jansky group and tubers from the F_5/F_6 RILs were planted in a greenhouse at the Plant Growth Facilities at Colorado State University, U.S.A. The temperature was set to 24°C and 18°C for day and night cycles respectively with a 16 h day length. Plants were grown in ProMix Bx General Purpose mix, fertilized with Osmocote Plus 15-9-12 (Scotts-MiracleGro, U.S.A.), and irrigated to saturation every other day until used for assays. Aphids and other pests were managed with Botaniguard ES and Molt-X (BioWorks, U.S.A.), Entrust SC (Corteva, U.S.A.), Distance IGR (Valent Biosicences, U.S.A.), Judo and Azatin (OHP Inc., U.S.A.), Avid 0.15EC (Syngenta, U.S.A.), and Compass (Bayer, U.S.A.) according to the manufacturer's recommendations. Under greenhouse conditions, approximately half of the RILs had deleterious phenotypes including poor growth, no tuber formation, and self-degeneration of tubers for multiplying and continuing the population. Therefore, only 53 RILs were considered suitable for comparative phenotypic analysis. LC-MS-grade water, analytical-grade methanol, acetonitrile (ACN), and hydrochloric acid were purchased from Fisher Chemicals (Thermo Fisher Scientific, U.S.A.) for metabolite extractions. *P. brasiliense* strain Pb1692 was used for all resistance experiments in this study. Nutrient broth (NB), agar, and skim milk powder were purchased from Difco Laboratories (Thermo Fisher Scientific, U.S.A.). Bacteria were grown at 30°C under continuous shaking conditions.

2.2 Extraction of metabolites

Stems from six-week-old potato plants from each RIL and parent line were harvested, flash frozen in liquid nitrogen, and ground to a crude powder using mortar and pestle. We performed two biological replicates of extraction from two different plant lots. Ground samples were lyophilized for at least 12 h (HarvestRight, U.S.A.). The freeze-dried stem tissues were then ground to a fine powder using a coffee grinder. For metabolite extraction, 2 ml of 70:30 methanol/water (vol/vol) was added to a 100 mg of tissue, agitated for 2 h at 4°C using a vortex, and then sonicated for 5 min at room temperature. The mixture was then centrifuged at 6,000 × g for 20 min at 4°C and the supernatant was transferred to a new vial and dried under a stream of nitrogen gas (Organomation Associates Inc., U.S.A.). The dry matter (the metabolite extract) was weighed and resuspended in sterile dH_2O.

2.3 Quorum sensing and exo-protease activity assays

Bacterial cultures were grown overnight in NB at 30°C under continuous shaking at 220 rpm. The cultures were centrifuged, and the cells were resuspended into sterile water. This cell suspension was used as a source of inoculum to test bacterial responses to metabolite extracts. Ten-milligrams of each metabolite extract were resuspended in 1 ml of sterile water and the metabolite extract was then inoculated with ~10^6 CFU of bacteria (calculated using optical density, O.D., measurements – OD value of 1 at 600 nm equivalent to 10^9 CFU). The metabolite plus bacteria suspensions were incubated for 15 h at 30°C under continuous shaking at 220 rpm. No difference in bacterial multiplication was observed (evaluated using O.D. measurements) (data not shown). These cultures were then centrifuged (8,000 × g, 5 min) to separate the supernatant from the Pb1692 cells. The supernatant was filter sterilized and used to measure quorum sensing activity (QS-A) via acyl homoserine lactone (AHL, using the reporter strain *Chromobacterium violaceum* - CV026) and exo-protease activity (EP-A). The quantitative metrics of these systems were performed in plate assays as previously described (Joshi et al., 2021b). Here, activity (A) is a unit of measure specific to each assay: pigmentation area for QS, and milk powder degradation area for EP. QS and EP inhibitory (I) activity was calculated as: I (%) = (DM1-A – Line-A)/DM1-A x 100, with DM1 as a susceptible control.

2.4 Non-targeted metabolomics

The metabolite extracts were analyzed with two independent platforms run in positive and negative mode for a total of four metabolomics data sets: reverse phase ultra high-performance liquid chromatography mass spectrometry (UHPLC-MS), and

hydrophilic interaction chromatography HILIC-MS). Together these platforms capture a wide range of molecules with different chemistry. For UHPLC-MS, samples were injected in a randomized order using 1 µl injection volume into a Waters Acquity UPLC system. Separation was achieved using a Waters Acquity UPLC CSH Phenyl Hexyl column (1.7 µM, 1.0 x 100 mm, part number 186009478), using a gradient from solvent A (water, 0.1% formic acid, 2 mM ammonium hydroxide) to solvent B (acetonitrile, 0.1% formic acid). Injections were made in 99% A, held at 99% A for 1 min, ramped to 98% B over 12 minutes, held at 98% B for 3 minutes, and then returned to starting conditions over 0.05 minutes and allowed to re-equilibrate for 3.95 minutes, with a 200 µl/min constant flow rate. The column and samples were held at 65°C and 6°C, respectively. The column eluent was infused into a Waters Xevo G2-XS Q-TOF-MS with an electrospray source in negative and positive mode (as independent runs), scanning 50-1200 m/z at 0.1 seconds per scan, alternating between MS (6 V collision energy) and MSE mode (15-30 V ramp). Calibration was performed using sodium formate with 1 ppm mass accuracy. The capillary voltage was held at 700 V, source temperature at 150°C, and nitrogen desolvation temperature at 600°C with a flow rate of 1000 L/hr. Quality control was performed in both phases and modes by running quality control samples after every 4 or 5 experimental samples. For HILIC-MS, separation was achieved using a Waters Acquity Premier BEH Amide column with built-in fit guard column (1.7 µM, 2.1 x 100 mm, part number 186009508), using a gradient from solvent B (95% acetonitrile, 5% water, 0.1% formic acid, 10 mM ammonium hydroxide) to solvent A (water, 0.1% formic acid, 10mM ammonium hydroxide). Injections were made in 90% B, held at 90% B for 0.5 minutes, ramped to 25% A over 6.50 minutes, ramped to 50% A over 2 minutes, ramped to 85% A over one minute, held at 85% A for 0.50 minutes, returned to starting conditions over one minute, and allowed to re-equilibrate for 3.50 minutes, with a 500 µl/min constant flow rate. The column and samples were held at 30°C and 6°C, respectively. The column eluent was infused into a Waters Xevo G2-XS Q-TOF-MS with an electrospray source in positive and negative ionization mode (as independent runs), scanning 50-1200 m/z at 0.1 seconds per scan, alternating between MS (6 V collision energy) and MSE mode (15-30 V ramp). Calibration was performed using sodium formate with 1 ppm mass accuracy. The capillary voltage was held at 700 V (positive ionization mode) or 1800 V (negative ionization mode), source temperature at 150°C, and nitrogen desolvation temperature at 600°C with a flow rate of 1000 L/hr.

For processing, data were converted from Waters.RAW to.mzML using Proteowizard MSConvert version 3.0.20154, and peak detection, detection, alignment, grouping, retention time correction, and peak filling was performed using XCMS in R (Smith et al., 2006). Deconvolution and normalization were performed using the RAMClust package in R (Broeckling et al., 2014). Interpretation of spectra was done using the R package Interpret MS Spectrum (Lai et al., 2018). Spectral clusters with their spectral abundance were exported to MS Excel for further analysis. The information was used to match the external databases and tools such as MS-Finder (Tsugawa et al., 2016), SIRIUS (Dührkop et al., 2019), HMDB (Wishart et al., 2022) and MassBank of North America for annotation.

2.5 Construction of linkage map and QTL mapping

Genotyping by sequencing (Elshire et al., 2011) was done on the genomic DNA of the parents and the RIL digested with restriction enzyme as mentioned in Jansky et al (Jansky et al., 2024). Raw reads were trimmed for quality and adapter sequences were trimmed using Trimmomatic (Bolger et al., 2014). High quality reads were aligned to the potato genome using BWA-MEM and SNPs were called using Freebayes. SNPs were filtered with a minimum quality of 20. Filtered SNP data was utilized for genetic linkage analysis using OneMap R for inbred based populations (Margarido et al., 2007). The marker positions were based on the reference genome DM_1-3_516_R44 – v6.1 (Pham et al., 2020). QTL identification was done with composite interval mapping in Windows QTL Cartographer v 2.5 011 (Wang et al., 2012). Mean values of QS-A, QS-I, EP-A, and EP-I across 8 replicates were used for QTL analysis. The sliding window for all traits was 1 cM. A forward and backward stepwise regression method with a probability of 0.1 and a window size of 5 cM were utilized to determine cofactors. LOD thresholds for significance was selected by using the thousand-permutation test to each data set ($p \leq 0.05$) (Churchill and Doerge, 1994). For each QTL, the 95% confidence interval was calculated using a 2-LOD support interval (van Ooijen, 1992).

2.6 Mapping genes associated with metabolites/metabolic pathways

Metabolites strongly associated with disease resistance were traced for their metabolic pathways and chromosomal location of pathway genes using plant metabolic pathway database PotatoCyc (SolCyc Biochemical Pathways) (Fernandez-Pozo et al., 2015). Associated genes in metabolic pathways were mapped to chromosomes of *Solanum tuberosum* group phujera DM1-3 (v4.03, id52025) using a visualization tool Phenogram (https://visualization.ritchielab.org/ Ritchie Lab, University of Pennsylvania).

2.7 Statistical analysis

RIL phenotypic data were analyzed with GraphPad Prism v10 (Dotmatics, Boston, MA, U.S.A.) including histograms by frequency distribution binning and establishing nonlinear fit lines. One-way Analysis of Variance (ANOVA) tests were conducted to compare lines with Tukey *post-hoc* tests on QS-A and EP-A data with a p threshold of 0.05, and Spearman's rank correlation was performed in Graphpad Prism v10. Data was tested for normality using the Kolmogorov-Smirnov test with a p threshold of 0.05. QS-I and EP-I were calculated as [DM1 activity – Line activity]/[DM1 activity] x 100%. Metabolomics data was analyzed using SIMCA v17.0 (Sartorius AG, Gottingen, Germany). Orthogonal Partial Least Square (OPLS) models were developed using the two-way Orthogonal Partial Least Square (O2PLS) workflow (Bouhaddani et al., 2016) by regressing the QS-A and EP-A data (two independent y's) against the metabolite data (x) in a single model and establishing

two components to explain joint (Component 1) and unique variation (Component 2). All multivariate data was z scaled. Cross-validation was performed using the 1/7th leave out approach and reported as predictive power ranging between 0-100% (Q^2). Univariate metabolite z transformations were performed by comparing metabolite abundances of lines to the mean of lines classified as "susceptible" based on QS-A and EP-A values that were not different from DM1, with a significance threshold of $z > 1.96$ corresponding to $p < 0.05$ indicating an association to resistance. All graphs were illustrated using Graphpad Prism.

3 Results

3.1 Exo-enzyme and quorum sensing inhibition were distributed as quantitative traits in the M6 x DM1 RIL population

In the M6 x DM1 RIL population generated from a cross of *S. tuberosum* line DM1-3 (diploid potato) to M6 (a diploid potato inbred line) we evaluated two specific metabolite-based mechanisms correlated with disease resistance, quorum sensing inhibition (QS-I) and exo-protease inhibition (EP-I) and compared them among RILs. These assays are compatible with tissues of varying morphologies because the assay is normalized with metabolite extractions.

Screening of parental and progeny metabolites for QS and EP activity (QS-A and EP-A, Figure 1A) showed that for QS-A, M6 and DM1 had 0.09 and 0.51 activity units respectively, with the QS-I of M6 at 82% of DM1. Of the 56 RIL lines, 3 lines exhibited transgressive segregation in the resistant direction (i.e., lower QS-A than the parent M6, QS-I between 83-92%), although these were not statistically different than M6 (ANOVA, Tukey *Post-hoc* $p > 0.05$). Ten RILs had greater QS_a than DM1, however, most of them were equally distributed between DM1 and M6. For EP-A, M6 and DM1 were at 0.09 and 0.51, corresponding to an EP-I of M6 at 85% of DM1. Eight RILs exhibited transgressive segregation in the resistant direction for EP, (EP-I between 87-100%) while no lines varieties transgressed DM1 in EP-related susceptibility.

Both QS-A and EP-A values were normally distributed (Kolmogorov-Smirnov tests, $p > 0.10$) and support QS-I and EP-I as quantitative resistance traits (Figure 1B). The relationship between QS-I and EP-I was evaluated using correlation analysis of the activity data (Figure 1C). The two traits were moderately

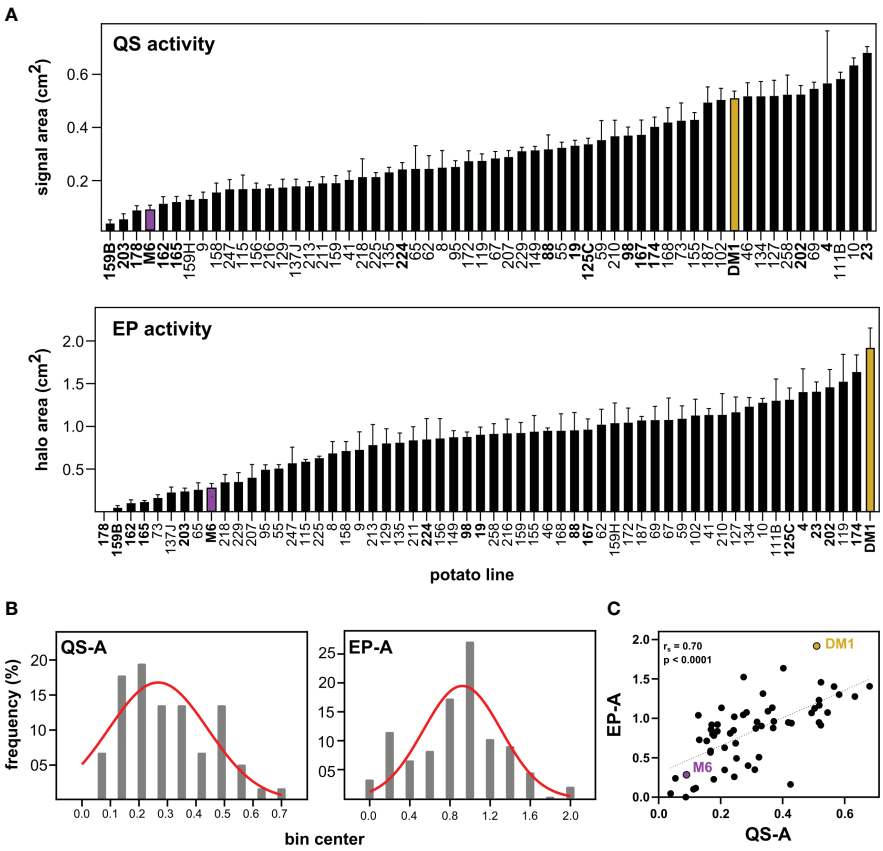


FIGURE 1
Effects of *Solanum* metabolite extracts on virulence-associated traits of *Pectobacterium*. **(A)** Metabolites were extracted from *S. chacoense* M6 (resistant), *S. tuberosum* DM1 (susceptible), or M6 x DM1 recombinant inbred lines (RILs) and tested for effects on quorum sensing activity (QS-A) and exo-protease activity (EP-A) of *Pectobacterium brasiliense* Pb1692. **(B)** Histogram distribution of activity data in the population fitted to a Gaussian curve. **(C)** Spearman's rank correlation of QS and EP activity, with regression line added as a visual aid. QS and EP activity data is presented as the mean ± SEM with n=8 replicates, with M6 (purple) and DM1 (brown) used as resistant and susceptible controls, respectively. RILs in bold text were chosen for metabolomics analysis.

correlated with Spearman's rank correlation $r_s = 0.70$ ($p < 0.0001$, $n = 52$), indicating that some, but not all metabolites affecting both QS and EP were either identical or were in the same extracts. For the remainder of this study, the RIL lines that were high in both QS-I and EP-I were denoted as "resistant lines", and all other lines were classified as "intermediate" or "susceptible".

3.2 QTL analysis identified genomic regions associated with disease resistance phenotypes

A genetic linkage map was constructed covering all the 12 chromosomes of potato (Figure S1). The genetic map had a total length of 1507 cM with an average marker interval of 0.72 cM. The largest linkage group was linkage group 1 which spanned 188.63 cM and contained 109 markers (Table S1). The length of each linkage group ranged from 85.04 cM to 188.63 cM and density ranged from 0.31 to 0.698 markers per cM. We identified 5 QTLs for QS-I (Table 1) that explained 8-28% of the variation. Two QTLs were identified in chromosome 5, while 1 QTL each was identified on chromosomes 2, 7, and 10 respectively (Figure 2). For protease activity, two QTLs were identified in chromosomes 3 and 5 that explained 14-19% of the phenotypic variation. Interestingly, the analysis of QS-I (vs. QS-A) identified two new QTLs including one on chromosome 11 that explained 15% of the phenotypic variation (Table 1). Similarly, for EP-I, we identified one QTL previously identified for EP-A and two new QTLs on chromosome 7 and 10 that explained 13-18% of the phenotypic variation. The 10,000 bp flanking region for each significant QTL was searched in the potato genome. For qQS-2-1 there were 54 genes, for qQS-5-1&2 there were 217 genes, for qQS-7-1 there were 39 genes, for qQS-10-1 there were 31 genes, for qPA-3-1 there were 249 genes, for qPA-5-1 there were 217 genes.

3.3 Metabolites associated with QS-I and EP-I

A comparative metabolomics experiment was performed to identify metabolites associated with QS-I and/or EP-I in RILs that inhibit these virulence traits. Because of the positive correlation between QS-I and EP-I, and that these traits are positively associated with disease resistance (Joshi et al., 2021b), a single subset of 15 lines was evaluated for metabolomics analysis with classifications of highly resistant, moderately resistant, or susceptible (5 lines from each group), along with the two parents. The four metabolomics platforms (UHPLC-MS, HILIC-MS, positive and negative mode) detected approximately 15,460 metabolites (Table S2). The 15,460 metabolites were regressed against QS-A or EP-A using orthogonal project to latent structures analysis, with both y variables included the same model (OPLS method, as an O2PLS design). The first model that included intermediates had a cumulative $R^2Y = 99\%$, however the model failed cross-validation with $Q^2 = 48\%$ (Figure 3A). A second OPLS model was constructed without data from lines without intermediate phenotypes achieved $R^2Y = 99\%$ and $Q^2 = 61\%$, supporting that metabolites profiles could sufficiently predict QS_a or EP_a (Figure 3B). In both models, two predictive components were generated that were linked to overall resistance (e.g. both QS_a and EP_a, joint variation, Component 1, ~95%) and a subset of resistance (e.g. either QS-A or EP-A, unique variation, Component 2, ~5%). The second model was subsequently analyzed for metabolites that met a correlated loadings threshold of 0.70 for Component 1, and these were determined to be metabolites associated with resistance. For Component 2, loadings values of > 0.10 were determined to be more associated with EP-A than QS-A and indicate trait bias, and loadings of < -0.10 indicate bias towards QS_a. Component 2 values in between -0.10 and 0.10 were considered associated with both traits.

TABLE 1 Summary of QTLs identified for quorum sensing and exo-protease activity/inhibition.

Trait	QTL code	Chromosome	Peak marker position (cM)	Peak LOD*	PVE (R^2)*
QS-A	qQS-A-2-1	2	93.11	2.61	0.10
QS-A	qQS-A-5-1	5	92.11	5.55	0.28
QS-A	qQS-A-5-2	5	98.11	5.39	0.25
QS-A	qQS-A-7-1	7	21.11	2.82	0.11
QS-A	qQS-A-10-1	10	60.48	2.58	0.09
QS-I	qQS-I-5-1	5	92.11	2.75	0.13
QS-I	qQS-I-7-1	7	19.51	3.75	0.16
QS-I	qQS-I-11-1	11	32.81	3.57	0.15
EP-A	qEP-A-3-1	3	122.49	3.76	0.19
EP-A	qEP-A-5-1	5	86.00	2.51	0.14
EP-I	qEP-I-3-2	3	124.51	5.48	0.26
EP-I	qEP-I-7-1	7	42.51	3.30	0.13
EP-I	qEP-I-10-1	10	77.21	4.24	0.18

*LOD, logarithm of the odds; PVE, variation explained.

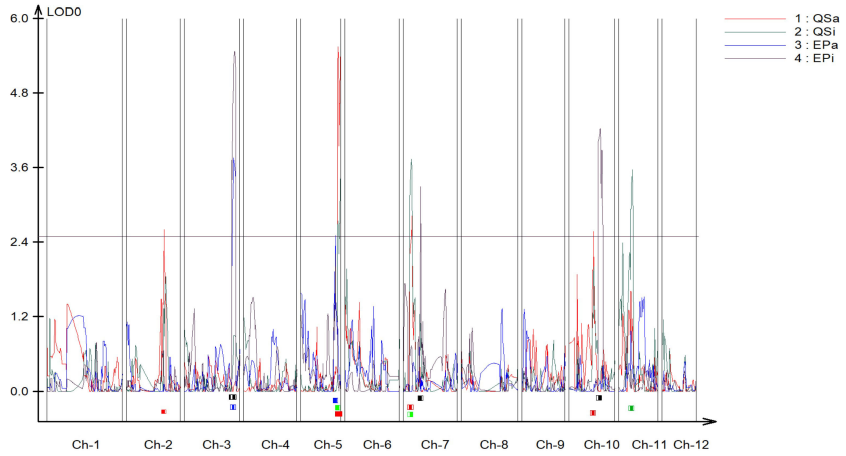


FIGURE 2
Quantitative trait locus (QTL) mapping of biochemical antivirulence traits in *S. tuberosum*. Resistance-associated traits included quorum sensing and exo-protease activity and inhibition data (QS-A, QS-I, EP-A, EP-I) each mapped independently using logarithm of the odds (LOD) scores as associations to genomic regions. Chr = linkage group (chromosome). QS-A (red): Quorum sensing activity, QS-I (green): % quorum sensing inhibition compared to susceptible DM1, EP-A (blue): Exo-protease activity, EP-I (black): % of exo-protease inhibition to susceptible DM1.

In total, 35 metabolites met the Component 1 threshold and were classified as associated with virulence inhibition. Of the 35 metabolites, the Component 2 analysis denoted that 13 metabolites were associated with both QS-I and EP-I, 10 metabolites with QS-I, and 12 metabolites with EP-I. The metabolite abundances were normalized using z transformation to analyze trends among each individual resistance and susceptible lines, with the intermediates added back to the analysis, and with DM1 set to 0 as the susceptible

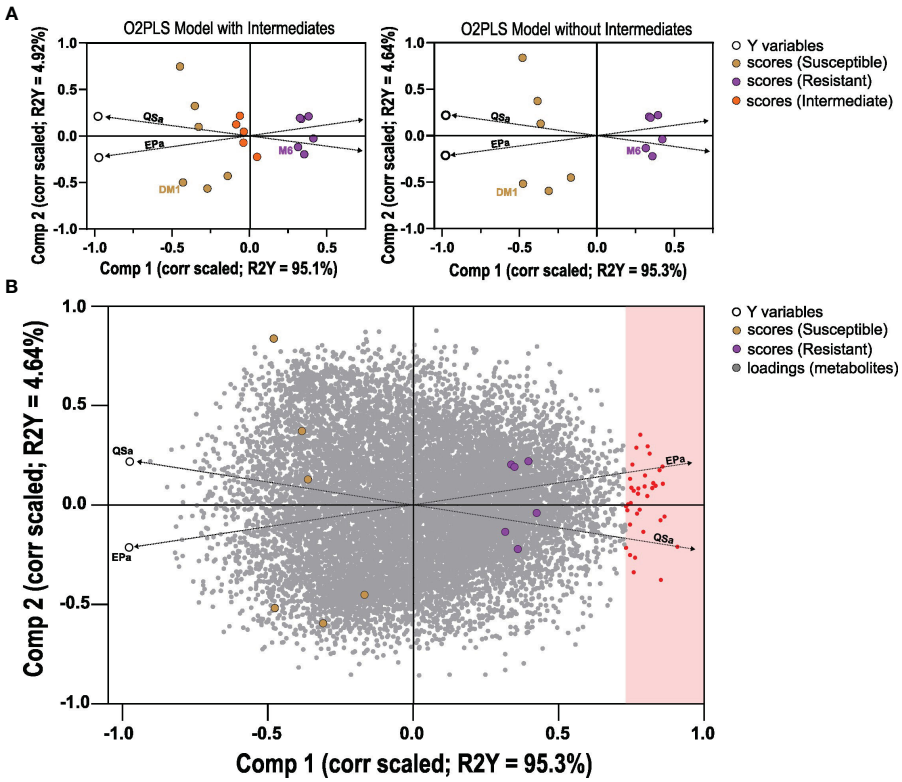


FIGURE 3
Orthogonal Projection to Latent Structures analysis of metabolites detected in potato RILs. Data was generated based on 15,460 metabolites (x) regressed against QS-A and EP-A (y) data in two O2PLS models: with 15 lines that includes intermediates, or 10 lines of only resistant and susceptible. **(A)** O2PLS scores plots, correlation scaled for 15 (left) or 10 lines (right), with the data points for the two Y variables QS-A and EP-A noted as a black circle. **(B)** O2PLS scores and loadings plot model of the 10 lines (scores) with metabolite loadings shown (gray dots), correlation scaled. Metabolites that met the joint variation (Component 1) threshold of 0.70 were considered associated with resistance. Dashed lines were manually added to highlight the direction of the Y variable vectors.

10.3389/fpls.2024.1336513

control (Figures 4A–C). All metabolites were confirmed to have higher mean z scores in the resistant lines compared to susceptible, supporting the OPLS model successfully characterized metabolites associated with QS-I and EP-I in this population. Interestingly, no metabolites exhibited a clear trend of high levels in resistant lines, medium levels in intermediates, and low levels in susceptible, confirming the OPLS model with intermediates that failed cross-validation. Some metabolites were higher in intermediates than resistant (e.g. LCneg_C4997), but often the intermediates were grouped with the susceptible (e.g. LCneg_2364). Further, only a few metabolites exhibited a presence/absence type of pattern in the data, denoted by excessively high z scores (e.g. HLneg_C0661, HLneg_C0518, HLneg_C0207), with most metabolites existing in both the resistant and susceptible lines, but at higher levels in the resistant lines. Two metabolites were higher in the resistant RILs than DM1, except for the parent M6 that was equal to or below DM1 (HLneg_C0184, LCneg_C2835).

The metabolites were annotated based on interpretation of mass spectra and cross-listing spectral data with metabolite databases (Table 2). Of the 35 metabolites, 19 were annotated and 16 were classified as unknowns with the inability to match to databases, or the inability to determine elemental formulas from the spectral data. Two alkaloids that are rather notable in *Solanum* spp. were the glycoalkaloid solacauline (HLneg_C0195) and the kukoamine N1, N5,N10-Tris-trans-p-coumaroylspermine (LCneg_C4997), both in the QS-I/EP-I classification. Three metabolites were annotated as terpenes and all trended in the EP-I direction, with no terpenes being associated with only QS-I: piperochromenoic acid (LCneg_C6339), demissine (HLneg_C0207), and parasiloxanthin (HLpos_C1110). Additionally, 2-C-methyl-D-erythritol-2,4-cyclodiphosphate (LCneg_C3974) is a terpene precursor metabolite and was biased towards EP-I. Two compounds were annotated as terpene conjugates, being alternate classes with a prenyl group attached to the base structure: isopentyl gentiobioside (two metabolite signals, LCneg_C6893, LCneg_6963, EP-I) and N6-(delta2-isopentenyl)-adenosine 5'-diphosphate (LCneg_C3220, a putative prenylated purine, EP-I). A third prenylated flavonoid was identified (albanin H, LCposC4899), although this was linked to QS-I, as well as the non-prenylated polyphenolic feruloylquinic acid (HLneg_C0184).

3.4 Association of disease resistance related metabolites with genetic variants

Candidate resistance metabolites that were consistently higher across all tested resistant lines were annotated and EP-I assigned to specific metabolic pathways and aligned to the chromosomal locations of pathway genes (Figure 5). These prominent pathways include alkaloid biosynthesis, zeatin biosynthesis, methylerythritol phosphate pathway (non-MVA terpene synthesis), glutathione redox reactions, 5-O-caffeoylquinic acid (phenolic) biosynthesis, and prenyl transferases. Notably, genes associated with these pathways were distributed across 12 distinct potato chromosomes.

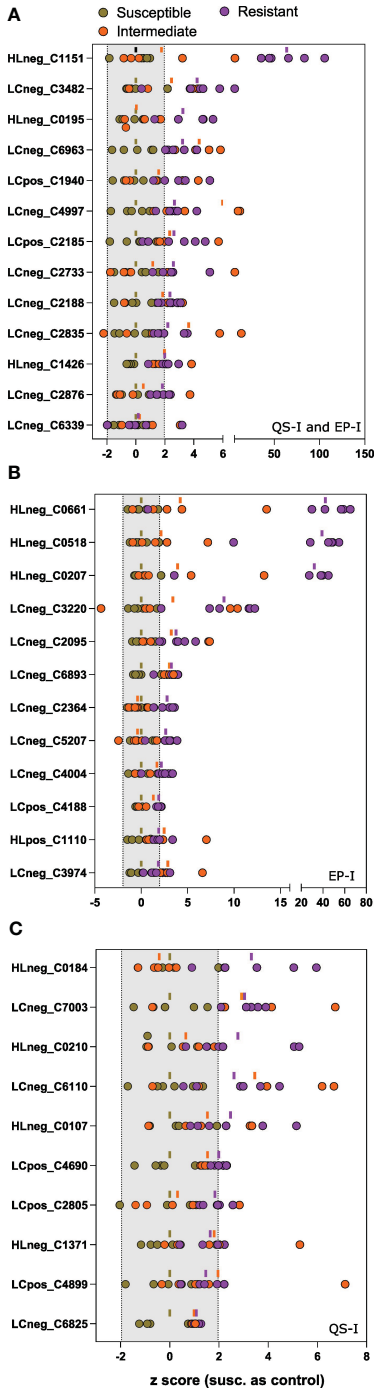


FIGURE 4
Univariate analysis of metabolites associated with inhibition of virulence traits. Z scores of each metabolite for each of the 15 lines (as dots) compared to the susceptible group as the control. Metabolites were determined to be associated with resistance based on an analysis of the OPLS loadings for **(A)** joint variation, QS-I and EP-I, or unique variation indicating trait bias for **(B)** EP-I and **(C)** QS-I. The dash above each z row indicates the mean z for the resistant (purple), intermediate (orange), or susceptible group (brown) for that metabolite. Z scores outside of the shaded region (-1.96 to 1.96) support statistical significance compared to the susceptible controls. Metabolites are reported as platform detected_compound C number.

10.3389/fpls.2024.1336513

TABLE 2 Metabolites associated with QS-I and EP-I in wild potato and resistant RILs.

Class	Sub-class	Metabolite	ID[a]	Mass[b]	error (ppm)	Corr[c]	Trait Bias[c]	Trait Corr[c]
Alkaloids	glycoalkaloids	solacauline	HLneg_C0195	M = 826.4706	1	0.77	-0.04	QS-I/EP-I
		demissine	HLneg_C0207	$[M+F]^-$= 1036.5593	10	0.85	0.18	EP-I
	kukoamines	N1,N5,N10-Tris-trans-p-coumaroylspermine	LCneg_C4997	M = 640.3067	30	0.83	0.10	QS-I/EP-I
Terpenes	intermediates	2-C-methyl-D-erythritol-2,4-cyclodiphosphate	LCneg_C3974	$[M+FA-H]^-$ = 320.9796	1	0.72	0.28	EP-I
	tetraterpenes	parasiloxanthin	HLpos_C1110	$[M+H]^+$ = 571.4609	17	0.77	0.29	EP-I
Saccharides	prenylated	isopentyl gentiobioside	LCneg_C6893	M = 412.1908	9	0.81	0.30	EP-I
		isopentyl gentiobioside	LCneg_C6963	M = 412.1908	9	0.86	0.11	EP-I
Phenolics	polyphenolics	feruloylquinic acid	HLneg_C0184	M = 368.1121	4	0.75*	-0.1	QS-I
	prenylated	albanin H	LCpos_C4899	$[M+2H]^{2+}$ = 421.1674	5	0.70	-0.34	QS-I
		piperochromenoic acid	LCneg_C6339	M = 340.2066	8	0.8*	0.05	QS-I/EP-I
Nucleotides	purines	diadenosine pentaphosphate	LCneg_C6110	$[M-H]^-$ = 914.9778	32	0.73	-0.21	QS-I
	prenylated	N6-(delta2-isopentenyl)-adenosine 5'-diphosphate	LCneg_C3220	$[M-H]^-$ = 491.038	51	0.81	0.26	EP-I
	pyrimidines	UDP-alpha-D-xylose	LCneg_C2876	M = 533.9952	65	.082	0.09	QS-I/EP-I
Lipids	glycerolipids	TG(62:0)	LCneg_C2188	$[M-H_20-H]^-$ = 969.9628	5	0.77	0.08	QS-I/EP-I
		TG(63:0)	LCpos_C1940	$[M+H+NH_4]^{2+}$ = 517.9981	15	0.75	0.09	QS-I/EP-I
		TG(58:0)	LCneg_C4004	M = 932.9485	37	0.75	0.13	EP-I
		DG-3-OH(40:5)	LCpos_C2185	$[M+Na]^+$ = 741.5279	0	0.76	0.07	QS-I/EP-I
		PE-NMe2(42:2)	HLneg_C0518	M = 855.6676	5	0.86	0.19	EP-I
		PE-NMe2(38:6)	LCpos_C2805	M = 791.5465	0	0.79	-0.13	QS-I
Unknown	phenolics	unknown phenolic	HLneg_C1151	M = 232.0754	1	0.86	-0.06	QS-I/EP-I
		unknown phenolic	HLneg_C1426	$[M-H]^-$ = 401.0705		0.85	-0.08	QS-I/EP-I
		unknown flavonoid glycoside	LCpos_C4690	$[M + K]^+$ = 821.1921	2	0.85	-0.38	QS-I
	glycosides	unknown glycoside	HLneg_C0210	M = 396.8958	9	0.76	-0.26	QS-I
	lipids	unknown lipid	LCneg_C3482	$[M-H]^-$ = 162.8387		0.77	0.06	QS-I/EP-I
	unknown	unknown	HLneg_C0107	M = 350.8118	5	0.75	-0.25	QS-I
		unknown	LCneg_C2364	M = 660.9885	2	0.83	0.11	EP-I
		unknown	LCneg_C2733	M = 806.9983	8	0.80	0.09	QS-I/EP-I
		unknown	HLneg_C1371	$[M-H]^-$ = 280.9837		0.76	-0.34	QS-I
		unknown	HLneg_C0661	$[M-H]^-$ = 448.8301		0.80	0.15	EP-I
		unknown	LCpos_C4188	$[M+H]^+$ = 515.036		0.75	0.20	EP-I
		unknown	LCneg_C2835	$[M-H]^-$ = 633.9719		0.78*	-0.02	QS-I/EP-I
		unknown	LCneg_C6825	$[M-H]^-$ = 644.9649		0.69	-0.32	QS-I

(Continued)

TABLE 2 Continued

Class	Sub-class	Metabolite	ID[a]	Mass[b]	error (ppm)	Corr[c]	Trait Bias[c]	Trait Corr[c]
		unknown	LCneg_C7003	$[M-H]^-$ = 703.9523		0.91	-0.21	QS-I
		unknown	LCneg_C2095	$[M-H]^-$ = 769.9478		0.78	0.35	EP-I
		unknown	LCneg_C5207	$[M-H]^-$ = 937.9673		0.72	0.3	EP-I

a: metabolomics platform (HILIC or reverse phase LC; positive or negative mode); b: inferred mass M based on InterpretMS algorithm or parent ion used for annotation, with unknowns reporting one assumed ion in the spectrum (e.g.$[M-H]^-$); c: corr = O2PLS Component 1 loading representing positive correlation value to QS-I/EP-I traits, known as joint predictive variation; trait bias = O2PLS Component 2 loading of bias correlation to QS-I (negative) or EP-I (positive) traits, with bias thresholds set at > or < 0.10; * = DM1 z score greater than M6.

A comparative analysis revealed overlaps between these genomic regions and the identified QTL regions, which encompass genes related to alkaloid synthesis, phenolic biosynthesis, peroxidases, and acylsugar acyltransferase.

4 Discussion

The resistance to necrotrophic bacterial diseases observed in wild *Solanum* species is multifactorial, with biomolecules playing a direct and indirect role in combating the disease. For chemical resistance, *S. chacoense* M6 has several sub-classes of compounds that work together to collectively reduce bacterial virulence and slow or completely prevent the development of disease. Therefore, there are several anti-virulence properties of plant molecules found on wild *Solanum*. We describe two sub-classes of these as antivirulence proteins (Joshi et al., 2022) and antivirulence metabolites (Joshi et al., 2021b). For proteins, these have been classified as protease inhibitors that affect bacterial exo-protease activity, motility, and cell morphology, but not quorum sensing. For metabolites, these molecules affect quorum sensing, motility, and exo-protease inhibition, but not morphology. This supports that the wild potato *S. chacoense* contains a mixture of molecules individually affecting sub-components of bacterial virulence, possibly with a synergy that culminates in broad phenotypic measures of host resistance, or less decay of plant tissues.

The phenomenon that multiple molecular compounds interact with multiple processes is consistent with the observation that resistance in *S. chacoense* is multi-genic (Zimnoch-Guzowska et al., 2000; Chung et al., 2017). Recently, RILs were developed by crossing DM1×M6, that allows use of genetic and high throughput metabolomic tools to explore soft rot and blackleg resistance in M6 potato (McCauley, 2021; Jansky et al., 2024). The RILs from the DM1×M6 were impossible to effectively compare for resistance in later generations (e.g., F_5/F_6) due to high phenotypic variability that may affect virulence assays, such as varying stem thickness, self-degenerating tubers, or inability to tuberize. In this study, we worked with a sub-set of the RIL population that allowed the study of metabolites, genes, and anti-virulence traits correlated with resistance. While the power of genetic associations is relatively low compared to a typical plant genetics study, we were still able to describe the distribution of resistance traits in this population.

Specifically, the two antivirulence traits were explored: the ability for a metabolite extract to inhibit quorum sensing or exo-protease activity (QS-I, EP-I). Importantly, these are quantitative

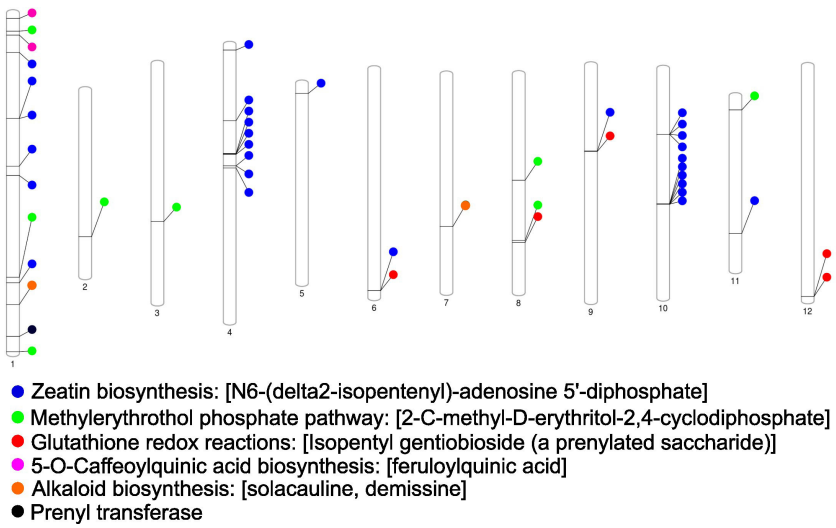


FIGURE 5
Genetic map of the chromosomes from the *Solanum tuberosum* group phujera DM1-3 (v4.03, id52025). The location of genes associated with soft rot and blackleg resistant metabolites is shown across potato chromosomes with individual color denoting specified biochemical pathways.

assays and therefore inhibition can be inferred as a percent reduction compared to a control, in this experiment with DM1 as the control. The distribution of the QS and EP activity data was normal among the RILs, with some evidence of potential transgressive segregation (Figure 1B). This supports that QS-I and EP-I are quantitatively inherited traits and therefore linked to multiple genes and/or metabolites, and that QS-I and EP-I are new traits that can be bred for in plants. Further, QS-I and EP-I were moderately positively correlated, indicating some metabolites or regulators impact both processes, but there are also metabolites that uniquely affect these sub-components of virulence. This is consistent with our previous observation that protease inhibitors affected some, but not all virulence processes in this system (Joshi et al., 2022).

The metabolomics analysis was performed on a subset of the RIL population. While screening the full RIL population would be a stronger design to associate metabolites with QS-I and EP-I, we were still able to create a predictive model that provided significant hits of metabolites associated with resistance. The OPLS method was used for its ability to work with mass spectrometry (x) data, and the normally distributed phenotypic (y) data as shown in Figure 1A. Further, we could integrate QS-I and EP-I concepts into a single model (similar to O2PLS (Bouhaddani et al., 2016)), in which the data could be divided into two subsets: joint variation (metabolites affecting both QS-I and EP-I), and individual variation (metabolites affecting either QS-I or EP-I, a virulence trait bias components). The output was a two-dimensional space that provided a score to each metabolite for each of the four options (Figure 3B). Interestingly, the model resulted in approximately 95% of the variation being explained by the joint component, and 5% for the trait bias component. This is inconsistent with the Spearman rank correlation data, which would support more variation on the trait bias component ($r_s = 0.70$), although the metabolomics data may be skewed with the 15 lines chosen for their effectiveness for both the QS-I and EP-I traits. This may also be due to the lack of intermediates in the final OPLS model; intermediates were a major component to decipher the distribution of QS-I or EP-I phenotypes in this population (Figure 1), but metabolomics analysis that included these intermediates resulted in poor models (Figure 3). One explanation is that the difficulty in the metabolomics analysis to decipher very minor differences in metabolites may lead to major differences in phenotypes in the intermediates. Therefore, these metabolomics data are most confident in the ability to discriminate the highly resistant RILs from the highly susceptible RILs, rather than subtle differences that occur in the intermediate RILs.

The metabolomics data showed that the highly resistant RILs were associated with a set of 35 metabolites that included a mixture of lipids, alkaloids, terpenes, phenolics, and several unknowns. Most lipids were determined to be membrane glycolipids or storage lipids. As lipid profiles are inherited, this class of compounds is expected to be indirectly related to resistance, where this profile is inherited from M6 to the resistant RILs, however not associated with resistance. Of the 35 resistance-related metabolites, 16 were 'known unknowns' (MSI levels 3-4 (Sumner et al., 2007)), and these are expected to occur in an experimental design that includes a wild plant species and progeny crossed with diploid potato. Regardless, this report includes mass spectral data for these compounds to continue to be tracked in subsequent studies.

Several compounds that were associated with resistance have known or predicted roles in resistance and virulence, and specifically for QS-I or EP-I. Three alkaloids were found to be associated with resistance, solacauline, demissine, and N1,N5,N10-Tris-trans-p-coumaroylspermine. Solacauline and demissine are steroidal alkaloids, a class of compounds largely unique to *Solanum* and *Liliaceae*, of which approximately 300 structures have been reported (Morillo et al., 2020). Steroidal alkaloids are nitrogenous sterols (a solanidane group) that are usually conjugated to 2-4 saccharides, with the saccharide component being critical to their bioactivity (Delbrouck et al., 2023). These compounds provide protection by disrupting cell membranes of pests and pathogens (Niño et al., 2009). The two main potato glycoalkaloids, chaconine and solanine, were not associated with resistance in this experiment. Solacauline and demissine have a similar solanidane backbone as chaconine and solanine, but both have unique saccharide components. Solacauline has a linear trisaccharide chain, compared to branched chain trisaccharides (chacotriose and solatriose) found in most glycoalkaloids (Shakya and Navarre, 2008). Glycoalkaloids may affect bacterial membranes (Yan et al., 2021), however given *S. chacoense* M6 extracts do not exhibit bactericidal activity (Joshi et al., 2021b), it is more likely to interact with bacterial proteins involved in virulence, and this interaction may be improved with the linear chain saccharide component. Further, demissine has a branched tetrasaccharide component, and this was biased towards EP-I, further supporting that the saccharide component of glycoalkaloids may be critical in determining their direct effects on bacterial processes and overall impact on resistance to these pathogens.

A second major class of metabolites associated with QS-I and EP-I are terpenes and terpene conjugates of other specialized metabolites. Terpene synthesis is linked to glycoalkaloid synthesis via the synthesis of sterols (glycoalkaloids are based on triterpenes). The terpene biosynthesis precursor 2-C-methyl-D-erythritol-2,4-cyclodiphosphate was associated with resistance in the RIL population, although biased toward EP-I. In these data, the common terpene-related trend was the prenylation of specialized metabolites. Prenylation is a modification to secondary metabolites that results in changes to the compounds solubility, and usually enhances function including antibacterial activity and enzyme inhibition, for example docking of the prenyl group in protein active sites (Botta et al., 2005; Yazaki et al., 2009; Chen et al., 2014; Wang et al., 2017). Therefore, the presence of prenyl side chains on specialized, exogenous plant molecules could directly affect bacterial proteins that regulate virulence pathways. The M6 resistance related metabolites included two prenylated phenolics (albanin H, piperochromenoic acid), a prenylated saccharide (isopentyl gentiobioside), and a prenylated purine (N6-(delta2-isopentenyl)-adenosine 5'-diphosphate, putative annotation). Gentiobiose is a saccharide common in unripe (green) tomato fruit (Dumville and Fry, 2003) and would therefore be expected in other *Solanum* spp. Prenylated gentiobiose has further been detected in tomato cell culture (De Rosa et al., 1996), and isopentyl gentiobioside was

previously found to be associated with resistance in *S. chacoense* (Joshi et al., 2021b). For effects on QS, some bacteria such as *Bacillus* have prenylated pheromones (e.g. ComX), with the prenyl group being essential to protein-binding function and overall diversity (Ansaldi et al., 2002; Okada et al., 2017). Prenylated phenolics have demonstrated QS-I properties (Paguigan et al., 2019), and in other metabolite-protein interactions, prenylated forms have stronger activity (Osorio et al., 2021). While *Pectobacterium* AHL pheromones are acylated and not prenylated, the side chains have similar molecular sizes, shapes, and lipophilic chemical properties to support this system a potential resistance mechanism by *S. chacoense*. Further, as with prenylated proteins, the prenylation of water-soluble metabolites such as saccharides and phenolics may focus their activity on cell membranes, where significant regulation of QS occurs (Joshi et al., 2021b).

Several associations between QS-I and EP-I resistance traits and individual antivirulence metabolites and metabolic pathways in *S.* chacoense were observed with genetic mapping. The overall SNP-QTL mapping approach was conducted using exo-protease activity inhibition (EP-A, EP-I), and quorum sensing inhibition (QS-A, QS-I) as a phenotype to pinpoint markers associated with these traits. Notably, protease inhibition demonstrated associations with markers located on chromosome 3, 5, 7, and 10. Similarly, AHL inhibition was linked to markers found on chromosome 2, 5, 7, 10, and 11. These markers encompass a wide range of regions within the potato chromosome, many of which are correlated with disease resistance metabolites and associated pathways, including acylsugar acyltransferase, alkaloid biosynthesis, phenolic biosynthesis, and peroxidases. Our findings are in line with recent findings where acylsugar metabolism, alkaloid biosynthesis, and phenolic biosynthesis is shown to be associated with plant-microbe interaction and disease resistance (Mandal et al., 2020; Song et al., 2023; Yuan et al., 2023). In summary, the integration of two robust approaches, genetic analysis, and metabolomics, has shed light on the biochemical pathways associated with soft rot and blackleg diseases resistance.

Data availability statement

The raw datasets presented in this study has been deposited in the online repository Metabolights: https://www.ebi.ac.uk/metabolights, accession number MTBLS8906. Data will be made available after review by the resource.

Author contributions

JJ: Writing – original draft, Writing – review & editing. DP: Writing – review & editing. AC: Writing – review & editing. EE: Writing – review & editing. AH: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the United States Department of Agriculture grant numbers 2017-51181-26827, COL00761, and the Colorado Potato Administrative Committee.

Acknowledgments

We thank S. Jansky for useful discussion, guidance, and generating resources critical to this 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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1336513/full#supplementary-material

References

Ansaldi, M., Marolt, D., Stebe, T., Mandic-Mulec, I., and Dubnau, D. (2002). Specific activation of the *Bacillus* quorum-sensing systems by isoprenylated pheromone variants. *Mol. Microbiol.* 44, 1561–1573. doi: 10.1046/j.1365-2958.2002.02977.x

Bethke, P. C., Halterman, D. A., Francis, D. M., Jiang, J., Douches, D. S., Charkowski, A. O., et al. (2022). Diploid potatoes as a catalyst for change in the potato industry. *Am. J. Potato Res.* 99, 337–357. doi: 10.1007/s12230-022-09888-x

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

Botta, B., Vitali, A., Menendez, P., Misiti, D., and Delle Monache, G. (2005). Prenylated flavonoids: pharmacology and biotechnology. *Curr. Med. Chem.* 12, 717–739. doi: 10.2174/0929867053202241

Bouhaddani, S. E., Houwing-Duistermaat, J., Salo, P., Perola, M., Jongbloed, G., and Uh, H.-W. (2016). Evaluation of O2PLS in omics data integration. *BMC Bioinf.* 17, S11. doi: 10.1186/s12859-015-0854-z

Broeckling, C. D., Afsar, F. A., Neumann, S., Ben-Hur, A., and Prenni, J. E. (2014). RAMClust: A novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal. Chem.* 86, 6812–6817. doi: 10.1021/ac501530d

Chen, X., Mukwaya, E., Wong, M.-S., and Zhang, Y. (2014). A systematic review on biological activities of prenylated flavonoids. *Pharm. Biol.* 52, 655–660. doi: 10.3109/13880209.2013.853809

Chung, Y. S., Kim, C., and Jansky, S. (2017). New source of bacterial soft rot resistance in wild potato (*Solanum chacoense*) tubers. *Genet. Resour. Crop Evol.* 64, 1963–1969. doi: 10.1007/s10722-017-0487-3

Churchill, G. A., and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963–971. doi: 10.1093/genetics/138.3.963

Davidsson, P. R., Kariola, T., Niemi, O., and Palva, E. T. (2013). Pathogenicity of and plant immunity to soft rot pectobacteria. *Front. Plant Sci.* 4, 191–191. doi: 10.3389/fpls.2013.00191

Delbrouck, J. A., Desgagné, M., Comeau, C., Bouarab, K., Malouin, F., and Boudreault, P. L. (2023). The therapeutic value of *Solanum* steroidal (Glyco) alkaloids: a 10-year comprehensive review. *Molecules* 28, 4957. doi: 10.3390/molecules28134957

De Rosa, S., De Giulio, A., and Tommonaro, G. (1996). Aliphatic and aromatic glycosides from the cell cultures of *Lycopersicon esculentum*. *Phytochemistry* 42, 1031–1034. doi: 10.1016/0031-9422(96)00083-0

Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A. A., Melnik, A. V., Meusel, M., et al. (2019). SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* 16, 299–302. doi: 10.1038/s41592-019-0344-8

Dumville, J. C., and Fry, S. C. (2003). Gentiobiose: a novel oligosaccharin in ripening tomato fruit. *Planta* 216, 484–495. doi: 10.1007/s00425-002-0869-3

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One* 6, e19379. doi: 10.1371/journal.pone.0019379

Fernandez-Pozo, N., Menda, N., Edwards, J. D., Saha, S., Tecle, I. Y., Strickler, S. R., et al. (2015). The Sol Genomics Network (SGN)–from genotype to phenotype to breeding. *Nucleic Acids Res.* 43, D1036–D1041. doi: 10.1093/nar/gku1195

Jansky, S. H., Charkowski, A. O., Douches, D. S., Gusmini, G., Richael, C., Bethke, P. C., et al. (2016). Reinventing potato as a diploid inbred line–based crop. *Crop Sci.* 56, 1412–1422. doi: 10.2135/cropsci2015.12.0740

Jansky, S. H., Chung, Y. S., and Kittipadukal, P. (2014). M6: A diploid potato inbred line for use in breeding and genetics research. *J. Plant Registrations* 8, 195–199. doi: 10.3198/jpr2013.05.0024crg

Jansky, S. H., Hamernik, A., and Endelman, J. B. (2024). Interspecific recombinant inbred lines for genetic mapping in potato. *Am. J. Potato Res.*

Joshi, J. R., Brown, K., Charkowski, A. O., and Heuberger, A. L. (2022). Protease inhibitors from *Solanum chacoense* inhibit *Pectobacterium* virulence by reducing bacterial protease activity and motility. *Mol. Plant-Microbe Interactions®* 35, 825–834. doi: 10.1094/MPMI-04-22-0072-R

Joshi, J. R., Khazanov, N., Charkowski, A., Faigenboim, A., Senderowitz, H., and Yedidia, I. (2021a). Interkingdom signaling interference: the effect of plant-derived small molecules on quorum sensing in plant-pathogenic bacteria. *Annu. Rev. Phytopathol.* 59, 153–190. doi: 10.1146/annurev-phyto-020620-095740

Joshi, J. R., Yao, L., Charkowski, A. O., and Heuberger, A. L. (2021b). Metabolites from wild potato inhibit virulence factors of the soft rot and blackleg pathogen *Pectobacterium brasiliense*. *Mol. Plant-Microbe Interactions®* 34, 100–109. doi: 10.1094/MPMI-08-20-0224-R

Lai, Z., Tsugawa, H., Wohlgemuth, G., Mehta, S., Mueller, M., Zheng, Y., et al. (2018). Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nat. Methods* 15, 53–56. doi: 10.1038/nmeth.4512

Lebecka, R., Śliwka, J., Grupa-Urbańska, A., Szajko, K., and Marczewski, W. (2021a). QTLs for potato tuber resistance to *Dickeya solani* are located on chromosomes II and IV. *Plant Pathol.* 70, 1745–1756. doi: 10.1111/ppa.13407

Lebecka, R., Wasilewicz-Flis, I., and Mańkowski, D. (2021b). Diploid potato germplasm with resistance to *Dickeya solani*. *Potato Res.* 64, 375–385. doi: 10.1007/s11540-020-09482-w

Ma, B., Hibbing, M. E., Kim, H. S., Reedy, R. M., Yedidia, I., Breuer, J., et al. (2007). Host range and molecular phylogenies of the soft rot enterobacterial genera *Pectobacterium* and *Dickeya*. *Phytopathology* 97, 1150–1163. doi: 10.1094/PHYTO-97-9-1150

Ma, X., Lofton, L., Bamberg, J., and Swingle, B. (2022). Identification of resistance to *Dickeya dianthicola* soft rot in *Solanum microdontum*. *Am. J. Potato Res.* 99, 58–68. doi: 10.1007/s12230-021-09859-8

Mandal, S., Ji, W., and Mcknight, T. D. (2020). Candidate gene networks for acylsugar metabolism and plant defense in wild tomato *Solanum pennellii*. *Plant Cell* 32, 81–99. doi: 10.1105/tpc.19.00552

Margarido, G. R. A., Souza, A. P., and Garcia, A. (2007). OneMap: software for genetic mapping in outcrossing species. *Hereditas* 144, 78–79. doi: 10.1111/j.2007.0018-0661.02000.x

McCauley, D. (2021). The world's first diploid recombinant inbred potato population. *CSA News* 66, 6–11. doi: 10.1002/csan.20602

Morillo, M., Rojas, J., Lequart,, Lamarti, A., and Martin, P. (2020). Natural and synthetic derivatives of the steroidal glycoalkaloids of *Solanum* genus and biological activity. *Nat. Prod. Chem. Res.* 8, 1–14. doi: 10.35248/2329-6836.20.8.371

Niño, J., Correa, Y. M., and Mosquera, O. M. (2009). Biological activities of steroidal alkaloids isolated from *Solanum leucocarpum*. *Pharm. Biol.* 47, 255–259. doi: 10.1080/13880200802434484

Okada, M., Sugita, T., and Abe, I. (2017). Posttranslational isoprenylation of tryptophan in bacteria. *Beilstein J. Organic Chem.* 13, 338–346. doi: 10.3762/bjoc.13.37

Osorio, M., Carvajal, M., Vergara, A., Butassi, E., Zacchino, S., Mascayano, C., et al. (2021). Prenylated flavonoids with potential antimicrobial activity: synthesis, biological activity, and in silico study. *Int. J. Mol. Sci.* 22, 5472. doi: 10.3390/ijms22115472

Paguigan, N. D., Rivera-Chávez, J., Stempin, J. J., Augustinović, M., Noras, A. I., Raja, H. A., et al. (2019). Prenylated diresorcinols inhibit bacterial quorum sensing. *J. Nat. Prod.* 82, 550–558. doi: 10.1021/acs.jnatprod.8b00925

Pham, G. M., Hamilton, J. P., Wood, J. C., Burke, J. T., Zhao, H., Vaillancourt, B., et al. (2020). Construction of a chromosome-scale long-read reference genome assembly for potato. *GigaScience* 9, 1–11. doi: 10.1093/gigascience/giaa100

Põllumaa, L., Alamäe, T., and Mäe, A. (2012). Quorum sensing and expression of virulence in pectobacteria. *Sensors (Basel)* 12, 3327–3349. doi: 10.3390/s120303327

Shakya, R., and Navarre, D. A. (2008). LC-MS analysis of solanidane glycoalkaloid diversity among tubers of four wild potato species and three cultivars (*Solanum tuberosum*). *J. Agric. Food Chem.* 56, 6949–6958. doi: 10.1021/jf8006618

Smith, C. A., Want, E. J., O'maille, G., Abagyan, R., and Siuzdak, G. (2006). XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* 78, 779–787. doi: 10.1021/ac051437y

Song, X., Mei, P., Dou, T., Liu, Q., and Li, L. (2023). Multi-omics analysis reveals the resistance mechanism and the pathogens causing root rot of *Coptis chinensis*. *Microbiol. Spectr.* 11, e04803–e04822. doi: 10.1128/spectrum.04803-22

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., et al. (2007). Proposed minimum reporting standards for chemical analysis, chemical analysis working group (CAWG), metabolomics standards initiative (MSI). *Metabolomics* 3, 211–221. doi: 10.1007/s11306-007-0082-2

Tsugawa, H., Kind, T., Nakabayashi, R., Yukihira, D., Tanaka, W., Cajka, T., et al. (2016). Hydrogen rearrangement rules: Computational MS/MS fragmentation and structure elucidation using MS-FINDER software. *Anal. Chem.* 88, 7946–7958. doi: 10.1021/acs.analchem.6b00770

van Ooijen, J. W. (1992). Accuracy of mapping quantitative trait loci in autogamous species. *Theoritical Appl. Genet.* 84, 803–811. doi: 10.1007/BF00227388

Vyska, M., Cunniffe, N., and Gilligan, C. (2016). Trade-off between disease resistance and crop yield: a landscape-scale mathematical modelling perspective. *J. R. Soc. Interface* 13, 20160451. doi: 10.1098/rsif.2016.0451

Wang, S., Basten, C. J., and Zeng, Z.-B. (2012). *Windows QTL Cartographer 2.5* (Raleigh, NC: Department of Statistics, North Carolina State University).

Wang, Q., Kuang, Y., He, J., Li, K., Song, W., Jin, H., et al. (2017). The prenylated phenolic natural product isoglycycoumarin is a highly selective probe for human cytochrome P450 2A6. *Eur. J. Pharmacol. Sci.* 109, 472–479. doi: 10.1016/j.ejps.2017.08.035

Wishart, D. S., Guo, A., Oler, E., Wang, F., Anjum, A., Peters, H., et al. (2022). HMDB 5.0: the human metabolome database for 2022. *Nucleic Acids Res.* 50, D622–D631. doi: 10.1093/nar/gkab1062

Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., et al. (2011). Genome sequence and analysis of the tuber crop potato. *Nature* 475, 189–195. doi: 10.1038/nature10158

Yan, Y., Li, X., Zhang, C., Lv, L., Gao, B., and Li, M. (2021). Research progress on antibacterial activities and mechanisms of natural alkaloids: a review. *Antibiotics (Basel)* 10, 1–30. doi: 10.3390/antibiotics10030318

Yazaki, K., Sasaki, K., and Tsurumaru, Y. (2009). Prenylation of aromatic compounds, a key diversification of plant secondary metabolites. *Phytochemistry* 70, 1739–1745. doi: 10.1016/j.phytochem.2009.08.023

Yuan, X., Sundin, G. W., Zeng, Q., Johnson, K. B., Cox, K. D., Yu, M., et al. (2023). Erwinia amylovora type III secretion system inhibitors reduce fire blight infection under field conditions. *Phytopathology®* 113, 2197–2204. doi: 10.1094/PHYTO-04-23-0111-SA

Zimnoch-Guzowska, E., Marczewski, W., Lebecka, R., Flis, B., Schäfer-Pregl, R., Salamini, F., et al. (2000). QTL analysis of new sources of resistance to *Erwinia carotovora* ssp. *atroseptica* in potato done by AFLP, RFLP, and resistance-gene-like markers. *Crop Sci.* 40, 1156–1167.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 01 May 2024
DOI 10.3389/fpls.2024.1394213

OPEN ACCESS

EDITED BY
Weicong Qi,
Jiangsu Academy of Agricultural Sciences (JAAS), China

REVIEWED BY
Chao Zhang,
Hebei Agricultural University, China
Xiaoxiao Zhang,
Guangxi University, China

*CORRESPONDENCE
Huiqing Wang
xjwhq1@163.com
Haifeng Gao
ghf20044666@163.com
Yue Li
liyue6905@126.com

RECEIVED 01 March 2024
ACCEPTED 18 April 2024
PUBLISHED 01 May 2024

CITATION
Lai H, Shen Y, Yang H, Fernando DWG, Ren C, Deng F, Lu Y, Sun N, Chen L, Li G, Wang H, Gao H and Li Y (2024) Comparative analysis of stripe rust resistance in seedling stage and *Yr* gene incidence in spring and winter wheat from Xinjiang, China. *Front. Plant Sci.* 15:1394213. doi: 10.3389/fpls.2024.1394213

Comparative analysis of stripe rust resistance in seedling stage and *Yr* gene incidence in spring and winter wheat from Xinjiang, China

Hanlin Lai[1,2], Yuyang Shen[2], Hong Yang[2], Dilantha W. G. Fernando[3], Chenrong Ren[4], Feifei Deng[2], Yi Lu[4], Na Sun[5], Li Chen[2], Guangkuo Li[2], Huiqing Wang[4]*, Haifeng Gao[2]* and Yue Li[1]*

[1]College of Life Science, Xinjiang Agricultural University, Urumqi, China, [2]Institute of Plant Protection, Xinjiang Academy of Agricultural Sciences/Key Laboratory of Integrated Pest Management on Crop in Northwestern Oasis, Ministry of Agriculture and Rural Affairs, Urumqi, Xinjiang, China, [3]Department of Plant Sciences, University of Manitoba, Winnipeg, MB, Canada, [4]Plant Protection Station of Xinjiang Uygur Autonomous Region, Urumqi, Xinjiang, China, [5]Institute of Agricultural Sciences of Ili Prefecture, Ili, Xinjiang, China

Background: Stripe rust, caused by the fungus *Puccinia striiformis* f.sp. *tritici* (*Pst*), poses a significant threat to global wheat production.

Objectives: This study aims to analyze the distribution of stripe rust resistance genes, characterize resistance phenotypes at the seedling stage of 137 spring and 149 winter wheat varieties in Xinjiang, China, and discern differences in resistance between spring and winter wheat varieties.

Design: We used various Pst races (CYR23, CYR29, CYR31, CYR32, CYR33, CYR34) to characterize seedling resistance of spring and winter wheat varieties and to correlate resistance to the presence of wheat resistance genes (*Yr5*, *Yr9*, *Yr10*, *Yr15*, *Yr17*, *Yr18*, *Yr26*, *Yr41*, *Yr80*, *Yr81*) using molecular markers.

Results: Among spring wheat varieties, 62, 60, 42, 26, 51, and 24 varieties exhibited resistance to CYR23, CYR29, CYR31, CYR32, CYR33, and CYR34, respectively, with four varieties resistant to all varieties. Among winter wheat varieties, 66, 32, 69, 26, 83, 40 varieties demonstrated resistance to CYR23, CYR29, CYR31, CYR32, CYR33, and CYR34, respectively, with four varieties resistant to all varieties. Molecular testing revealed that, in spring wheat, 2, 17, 21, 61, 10, 0, 10, 79, and 32 varieties carried *Yr9*, *Yr10*, *Yr15*, *Yr17*, *Yr18*, *Yr26*, *Yr41*, *Yr80*, and *Yr81* genes, respectively. In winter wheat, 40, 20, 7, 143, 15, 1, 6, 38, and 54 varieties carried *Yr9*, *Yr10*, *Yr15*, *Yr17*, *Yr18*, *Yr26*, *Yr41*, *Yr80*, and *Yr81* genes, respectively. Notably, winter wheat exhibited a significantly higher resistance frequency than spring wheat, particularly in the incidence of *Yr9*, *Yr10*, *Yr17*, *Yr18*, and multi-gene combinations.

Conclusion: In summary, this study provides information on seedling stage resistance to stripe rust 286 Xinjiang wheat varieties, elucidates the distribution of resistance genes in this population, and offers a mechanistic basis for breeding durable resistance in wheat. varieties from Xinjiang.

KEYWORDS

wheat variety, disease resistance, molecular detection, *Puccinia striiformis* f. sp. *tritici*, Xinjiang

1 Introduction

China is the world's largest producer of wheat, generating 128 million metric tons per annum (Zhao and Kang, 2023). In China, wheat is cultivated across multiple environmentally unique and geographically isolated regions, giving rise to multiple strains of the parasitic fungus, *Puccinia striiformis* f. sp. *tritici* (*Pst*), the causal agent of wheat stripe rust (Zeng and Luo, 2006; Wan et al., 2007; Zhan et al., 2016; Ma et al., 2023). *Pst* is an ancient airborne pathogen that specifically colonizes wheat during its asexual life cycle (Brown and Hovmoller, 2002; Schwessinger, 2017; Li et al, 2021; Zhao et al., 2013), and *Pst* urediniospores are dispersed via wind, resulting in specific races localizing to certain regions (Yao et al., 2019a). Virulent strains of *Pst* cause economically devasting stripe rust outbreaks, which resulted in an average yield loss of 1.54 million metric tons per year from 2000 to 2018 in China alone (Zhang et al., 2022). Wheat stripe rust has been responsible for several pandemics in China, most recently in 2017, which resulted in massive yield losses and significant economic impacts (Shen et al., 2002; Wan et al., 2007; Chen et al., 2009; Ma, 2018).

Local varieties of wheat are often referred to as landrace varieties. These are crop composite populations that have been adapted to specific climatic and geographic conditions of a particular region over a long period of cultivation, and they are relatively genetically stable (Shen et al., 2002). Xinjiang Province is found far northwest of China, bordering Tibet, and is classified as an independent epidemic area of wheat stripe rust (Zeng and Luo, 2006). In China, wheat stripe rust has the characteristics of high epidemic frequency, wide occurrence range, and severe damage to wheat production (Jiang et al., 2022). Xinjiang wheat appears to have originated through natural hybridization between Polish and common wheat, suggesting that local wheat was likely introduced and selectively bred within the region itself rather than through the domestication of wild wheat (Betts et al., 2014). Furthermore, Xinjiang's distinctive climate allows for the cultivation of both oversummer and overwinter wheat, providing year-round hosts for *Pst*. Xinjiang is also home to Chinese *Berberis*, or the *barberry* plant, the host required for the sexual reproduction phase of the *Pst* lifecycle, indicating that Xinjiang supports significant diversity in *Pst* races (Zhuang, 2019). To protect wheat crops within Xinjiang, it is critical to determine the resistance distribution of local wheat cultivars to prevalent races of *Pst*. Additionally, as wheat varieties expressing a single resistance gene are cultivated over time, selective pressure has driven pathogen evolution to escape gene-for-gene resistance, leading to obsolete resistance genes (Kang et al., 2015; Li et al., 2020). Therefore, Xinjiang's unique local landraces represent valuable genetic sources for potential resistance genes against *Pst* that can be bred into major cultivars (Mujeeb-Kazi et al., 2013; Chen et al., 2016, Chen et al., 2018; Yao et al., 2019b; Dai et al., 2020).

Despite the effective use of fungicides to combat recent epidemics, deployment of R-genes against virulent *Pst* races remains the most economically and environmentally viable approach (Zhao and Kang, 2023). Currently, the dominant *Pst* races in China, designated with the CYR (Chinese yellow rust) prefix, are CYR32, CYR33, and CYR34 (Zhao and Kang, 2023). Over the last century, the *Yr* wheat resistance genes *Yr9*, *Yr10*, *Yr17*, and *Yr26* have been extensively used in breeding for wheat stripe rust resistance (Han et al., 2010). With the emergence of new virulent races, such as CYR29, CYR31, CYR32, and CYR34, resistance conveyed by *Yr9*, *Yr10*, *Yr17*, and *Yr26* has been revealed (Han et al., 2010). Some all-stage resistance (ASR) genes, such as *Yr5* and *Yr15*, are still effective against prevalent races of *Pst*, but are rarely used in breeding (Zeng et al., 2014). The *Yr18* gene confers non-race specific resistance to stripe rust and slows infection time and spore production, resulting in adult plant resistance (APR) (Singh et al., 2000). Presently, the main race types prevalent in Ili prefecture, the largest wheat-producing region of Xinjiang, are CYR34 and Su 11-1 (Chen et al., 2023). It has been shown that *Yr5* and *Yr15* are highly resistant to stripe rust (Zhou et al., 2023), making the detection of these resistance genes essential for the prevention and control of wheat stripe rust in Xinjiang.

With the expansion of molecular marker technologies, an increasing number of molecular markers are being used in the breeding of disease-resistant wheat varieties. Currently, the major molecular markers based on genomic DNA molecular polymorphisms include sequence tagged sites (STS), simple sequence repeats (SSR), and kompetitive allele-specific polymerase chain reactions (KASP) (Song et al., 2023). By detecting linked molecular markers, disease resistance genes in wheat can be efficiently detected, either directly or indirectly. Identification of

cultivars containing effective resistance genes permits selective breeding for robust resistance and enables the stacking of multiple genes to establish long-lasting disease-resistant varieties (Luo et al., 2023). Presently, 83 stripe rust resistance genes (*Yr1-Yr83*) have been discovered and conclusively named, and more than 300 genes or QTL have been identified and temporarily named (McIntosh et al., 2009; Li et al., 2020).

A previous study detected a high frequency of *Yr9* in wheat cultivars of the Huang-Huai region, along with significant expression of *Yr18* in Huang-Huai landraces (Huang et al., 2020). Alternatively, the main wheat varieties in southwest China exhibit a high frequency of polygene polymerization (Xi et al., 2021). Landraces in northwest China display high frequency of *Yr9*, which has been widely used for breeding, as well as varieties with multiple *Yr* aggregates (Wang et al., 2018). The ASR genes *Yr5* and *Yr15* are generally absent from wheat panels in China, where only a few reports identified *Yr15* (Li et al., 2016). Regions adjacent to Xinjiang, such as Pakistan and Kazakhstan, have the highest distribution frequency of *Yr18* and *Yr10*, which convey resistance to most races of stripe rust (Sobia et al., 2010; Kokhmetova et al., 2021). The frequencies of *Yr26* and *Yr10* in Indian wheat are 69.2% and 50%, respectively, with some highly resistant varieties containing stacks of 15 *Yr* genes (Rani et al., 2019).

Here, we perform a systematic study of phenotypic variation and molecular characterization of stripe rust resistance of winter and spring wheat verities in Xinjiang. We employ a panel of 149 winter wheat varieties and 137 spring wheat varieties in Xinjiang for the characterization of stripe rust resistance against six races at the seedling stage. We further correlate resistance phenotypes to the prevalence of 10 stripe rust resistance genes, as identified with appropriate molecular markers. While the distribution of wheat stripe rust resistance genes in Xinjiang has been recently evaluated, this study is the first to directly assess resistance phenotypes in tandem with the identification of resistance genes (Zhang et al., 2023). This work provides valuable resources for the identification of breeding targets for enhancing wheat stripe rust resistance from the pool of Xinjiang wheat.

2 Materials and methods

2.1 Materials

A total of 286 wheat varieties, including 137 spring and 149 winter varieties, were tested. The Spring Wheat Breeding Team of the Grain Crops Research Institute of the Xinjiang Academy of Agricultural Sciences provided the 137 spring wheat varieties. The Wheat Breeding Team of the Agricultural Research Institute of the Ili Region contributed 55 YINONG winter wheat varieties. The Winter Wheat Breeding Team of the Grain Crops Research Institute of the Xinjiang Academy of Agricultural Sciences provided 94 wheat varieties of the winter wheat varieties. Associate Professor Zhan Gangming from the College of Plant Protection, North Agriculture and Forestry University of Science and Technology, provided the control materials, including Mingxian 169 and single gene line materials.

2.2 Identification of seedling disease resistance

Phenotyping of seedling diseases was carried out in climatic chambers at the Institute of Plant Protection, Xinjiang Academy of Agricultural Sciences, Xinjiang, China. In brief, 10–15 seeds from each wheat variety were sown in small pots and cultivated indoors until the one-leaf-one-heart stage. Seedlings were inoculated with *Pst* races combined with e-fluoridized solution using a pipetting gun set to 5 μL and then maintained in dark conditions for 24 h (10°C). Afterward, seedlings were transferred to climatic chambers (12-hr light/12-hr dark). When symptoms were fully visible in the control group, Mingxian 169, the infection type (IT) was assessed and classed as high resistance (IT: 0–3), moderate resistance (IT: 4–5), or high susceptibility (IT: 6–9).

2.3 Molecular detection of stripe rust resistance genes

Appropriate amounts of wheat leaf tissue were collected, and a modified CTAB method was used for the extraction of genomic DNA (Allen et al., 2006). The quality and quantity of the extracted genomic DNA was determined using a NanoDrop 610 spectrophotometer (ThermoScientific, Wilmington, DE, USA), and the DNA was diluted to 100 ng/μL and stored at -20°C. Polymerase chain reaction (PCR) amplification using specific primers for SSR, STS, and SNP molecular markers associated with *Yr5*, *Yr9*, *Yr10*, *Yr15*, *Yr17*, *Yr18*, *Yr26*, *Yr41*, *Yr80*, and *Yr81* was performed, followed by agarose gel electrophoresis on 1–2% agarose gel. PCR reactions consisted of 2× Easy *Taq* PCR mix (12.5 μL), forward primer (F: 1 μL), and reverse primer (R: 1 μL). PCR reactions involved a 4-minute predenaturation at 94°C, followed by 35 cycles of denaturation at 94°C for 30 seconds, annealing at 55–65°C for 30 seconds, and extension at 72°C for 30–60 seconds. KASP-SNP molecular marker PCR reactions consisted of 2× KASP mix (170 μL), forward primer 1 (F1: 0.51 μL), forward primer 2 (F2: 0.51 μL), and reverse primer (R: 1.36 μL). Subsequently, 1 μL of DNA template was added to each reaction. Primers were synthesized by Shanghai Bioengineering. Refer to Table 1 for primer sequences.

2.4 Data analysis

UpSetR (version 1.4.0) package in the R environment was used for gene combination analysis, and ggplot2 (version 3.4.4) and venn (version 1.11) packages were used for data processing and graph generation, respectively.

3 Results

3.1 Resistance of spring and winter wheat seedlings to different strip rust races

The number of varieties of spring wheat resistant to CYR23, CYR29, CYR31, CYR32, CYR33, and CYR34 amounted to 62

TABLE 1 Primers used in this study.

Gene	Marker Type	Primer Name	Primer Sequence (5′–3′)	Annealing temperature (°C)	Reference
Yr5	SSR	Wmc175	GCTCAGTCAAACC GCTACTTCT	57	(Chen et al., 2003)
			CACTACTCCAA TCTATCGCCGT		
	KASP	Yr5F	GAAGGTGACCAAGTTCATGC TGCGCCCCTTTT CGAAAAAATA	touchdown PCR	(Marchal et al., 2018)
		Yr5H	GAAGGTCGGAG TCAACGGATTC TAGCATCAAACAAGCTAAATA		
		Yr5R	ATGTCGAAAT ATTGCATAACATGG		
Yr9	STS	H2O	GTTGTAAGGGAGCTCGAGCTG	57	(Liu et al., 2008)
			GTTGGGCAGA AAGGTCGACATC		
	RAPD	AF1/AF4	GGAGACATCATGAAACATTTG	58	(Francis et al., 1995)
			CTGTTGTTGGGCAGAAAG		
Yr10	AFLP	Sc200	CTGCAGAGTGACATCATACA	60	(Shao et al., 2001)
			TCGAACTAGTAGATGCTGGC		
	SSR	Xpsp3000	GCAGACCTGTGTCATTGGTC	57	(Wang et al., 2002)
			GATATAGTGGCAG CAGCAGGATAC		
Yr15	SSR	Barc8	GCGGGAAT CATGCATAGGAAAACAGAA	57	(Peng et al., 2000)
			GCGGGGG CGAAACATACACATAAAAACA		
	SSR	Xgwm413	TGCTTGTCTAGATTGCTTGGG	60	
			GATCGTCTCGTCCTTGGCA		
Yr17	SCAR	VENTRIUP-LN2	AGGGGCTACTGACCAAGGCT	58	(Jia et al., 2010)
			TGCAGCTACA GCAGTATGTACACAAAA		

(Continued)

10.3389/fpls.2024.1394213

TABLE 1 Continued

Gene	Marker Type	Primer Name	Primer Sequence (5′–3′)	Annealing temperature (°C)	Reference
Yr18	Cssfr1	L34DINT9F	TTGATGAAAC CAGTTTTTTTTCTA	57	(Lagudah et al., 2009)
		L34PLUSR	GCCATTTA ACATAATCATGATGGA		
		L34SPF	GGGAGCA TTATTTTTTTCCATCATG	57	
		L34DINT13R2	ACTTTCCTGAAAAT AATACAAGCA		
	KASP	Lr34-KASP-E11	GAAGGTGACCAAG TTCATGCTGGGAG CATTATTTTTT TCCATCA	touchdown PCR	(Fang et al., 2020)
			GAAGGTCGGA GTCAACGGATT GGGAGCATTATT TTTTTCCATCT		
			AGCGAAT CCAGTATGGAAAT		
Yr26	STS	Xwe173	GGGACAA GGGGAGTTGAAGC	61	(Wang et al., 2008)
			GAGAGTTC CAAGCAGAACAC		
Yr41	STS	BE446068F	ATGGCTT GGTTTCCCTTTTT	59	(Zhang, 2016)
			TATCAAG CTCGCTCGGCTAA		
Yr80	KASP	KASP_53113	GAAGGTGA CCAAGTTC ATGCTTGTA CAATGACTC CTCGACTAACA	touchdown PCR	(Nsabiyera et al., 2018)
			GAAGGTCGGA GTCAACGGATT TGTACAAT GACTCCTC GACTAACG		

(Continued)

10.3389/fpls.2024.1394213

TABLE 1 Continued

Gene	Marker Type	Primer Name	Primer Sequence (5′–3′)	Annealing temperature (°C)	Reference
			GCCACGCAAT ATCACCATCG		
Yr81	KSAP	KASP_3077	GAAGGTGACCA AGTTCATGCTAT TCCAAAGTAATT GGCAACAGGTTCA	touchdown PCR	(Gessese et al., 2019)
			GAAGGTCGG AGTCAACGGATT CCAAAGTAATT GGCAACAGGTTCG		
			TGTGGAGC GTGACAATGA GGAAGTT		

(45.26%), 60 (43.8%), 44 (32.12%), 25 (18.25%), 50 (36.60%), and 25 (18.25%), respectively. Notably, Xinchun No. 32, Xinchun No. 51, Liangchun1723, and Liangchun1817 were resistant to all 6 races, accounting for 2.92% of all resistant varieties (Figure 1A; Supplementary Table S1).

We identified 62 (44.3%), 32 (21.48%), 70 (46.89%), 26 (17.45%), 84 (56.38%), and 40 (26.85%) winter wheat varieties resistant to CYR23, CYR29, CYR31, CYR32, CYR33, and CYR34, respectively. Among these, varieties 2014-132-4-5, 2014-129-13-9, 6239, and 6444 were resistant to all 6 races, accounting for 2.68% of all resistant varieties (Figure 1B; Supplementary Table S2).

Presently, the predominant races in Xinjiang are CYR32, CYR33, and CYR34. A detailed analysis of these races has revealed that 11 spring wheat varieties exhibited resistance to CYR33 and CYR34, including Xinchun No. 12, Xinchun No. 14, and Xinchun No. 16. Four varieties, including Xinchun No. 34, Xinchun No. 39, and Liangchun1934, exhibited resistance to CYR33 and CYR32. Another four varieties, including 2016, Liangchun1758, and Liangchun1832, demonstrated resistance to CYR32 and CYR34. Eight varieties, including Xinchun No. 29, Heli1881, and Liangchun547, exhibited resistance to all three races.

Among the winter wheat varieties, 24 varieties, including Zhaonong147, 2014-132-10-10, and 2014-12-3-1, demonstrated resistance to both CYR33 and CYR34. Twelve varieties exhibited resistance to CYR33 and CYR32, including 2014-129-1-6, Pin I-1, and Pin I-8. Additionally, 2 varieties, namely Jindong008 and Tiandong33, exhibited resistance to both CYR32 and CYR34, whereas 7 varieties, including 2014-132-10-6, 6222, and 6238, demonstrated resistance to all three races. Overall, Xinjiang wheat exhibits the highest frequency of resistance to CYR33 among the Su 11 taxa. Previous surveys have also revealed the highest proportion of Su 11 pathogenic taxa of stripe rust (Chen et al., 2023; Ma et al., 2023). Consequently, Xinjiang wheat possesses resistance to the currently prevalent races.

3.2 Comparison of disease resistance between spring and winter wheat

Our results indicate that the resistance of spring and winter wheat to CYR23 and CYR32 was similar. While the probability of spring wheat possessing resistance to CYR29 was significantly higher than that of winter wheat (Figure 2), the probability of winter wheat possessing resistance to CYR31, CYR33, and CYR34 was significantly higher than that of spring wheat (Figure 2).

3.3 Molecular detection of resistance genes in spring wheat

Molecular markers closely linked to the *Yr5*, *Yr9*, *Yr10*, *Y15*, *Yr17*, *Yr18*, *Yr26*, *Yr41*, *Yr80*, and *Yr81* genes were used to identify potential *Yr* genes in our spring wheat panel (Figure 3; Supplementary Table S1). Spring wheat varieties Xinchun No. 33 and 2020J/54 were identified as potential carriers of *Yr9*. The *Yr10*-linked marker amplified the target band in 17 (12.41%) varieties. We identified 21 (15.32%) varieties potentially carrying *Yr15*. The

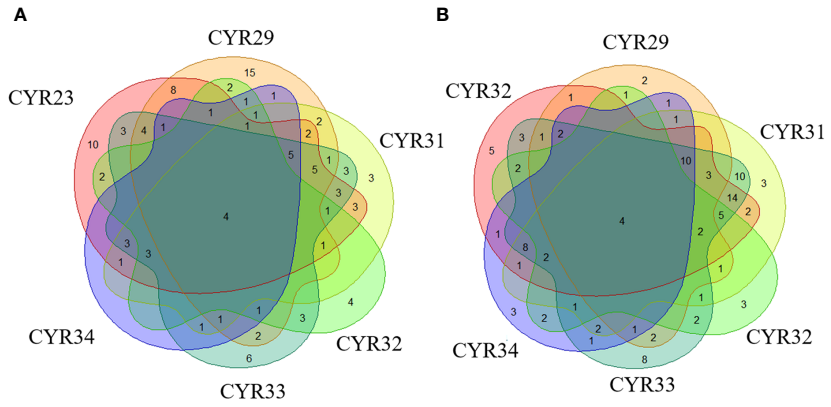


FIGURE 1
Identification of resistance against six *Pst* races in spring and winter wheat landraces from Xinjiang. **(A)** Depicts the number of varieties resistant to different races in spring wheat, and **(B)** Depicts the number of varieties resistant to different races in winter wheat.

VENTRIUP-LN2 marker, which is closely linked to *Yr17*, amplified the target band in 61 (44.52%) varieties. The KASP marker, which is closely linked to *Yr18*, amplified the target band in 10 (7.30%) varieties (Figure 3C). BE446068F, which is closely linked to *Yr41*, was amplified in 10 (7.30%) varieties. The KASP marker, closely linked to *Yr80*, was amplified in 79 (57.66%) varieties (Figure 3A), while the KASP marker, closely linked to *Yr81*, was amplified in 32 (23.36%) varieties (Figure 3B). *Yr26* and *Yr5* was not detected in all spring wheat varieties. *Yr17* and *Yr80* genes exhibited the highest frequency of distribution in spring wheat.

3.4 Molecular detection of resistance genes in winter wheat

In the winter wheat panel, 40 (26.85%) varieties were identified as potential carriers of *Yr9*. The *Yr10* marker amplified the target band in 20 (13.42%) varieties. We identified 7 (4.7%) varieties potentially carrying *Yr15*. The VENTRIUP-LN2 marker amplified the target band in 143 (95.97%) varieties. The KASP marker, which is closely linked to *Yr18*, amplified the target band in 15 (10.03%) varieties (Figure 3C). Furthermore, 2014-132-7 was found to harbor the *Yr26* coding region, where the closely linked WE173 marker amplified the target band. BE446068F was amplified in 6 (4.03%) varieties. The KASP marker, closely linked to *Yr80*, was amplified in 38 (25.50%) varieties (Figure 3A). Furthermore, 54 (36.24%) varieties amplified the KASP marker, which is closely linked to *Yr81* (Figure 3B; Supplementary Table S2). Notably, as observed for spring wheat varieties, *Yr17* and *Yr80* also showed the highest frequency of distribution in winter wheat.

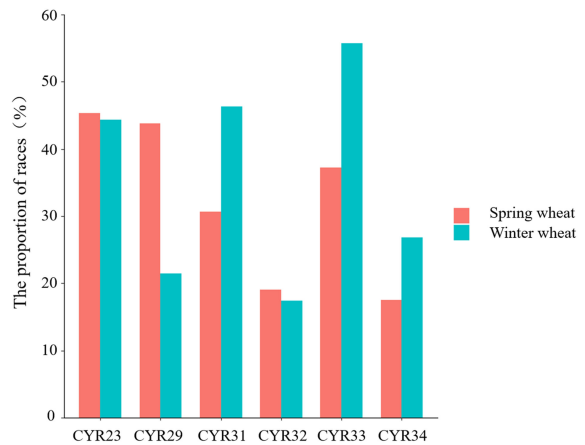


FIGURE 2
Proportion of spring and winter wheat varieties resistant to six *Pst* races.

3.5 Distribution of *Yr* genes in spring and winter wheat

The results of our *Yr* gene distribution analysis are summarized in Figure 4. Specifically, we observed that the proportions of winter wheat expressing *Yr9*, *Yr17*, *Yr18*, and *Yr81* were 26.85%, 95.97%, 10.07%, and 36.24%, respectively. These genes were significantly more prevalent in winter than in spring wheat, whereas the proportion of winter wheat varieties carrying *Yr10* (13.42%) was slightly higher than spring wheat varieties. The proportions of *Yr15* (15.32%) and *Yr41* (7.3%) in spring wheat were slightly higher than in winter wheat, whereas the proportion of *Yr80* (57.66%) in spring wheat was significantly higher than in winter wheat.

3.6 Multi-gene combination analysis

To further investigate the distribution of resistance genes in different wheat varieties, we assessed the number of varieties containing aggregates of 2, 3, 4, and 5 *Yr* genes in spring and winter wheat (Figure 5). The number of multi-gene combined varieties of winter wheat was higher than that of spring wheat. In spring wheat, *Yr17+Yr80* was found in 29 of the 137 varieties. However, there were no varieties with 5 or more *Yr* genes (Figure 5A). Similarly, *Yr17+Yr81* was found in 25 of the 149

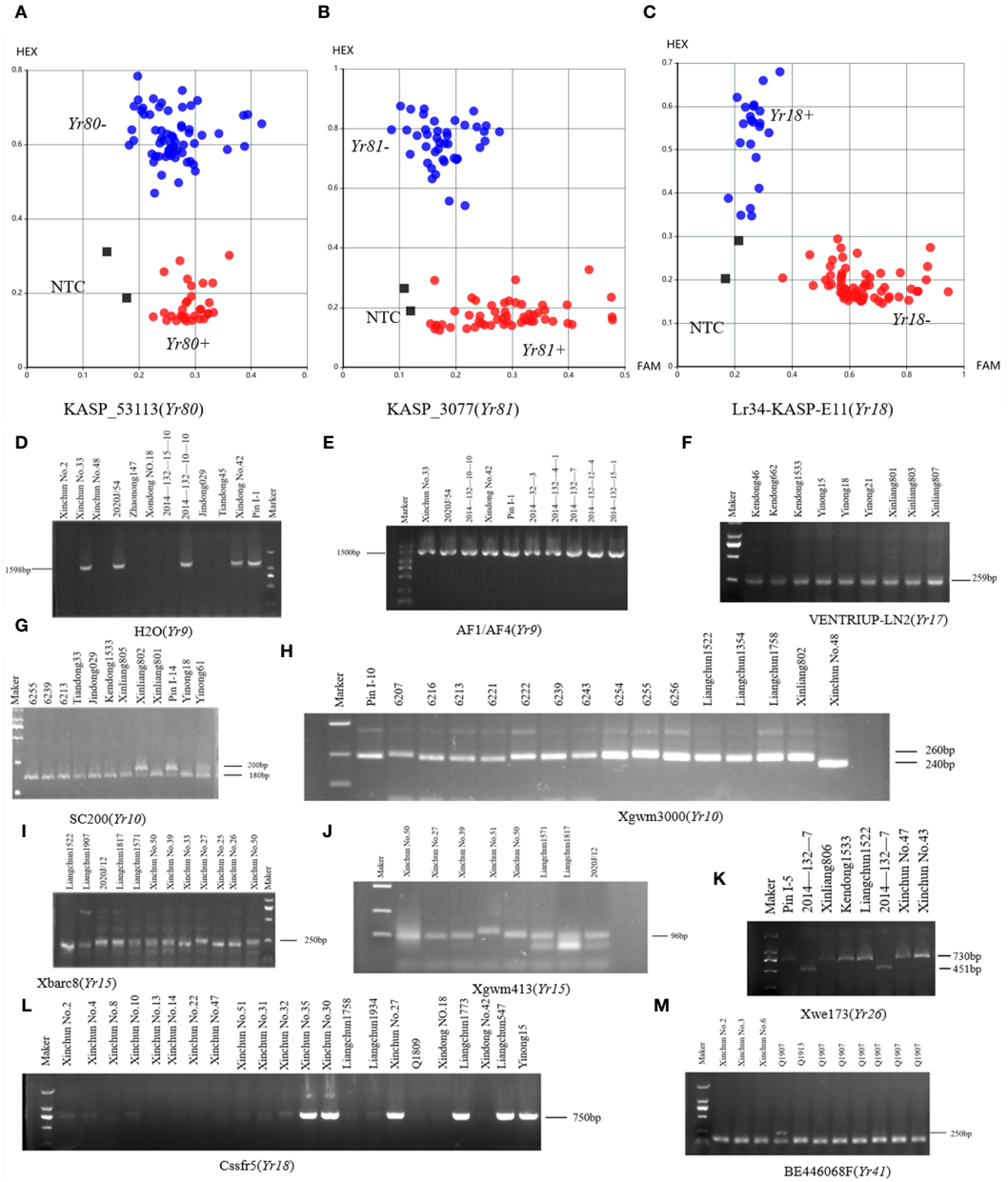


FIGURE 3
Results of PCR amplification of wheat *Yr* gene. **(A–C)** KASP cluster plots of *Yr80*, *Yr81*, and *Yr18* genotyping. **(D–M)** Electrophoretograms of the primers for *Yr9*, *Yr10*, *Yr15*, *Yr17*, *Yr18*, *Yr26*, and *Yr41* resistance genes.

winter wheat varieties. Notably, we identified an aggregate in Pin I-9 containing *Yr17*, *Yr81*, *Yr9*, *Yr80*, and *Yr18* (Figure 5B).

4 Discussion

We assessed disease resistance phenotypes against 6 races of *Pst* in a wheat panel that contained 286 spring and winter wheat varieties found in Xinjiang. Our results show that spring wheat has strong resistance to early popular races CYR23 and CYR29 but weak resistance to currently popular races CYR31, CYR32, CYR33, and CYR34. Only 24 varieties were resistant to CYR34. The resistance of winter wheat to predominant races was generally stronger than that of spring wheat. The number of winter wheat varieties resistant to CYR33 was significantly higher than that of other races, followed by resistance to CYR31 and CYR34, two of the most virulent *Pst* races (Chen et al., 2023; Ma et al., 2023). As the planting area of winter wheat in Xinjiang has expanded, host plants for *Pst* are maintained through summer and winter, allowing the pathogen to complete its annual cycle. This makes it particularly important to select winter wheat varieties with high resistance to stripe rust at the seedling stage, which cannot only reduce stripe rust in autumn plantings of winter wheat, but also prevent propagation of the disease in spring wheat.

To identify potential resistance loci in wheat, it is critical to assess resistance at the seedling stage in a climate chamber, followed

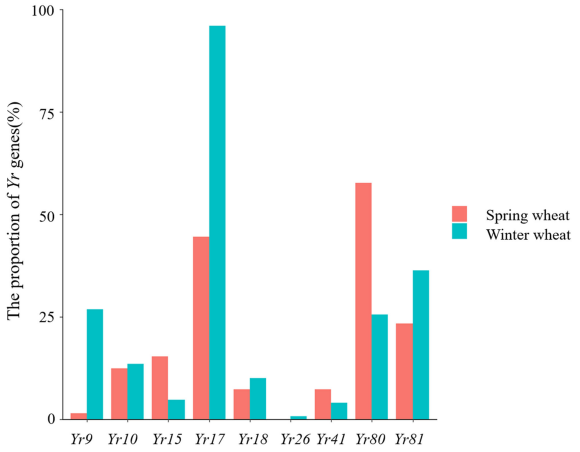


FIGURE 4
Proportion of *Yr* genes detected in winter and spring wheat varieties.

by field experiments with artificial inoculations. This conventional method is limited by natural conditions and requires multiple generations to produce resistant cultivars. Molecular marker-assisted selection (MAS) is an effective tool to guide breeding for heritable resistance traits in breeding cultivars (Song et al., 2023). Here, we applied MAS to assess the distribution of wheat resistance genes in varieties from Xinjiang, to better understand endogenous resistance, and to identify potential loci that could be bred into agricultural lines for improved resistance.

The resistance genes *Yr5* and *Yr15* have been reported to convey strong all-stage resistance to the races of stripe rust in China (Zeng et al., 2014). However, *Yr5* was not detected in our panel, and *Yr15* was identified in only a few varieties (Pin I-6, 1-8-2, Liangchun1817). Among wheat varieties that carried *Yr15*, only Liangchun1817 showed resistance to all races at the seedling stage, which may be due to false *Yr15* positives in the susceptible varieties. Based on a previous study, *Yr5* is rarely used in wheat breeding in China, and *Yr15* has only been detected in one study (Zeng et al., 2014). However, Kazakhstan, which is adjacent to Xinjiang, has high frequencies of *Yr5* and *Yr15*, suggesting active breeding efforts towards *Yr5*- and *Yr15*-positive cultivars (Kokhmetova et al., 2021).

The highly virulent race CYR34 has overcome resistance previously conveyed by *Yr26* (Han et al., 2010). Based on a previous study of wheat in Xinjiang, the frequency of *Yr5* is approximately 100% in spring and winter wheat, and that of *Yr26* is 84.15% (Zhang et al., 2023). This is in contrast to our results, where we did not identify *Yr5* in either wheat population, and similarly no *Yr26* gene in spring wheat and only one *Yr26* gene in a winter wheat line (2014-132-7). This may be due to the use of different molecular markers, and possibly false positives or negatives. We did identify one winter wheat variety with an aggregate containing *Yr26*, *Yr9*, and *Yr17* that displayed strong resistance to CYR34, indicating that this polygene combination can supplement *Yr26*-mediated resistance against CYR34. In Indian wheat, *Yr26* and *Yr10* have high frequencies in the distribution of resistance genes (Rani et al., 2019). While resistance gene distributions in wheat from Xinjiang were similar to those from the rest of China, there were notable differences between Xinjiang and neighboring India and Kazakhstan. This is expected, given the completely different cultivated varieties based on climate and breeder preference. We suggest that a comparative analysis of wheat *Yr* genes and phenotypes in different regions be performed to reveal the breeding trajectories of different regions.

Yr18 is located on chromosome 7D and codes for an APR gene that slows *Pst* infections in the adult stage. Additionally, *Yr18* conveys resistance to leaf rust, stem rust, and powdery mildew (Singh et al., 2000). While the frequency of *Yr18* detection in Chinese wheat

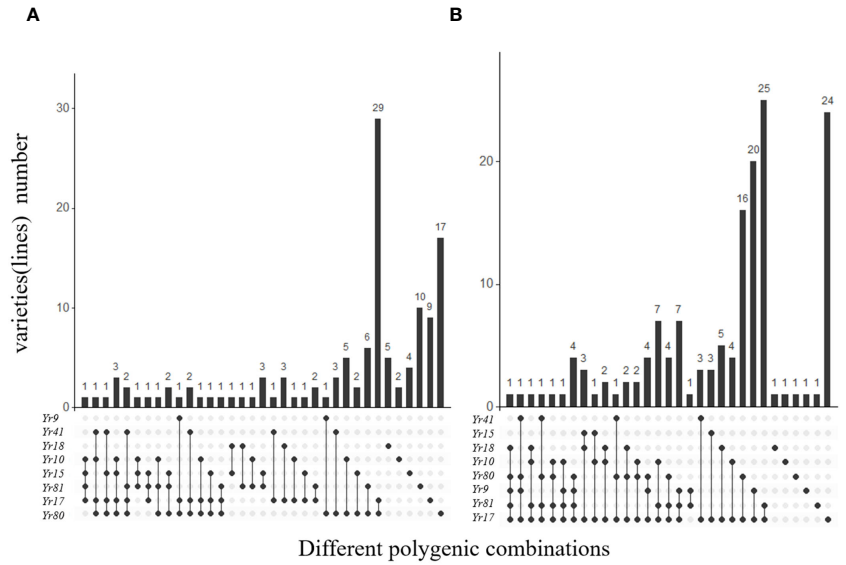


FIGURE 5
Comparative analysis of *Yr* genes found in 137 spring and 149 winter wheat varieties from Xinjiang. **(A)** Gives the combinations of different *Yr* genes in spring wheat, and **(B)** Gives the combinations of different *Yr* genes in winter wheat.

landraces is generally high, *Yr18* is largely absent from Xinjiang wheat. This could be attributed to the infrequent use of wheat landraces in Xinjiang wheat breeding. Therefore, combination of *Yr18* with other ASR and APR genes may significantly improve the stripe rust resistance of Xinjiang wheat varieties.

Yr5, *Yr10*, *Yr15*, *Yr26*, and *Yr18* genes exhibit partial resistance to stripe rust and occur infrequently in Xinjiang wheat varieties. *Yr26* and *Yr10* genes have been overcome by CYR34, yet recent research indicates that the predominant stripe rust population in Xinjiang belongs to the Su 11 pathogenic type (Chen et al., 2023; Ma et al., 2023). *Yr26* and *Yr10* retain their efficacy in controlling stripe rust in Xinjiang. Wheat varieties in Xinjiang carrying *Yr10* and *Yr26*, such as 2014-129-1-6, Kendong47, and 2014-132-7, can effectively manage the prevalent stripe rust type in the region. Given their strong resistance to Xinjiang stripe rust, these genes can be incorporated into future breeding programs for Xinjiang wheat varieties.

With the source variation of stripe rust and the emergence of new virulence factors, many single resistance genes have been overcome. In a previous analysis of wheat resistance genes, polymerization or stacking of multiple genes has been shown to effectively improve resistance (Zhang et al., 2023). In this study, we found that most of the wheat varieties in Xinjiang contained two or more *Yr* genes. We found high prevalence of *Yr80*, *Yr81* and all-stage resistant *Yr17* in both spring and winter wheat, mainly as components of polygene doublets of *Yr17+Yr80* and *Yr17+Yr81*, as well as some other combinations. The results of wheat resistance to stripe rust in Sichuan, the Huang-Huai-Hai region, and Yunnan were similar to ours, indicating that polygene polymerization, specifically *Yr17+Yr81*, can improve the resistance to wheat stripe rust (Li et al., 2016; Wang et al., 2018; Guan et al., 2020; Huang et al., 2020; Xi et al., 2021). *Yr17* from the 2NS/2AL translocation line of *Aegilops* has all-stage resistance and the distribution frequency in spring and winter wheat, however, all the prevalent *Pst* races have overcome *Yr17*-mediated resistance. Since the emergence of CYR29, only wheat varieties carrying *Yr9* alone have lost their resistance. In this study, we found that wheat varieties with *Yr9+Yr17* showed strong resistance to wheat stripe rust races. Notably, we identified two *Yr9+Yr17* winter wheat lines, 2014-132-4-5 and 2014-129-13-9, that displayed seedling resistance to all 6 races, indicating that the combination of *Yr9* and *Yr17* can effectively defend against all prevalent *Pst* races.

5 Conclusions

Our comprehensive analysis of 286 varieties of spring and winter wheat in Xinjiang provides valuable insights into the dynamics of stripe rust resistance. Spring wheat demonstrates robust resistance to early races CYR23 and CYR29 but shows weak resistance against contemporary races CYR31, CYR32, CYR33, and CYR34, with only 24 varieties displaying resistance to CYR34. Conversely, winter wheat exhibits stronger resistance, particularly against CYR33, CYR31, and CYR34, highlighting the importance of selecting resistant winter wheat varieties to curtail stripe rust propagation into spring wheat.

Utilizing MAS, we identify the distribution of key resistance genes, revealing variations in the prevalence of genes such as *Yr5*, *Yr15*, *Yr26*, and *Yr18* compared to previous studies. The absence of *Yr5* and limited detection of *Yr15* suggest regional disparities in breeding priorities, as seen in neighboring Kazakhstan. The emergence of highly virulent CYR34, overcoming *Yr26*-mediated resistance, underscores the evolving nature of stripe rust and the need for adaptable resistance strategies.

Our findings emphasize the significance of polygene combinations in enhancing resistance. Notably, the combination of *Yr17* and *Yr81* demonstrates effectiveness in resisting stripe rust, as observed in the prevalence of *Yr17+Yr80* and *Yr17+Yr81* combinations. Furthermore, the synergy between *Yr9* and *Yr17* proves highly effective against all prevalent *Pst* races, offering promising avenues for breeding resilient wheat varieties.

The absence of *Yr18* in Xinjiang wheat, despite its widespread presence in Chinese landraces, underscores the potential for improved resistance through strategic gene combinations. We propose that combining *Yr18* with other APR genes could significantly enhance stripe rust resistance in Xinjiang wheat. In conclusion, our study contributes crucial data for breeding durable and resilient wheat varieties in Xinjiang, offering insights into regional resistance dynamics and highlighting the efficacy of polygene combinations in mitigating the impact of evolving stripe rust races. These findings can guide future breeding strategies and foster international collaboration for a more comprehensive understanding of wheat resistance across diverse regions.

Data availability statement

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

Author contributions

HL: Writing – original draft, Conceptualization, Data curation, Formal analysis. YS: Conceptualization, Writing – original draft. HY: Formal analysis, Writing – original draft. DF: Data curation, Formal analysis, Writing – original draft. CR: Conceptualization, Data curation, Writing – original draft. FD: Data curation, Methodology, Writing – review & editing. YiL: Formal analysis, Funding acquisition, Writing – review & editing. NS: Project administration, Validation, Writing – review & editing. LC: Investigation, Methodology, Writing – review & editing. GL: Funding acquisition, Resources, Writing – review & editing. HW: Funding acquisition, Writing – review & editing. HG: Funding acquisition, Writing – review & editing. YuL: Methodology, Project administration, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This project

was supported by the Science and Technology Assistance Project of Xinjiang Uygur Autonomous Region (2022E02070); Regional Coordinated Innovation Project, Shanghai Cooperation Organization Science and Technology Partnership Program (2023E01015); Natural Science Foundation of Xinjiang Uygur Autonomous Region (2022D01A135); Key Laboratory of Integrated Pest Management on Crop in Northwestern Oasis (KFJJ202104); Third Tianshan Talent Plan of Xinjiang Uygur Autonomous Region (2022TSYCCX0081); Earmarked Fund for XJARS-01; Science and Technology Major Project of Xinjiang Uygur Autonomous Region (2023A02009); Urumqi Integrated Experimental Station of China Agriculture Research System for Wheat (CARS-03-88); and Project of Fund for Stable Support to Agricultural Sci-Tech Renovation (XJNKYWDZC-2023004-1).

Acknowledgments

We would like to thank A&L Scientific Editing (www.alpublish.com) for the linguistic assistance during the preparation of this manuscript.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1394213/full#supplementary-material

References

Allen, G. C., Flores-Vergara, M. A., Krasynanski, S., Kumar, S., and Thompson, W. F. (2006). A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* 1, 2320–2325. doi: 10.1038/nprot.2006.384

Betts, A., Jia, P. W., and Dodson, J. (2014). The origins of wheat in China and potential pathways for its introduction: as review. *Quatern Int.* 348, 158–168. doi: 10.1016/j.quaint.2013.07.044

Brown, J. K., and Hovmoller, M. S. (2002). Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science.* 297, 537–541. doi: 10.1126/science.1072678

Chen, L., Awais, M., Yang, H., Shen, Y. Y., Li, G. K., Gao, H. F., et al. (2023). Races CYR34 and Suwon11-1 of *Puccinia striiformis* f. sp. *tritici* Played an Important Role in Causing the Stripe Rust Epidemic in Winter Wheat in Yili, Xinjiang, China. *J. Fungi.* 9, 436. doi: 10.3390/jof9040436

Chen, Q., Song, J., Du, W. P., Xu, L. Y., Jiang, Y., Zhang, J., et al. (2018). Phylogenetic analyses of four Chinese endemic wheat landraces based on two single copy genes. *Cereal Res. Commun.* 46, 191–200. doi: 10.1556/0806.46.2018.01

Chen, Q., Song, J., Du, W. P., Xu, L. Y., and Yu, G. R. (2016). Possible origin of *Triticum* petropavlovskyi based on cytological analyses of crosses between T. petropavlovskyi and tetraploid, hexaploid, and synthetic hexaploid (SHW-DPW) wheat accessions. *Span J. Agric. Res.* 14, e0713–e0713. doi: 10.5424/sjar/2016144-8476

Chen, X., Soria, M. A., Yan, G., Sun, J., and Dubcovsky, J. (2003). Development of sequence tagged site and cleaved amplified polymorphic sequence markers for wheat stripe rust resistance gene *Yr5*. *Crop Sci.* 43, 2058–2064. doi: 10.2135/cropsci2003.2058

Chen, W. Q., Wu, L. R., Liu, T. G., Xu, S. C., Jin, S. L., Peng, Y. L., et al. (2009). Race dynamics, diversity, and virulence evolution in *Puccinia striiformis* f. sp. *tritici*, the causal agent of wheat stripe rust in China from 2003 to 2007. *Plant Dis.* 93, 1093–1101. doi: 10.1094/PDIS-93-11-1093

Dai, S. F., Xu, D. Y., Yan, Y. L., Wen, Z. J., Zhang, J. B., Chen, H. X., et al. (2020). Characterization of high-and low-molecular-weight glutenin subunits from Chinese Xinjiang wheat landraces and historical varieties. *J. Food Sci. Technol.* 57, 3823–3835. doi: 10.1007/s13197-020-04414-5

Fang, T. L., Lei, L., Li, G. Q., Powers, C., Hunger, R. M., Carver, B. F., et al. (2020). Development and deployment of KASP markers for multiple alleles of *Lr34* in wheat. *Theor. Appl. Genet.* 133, 2183–2195. doi: 10.1007/s00122-020-03589-x

Francis, H. A., Leitch, A. R., and Koebner, R. M. D. (1995). Conversion of a RAPD-generated PCR product, containing a novel dispersed repetitive element, into a fast and robust assay for the presence of rye chromatin in wheat. *Theor. Appl. Genet.* 90, 636–642. doi: 10.1007/BF00222127

Gessese, M., Bariana, H., Wong, D., Hayden, M., and Bansal, U. (2019). Molecular mapping of stripe rust resistance gene *Yr81* in a common wheat landrace Aus27430. *Plant Dis.* 103, 1166–1171. doi: 10.1094/PDIS-06-18-1055-RE

Guan, F. N., Long, L., Yao, F. J., Wang, Y. Q., Jiang, Q. T., Kang, H. Y., et al. (2020). Evaluation of resistance to stripe rust and molecular detection of important known *Yr* gene (s) of 152 Chinese wheat landraces from the Huang-huai-hai. *Scientia Agricultura Sin.* 53, 3629–3637. doi: 10.3864/J.issn.0578-1752.2020.18.001

Han, D. J., Wang, Q., Zhang, L., Wei, G., Zeng, Q., Zhao, J., et al. (2010). Evaluation of resistance of current wheat cultivars to stripe rust in northwest China, north China and the middle and lower reaches of Changjiang river epidemic area. *Scientia Agricultura Sinica.* 43, 2889–2896. doi: 10.3864/j.issn.0578-1752.2010.14.007

Huang, L., Xiao, X. Z., Liu, B., Gao, L., Gong, G. S., Chen, W. Q., et al. (2020). Identification of stripe rust resistance genes in common wheat cultivars from the Huang-Huai-Hai region of China. *Plant Dis.* 104, 1763–1770. doi: 10.1094/PDIS-10-19-2119-RE

Jia, J. Q., Lie, M. P., Liu, C., Li, G. R., and Yang, Z. J. (2010). Exploitation and application of a new SCAR marker linked to strip rust resistance gene *Yr17* in wheat. *J. Triticeae Crops.* 30, 11–16doi: 10.09-1041(2010)01-0011-06

Jiang, Q., Wang, H. L., and Wang, H. G. (2022). Two new methods for severity assessment of wheat stripe rust caused by *Puccinia striiformis* f. sp. *tritici*. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1002627

Kang, Z. S., Wang, X. J., Zhao, J., Tang, C. L., and Huang, L. L. (2015). Advances in research of pathogenicity and virulence variation of the wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici*. *Sci. Agric. Sin.* 48, 3439–3453. doi: 10.3864/j.issn.0578-1752.2015.17.011

Kokhmetova, A., Rsaliyev, A., Malysheva, A., Atishova, M., Kumarbayeva, M., and Keishilov, Z. (2021). Identification of stripe rust resistance genes in common wheat cultivars and breeding lines from Kazakhstan. *Plants.* 10, 2303. doi: 10.3390/plants10112303

Lagudah, E. S., Krattinger, S. G., Herrera-Foessel, S., Singh, R. P., Huerta-Espino, J., Spielmeyer, W., et al. (2009). Gene-specific markers for the wheat gene *Lr34/Yr18/Pm38* which confers resistance to multiple fungal pathogens. *Theor. Appl. Genet.* 119, 889–898. doi: 10.1007/s00122-009-1097-z

Li, J. B., Dundas, I., Dong, C. M., Li, G. R., Trethowan, R., Yang, Z. J., et al. (2020). Identification and characterization of a new stripe rust resistance gene *Yr83* on rye chromosome 6R in wheat. *Theor. Appl. Genet.* 133, 1095–1107. doi: 10.1007/s00122-020-03534-y

Li, Q., Wang, B. T., Chao, K. X., Guo, J., Song, J. R., Yue, W. Y., et al. (2016). Molecular detection of stripe rust resistance gene (s) in 115 wheat cultivars (lines) from the Yellow and Huai River valley wheat region. *J. Phytopathol.* 164, 946–958. doi: 10.1111/jph.12515

Li, S. N., Wen, C., Tian, X. X., Yao, L., Huang, L. L., Kang, Z. S., et al. (2021). Identification of eight *Berberis* species from the Yunnan-Guizhou plateau as aecial hosts for *Puccinia striiformis* f. sp. *tritici*, the wheat stripe rust pathogen. *J. Integr. Agr.* 20, 1563–1569. doi: 10.1016/S2095-3119(20)63327-5

Liu, C., Yang, Z. J., Li, G. R., Zeng, Z. X., Zhang, Y., Zhou, J. P., et al. (2008). Isolation of a new repetitive DNA sequence from Secale africanum enables targeting of Secale chromatin in wheat background. *Euphytica.* 159, 249–258. doi: 10.1007/s10681-007-9484-5

Luo, K., He, D. J., Guo, J., Li, G. W., Li, B. L., and Chen, X. L. (2023). Molecular advances in breeding for durable resistance against pests and diseases in wheat: opportunities and challenges. *Agronomy.* 13, 628. doi: 10.3390/agronomy13030628

Ma, Z. H. (2018). Researches and control of wheat stripe rust in China. *J. Plant protection.* 45, 1–6. doi: 10.13802/j.cnki.zwbhxb.2018.2018900

Ma, J. B., Awais, M., Chen, L., Yang, H., Lai, H. L., Shen, Y., et al. (2023). Identification of *Puccinia striiformis* races from the spring wheat crop in Xinjiang, China. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1273306

Marchal, C., Zhang, J. P., Zhang, P., Fenwick, P., Steuernagel, B., Adamski, N. M., et al. (2018). BED-domain-containing immune receptors confer diverse resistance spectra to yellow rust. *Nat. plants.* 4, 662–668. doi: 10.1038/s41477-018-0236-4

McIntosh, R. A., Dubcovsky, J., Rogers, W. J., Morris, C., Appels, R., Xia, X. C., et al. (2009). Catalogue of gene symbols for wheat: 2009 supplement. *Annu. wheat newsletter.* 55, 256–278.

Mujeeb-Kazi, A., Kazi, A. G., Dundas, I., Rasheed, A., Ogbonnaya, F., Kishii, M., et al. (2013). Genetic diversity for wheat improvement as a conduit to food security. *Adv. Agron.* 122, 179–257. doi: 10.1016/B978-0-12-417187-9.00004-8

Nsabiyera, V., Bariana, H. S., Qureshi, N., Wong, D., Hayden, M. J., and Bansal, U. K. (2018). Characterisation and mapping of adult plant stripe rust resistance in wheat accession Aus27284. *Theor. Appl. Genet.* 131, 1459–1467. doi: 10.1007/s00122-018-3090-x

Peng, J. H., Fahima, T., Röder, M. S., Huang, Q. Y., Dahan, A., Li, Y. C., et al. (2000). High-density molecular map of chromosome region harboring stripe-rust resistance genes *YrH52* and *Yr15* derived from wild emmer wheat, *Triticum dicoccoides*. *Genetica.* 109, 199–210. doi: 10.1023/A:1017573726512

Rani, R., Singh, R., and Yadav, N. R. (2019). Evaluating stripe rust resistance in Indian wheat genotypes and breeding lines using molecular markers. *Cr Biol.* 342, 154–174. doi: 10.1016/j.crvi.2019.04.002

Schwessinger, B. (2017). Fundamental wheat stripe rust research in the 21st century. *New Phytol.* 213, 1625–1631. doi: 10.1111/nph.14159

Shao, Y. T., Niu, Y. C., Zhu, L. H., Zhai, W. X., Xu, S. C., and Wu, L. R. (2001). Identification of an AFLP marker linked to the stripe rust resistance gene *Yr10* in wheat. *Chin. Sci. Bull.* 46, 1466–1468. doi: 10.1007/BF03187033

Shen, Y. H., Wang, H. Q., Yang, T. Y., Zhang, H. G., and Huang, X. G. (2002). Evolvement of genetic diversity of spring wheat varieties in Gansu and Qinghai Provinces. *Acta Botanica Boreali-occidentalia Sinica.* 22, 731–740. doi: 10.1023/A:1017573726512

Singh, R. P., Nelson, J. C., and Sorrells, M. E. (2000). Mapping *Yr28* and other genes for resistance to stripe rust in wheat. *Crop Sci.* 40, 1148–1155. doi: 10.2135/cropsci2000.4041148x

Sobia, T., Muhammad, A., and Chen, X. M. (2010). Evaluation of Pakistan wheat germplasms for stripe rust resistance using molecular markers. *Sci. China Life Sci.* 53, 1123–1134. doi: 10.1007/s11427-010-4052-y

Song, L. Q., Wang, R. H., Yang, X. J., Zhang, A. M., and Liu, D. C. (2023). Molecular markers and their applications in marker-assisted selection (MAS) in bread wheat (*Triticum aestivum L.*). *Agriculture.* 13, 642. doi: 10.3390/agriculture13030642

Wan, A. M., Chen, X. M., and He, Z. H. (2007). Wheat stripe rust in China. *Aust. J. Agr Res.* 58, 605–619. doi: 10.1071/AR06142

Wang, S. H., Gong, K. Y., Chu, B. Y., Sun, Q. Y., Luo, Y., and Ma, Z. H. (2018). Molecular detection of stripe rust resistance gene (s) in 100 wheat cultivars (lines) from Sichuan Province in China. *Acta Phytopathologica Sinica.* 48, 195–206. doi: 10.13926/j.cnki.apps.000163

Wang, L. F., Ma, J. X., Zhou, R. H., Wang, X. M., and Jia, J. Z. (2002). Molecular tagging of the yellow rust resistance gene *Yr10* in common wheat, PI 178383 (*Triticum aestivum L.*). *Euphytica.* 124, 71–73. doi: 10.1023/A:1015689817857

Wang, C. M., Zhang, Y. P., Han, D. J., Kang, Z. S., Li, G. P., Cao, A. Z., et al. (2008). SSR and STS markers for wheat stripe rust resistance gene *Yr26*. *Euphytica.* 159, 359–366. doi: 10.1007/s10681-007-9524-1

Xi, L., Wang, Y., Yang, X., Zhu, W., Chen, G. Y., Wang, Y., et al. (2021). Evaluation of resistance to stripe rust and molecular detection of resistance gene(s) in 243 common wheat landraces from the Yunnan Province. *Scientia Agricultura Sinica.* 54, 684–695. doi: 10.3864/j.issn.0578-1752.2021.04.002

Yao, Q., Huang, S., Yan, J., Hou, L., Guo, Q., Zhao, J., et al. (2019a). Investigation and identification of *barberry* as alternate hosts for *Puccinia striiformis* f. sp. *tritici* in eastern Qinghai. *Acta Phytopathologica Sinica.* 49, 370–378. doi: 10.13926/j.cnki.apps.000290

Yao, F. J., Zhang, X. M., Ye, X. L., Li, J., Long, L., Yu, C., et al. (2019b). Characterization of molecular diversity and genome-wide association study of stripe rust resistance at the adult plant stage in Northern Chinese wheat landraces. *BMC Genet.* 20, 1–16. doi: 10.1186/s12863-019-0736-x

Zeng, Q. D., Han, D. J., Wang, Q. L., Yuan, F. P., Wu, J. H., Zhang, L., et al. (2014). Stripe rust resistance and genes in Chinese wheat cultivars and breeding lines. *Euphytica.* 196, 271–284. doi: 10.1007/s10681-013-1030-z

Zeng, S., and Luo, Y. (2006). Long-distance spread and interregional epidemics of wheat stripe rust in China. *Plant Dis.* 90, 980–988. doi: 10.1094/PD-90-0980

Zhan, G. M., Wang, F. P., Wan, C. P., Han, Q. M., Huang, L. L., Kang, Z. S., et al. (2016). Virulence and molecular diversity of the *Puccinia striiformis* f. sp. *tritici* population in Xinjiang in relation to other regions of western China. *Plant Dis.* 100, 99–107. doi: 10.1094/PDIS-11-14-1142-RE

Zhang, Z. J. (2016). *Fine mapping of a stripe rust resistance gene* Yr41 *and the function of two genes induced by* Pstriiformis *W. @ in Wheat. PHD Dissertation* (Yaan, Chain: Sichuan Agricultural University).

Zhang, Q. Q., Men, X. Y., Hui, C., Ge, F., and Ouyang, F. (2022). Wheat yield losses from pests and pathogens in China. *Agr Ecosyst. Environ.* 326, 107821. doi: 10.1016/j.agee.2021.107821

Zhang, M. H., Saimi, A., Liu, Q., Ma, Z. Y., and Chen, J. (2023). The detection of *yr* genes in xinjiang wheat cultivars using different molecular markers. *Int. J. Mol. Sci.* 24, 13372. doi: 10.3390/ijms241713372

Zhao, J., and Kang, Z. S. (2023). Fighting wheat rusts in China: a look back and into the future. *Phytopathol. Res.* 5, 6. doi: 10.1186/s42483-023-00159-z

Zhao, J., Wang, L., Wang, Z. Y., Chen, X. M., Zhang, H. C., Yao, J. N., et al. (2013). Identification of eighteen *Berberis* species as alternate hosts of *Puccinia striiformis* f. sp. *tritici* and virulence variation in the pathogen isolates from natural infection of *barberry* plants in China. *Phytopathology.* 103, 927–934. doi: 10.1094/PHYTO-09-12-0249-R

Zhou, A. H., Wang, J., Chen, X. M., Xia, M. H., Feng, Y. X., Ji, F., et al. (2023). Virulence Characterization of *Puccinia striiformis* f. sp. *tritici* in China Using the Chinese and *Yr* single-gene Differentials. *Plant Dis.* 108, 671–683. doi: 10.1094/PDIS-08-23-1524-RE

Zhuang, H. (2019). Identification of three *Berberis* species as potential alternate hosts for *Puccinia striiformis* f. sp. *tritici* in wheat-growing regions of Xinjiang, China. *J. Integr. Agr.* 18, 2786–2792. doi: 10.1016/S2095-3119(19)62709-7

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 08 July 2024
DOI 10.3389/fpls.2024.1400213

OPEN ACCESS

EDITED BY
Weicong Qi,
Jiangsu Academy of Agricultural Sciences (JAAS), China

REVIEWED BY
Honghong Wu,
Huazhong Agricultural University, China
Rui Shi,
North Carolina State University, United States
Tengfang Ling,
Korea Research Institute of Bioscience and Biotechnology (KRIBB), Republic of Korea

*CORRESPONDENCE
Jun Yang
yangjun@ztri.com.cn
Sheng Xu
xusheng@cnbg.net

RECEIVED 13 March 2024
ACCEPTED 11 June 2024
PUBLISHED 08 July 2024

CITATION
Wu M, Li Y, Liu Z, Xia L, Xiang Y, Zhao L, Yang X, Li Z, Xie X, Wang L, Wang R, Xu S and Yang J (2024) Genome-wide identification of the *CAD* gene family and functional analysis of putative *bona fide CAD* genes in tobacco (*Nicotiana tabacum* L.).
Front. Plant Sci. 15:1400213.
doi: 10.3389/fpls.2024.1400213

Genome-wide identification of the *CAD* gene family and functional analysis of putative *bona fide CAD* genes in tobacco (*Nicotiana tabacum* L.)

Mingzhu Wu[1], Yijun Li[1], Zhengtai Liu[2,3], Lin Xia[1], Yiyu Xiang[2,3], Lijie Zhao[1], Xiaobei Yang[1], Zefeng Li[1], Xiaodong Xie[1], Lin Wang[4], Ren Wang[2,3], Sheng Xu[2,3]* and Jun Yang[1]*

[1]China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute of China National Tobacco Corporation (CNTC), Zhengzhou, China, [2]Nanjing University of Chinese Medicine, Nanjing, China, [3]Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China, [4]College of Life Science, Henan Agricultural University, Zhengzhou, China

Cinnamyl alcohol dehydrogenase (CAD) plays a crucial role in lignin biosynthesis, and the gene family encoding various CAD isozymes has been cloned and characterized in numerous plant species. However, limited information regarding the *CAD* gene family in tobacco is currently available. In this study, we identified 10 *CAD* genes in *Nicotiana tabacum*, four in *N. tomentosiformis*, and six in *N. sylvestris*. The nucleotide and amino acid sequences of these tobacco CADs demonstrate high levels of similarity, whereas the putative protein sequences conservatively possessed two Zn^{2+} binding motifs and an NADP(H) cofactor binding motif. Both NtCAD1 and NtCAD2 had conservative substrate binding sites, similar to those possessed by *bona fide* CADs, and evidence from phylogenetic analysis as well as expression profiling supported their role as *bona fide* CADs involved in lignin biosynthesis. NtCAD1 has two paralogous genes, *NtCAD1–1* and *NtCAD1–2*. Enzyme activity analysis revealed that NtCAD1–1 and NtCAD1–2 had a high affinity to coniferyl aldehyde, *p*-coumaryl aldehyde, and sinapyl aldehyde, whereas NtCAD2 preferred coniferyl aldehyde and *p*-coumaryl aldehyde as substrates. The kinetic parameter assay revealed that NtCAD1–2 functions as the most efficient enzyme. Downregulation of both NtCAD1–1 and NtCAD1–2 resulted in reddish-brown stems without significant changes in lignin content. Furthermore, NtCAD1–1, NtCAD1–2, and NtCAD2 showed distinct expression patterns in response to biotic and abiotic stresses, as well as different phytohormones. Our findings suggest that NtCAD1–1 and NtCAD1–2 are involved in lignin biosynthesis, with NtCAD1–2 also participating in both biological and abiotic stresses, whereas NtCAD2 plays a distinct role mainly in responding to biological and abiotic stresses in tobacco.

KEYWORDS

cinnamyl alcohol dehydrogenase, lignin, enzymatic assay, stresses, tobacco

1 Introduction

Lignin, a vital structural component of plant cell walls, plays a pivotal role in plant development, water transport, mechanical support of the cell wall, defense against pathogens and insects, and enhancement of resistance to abiotic stress (Cesarino, 2019). Lignin synthesis is a complex process involving numerous enzymes and intermediates (Liu et al., 2018; Vanholme et al., 2019). Cinnamyl alcohol dehydrogenase (CAD, EC 1.1.1.195), discovered in 1973 (Gross et al., 1973), is an NADPH-dependent enzyme that catalyzes the final step in the reduction of three hydroxycinnamaldehydes (sinapyl aldehyde, *p*-coumaryl aldehyde, and coniferyl aldehyde) to their corresponding hydroxycinnamyl alcohols, with zinc acting as a cofactor (O'malley et al., 1992; Kim et al., 2004). This enzymatic reaction is indispensable for generating monomeric lignin precursors and contributes significantly to the diversity of lignin compositions. The primary constituents of aromatic lignin polymers commonly observed in angiosperms are sinapyl alcohols, *p*-coumaryl alcohols, and coniferyl alcohols. These monolignols undergo polymerization to yield syringyl (S), *p*-hydroxyphenyl (H), and guaiacyl (G) lignin (Li et al., 2008a). CAD and its homologs have been identified in multigene families of various plant species. Specifically, melon and mulberry each have 5 members (Jin et al., 2014; Chao et al., 2022), *Brachypodium distachyon* and oil palm each have 7 members (Bukh et al., 2012; Yusuf et al., 2022), *Arabidopsis* and poplar each have 9 members (Kim et al., 2004; Barakat et al., 2009; Chao et al., 2014), rice has 12 members (Tobias and Chow, 2005), sorghum has 14 members (Saballos et al., 2009), and pomegranate has 25 members (Hu et al., 2022).

CAD-deficient plants display a brown midrib phenotype, characterized by a distinct reddish-brown color and altered lignin biosynthesis. This phenotypic variation has been observed in several plant species, including rice (Zhang et al., 2006), maize (Halpin et al., 1998; Kolkman et al., 2023), sorghum (Sattler et al., 2009; Rivera-Burgos et al., 2019), poplar (Baucher et al., 1996), and mulberry (Chao et al., 2022). This change in stem color is attributed to the accumulation or enrichment of coniferyl aldehydes (Li et al., 2008b). Downregulation of *CAD* expression has varying effects on lignin content across different plant species. For example, in a maize *bm1* mutant with a mutation in the *ZmCAD2* regulator, there was a 10%–20% reduction in lignin content, along with a decrease in the S/G ratio (Guillaumie et al., 2007). Similarly, transgenic tall fescue plants with downregulated *CAD* expression exhibited a 28% decrease in lignin content and S/G ratio compared with control plants (Chen et al., 2003). In alfalfa, decreased activity of the CAD enzyme resulted in a lower S/G ratio whereas lignin content remained unchanged (Baucher et al., 1999). Similarly, transgenic *Arabidopsis* and tobacco plants with reduced *CAD* expression showed only a slight decrease in lignin content (Bernard Vailhé et al., 1998; Kim et al., 2004).

CADs in angiosperm species are classified into distinct classes based on phylogenetic tree analysis (Ma, 2010; Bukh et al., 2012; Jin et al., 2014; Hu et al., 2022; Yusuf et al., 2022). Comprehensive evolutionary analyses of CADs from diverse plant species have unequivocally revealed the involvement of *bona fide CAD* genes in monolignol biosynthesis (Guo et al., 2010). Notably, AtCAD4 and AtCAD5 of *Arabidopsis* (Kim et al., 2004), EgCAD2 of Eucalyptus (Goffner et al., 1992), OsCAD2 of rice (Hirano et al., 2012), and ZmCAD2 of *Zea mays* (Liu et al., 2021) play key roles in this process. Furthermore, other CAD homologs appear to play important roles in plant defense against both abiotic and biotic stresses (Barakat et al., 2010; Kim et al., 2010; Park et al., 2018).

Tobacco is an important crop that serves as both a model plant for scientific research and an economic crop. Individual *CAD* genes in tobacco have been extensively characterized in terms of mutant phenotypes, encoded protein activity, and lignin content (Halpin et al., 1992; Chabannes et al., 2001; Damiani et al., 2005; Weiller et al., 2020). However, studies on the expression patterns of *NtCAD* genes in tobacco remain limited. With the completion of tobacco genome sequencing, it now offers great convenience for the comprehensive analysis of the *CAD* gene family at a genomic scale. In this study, we identified 20 putative *CAD* genes from allotetraploid tobacco (*N. tabacum*) and two diploid tobacco species (*N. tomentosiformis* and *N. sylvestris*). Simultaneously, we investigated the genomic location, intron/exon organization, evolutionary relationships among members of the tobacco *CAD* gene family, and expression patterns across different tissues. Furthermore, systematic functional analysis was conducted on three putative *bona fide* NtCADs from *N. tabacum*. The results presented here provide *in vivo* and *in vitro* evidence to elucidate the function of *bona fide CAD* genes in regulating monolignol composition and stress responses in tobacco plants.

2 Materials and methods

2.1 Plant materials and growth conditions

The tobacco cultivar K326 (*N. tabacum*) obtained from the Yunnan Academy of Tobacco Agricultural Sciences (Kunming, China) was used in all experiments unless otherwise specified. K326 seeds were subjected to surface sterilization with 10% NaClO for 10 min, followed by thorough rinsing with distilled water multiple times. Subsequently, the seeds were germinated and cultivated in plastic pots under standard conditions (daily: 28 ± 1°C for 16 h under 200 μmol m^{-2} s^{-1} light, followed by 23 ± 1°C for 8 h of darkness) until reaching a stage with five to six true leaves (approximately 2 months old). Afterward, they were transplanted into open fields and grown for another 2 months. Then, various plant tissues, including young (leaf number 15, counted from bottom) and senescent (leaf number 5, counted from bottom) tobacco leaves, the veins of the young and senescent leaves, lateral and fibrous roots, stems (from leaf number 9 to leaf number 10, counted from bottom), and axillary and flower buds, were collected for RNA extraction at the flowering stage.

Four-week-old tobacco seedlings (with three to five leaves) were subjected to various abiotic and biotic stresses under identical growth conditions. These stresses included cold treatment (4°C, 24 h), heat treatment (40°C, 2 h), a 10-day drought stress, and a 5-day dark treatment. The conidia of *E. cichoracearum* DC., extracted from four severely infected tobacco leaves using 10 mL of sterile distilled water containing one drop of Tween 20 (Sigma-Aldrich, St.

Louis, USA), were subsequently inoculated onto the tobacco seedlings using the spray method. After the 10-day treatment, leaves were collected from the seedlings for further analysis. For abscisic acid (ABA), methyl jasmonate (MeJA), and salicylic acid (SA) treatments, the roots of the seedlings were immersed in solutions of 10 μM ABA, 50 μM MeJA, or 100 μM SA in 1/2 Hoagland solution for 2 h. As a control, seedlings were treated with only 1/2 Hoagland solution.

2.2 Phylogenetic and gene structure analyses

CAD sequences of other plant species were retrieved from the NCBI GenBank database using the keyword "cinnamyl alcohol dehydrogenase" as the query. DNAMAN (version 6.0) and ClustalX (version 1.83) were used for multiple alignments of *CAD* nucleotide and deduced amino acid sequences, respectively, with default gap penalties. A phylogenetic tree of CAD amino acid sequences was constructed using MEGA 7.0 with a neighbor-joining algorithm.

2.3 RNA extraction and cDNA synthesis

The SuperPure Plant Poly RNA Kit (Codon Biotechnology Co., Beijing, China) was used to extract total RNA from each biological sample, following the manufacturer's instructions. RNase-free DNase I (TaKaRa Bio Inc., Dalian, China) was used to eliminate DNA contamination during the extraction process. A NanoDrop 2000 instrument (Thermo Fisher Scientific, USA) was used to assess the quality and concentration of extracted RNA. For first-strand cDNA synthesis, 1 μg of total RNA was subjected to PrimeScript™ RT Reagent Kit (TaKaRa, Bio Inc., Dalian, China) with random primers.

2.4 Quantitative real-time PCR analysis

Gene expression levels in each sample were measured using quantitative real-time PCR (qRT-PCR) with the TB Green® Premix Ex Taq™ II Reagent (TaKaRa, Bio Inc., Dalian, China) and analyzed using a CFX96 Touch Real-Time PCR System (Bio-Rad, USA). Gene-specific primers were designed and synthesized (Supplementary Table 1), and the reference gene *L25* was used for normalization (Schmidt and Delaney, 2010). Relative expression levels of each gene in treated tobacco samples were determined by comparison with their corresponding control samples at specific time points or under specific conditions after normalization to *L25* transcript levels according to $2^{-\Delta\Delta Ct}$ method.

2.5 Cloning of tobacco *bona fide CAD* genes

PCR was used to amplify the open reading frames (ORFs) of NtCAD1–1, NtCAD1–2, and NtCAD2 from root tissue cDNA. Gene-specific primers (Supplementary Table 3) were employed for amplification, and a three-step program with a gradient annealing temperature ranging from 50°C to 60°C was utilized. The PCR products were purified by gel electrophoresis and cloned into the pMD-19T vector (TaKaRa Bio Inc., Dalian, China) for sequencing (Liuhe Huada Gene Technology Co., Ltd., Beijing, China).

2.6 Prokaryotic expression and purification of NtCADs

NtCAD1–1, NtCAD1–2, and NtCAD2 were cloned into the pET-28a (+) plasmid containing 6×His-tagged proteins using the ClonExpress Ultra One Step Cloning Kit (Vazyme Biotech Co., Ltd., Nanjing, China). After sequence confirmation, E. coli BL21 (DE3) cells were transformed with the resulting plasmids: pET-28(a)-NtCAD1–1, pET-28(a)-NtCAD1–2, and pET-28(a)-NtCAD2. The transformed cells were then cultured in LB media supplemented with kanamycin until reaching an optical density at 600 nm (OD600) of approximately 0.6 to 0.8 at 37°C. Protein expression was induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.2 mM and incubating at 15°C for 12 h. Cells were then harvested through centrifugation at 4,000 g and 4°C for 15 min and stored at −20°C until further use. His-tagged NtCADs were purified using a previously published method (Sun et al., 2018), followed by SDS-PAGE analysis and quantification using the Bradford Protein Assay Kit (Yeasen Biotechnology Co., Ltd., Shanghai, China).

2.7 Measurement of NtCAD activity

Enzymatic activity assays and kinetic determinations were performed according to established methods described in previous studies (O'malley et al., 1992; Chao et al., 2014). NtCAD1–1, NtCAD1–2, and NtCAD2 catalyzed three different substrates (coniferyl aldehyde, *p*-coumaryl aldehyde, and sinapyl aldehyde, respectively) obtained from Yuanye Biotechnology Co., Ltd. (Shanghai, China). The reduction in substrate concentration was measured using a SpectraMax Plus Microplate Reader (Molecular Devices, Shanghai, China). K_m and V_{max} values were determined by fitting the experimental data to the Michaelis–Menten equation using OriginPro 8.5 software (OriginLab Corporation, USA).

2.8 Structural modeling and docking analysis of NtCADs

The NtCAD1–1, NtCAD1–2, and NtCAD2 sequences were analyzed using SWISS-MODEL to search for templates and model their 3D structures, following established methods (Waterhouse et al., 2018). The crystal structure of AtCAD5 was used as the template to create the model. Subsequently, we obtained the 3D structures of coniferyl aldehyde and *p*-coumaryl aldehyde from the PubChem website (https://pubchem.ncbi.nlm.nih.gov/).

The PyMOL software was utilized for processing protein, including removing water molecules and adding hydrogen atoms. PyRx-0.8 (https://sourceforge.net/projects/pyrx/) was employed for molecular docking. Discovery Studio 4.5 (Accelrys Inc., San Diego, CA, USA) was used for the visualization of the 3D models and molecular docking structures.

2.9 Obtaining *NtCAD1* and *NtCAD2* VIGS transgenic tobacco

Transgenic tobacco plants with reduced *NtCAD1–1*, *NtCAD1–2*, and *NtCAD2* expression were created using virus-induced gene silencing (VIGS). The In-Fusion® Snap Assembly Master Mix from TaKaRa Bio Inc., Dalian, China, was used to insert interference fragments that target *NtCAD1–1*, *NtCAD1–2*, and *NtCAD2* into the pTRV2 vector. All primers used to construct the plasmids contained specific sequences corresponding to the genes of interest, along with a 20-bp recombination arm sequence derived from pTRV2 (Supplementary Table 2). Following purification of the fragments and Pst I digestion of the pTRV2 plasmids, ligation was performed using an infusion enzyme, according to the manufacturer's instructions. After ligation, the products were transformed into *E. coli* DH5α and confirmed by sequencing analysis. The recombinant plasmids pTRV2-NtCAD1–1, pTRV2-NtCAD1–1, pTRV2-NtCAD2, pTRV1, and pTRV2 (negative control) were transformed into *Agrobacterium tumefaciens* LBA4404 using the freeze-thaw method, as described previously (Cowan et al., 2023). The knockdown efficiencies of the three NtCADs were evaluated using qRT-PCR, and their expression levels were compared with those of the control group.

2.10 Determination of lignin content

Lignin content was determined using a Lignin Content Assay Kit (Sangon Biotech Co., Ltd., Shanghai, China), following the manufacturer's instructions. The determination of *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) lignin-derived monomers was carried out using established methods from previous studies (Besseau et al., 2007) by using a Thermo gas chromatograph-mass spectrometry (GC-MS) technology (TRACE 1310-ISQ, Thermo, USA). A Phloroglucinol assay kit (Yuanye Biotechnology Co., Ltd., Shanghai, China) was used to stain the cross-sectioned stems, which were obtained from the reddish-brown part of the stems or from the same part of other treatments of tobacco (*Nicotiana benthamiana*), in accordance with the manufacturer's instructions. Microscopic examination of the stained sections was performed using a Leica CH-9435 microscope (Heerbrugg, Germany), and a digital camera was used for image capture.

2.11 Statistical analysis

The results presented in this study are expressed as the means ± standard deviation (SD) of at least four independent experiments for each treatment. Statistical analysis was conducted using Duncan's multiple test ($P < 0.05$) or *t*-test ($P < 0.05$ or $P < 0.01$).

3 Results

3.1 Identification of *CAD* gene family members in tobacco

A search was conducted in the China Tobacco Genome Database v3.0 (unpublished data) using the keyword "cinnamyl alcohol dehydrogenase" as the query to identify putative *CAD* family genes in *N. tabacum*. Subsequently, BLAST searches were conducted using *Arabidopsis* CAD protein sequences as query proteins. After excluding lower-identity sequences, 10 *NtCAD* genes were obtained (Table 1); these *NtCAD* genes exhibited variable genomic DNA sizes ranging from 1,808 bp to 19,306 bp, with corresponding cDNA sequence sizes ranging from 945 bp to 1,407 bp. The predicted amino acids ranged from 315 to 469 (molecular weights ranging from 34.29 kDa to 51.81 kDa), with isoelectric point values varying between 7.03 and 8.07. Moreover, we also characterized four and six *CAD* genes in the genomes of *N. tomentosiformis* and *N. sylvestris*, respectively.

The DNAMAN algorithm was used to calculate the consistency of the 10 *NtCAD* coding and amino acid sequences. As presented in Table 2, *NtCAD* genes were classified into four pairs exhibiting high similarity (amino acid sequence identities greater than 72.92%), except Ntab0704510 and Ntab0181830, showing that most of the *NtCAD* genes have two potential paralogs. Furthermore, inter- and intraspecies sequence alignments of NtomCADs and NsylCADs were also performed. It showed that four pairs shared significant levels of sequence identities (60.89%–97.53% at the amino acid level) in interspecies alignment (Supplementary Table 3). Subsequently, we designated these 10 *NtCAD* genes in *N. tabacum*, and their putative orthologous genes in *N. sylvestris* and *N. tomentosiformis* (Table 3). These results suggest that both diploid genomes possess one ortholog for each *NtCAD* gene, except *NtCAD2* and *NtCAD4* in *N. tomentosiformis*. These homologous relationships were validated by subsequent phylogenetic analyses (Figure 1).

3.2 Alignment and chromosomal distributions of tobacco *CADs* across genomes

Alignment analysis demonstrated that all NtCADs exhibited a high degree of amino acid sequence conservation compared with the *bona fide Arabidopsis* CADs (AtCAD4 and AtCAD5). They shared identical residues at the critical Zn1 catalytic center (C47, H69, and C163) and displayed characteristic motifs commonly found in CAD proteins (Figure 2). Consistent findings were obtained from the analysis of CAD sequences in the two diploid tobacco plants (Supplementary Figure 1). The conserved motifs identified in all CADs included a GHE(X)2G(X)5G(X)2V motif and a GD(X)10C(X)2C(X)2C(X)7C motif, which are involved in the

10.3389/fpls.2024.1400213

TABLE 1 Characteristics of *CAD* genes in tobacco.

	Gene ID	gDNA size (bp)	CDS size (bp)	Number of exons	Amino acid (aa)	Molecular weight (kD)	Isoelectric point
Nicotiana tabacum	Ntab0704510	3,069	1,071	4	357	38.99	5.62
	Ntab0152820	1,808	1,065	6	355	38.69	7.47
	Ntab0529780	3,717	1,071	5	357	38.91	6.05
	Ntab0420320	4,381	1,092	5	364	39.62	7.29
	Ntab0336770	2,496	1,080	6	360	39.52	7.43
	Ntab0554550	1,965	1,080	6	360	39.54	7.87
	Ntab0262430	4,304	1,071	5	357	38.80	6.04
	Ntab0181830	5,510	1,080	6	360	39.50	7.31
	Ntab0868330	7,800	1,407	8	469	51.81	8.07
	Ntab0095940	19,306	945	4	315	34.29	7.03
Nicotiana tomentosiformis	Ntom0118530	1,234	1,068	7	355	38.48	7.05
	Ntom0224390	1,634	1,095	3	364	39.51	6.59
	Ntom0265640	3,741	1,074	5	357	38.89	5.84
	Ntom0311040	1,965	1,083	6	360	39.54	7.87
Nicotiana sylvestris	Nsyl0054440	4,381	1,095	5	364	39.62	7.30
	Nsyl0093350	2,496	1,083	6	360	39.52	7.43
	Nsyl0290350	3,213	1,074	4	357	38.99	5.62
	Nsyl0426890	11,748	1,683	9	560	62.25	4.65
	Nsyl0449860	4,332	1,074	5	357	38.78	6.05
	Nsyl0459450	8,277	1,089	6	362	39.90	7.34

binding of catalytic Zn^{2+} and structural Zn^{2+}, respectively. In addition, the GLGGV(L)G motif plays a role in NADP(H) cofactor binding. We also observed that NtCAD1–1, NtCAD1–2, NtCAD2, and their homologous genes (Nsyl0449860, Ntom0265640, and Nsyl0290350) exhibit a high degree of conservation in 11 proposed amino acids (T49, Q53, L58, M60, W119, V276, P286, M289, L290, F299, and I300) within the substrate-binding sites of *bona fide* CADs (Kim et al., 2004). In addition, we also found that NtCAD1–1, NtCAD1–2, and NtCAD2 all have three conserved amino acids (Thr49, His52, and Asp57) that play an important role in the reduction process of AtCAD5 (Youn et al., 2006). Conversely, other CADs displayed varying amino acid substitutions at the proposed residues (Figure 2; Supplementary Figure 1).

To determine the distribution of *CAD* genes on chromosomes, we retrieved physical maps of *CAD* genes from the China Tobacco Genome Database v3.0. Subsequently, we created simplified maps including information on chromosome length, start/end sites, and the numbers of *CAD* genes. As shown in Figure 3, chromosomes 1, 3, 6, 10, 11, 12, 14, 19, and 24 each harbored one *NtCAD* gene; additionally, one *NtCAD* gene was found to be located on a scaffold. Furthermore, in *N. sylvestris*, there are six *NsylCAD* genes, which were located on chromosomes 1, 2, 6, 7, 8, and 9, respectively (Supplementary Figure 2A), whereas four *NtomCAD* genes were located on chromosomes 5, 7, 10, and 12 (Supplementary Figure 2B) in *N. tomentosiformis*.

3.3 Intron-exon structure of tobacco *CAD* genes

To explore the origins of the identified tobacco *CAD* genes, the exon/intron structure of each gene was analyzed. As depicted in Figure 4, there was significant variation in intron length among NtCADs. Most CAD introns ranged from 80 bp to 2,000 bp, with exceptions in *NtCAD3–1* and *NtCAD6–2*, which had introns exceeding 2,000 bp. Similarly, diploid tobacco exhibited variations in intron length. *Nsyl0426890* and *Nsyl0459450* both have one intron exceeding 3,000 bp (Supplementary Figure 3). Moreover, the number of introns varied among the 10 *NtCADs*. For example, there are four introns in *NtCAD1–1*, *NtCAD1–2*, and *NtCAD6–1*, respectively, whereas only three introns were observed in *NtCAD2* and *NtCAD6–2*. In addition, *NtCAD3–2*, *NtCAD4*, and *NtCAD5* had five introns each, and *NtCAD3–1* had seven introns. Most *CAD* genes in diploid tobacco genomes exhibited a consistent number of introns with their corresponding orthologs in *N. tabacum*; however,

TABLE 2 Identity matrix of predicted *NtCAD* genes and their coding amino acid in *N. tabacum*.

		Amino acid identity									
		Ntab0262430	Ntab0529780	Ntab0704510	Ntab0336770	Ntab0554550	Ntab0181830	Ntab0420320	Ntab0095940	Ntab0868330	Ntab0152820
Nucleotide identity	Ntab0262430		94.96	82.91	39.39	39.39	46.11	50.83	43.85	35.38	47.49
	Ntab0529780	96.55		85.43	39.66	39.66	45.66	50.56	43.85	35.17	47.21
	Ntab0704510	83.52	84.54		37.15	37.15	43.33	47.78	41.34	33.05	44.13
	Ntab0336770	55.40	55.49	54.80		96.67	63.06	46.30	45.45	30.87	40.39
	Ntab0554550	56.05	55.77	55.72	96.12		63.33	46.03	45.15	31.08	40.67
	Ntab0181830	56.09	55.35	52.91	76.64	76.92		56.16	47.65	35.10	46.26
	Ntab0420320	55.20	54.52	54.84	59.02	58.83	60.38		80.00	39.71	52.33
	Ntab0095940	49.68	49.12	49.49	51.84	51.75	52.58	80.06		32.91	43.98
	Ntab0868330	42.57	42.99	41.23	42.38	42.45	42.05	43.23	38.06		72.92
	Ntab0152820	56.55	56.73	53.95	55.57	56.12	55.20	55.92	50.37	73.26	

The values in the boxes indicate a higher level of similarity.

10.3389/fpls.2024.1400213

TABLE 3 CAD orthologs among *N. tabacum*, *N. sylvestris*, and *N. tomentosiformis*.

Protein symbo	Protein symbo	*Nicotiana tabacum*	*Nicotiana sylvestris*	*Nicotiana tomentosiformis*
NtCAD1	NtCAD1–1	Ntab0529780	Nsyl0449860	Ntom0265640
	NtCAD1–2	Ntab0262430		
NtCAD2	NtCAD2	Ntab0704510	Nsyl0290350	NO
NtCAD3	NtCAD3–1	Ntab0868330	Nsyl0426890	Ntom0118530
	NtCAD3–2	Ntab0152820		
NtCAD4	NtCAD4	Ntab0181830	Nsyl0459450	NO
NtCAD5	NtCAD5–1	Ntab0554550	Nsyl0093350	Ntom0311040
	NtCAD5–2	Ntab0336770		
NtCAD6	NtCAD6–1	Ntab0420320	Nsyl0054440	Ntom0224390
	NtCAD6–2	Ntab0095940		

exceptions were noted for Nsyl0426890, Nsyl0054440, and their corresponding homologs, Ntom0118530 and Ntom0224390 (Figure 4; Supplementary Figure 3).

3.4 Phylogenetic analysis of NtCAD proteins

To better assess the evolutionary relationships among CAD proteins in tobacco and other plant species, 65 full-length CAD amino acid sequences from 13 distinct species were analyzed. The phylogenetic classification revealed that CAD proteins could be grouped into three subfamilies (Figure 1). Group I comprises CAD1 and CAD2, along with *bona fide* CAD proteins like AtCAD4, AtCAD5 (Kim et al., 2004). NtCAD3 was classified into group IIa, along with AtCAD1, OsCAD1, and OsCAD4, which have been reported to exhibit no detectable CAD activity (Kim et al., 2004; Park et al., 2018). Group IIb comprises NtCAD4, NtCAD5, and OsCAD6, which was demonstrated to have CAD activity toward hydroxycinnamaldehydes (Park et al., 2018). Additionally, CAD6, classified under group III, is associated with AtCAD2, AtCAD3, and AtCAD9, which participate in plant defense mechanisms or function as functionally redundant *CAD* genes (Kim et al., 2004). Consistent with previous findings, highly similar CAD protein sequences clustered together within the same subgroup.

3.5 Expression analysis of tobacco *CAD* genes

The physiological functions of genes can often be partially understood by analyzing their expression patterns. Therefore, qRT-PCR analysis was conducted to examine the expression levels of *NtCAD* genes in different tobacco tissues. As shown in Figure 5, *NtCAD1–1* and *NtCAD1–2* exhibit similar expression patterns, with higher expression levels were observed in lignified tissues such as roots and stems (Figures 5A, B). *NtCAD2* is mainly expressed in young leaf veins, fibrous roots, and flower buds (Figure 5C), suggesting probable tissue-specific functions for *NtCAD1* and *NtCAD2*. However, other *NtCAD* genes showed different expression preferences; for example, *NtCAD3–1* is preferentially expressed in roots (Figure 5D), *NtCAD3–2*, *NtCAD4*, *NtCAD6–1*, and *NtCAD6–2* in leaves (Figures 5F, I, J), as well as *NtCAD5–2* in stems (Figure 5H). In addition, the fold change of *NtCAD5* gene expression exceeds or is similar to that of *NtCAD1* gene expression in both leaves and roots. However, *NtCAD1–1* and *NtCAD1–2*, which belong to the *bona fide CAD* phylogenetic group, had the highest and second-highest expression levels in the roots and stems, respectively (Figure 5K). These two genes were expressed 11 to 25 times higher than *NtCAD2*, showing the mRNA transcript abundance of them was richer in such tissues, and implying that *NtCAD1–1* and *NtCAD1–2* are likely playing important roles in tobacco lignification.

3.6 Biochemical characterization of the recombinant NtCAD1 and NtCAD2 enzymes

Recombinant NtCAD1–1, NtCAD1–2, and NtCAD2 proteins were then successfully purified at concentrations ranging from 0.3 mg/mL to 0.7 mg/mL, with calculated theoretical molecular weights of 40.33 kDa, 40.22 kDa, and 40.42 kDa respectively. The successful expression of the fusion protein comprising NtCADs was confirmed by SDS-PAGE and western blot analyses (Supplementary Figures 4, 5). These results suggest that the purified NtCADs exhibited slight deviations from their predicted molecular weights (Table 1), potentially due to factors such as fusion expression with the His-tag, posttranslational modifications, or experimental variations.

To assess the pH dependence and temperature sensitivity of the reduction activity of NtCAD1–1, NtCAD1–2, and NtCAD2, we conducted *in vitro* experiments using coniferyl aldehyde as the substrate (Figures 6A–C). The optimal pH values for the reduction activity were determined to be 7.0, 7.0, and 3.0 for NtCAD1–1,

NtCAD1–2, and NtCAD2, respectively. Furthermore, the optimum temperatures were found to be 37°C, 37°C, and 30°C for NtCAD1–1, NtCAD1–2, and NtCAD2, respectively (Figures 6D–F).

The kinetic parameters of the reactions catalyzed by recombinant NtCAD1–1, NtCAD1–2, and NtCAD2 were determined using three distinct substrates: coniferyl aldehyde, *p*-coumaryl aldehyde, and sinapyl aldehyde (Table 4). All three enzymes exhibited NADPH-dependent reductase activity toward the assessed substrates. The K_m values of the detected substrates indicated a closer proximity in value between NtCAD1–1 and NtCAD1–2 (2.95 μM–4.17 μM for NtCAD1–1 and 0.71 μM–2.00 μM for NtCAD1–2, Table 4). However, the K_m values of NtCAD2 were significantly higher than those of NtCAD1–1 and NtCAD1–2, particularly for the sinapyl aldehydes (Table 4). In contrast, lower V_{max} values were observed for NtCAD1–1 than for NtCAD1–2 or NtCAD2 (Table 4). As a result, the catalytic efficiencies (K_{cat}/K_m values) of NtCAD1–2 were found to be more than 6 to 178 times higher than those of NtCAD1–1 and NtCAD2, suggesting that NtCAD1–2 exhibits significantly higher catalytic activity than NtCAD1–1 or NtCAD2.

3.7 Predicted protein structures of NtCAD1 and NtCAD2

Homology models of NtCAD1–1, NtCAD1–2, and NtCAD2 were constructed and evaluated. The modeled structures of NtCAD1–1, NtCAD1–2, and NtCAD2 did not change significantly (Supplementary Figure 6) and clearly showed properly sized binding pockets for NADPH and the cinnamyl aldehyde substrate (Figure 7; Supplementary Figure 6). The size of the substrate-binding pocket in each NtCAD is similar, and the residues involved in interacting with coniferyl aldehydes were conserved. The amino acids at positions 69, 95, and 121 play a crucial role in assisting other amino acids to secure the substrate by interacting with its Pi bond in the benzene ring (Figure 7). However, the docking conformations of substrates with NtCAD1 and NtCAD2 were different. The Tyr-121 of the benzene side chain in NtCAD2 was replaced with Cys-121 in NtCAD1–1 and NtCAD1–2, resulting in a reduced limiting effect on the substrate and exposing the cavity, further expanding the reaction cavity. When the substrate binds to the protein, the benzene ring plane reverses outward, and the oxymethyl group on the benzene ring extends into the interior of the reaction cavity. This conformation may allow the substrate to bind more stably to the reaction cavity and provide broad spectrum conditions for the substrate's oxymethyl side chain.

3.8 The effects of downregulation of *NtCAD1* and *NtCAD2* on lignin contents

Downregulation of *NtCAD1* and *NtCAD2* was achieved by VIGS performance. Reddish-brown stems were observed in the VIGS-NtCAD1–1 and VIGS-NtCAD1–2 transgenic tobacco plants (Figures 8B, C). It is worth noting that only a small portion of the stems in VIGS-NtCAD1–1 transgenic tobacco plants exhibited red coloration in comparison with the VIGS-NtCAD1–2 transgenic tobacco. This observation aligns with the fact that NtCAD1–1 enzyme activity is lower than that of NtCAD1–2. The extent of downregulation of *NtCAD1–1*, *NtCAD1–2*, and *NtCAD2* in transgenic tobacco was quantified using qRT-PCR (Figures 8E–G). Significant downregulation of *NtCAD1–1* was observed in leaves and stems. Despite the significant decrease in *NtCAD1–1* expression (ranging from 43.07% to 70.55% in leaves and 62.74% to 79.14% in stems), there was no noticeable change in lignin content in these tissues (Figure 8H). Moreover, the downregulation of *NtCAD1–2* and *NtCAD2* in leaves and stems did not significantly affect lignin content (Figures 8I, J). Phloroglucinol staining was used to assess the lignin content in the secondary xylem of stems, also revealing no significant decrease (Figures 8A–D). Subsequently, the contents of *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) lignin-derived monomers in *NtCADs*-silenced plants were also detected by using GC-MS. As shown in Table 5, the lignin consists of a high content of the S unit when compared with the levels of the H and G units in whether control or *NtCAD*-silenced plants. In addition, the contents of both G and S monomers were significantly decreased when the *NtCAD* genes were silenced. However, the S/G ratio only decreased in *NtCAD1*-silenced plants.

3.9 Diverse roles of *NtCAD1* and *NtCAD2* in response to abiotic and biotic stresses

Previous studies have indicated that *CAD* homologs play diverse roles in various plant biological processes, including their involvement in stress responses (Chao et al., 2022). To further investigate the potential role of *NtCAD1–1*, *NtCAD1–2*, and *NtCAD2* involved in plant regulatory networks, the 1,500-bp-length promoter regions upstream of the start codon (ATG) of each gene was analyzed. Subsequently, we analyzed these promoter sequences for cis-regulatory elements using the PlantCARE database (Lescot et al., 2002) and identified several cis-elements associated with plant growth and development (e.g., light responsiveness), biotic and abiotic stresses (e.g., cold stress, anaerobic induction), and phytohormone responsiveness including ABA, MeJA, and SA (Supplementary Table 4). In addition to conserved *cis*-elements shared with *NtCAD1–1* and *NtCAD1–2*, *NtCAD2* also possesses distinctive motifs, such as TC-rich repeats, which have been implicated in defense mechanisms and stress response (Zhang et al., 2021).

The expression levels of *NtCAD* genes in response to abiotic and biotic stresses were subsequently determined. Notably, *NtCAD1–1* exhibited significant downregulation or minor effects in response to all tested biotic and abiotic stresses (Figures 9A, B, G). Distinct response patterns were observed for *NtCAD1–2* and *NtCAD2* in the leaves and roots (Figures 9C–F). Specifically, under heat and dark stress conditions, the expression of *NtCAD1–2* was significantly induced in roots but decreased in leaves (Figures 9C, D). Conversely, when exposed to cold or heat stress conditions, the expression of *NtCAD2* was significantly increased only in the roots; however, it displayed either significant downregulation or minor

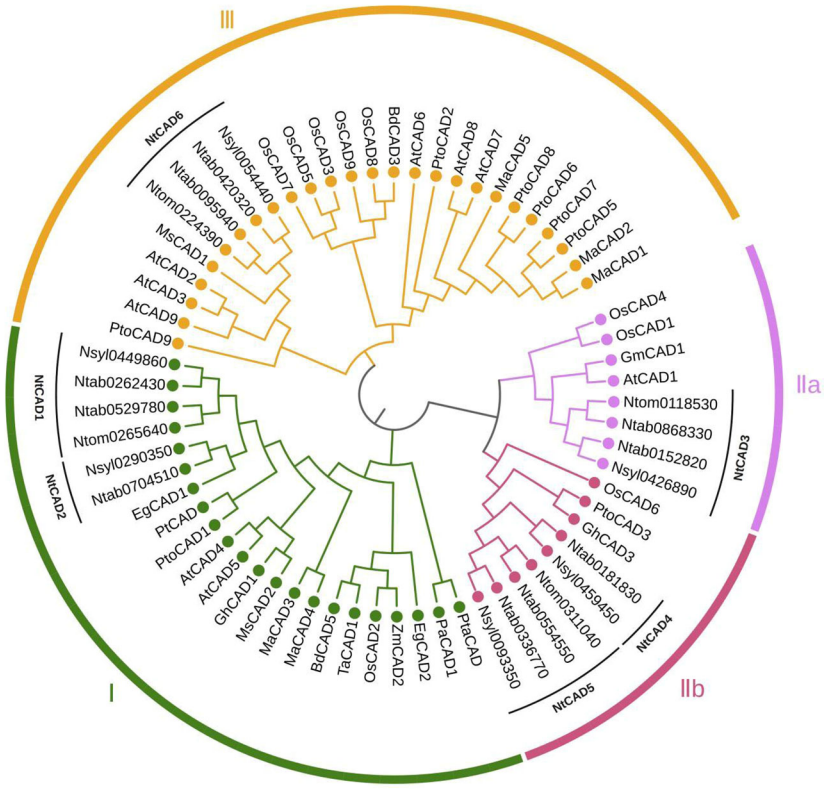


FIGURE 1
Phylogenetic analysis of tobacco CAD isoforms and other plant CAD homologs. The corresponding plant species and isozymes investigated in this study include tobacco (Ntab, Ntom, and Nsyl); *Arabidopsis thaliana*, AtCAD1–9 (accession number: Q9CAI3, Q9SJ25, Q9SJ10, P48523, O49482, O65621, Q02971, Q02972, P42734); *Oryza sativa*, OsCAD1–9 (Q8H859, Q6ZHS4, Q337Y2, Q2R114, Q0J6T3, Q7XWU3, Q0JA75, Q6ERX1, Q10PS6); *Brachypodium distachyon*, BdCAD3 (AFK80371), BdCAD5 (AFK80372); *Triticum aestivum*, TaCAD1 (GU563724); *Elaeis guineensis*, EgCAD1 (UTE99576), EgCAD (XP_010943210); *Medicago sativa*, MsCAD1 (O82515), MsCAD2 (P31656); *Morus alba*, MaCAD1–5 (UZH97791, UZH97792, UZH97793, UZH97794, UZH97795); *Picea abies*, PaCAD1 (CAA05097); *Populus tremuloides*, PtCAD (AAF43140); *Pinus taeda*, PtaCAD (CAA86073); *Populus tomentosa*, PtoCAD1, 2, 3, 5, 6, 7, 8, 9 (AGU43755, AGU43756, AGU43757, AGU43758, AGU43754, AHX58273, AGU43750, AGU43751); *Zea mays*, ZmCAD2 (CAA74070); *Glycine max*, GmCAD1 (XP_003543132); *Gossypium hirsutum*, GhCAD1 (ABZ01817).

effects in leaves under the same stress conditions. With *Erysiphe cichoracearum* DC infection, diseased leaves exhibited a significant increment in *NtCAD1–2* expression compared with healthy leaves, whereas *NtCAD1–1* and *NtCDAD2* exhibited notable decreases (Figure 9G). Meanwhile, the *NtCAD*-silenced transgenic tobacco plants (VIGS-NtCAD1–1, VIGS-NtCAD1–2, and VIGS-NtCAD2) were then inoculated with *E. cichoracearum* DC. for 12 days. As expected, *E. cichoracearum* DC mycelium and spores were observed in all the inoculated plants. However, the symptoms were more severe in the *NtCAD*-silenced (especially for *NtCAD1–2*-silenced) transgenic tobacco than in the control plants. Similar results were also observed in withered leaves when inoculated with *E. cichoracearum* DC (Supplementary Figure 7).

Further investigations were conducted to analyze the expression patterns of these genes following ABA, MeJA, and SA treatment, considering the presence of specific motifs within their promoter regions known to be responsive to these treatments (Supplementary Table 4). Both *NtCAD1–1* and *NtCAD1–2* were significantly downregulated under ABA, MeJA, and SA treatments. Interestingly, the expression of *NtCAD2* was upregulated under both ABA and SA treatments (Figure 9H). These findings highlight the functional divergence of these three *NtCADs* in response to biotic and abiotic stresses in tobacco.

4 Discussion

The final step of the monolignol biosynthesis pathway is critically influenced by CAD, which is primarily responsible for lignin deposition and the formation of lignin components in secondary cell walls (Özparpucu et al., 2018). Comparative genomic analysis has been used to investigate multiple *CAD* gene families across diverse plant species, including *Arabidopsis*, rice, poplar, and oil palm. In this study, we identified 10 *CAD* genes in *N. tabacum*, 4 in *N. tomentosiformis*, and 6 in *N. sylvestris*. These tobacco CAD members exhibited significant similarities in their nucleotide and amino acid sequences, leading us to name them CAD1-CAD6 based on their unique molecular signatures. We further analyzed the genetic structure, evolutionary relationships, and expression patterns of NtCADs in different tobacco tissues. Additionally, enzyme activity assays and structural modeling were performed to characterize the putative *bona fide* NtCADs.

10.3389/fpls.2024.1400213

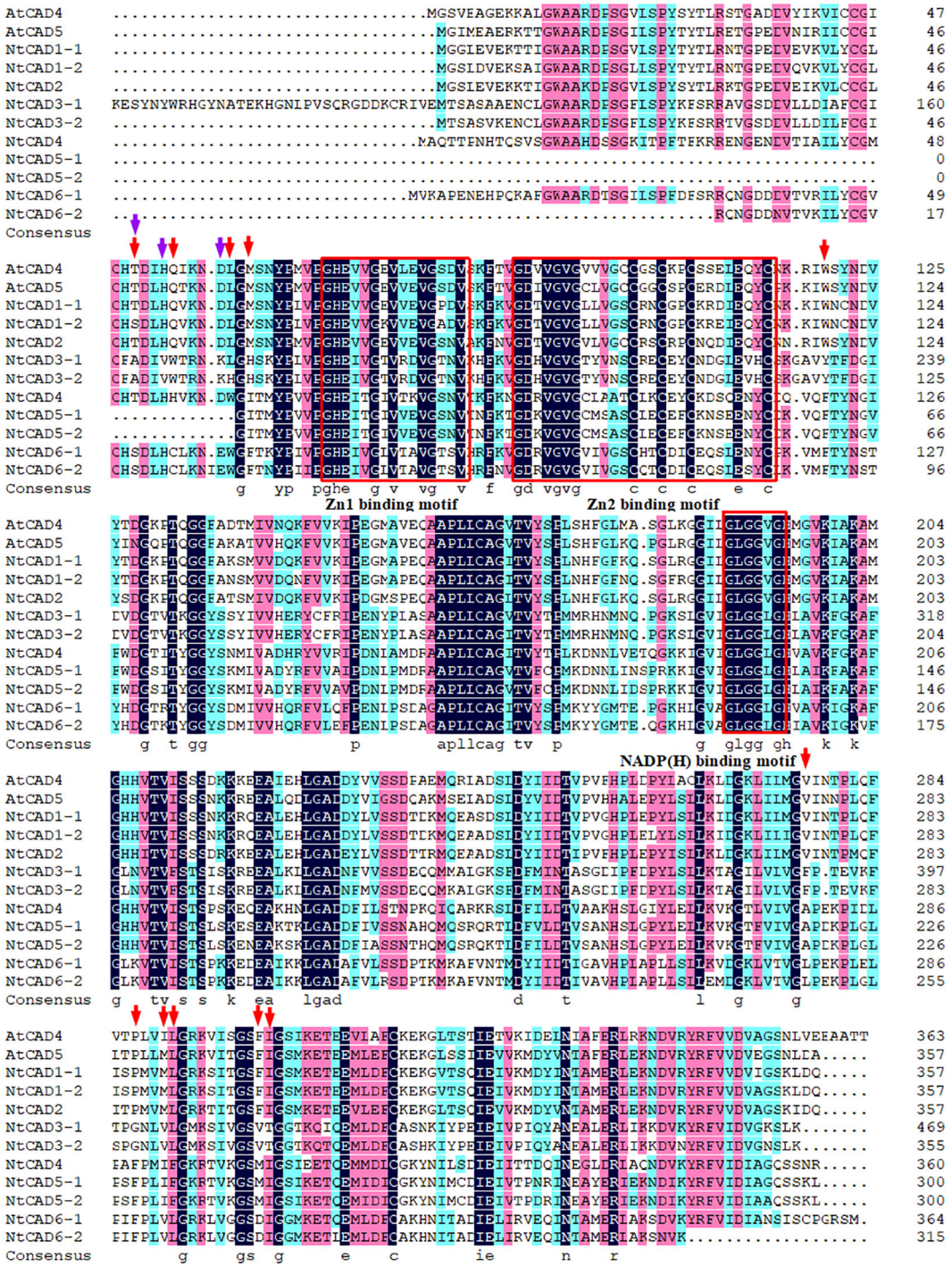


FIGURE 2
Alignment of NtCADs and *bona fide* Arabidopsis CADs. The *bona fide* AtCADs AtCAD4 (At3g19450) and AtCAD5 (At4g34230) were used as reference sequences. Zn-binding motifs are highlighted in red boxes, and the NADP(H)-binding motif is also indicated. The red and purple arrows denote the presence of conserved amino acids.

Furthermore, we investigated the expression patterns of these three *bona fide* NtCADs under various biological and abiotic stresses, as well as in response to phytohormones. Therefore, this comprehensive functional analysis provides insights into the potential roles of *bona fide* NtCADs in various growth and developmental processes in tobacco.

4.1 Evolutionary conservation and divergence among tobacco *CAD* genes

By conducting a comprehensive analysis of the *CAD* gene family in tobacco, including examining conserved motifs, exon/intron gene structures, and phylogeny, researchers can derive generalizations and

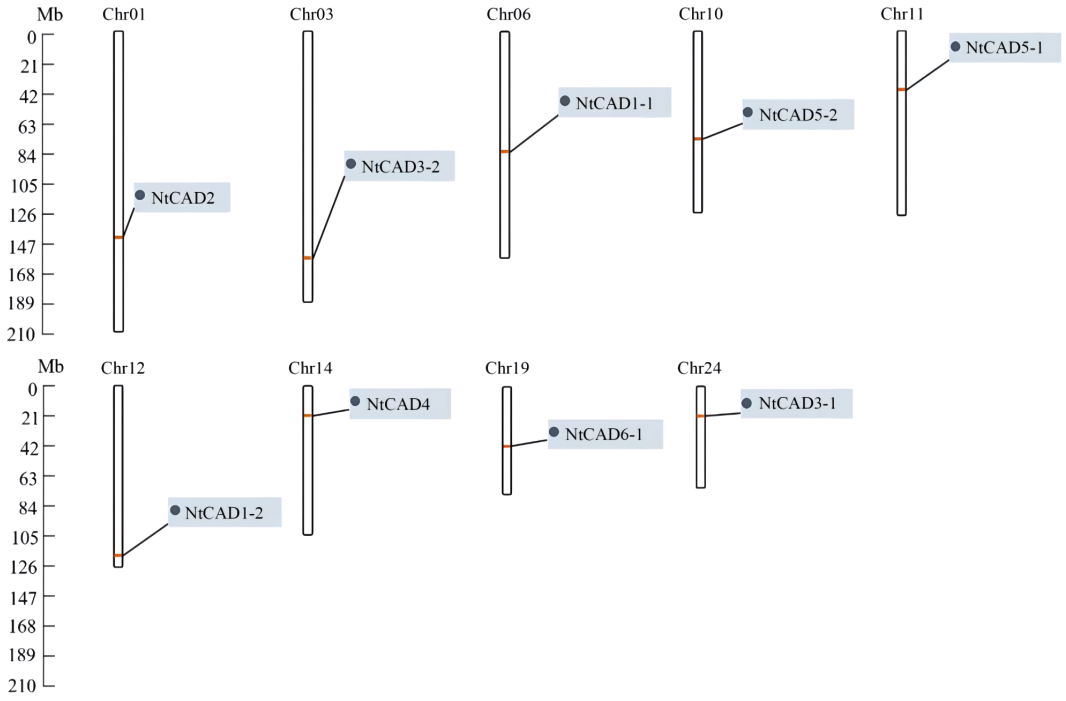


FIGURE 3
Localization of *CAD* gene family members on chromosomes in *Nicotiana tabacum*.

predictions regarding genetic and evolutionary relationships among uncharacterized members of this gene family. Moreover, this analysis facilitates the prediction of their potential functions. In this study, we conducted a phylogenetic analysis of CAD homologs from three tobacco species and compared them with those from 13 other plant species. Consistent with previous investigations (Barakat et al., 2009), our analysis classified tobacco CADs into three distinct clusters (Figure 1). CAD group I comprises *bona fide* CADs involved in lignin biosynthesis (Guo et al., 2010; Hirano et al., 2012; Liu et al., 2021), encompassing sequences from gymnosperm clades, monocots, and eudicots. Groups II and III contain sequences from both monocots and eudicots. This suggests that the development of group II and group III occurred in the precursor of angiosperms, or at least before the divergence of monocots and dicots. Our findings are consistent with those of previous analyses that classified the *CAD* genes in *Populus* into three classes, with gymnosperm sequences clustering in group I alongside monocots and eudicots (Barakat et al., 2009). However, the phylogenetic tree obtained in this study differs from a previously

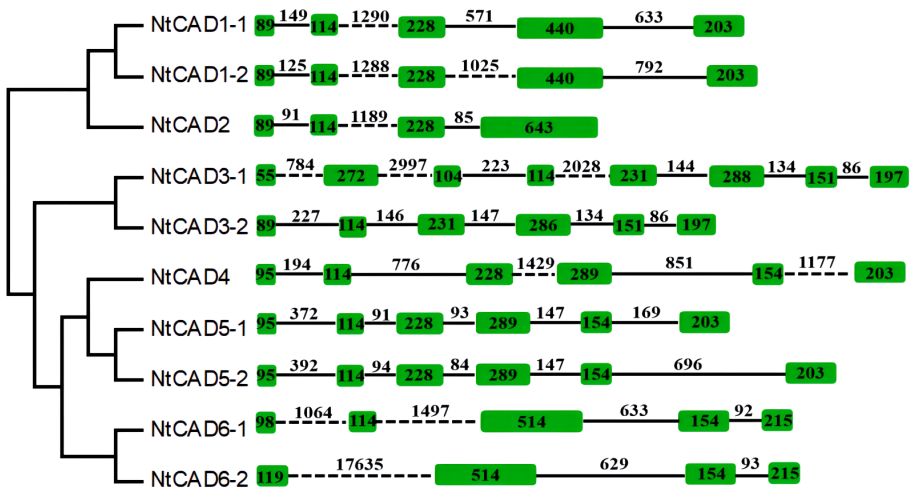


FIGURE 4
Analysis of the exon/intron structural organization of the 10 *NtCAD* genes. Introns and exons are indicated by solid lines and open boxes, respectively.

10.3389/fpls.2024.1400213

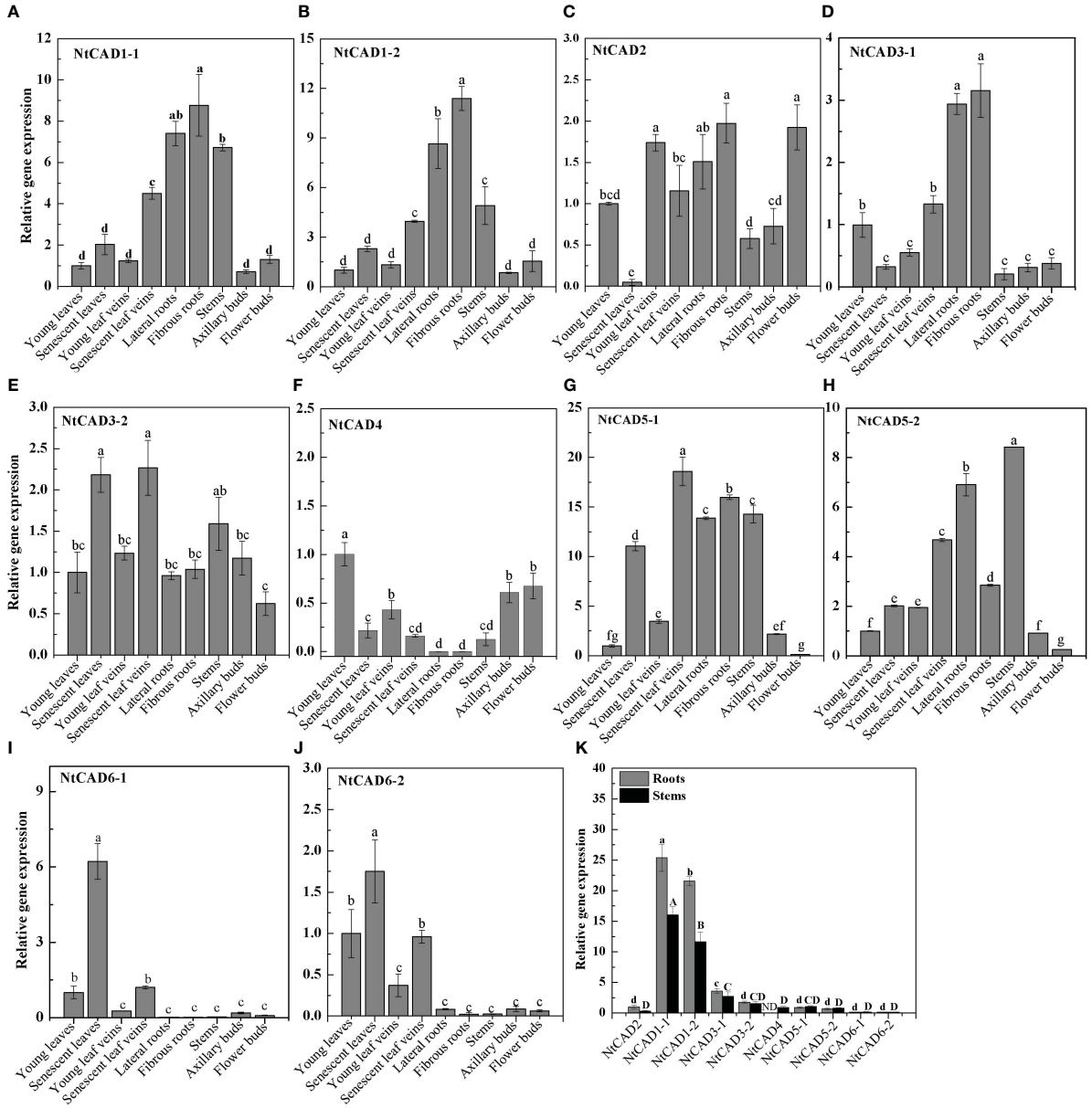


FIGURE 5
Expression profiles of *NtCAD* genes in different tobacco tissues at the flowering stage. Different tobacco tissues including young and senescent leaves, young and senescent leaf veins, lateral and fibrous roots, stems, and axillary and flower buds were samples from 4-month-old tobacco with half of the blooming flowers. Afterward, the expression levels of *NtCAD* genes **(A–J)** were detected by qRT-PCR. In addition, relative expression levels of *NtCAD* genes in roots and stems **(K)** were at the flowering stage. The tobacco *L25* gene was used as the internal control. Bars labeled with different letters indicate significant differences at $P < 0.01$ based on Duncan's multiple test.

published tree, which classifies CADs into two or five main groups (Tobias and Chow, 2005; Hu et al., 2022), as well as from a previously published *Arabidopsis* tree, which demonstrates a distinct grouping of gymnosperm CAD sequences (Raes et al., 2003). This discrepancy observed between our phylogenetic tree and those previously published (Raes et al., 2003; Tobias and Chow, 2005; Hu et al., 2022) can be attributed to our incorporation of a more comprehensive range of species in this investigation.

Previous studies have demonstrated that green algae CADs typically possess seven to eight introns, whereas most land plants CADs exhibit fewer than six introns, such as *Populus*, watermelon, and pomegranate (Barakat et al., 2009; Guo et al., 2010; Jin et al., 2014; Hu et al., 2022). This finding is consistent with previous research indicating the complex structure of early eukaryotic genes (Roy and Gilbert, 2005). In this study, the structural analysis of isolated tobacco *CAD* genes revealed a varying number of introns, ranging from two to eight per gene (Figure 4; Supplementary Figure 3). However, most *CAD* genes in tobacco, like those in other land plants, typically possess four or five introns. The prevalence of low rates of intron gain and reduction in eukaryotic evolutionary processes (Roy and Gilbert, 2005; Roy and Penny, 2007) leads to the hypothesis that *CAD* genes with fewer introns are subjected to stronger selective pressure.

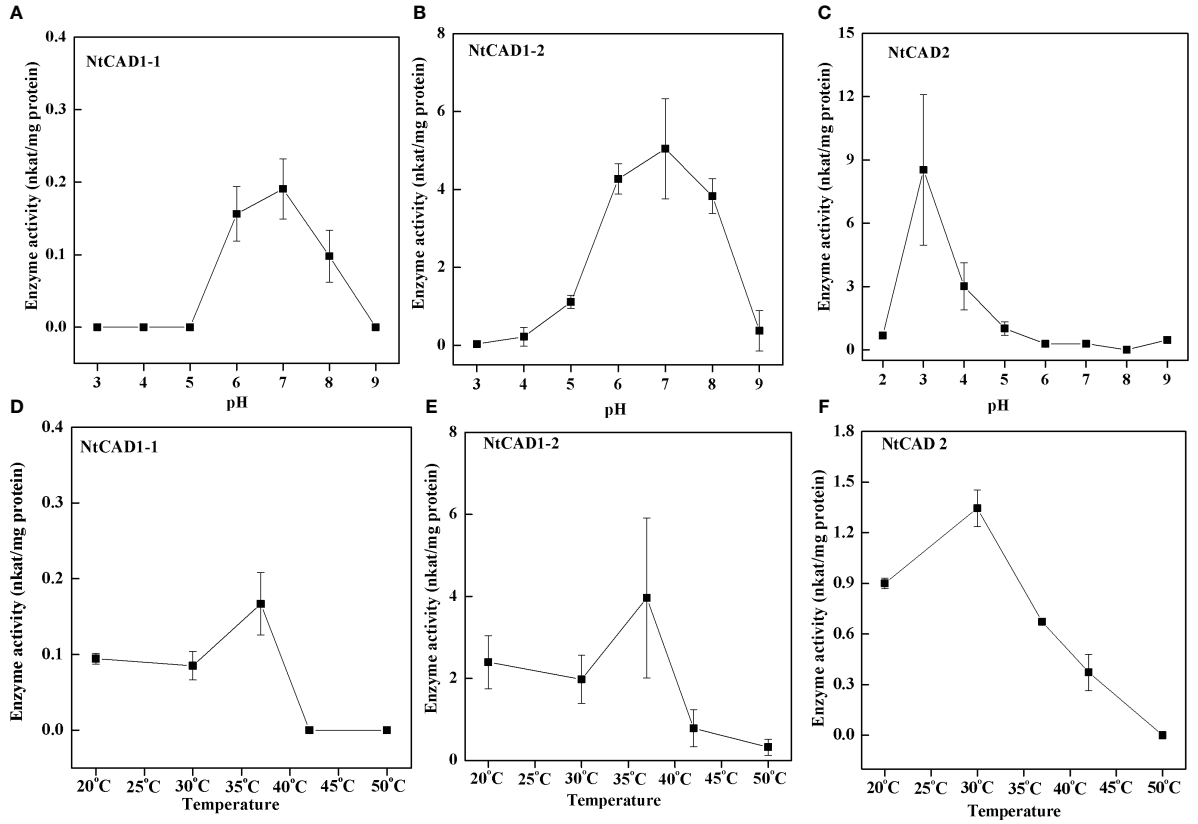


FIGURE 6
Optimal pH and temperature conditions for NtCAD1 and NtCAD2 utilizing coniferyl aldehyde as a substrate. **(A–C)**, pH profiles of NtCAD1–1 **(A)**, NtCAD1–2 **(B)**, and NtCAD2 **(C)** activities; **(D–F)**, Temperature profiles of NtCAD1–1 **(D)**, NtCAD1–2 **(E)**, and NtCAD2 **(F)**.

4.2 NtCAD1 is the main CAD involved in lignin biosynthesis in tobacco

Typically, one or two *bona fide* CADs are identified to fulfill essential functions in lignification (Kim et al., 2004; Tobias and Chow, 2005; Ma, 2010; Chao et al., 2014). Kim et al. (2004) reported that six AtCADs exhibit *in vitro* activity against hydroxycinnamaldehydes; however, only two *bona fide* AtCADs, AtCAD4 and AtCAD5, were identified as the predominant CADs involved in lignification. In *Populus*, PtrSAD, PtoCAD1, PtoCAD2, and PtoCAD8 demonstrate *in vitro* activity against hydroxycinnamaldehydes (Bomati and Noel, 2005; Chao et al., 2014); however, only PtoCAD1, which is classified within the *bona fide* clade, participates in the process of lignification. Similarly, in mulberry, MaCAD1, MaCAD2, MaCAD3/4, and MaCAD5 demonstrate activity against hydroxycinnamaldehydes *in vitro*. Among these, MaCAD3/4, classified as a *bona fide* member of

TABLE 4 Kinetic analysis of NtCAD1 and NtCAD2.

Protein symbol	Substrate	K_m (μM)	V_{max} (nkat mg^{-1} protein)	K_{cat} (s^{-1})	K_{cat}/K_m (μM^{-1} s^{-1})
NtCAD1–1	Coniferyl aldehyde	2.95	0.16	0.007	0.002
	p-Coumaryl aldehyde	4.17	0.11	0.004	0.001
	Sinapyl aldehyde	4.01	0.15	0.006	0.001
NtCAD1–2	Coniferyl aldehyde	1.00	5.83	0.23	0.230
	p-Coumaryl aldehyde	2.00	6.79	0.27	0.135
	Sinapyl aldehyde	0.71	4.76	0.19	0.268
NtCAD2	Coniferyl aldehyde	4.97	2.48	0.10	0.020
	p-Coumaryl aldehyde	7.95	4.17	0.17	0.021
	Sinapyl aldehyde	18.75	3.74	0.15	0.008

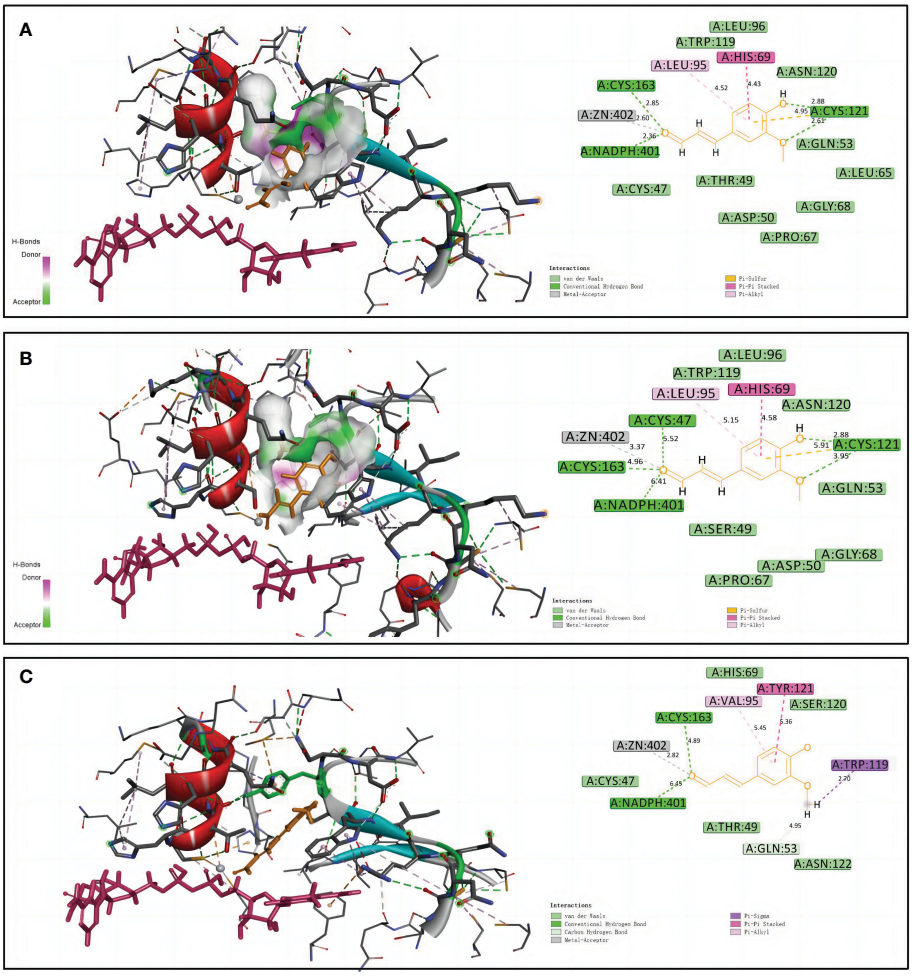


FIGURE 7
Structural modeling and docking analysis of NtCAD1–1, NtCAD1–2, and NtCAD2. Docking analysis was performed for NtCAD1–1 **(A)**, NtCAD1–2 **(B)**, and NtCAD2 **(C)** with coniferyl aldehyde as a ligand. Coniferyl aldehyde is depicted in orange, NADPH is represented in red, and the Zn atoms are illustrated in gray.

the CAD family, is considered the primary CAD involved in lignification (Chao et al., 2014). In the present study, NtCAD1–1, NtCAD1–2, and NtCAD2, belonging to group I, emerged as the most promising *bona fide* CAD candidates in tobacco. NtCAD1–1 and NtCAD1–2 showed the highest and second-highest expression (11- to 25-fold higher than NtCAD2) among *NtCADs* in roots and stems, which accumulate lignin (Figure 5). Interestingly, similar expression patterns have been observed for other *bona fide* CADs, such as rice *OsCAD2*, which also demonstrates the highest expression levels in both roots and stems (Tobias and Chow, 2005). In contrast, wheat *bona fide TaCAD1* exhibits high expression exclusively in stem tissues (Ma, 2010). These findings suggest that *NtCAD1–1* and *NtCAD1–2* may play significant roles in lignification.

Enzymatic activity analysis confirmed the hypothesis that *NtCAD1* is the predominant *CAD* gene in tobacco. All three recombinant proteins exhibited significant *in vitro* activity against hydroxycinnamaldehydes (Table 4). Notably, NtCAD1–2 displayed catalytic activity ranging from 115 to 268 times higher than that of NtCAD1–1 (Table 4). These findings are consistent with those of previous studies in *Arabidopsis*, where AtCAD5 showed approximately 270 times higher catalytic activity than AtCAD4 (Kim et al., 2004). However, the catalytic efficiency of NtCAD1–2 was comparable with well-established CADs from other plant species, such as rice and mulberry (Hirano et al., 2012; Chao et al., 2022). The enzymatic conversion of coniferyl aldehyde to coniferyl alcohol using expressed and purified NtCAD1–1, 1–2, and 2 proteins showed significant pH dependence (Figure 6), with optimal activity observed at pH 7.0, 7.0, and 3.0 respectively. The difference in the optimal pH between BdCAD3 and BdCAD5 in *Brachypodium distachyon* can be partially attributed to the variation in the motif at positions 57–58 (Bukh et al., 2012). BdCAD3 possessed an EW motif, whereas BdCAD5 possessed an HL motif. The residue in AtCAD5 was identified as Asp, which is potentially crucial for the catalytic mechanism (Youn et al., 2006; Saathoff et al., 2011). CADs possessing the HL or DL motif and demonstrating notable catalytic activity toward monolignols are linked to cell wall lignification (Saathoff et al., 2011). However, the presence of a DL motif in all three NtCADs implies that additional residues may contribute to regulating optimum catalytic pH. Previous studies indicated that the diverse catalytic capabilities and functions of

10.3389/fpls.2024.1400213

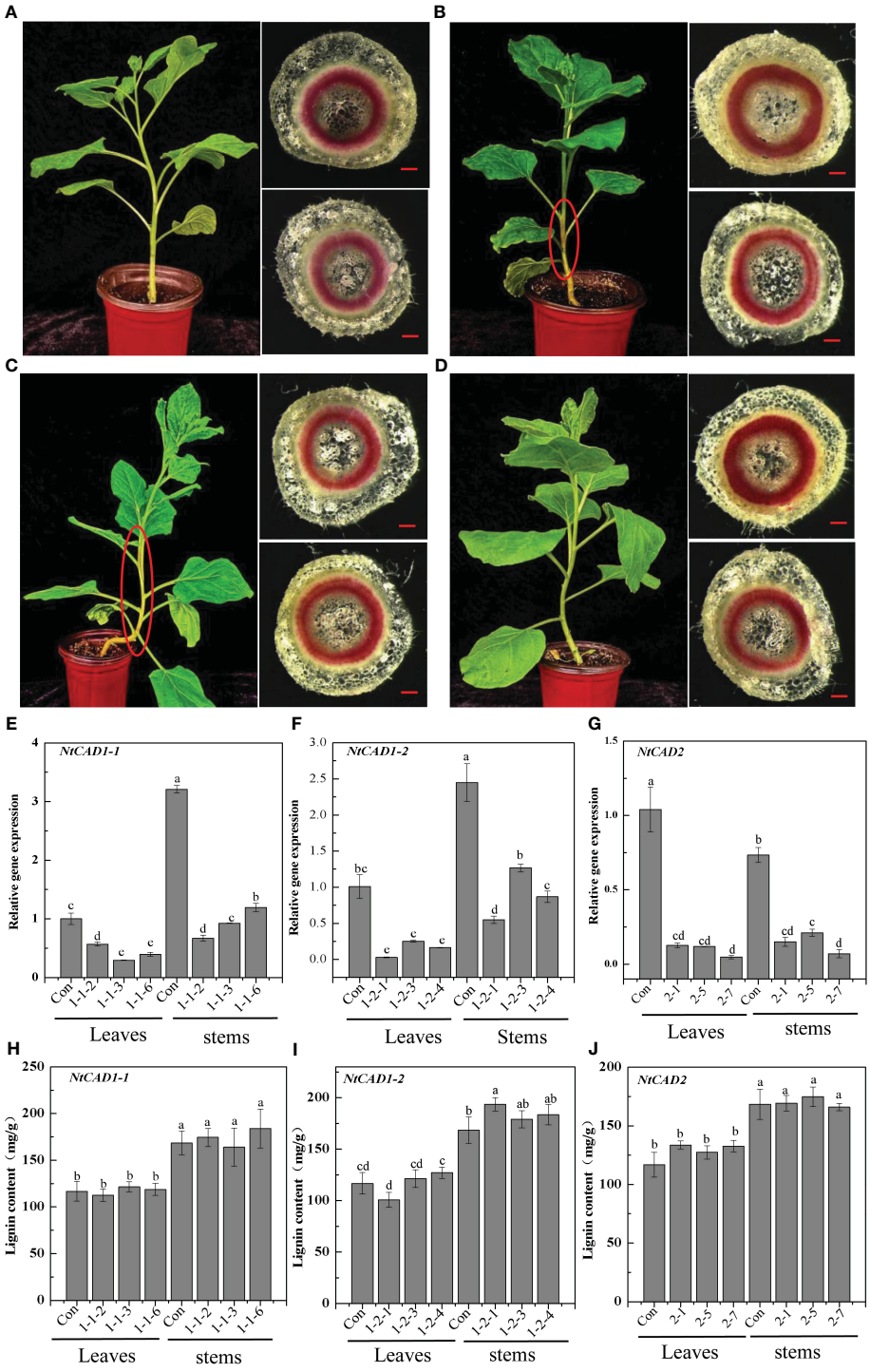


FIGURE 8
The downregulation of *NtCAD1–1*, *NtCAD1–2*, and *NtCAD2* does not exert any significant impact on the lignin content. **(A–D)**. Plant growth conditions and lignin phloroglucinol staining of transgenic tobacco with downregulated *NtCAD1–1*, *NtCAD1–2*, and *NtCAD2* using VIGS. Representative images of the Con (pTRV2, A), VIGS-NtCAD1–1 **(B)**, VIGS-NtCAD1–2 **(C)**, and VIGS-NtCAD2 **(D)** groups are shown. The scale bars indicate 40 μm; the red-brown stem is highlighted by a red circle in **(B, C)**; the expression levels of NtCAD1–1 **(E)**, NtCAD1-2 **(F)**, and NtCAD1–1 **(G)** in transgenic tobacco leaves and stems were assessed, whereas the lignin content of NtCAD1–1 **(H)**, NtCAD1-2 **(I)**, and NtCAD2 **(J)** was measured in both transgenic tobacco leaves and stems. Bars denoted by different letters indicate significant differences at $P < 0.05$, as determined by Duncan's multiple-range test.

TABLE 5 Gas chromatography-mass spectrometry identification and levels of the lignin monomers accumulated in stems of *NtCAD1* and *NtCAD2*-deficient tobaccos. Data are means ± SD of at least four independent samples.

	H+G+S (mg/g DW)	G (mg/g DW)	H (mg/g DW)	S (mg/g DW)	S/G ratio
Con	80.53 ± 1.93	16.61 ± 0.32	0.09 ± 0.00	63.83 ± 1.65	3.84 ± 0.05
NtCAD1–1-1	61.28 ± 2.15 **	15.69 ± 0.45 **	0.06 ± 0.01	45.52 ± 0.88 **	2.90 ± 0.03 **
NtCAD1–1-2	65.46 ± 1.85 **	16.01 ± 0.23 **	0.06 ± 0.01	49.38 ± 1.56 **	3.08 ± 0.07 **
NtCAD1–1-3	62.45 ± 2.47 **	15.54 ± 0.21 **	0.06 ± 0.00	46.84 ± 1.32 **	3.01 ± 0.01 **
NtCAD1–2-1	62.45 ± 1.18 **	13.97 ± 0.27 **	0.05 ± 0.00	48.43 ± 1.96 **	3.46 ± 0.14 *
NtCAD1–2-2	62.25 ± 1.45 **	13.88 ± 0.32 **	0.05 ± 0.01	48.31 ± 2.01 **	3.47 ± 0.11 **
NtCAD1–2-3	60.31 ± 2.01 **	13.41 ± 0.47 **	0.06 ± 0.00	46.84 ± 2.00 **	3.49 ± 0.05 **
NtCAD2–1	70.16 ± 3.21 **	14.84 ± 0.28 **	0.06 ± 0.00	55.27 ± 1.47 **	3.72 ± 0.14
NtCAD2–2	64.28 ± 1.25 **	13.50 ± 0.35 **	0.06 ± 0.00	50.71 ± 1.14 **	3.75 ± 0.12
NtCAD2–3	69.03 ± 2.14 **	14.29 ± 0.24 **	0.06 ± 0.01	54.67 ± 1.05 **	3.82 ± 0.09

Asterisks within columns are significantly different in comparison with corresponding Con samples at $P < 0.05$ and $P < 0.01$ (t test).

various CAD- and CAD-like enzymes in plants are not solely dictated by substrate-binding sites (Bukh et al., 2012). Similarly, we observed that despite the conservation of residues in the substrate-binding site of both NtCAD1–1 and NtCAD1–2, a single residue alteration has the potential to modulate enzymatic catalytic activity by inducing changes in the orientation of the substrate benzene rings (Figure 7).

Knocking down the *NtCAD1–1*, *NtCAD1–2*, and *NtCAD2* genes resulted in only a minimal or negligible reduction in lignin content (Figures 8H–J). Additionally, reddish-brown stems were observed in plants with knocked-down NtCAD1–1 or NtCAD1–2 (Figures 8B, C). This finding is consistent with previous research conducted on tobacco, where CAD downregulation plants exhibited a distinct red coloration of the xylem but showed only a slight decrease in lignin content (Stewart et al., 1997; Bernard Vailhé et al., 1998). Earlier studies provided evidence suggesting that the conversion of cinnamaldehyde to cinnamyl alcohols is not the rate-limiting step in lignin biosynthesis (Higuchi et al., 1994). Furthermore, disrupting the expression of AtCAD5 had minimal impact on overall lignin deposition (Kim et al., 2004). The same phenomenon was observed in alfalfa, wherein the downregulation of the CAD enzyme did not result in any significant alteration in lignin content (Baucher et al., 1999). This is primarily attributed to the compensatory action of alternative phenolic compounds and other CAD homologs (Guo et al., 2010). However, some studies have demonstrated that suppressing CAD leads to alterations in the lignin composition of transgenic plants while leaving the lignin content unaffected (Baucher et al., 1999; Fornalé et al., 2012; Trabucco et al., 2013). Therefore, the red stem phenotype generated by the knockdown of NtCAD1–1 or NtCAD1–2 (Figures 8B, C) may be attributed to the accumulation or enrichment of coniferyl aldehydes (Li et al., 2008b). Considering the high level of gene expression and enzyme activity of NtCAD1 coupled with concomitant phenotypic alterations, it is postulated that NtCAD1 may function as the main CADs in tobacco.

4.3 *Bona fide* NtCADs exhibit distinct expression patterns in response to abiotic and biotic stresses

Previous research indicates that certain *bona fide* CADs are not only involved in lignin biosynthesis but also actively participate in responding to both abiotic and biotic stresses. In flax, LuCAD1 and LuCAD2, classified as *bona fide* members of class I, have been found to respond to drought and cold stress, except for their role in lignification of maturing stems (Preisner et al., 2018). Similarly, the *bona fide* OsCAD2 in rice is induced by both biotic and abiotic stresses like pathogen infection and UV irradiation, implying that OsCAD2 contributes to developmental lignification and stress responses (Park et al., 2018). Our findings align with these studies, indicating that NtCAD1–2 is significantly upregulated in response to heat, dark stress, and pathogen infection. Furthermore, NtCAD2 demonstrates enhanced expression levels in roots under cold and heat stress conditions and in response to ABA and SA treatments, both pivotal in biotic and abiotic stress responses (Wu et al., 2019; Yu et al., 2019). However, we also observed a negative correlation between NtCAD1–1 expression and responses to abiotic and biotic stresses, as well as phytohormones. Similar to the mulberry findings reported by Chao et al. (2022), *bona fide* MaCAD3/4 expression was negatively associated with responses to abiotic and biotic stresses. These findings suggest that NtCAD1–2 functions as the main CAD, also participating in both abiotic and biotic stresses, whereas NtCAD2 may play a significant role in the response to various stresses.

5 Conclusion

In this study, 10 *CAD* genes were identified in *N. tabacum*, 4 in *N. tomentosiformis*, and 6 in *N. sylvestris*. Through genetic evolutionary analysis, we identified three putative *bona fide* *NtCAD* genes from *N. tabacum*. Gene expression assays

10.3389/fpls.2024.1400213

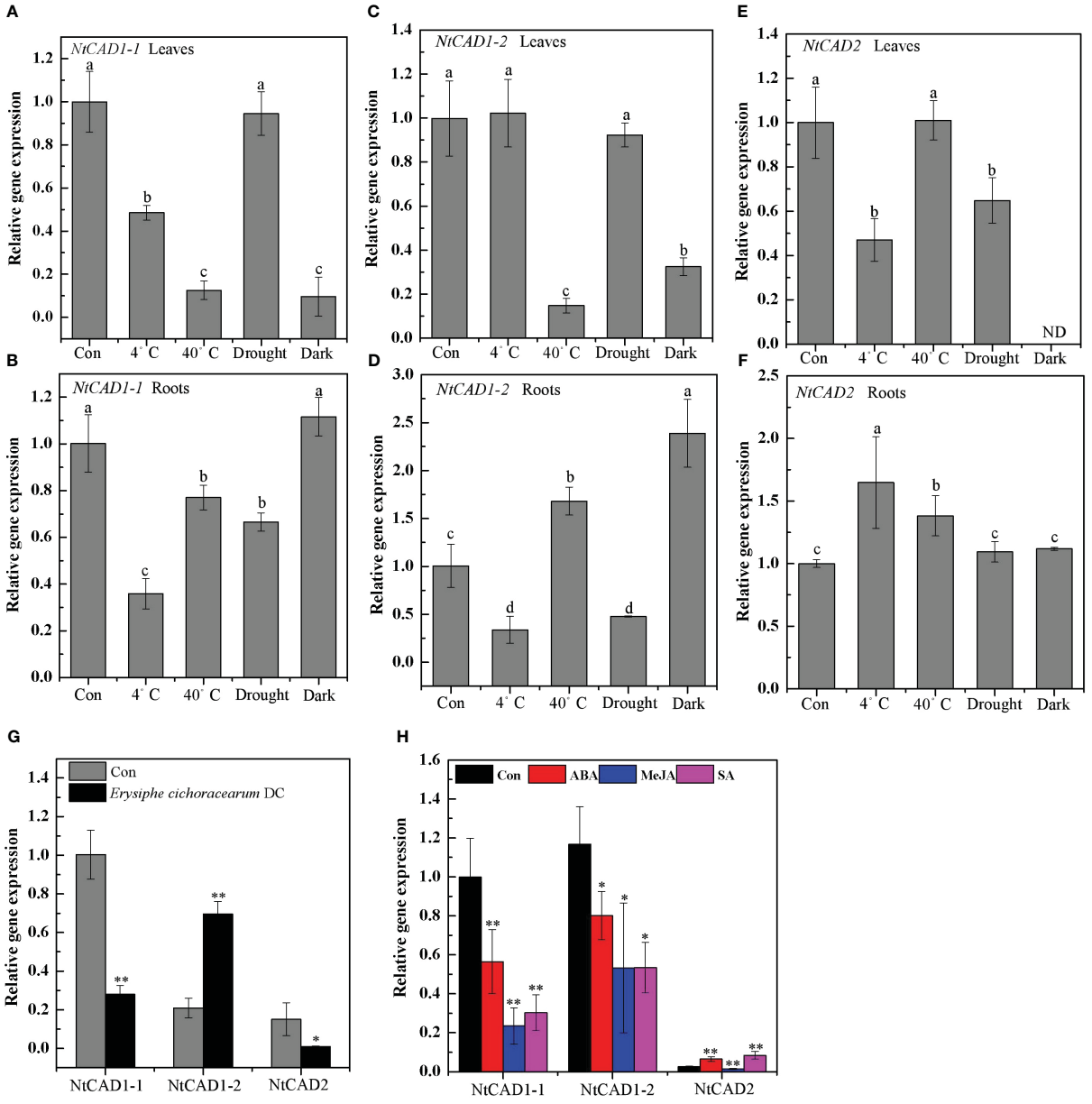


FIGURE 9
The expression profiles of NtCAD1–1, NtCAD1–2, and NtCAD2 in response to cold (4°C), heat (40°C), drought, dark stress **(A–F)**, fungus infection **(G)**, and ABA, MeJA, and SA treatments. The expression levels of the *NtCAD* genes were related to the internal control *L25* gene. The data represent means ± SD of at least four independent measurements obtained from different experiments. Significant differences between groups were determined using Duncan's multiple test, with bars labeled with different letters **(A–F)** indicating significance at $P < 0.05$. Additionally, bars marked with asterisks **(G, H)** demonstrate significant differences compared to the control (Con) samples based on Student's *t*-test ($P < 0.05$, $P < 0.01$).

conducted across different tobacco tissues provided compelling evidence supporting the potential involvement of NtCAD1–1 and NtCAD1–2 in lignin biosynthesis as they exhibited high expression levels in roots and stems. Enzyme activity analysis revealed that both NtCAD1 and NtCAD2 exhibited *in vitro* activity against hydroxycinnamaldehydes, with the highest efficiency observed for NtCAD1–2. Downregulation of NtCAD1–1 and NtCAD1–2 led to reddish-brown stems without significant altering lignin content; this observation can be attributed to the accumulation or enrichment of coniferyl aldehydes. Further investigations showed that these three NtCADs exhibit differential responses to biotic/abiotic stresses and phytohormones; notably, significant decreases or minor effects were observed for all tested conditions on the expression of NtCAD1–1 whereas heat stress, dark stress, and pathogen infection induced the expression of NtCAD1–2; conversely, cold/heat stress along with ABA and SA treatments induced the expression of NtCAD2. In summary, our findings provide a comprehensive analysis of NtCADs in tobacco and highlight the potential role of *NtCAD1* and *NtCAD2* in lignin biosynthesis. However, further investigation is required to elucidate the precise mechanism underlying the unaltered lignin content following NtCAD1 downregulation.

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

MW: Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. YL: Investigation, Methodology, Validation, Writing – review & editing. ZTL: Investigation, Software, Writing – review & editing. LX: Investigation, Software, Writing – review & editing. YX: Investigation, Methodology, Validation, Writing – review & editing. LZ: Data curation, Investigation, Methodology, Validation, Writing – review & editing. XY: Data curation, Investigation, Methodology, Writing – review & editing. ZFL: Investigation, Software, Writing – review & editing. XX: Formal analysis, Investigation, Writing – review & editing. LW: Investigation, Methodology, Writing – review & editing. RW: Formal analysis, Investigation, Methodology, Writing – review & editing. SX: Writing – review & editing. JY: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The present study received financial support from the major scientific and technological projects of CNTC (110202101043(JY-20)), the Science and Technology Project of Henan Province (Grant No. 222102110164), and the Natural Science Foundation of Henan Province (Grant No. 232300420032).

Acknowledgments

We express our heartfelt gratitude to all colleagues who have dedicated their unwavering efforts to the successful execution of these experiments.

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.2024.1400213/full#supplementary-material

References

Barakat, A., Bagniewska-Zadworna, A., Choi, A., Plakkat, U., DiLoreto, D. S., Yellanki, P., et al. (2009). The cinnamyl alcohol dehydrogenase gene family in Populus: phylogeny, organization, and expression. *BMC Plant Biol.* 9, 26. doi: 10.1186/1471-2229-9-26

Barakat, A., Bagniewska-Zadworna, A., Frost, C. J., and Carlson, J. E. (2010). Phylogeny and expression profiling of CAD and CAD-like genes in hybrid Populus (*P. deltoides* × *P. nigra*): evidence from herbivore damage for subfunctionalization and functional divergence. *BMC Plant Biol.* 10, 100. doi: 10.1186/1471-2229-10-100

Baucher, M., Bernard-Vailhé, M. A., Chabbert, B., Besle, J. M., Opsomer, C., Van Montagu, M., et al. (1999). Down-regulation of cinnamyl alcohol dehydrogenase in transgenic alfalfa (*Medicago sativa* L.) and the effect on lignin composition and digestibility. *Plant Mol. Biol.* 39, 437–447. doi: 10.1023/A:1006182925584

Baucher, M., Chabbert, B., Pilate, G., Van Doorsselaere, J., Tollier, M. T., Petit-Conil, M., et al. (1996). Red xylem and higher lignin extractability by down-regulating a cinnamyl alcohol dehydrogenase in Poplar. *Plant Physiol.* 112, 1479–1490. doi: 10.1104/pp.112.4.1479

Bernard Vailhé, M. A., Besle, J. M., Maillot, M. P., Cornu, A., Halpin, C., and Knight, M. (1998). Effect of down-regulation of cinnamyl alcohol dehydrogenase on cell wall composition and on degradability of tobacco stems. *J. Sci. Food Agric.* 76, 505–514. doi: 10.1002/(SICI)1097-0010(199804)76:4<505::AID-JSFA981>3.0.CO;2-M

Besseau, S., Hoffmann, L., Geoffroy, P., Lapierre, C., Pollet, B., and Legrand, M. (2007). Flavonoid accumulation in Arabidopsis repressed in lignin synthesis affects auxin transport and plant growth. *Plant Cell.* 19, 148–162. doi: 10.1105/tpc.106.044495

Bomati, E. K., and Noel, J. P. (2005). Structural and kinetic basis for substrate selectivity in *Populus tremuloides* sinapyl alcohol dehydrogenase. *Plant Cell.* 17, 1598–1611. doi: 10.1105/tpc.104.029983

Bukh, C., Nord-Larsen, P. H., and Rasmussen, S. K. (2012). Phylogeny and structure of the cinnamyl alcohol dehydrogenase gene family in *Brachypodium distachyon*. *J. Exp. Bot.* 63, 6223–6236. doi: 10.1093/jxb/ers275

Cesarino, I. (2019). Structural features and regulation of lignin deposited upon biotic and abiotic stresses. *Curr. Opin. Biotechnol.* 56, 209–214. doi: 10.1016/j.copbio.2018.12.012

Chabannes, M., Barakate, A., Lapierre, C., Marita, J. M., Ralph, J., Pean, M., et al. (2001). Strong decrease in lignin content without significant alteration of plant development is induced by simultaneous down-regulation of cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD) in tobacco plants. *Plant J.* 28, 257–270. doi: 10.1046/j.1365-313X.2001.01140.x

Chao, N., Huang, S., Kang, X., Yidilisi, K., Dai, M., and Liu, L. (2022). Systematic functional characterization of cinnamyl alcohol dehydrogenase family members revealed their functional divergence in lignin biosynthesis and stress responses in mulberry. *Plant Physiol. Biochem.* 186, 145–156. doi: 10.1016/j.plaphy.2022.07.008

Chao, N., Liu, S. X., Liu, B. M., Li, N., Jiang, X. N., and Gai, Y. (2014). Molecular cloning and functional analysis of nine cinnamyl alcohol dehydrogenase family members in *Populus tomentosa*. *Planta* 240, 1097–1112. doi: 10.1007/s00425-014-2128-9

Chen, L., Auh, C. K., Dowling, P., Bell, J., Chen, F., Hopkins, A., et al. (2003). Improved forage digestibility of tall fescue (*Festuca arundinacea*) by transgenic down-regulation of cinnamyl alcohol dehydrogenase. *Plant Biotechnol. J.* 1, 437–449. doi: 10.1046/j.1467-7652.2003.00040.x

Cowan, G., MacFarlane, S., and Torrance, L. (2023). A new simple and effective method for PLRV infection to screen for virus resistance in potato. *J. Virol. Methods* 315, 114691. doi: 10.1016/j.jviromet.2023.114691

Damiani, I., Morreel, K., Danoun, S., Goeminne, G., Yahiaoui, N., Marque, C., et al. (2005). Metabolite profiling reveals a role for atypical cinnamyl alcohol dehydrogenase CAD1 in the synthesis of coniferyl alcohol in tobacco xylem. *Plant Mol. Biol.* 59, 753–769. doi: 10.1007/s11103-005-0947-6

Fornalé, S., Capellades, M., Encina, A., Wang, K., Irar, S., Lapierre, C., et al. (2012). Altered lignin biosynthesis improves cellulosic bioethanol production in transgenic maize plants down-regulated for cinnamyl alcohol dehydrogenase. *Mol. Plant* 5, 817–830. doi: 10.1093/mp/ssr097

Goffner, D., Joffroy, I., Grima-Pettenati, J., Halpin, C., Knight, M. E., Schuch, W., et al. (1992). Purification and characterization of isoforms of cinnamyl alcohol dehydrogenase from *Eucalyptus* xylem. *Planta* 188, 48–53. doi: 10.1007/BF01160711

Gross, G. G., Stöckigt, J., Mansell, R. L., and Zenk, M. H. (1973). Three novel enzymes involved in the reduction of ferulic acid to coniferyl alcohol in higher plants: ferulate: CoA ligase, feruloyl-CoA reductase and coniferyl alcohol oxidoreductase. *FEBS Lett.* 31, 283–286. doi: 10.1016/0014-5793(73)80123-1

Guillaumie, S., Pichon, M., Martinant, J. P., Bosio, M., Goffner, D., and Barrière, Y. (2007). Differential expression of phenylpropanoid and related genes in brown-midrib *bm1*, *bm2*, *bm3*, and *bm4* young near-isogenic maize plants. *Planta* 226, 235–250. doi: 10.1007/s00425-006-0468-9

Guo, D. M., Ran, J. H., and Wang, X. Q. (2010). Evolution of the cinnamyl/sinapyl alcohol dehydrogenase (CAD/SAD) gene family: the emergence of real lignin is associated with the origin of *bona fide* CAD. *J. Mol. Evol.* 71, 202–218. doi: 10.1007/s00239-010-9378-3

Halpin, C., Holt, K., Chojecki, J., Oliver, D., Chabbert, B., Monties, B., et al. (1998). Brown-midrib maize (*bm1*)-a mutation affecting the cinnamyl alcohol dehydrogenase gene. *Plant J.* 14, 545–553. doi: 10.1046/j.1365-313X.1998.00153.x

Halpin, C., Knight, M. E., Grima-Pettenati, J., Goffner, D., Boudet, A., and Schuch, W. (1992). Purification and characterization of cinnamyl alcohol dehydrogenase from tobacco stems. *Plant Physiol.* 98, 12–16. doi: 10.1104/pp.98.1.12

Higuchi, T., Ito, T., Umezawa, T., Hibino, T., and Shibata, D. (1994). Redbrown colour of lignified tissues of transgenic plants with antisense CAD gene: Wine-red lignin from coniferaldehyde. *J. Biotechnol.* 37, 151–158. doi: 10.1016/0168-1656(94)90006-X

Hirano, K., Aya, K., Kondo, M., Okuno, A., Morinaka, Y., and Matsuoka, M. (2012). OsCAD2 is the major CAD gene responsible for monolignol biosynthesis in rice culm. *Plant Cell Rep.* 31, 91–101. doi: 10.1007/s00299-011-1142-7

Hu, L., Zhang, X., Ni, H., Yuan, F., and Zhang, S. (2022). Identification and functional analysis of CAD gene family in pomegranate (*Punica granatum*). *Genes (Basel)* 14, 26. doi: 10.3390/genes14010026

Jin, Y., Zhang, C., Liu, W., Qi, H., Chen, H., and Cao, S. (2014). The cinnamyl alcohol dehydrogenase gene family in melon (*Cucumis melo* L.): bioinformatic analysis and expression patterns. *PloS One* 9, e101730. doi: 10.1371/journal.pone.0101730

Kim, S. J., Kim, M. R., Bedgar, D. L., Moinuddin, S. G., Cardenas, C. L., Davin, L. B., et al. (2004). Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 101, 1455–1460. doi: 10.1073/pnas.0307987100

Kim, Y. H., Bae, J. M., and Huh, G. H. (2010). Transcriptional regulation of the cinnamyl alcohol dehydrogenase gene from sweet potato in response to plant developmental stage and environmental stress. *Plant Cell Rep.* 29, 779–791. doi: 10.1007/s00299-010-0864-2

Kolkman, J. M., Moreta, D. E., Repka, A., Bradbury, P., and Nelson, R. J. (2023). Brown midrib mutant and genome-wide association analysis uncover lignin genes for disease resistance in maize. *Plant Genome* 16, e20278. doi: 10.1002/tpg2.20278

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

Li, X., Weng, J. K., and Chapple, C. (2008a). Improvement of biomass through lignin modification. *Plant J.* 54, 569–581. doi: 10.1111/j.1365-313X.2008.03457.x

Li, X., Yang, Y., Yao, J., Chen, G., Li, X., Zhang, Q., et al. (2008b). FLEXIBLE CULM 1 encoding a cinnamyl-alcohol dehydrogenase controls culm mechanical strength in rice. *Plant Mol. Biol.* 69, 685–697. doi: 10.1007/s11103-008-9448-8

Liu, Q., Luo, L., and Zheng, L. (2018). Lignins: biosynthesis and biological functions in plants. *Int. J. Mol. Sci.* 19, 335. doi: 10.3390/ijms19020335

Liu, X., Van Acker, R., Voorend, W., Pallidis, A., Goeminne, G., Pollier, J., et al. (2021). Rewired phenolic metabolism and improved saccharification efficiency of a Zea mays *cinnamyl alcohol dehydrogenase 2* (*zmcad2*) mutant. *Plant J.* 105, 1240–1257. doi: 10.1111/tpj.15108

Ma, Q. H. (2010). Functional analysis of a cinnamyl alcohol dehydrogenase involved in lignin biosynthesis in wheat. *J. Exp. Bot.* 61, 2735–2744. doi: 10.1093/jxb/erq107

O'malley, D. M., Porter, S., and Sederoff, R. R. (1992). Purification, characterization, and cloning of cinnamyl alcohol dehydrogenase in loblolly pine (*Pinus taeda* L.). *Plant Physiol.* 98, 1364–1371. doi: 10.1104/pp.98.4.1364

Özparpucu, M., Gierlinger, N., Burgert, I., Van Acker, R., Vanholme, R., Boerjan, W., et al. (2018). The effect of altered lignin composition on mechanical properties of CINNAMYL ALCOHOL DEHYDROGENASE (CAD) deficient poplars. *Planta* 247, 887–897. doi: 10.1007/s00425-017-2828-z

Park, H. L., Kim, T. L., Bhoo, S. H., Lee, T. H., Lee, S. W., and Cho, M. H. (2018). Biochemical characterization of the rice cinnamyl alcohol dehydrogenase gene family. *Molecules* 23, 2659. doi: 10.3390/molecules23102659

Preisner, M., Wojtasik, W., Kostyn, K., Boba, A., Czuj, T., Szopa, J., et al. (2018). The cinnamyl alcohol dehydrogenase family in flax: Differentiation during plant growth and under stress conditions. *J. Plant Physiol.* 221, 132–143. doi: 10.1016/j.jplph.2017.11.015

Raes, J., Rohde, A., Christensen, J. H., Van de Peer, Y., and Boerjan, W. (2003). Genome-wide characterization of the lignification toolbox in Arabidopsis. *Plant Physiol.* 133, 1051–1071. doi: 10.1104/pp.103.026484

Rivera-Burgos, L. A., Volenec, J. J., and Ejeta, G. (2019). Biomass and bioenergy potential of brown midrib sweet sorghum germplasm. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01142

Roy, S. W., and Gilbert, W. (2005). Complex early genes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1986–1991. doi: 10.1073/pnas.0408355101

Roy, S. W., and Penny, D. (2007). Patterns of intron loss and gain in plants: intron loss-dominated evolution and genome-wide comparison of *O. sativa* and *A. thaliana*. *Mol. Biol. Evol.* 24, 171–181. doi: 10.1093/molbev/msl159

Saathoff, A. J., Tobias, C. M., Sattler, S. E., Haas, E., Twigg, P., and Sarath, G. (2011). Switchgrass contains two cinnamyl alcohol dehydrogenases involved in lignin formation. *Bioenerg. Res.* 4, 120–133. doi: 10.1007/s12155-010-9106-2

Saballos, A., Ejeta, G., Sanchez, E., Kang, C., and Vermerris, W. (2009). A Genomewide analysis of the cinnamyl alcohol dehydrogenase family in sorghum [*Sorghum bicolor* (L.) Moench] identifes *SbCAD2* as the *Brown midrib6* Gene. *Genetics* 181, 783–795. doi: 10.1534/genetics.108.098996

Sattler, S. E., Saathoff, A. J., Haas, E. J., Palmer, N. A., Funnell-Harris, D. L., Sarath, G., et al. (2009). A nonsense mutation in a cinnamyl alcohol dehydrogenase gene is responsible for the Sorghum *brown midrib6* phenotype. *Plant Physiol.* 150, 584–595. doi: 10.1104/pp.109.136408

Schmidt, G. W., and Delaney, S. K. (2010). Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. *Mol. Genet. Genomics* 283, 233–241. doi: 10.1007/s00438-010-0511-1

Stewart, D., Yahiaoui, N., McDougall, G. J., Myton, K., Marque, C., Boudet, A. M., et al. (1997). Fourier-transform infrared and Raman spectroscopic evidence for the incorporation of cinnamaldehydes into the lignin of transgenic tobacco (*Nicotiana tabacum* L.) plants with reduced expression of cinnamyl alcohol dehydrogenase. *Planta* 201, 311–318. doi: 10.1007/s004250050072

Sun, B., Wang, P., Wang, R., Li, Y., and Xu, S. (2018). Molecular Cloning and Characterization of a meta/para-O-Methyltransferase from *Lycoris aurea*. *Int. J. Mol. Sci.* 19, 1911. doi: 10.3390/ijms19071911

Tobias, C. M., and Chow, E. K. (2005). Structure of the cinnamyl-alcohol dehydrogenase gene family in rice and promoter activity of a member associated with lignification. *Planta* 220, 678–688. doi: 10.1007/s00425-004-1385-4

Trabucco, G. M., Matos, D. A., Lee, S. J., Saathoff, A. J., Priest, H. D., Mockler, T. C., et al. (2013). Functional characterization of cinnamyl alcohol dehydrogenase and caffeic acid O-methyltransferase in *Brachypodium distachyon*. *BMC Biotechnol.* 13, 61. doi: 10.1186/1472-6750-13-61

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

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, W296–W303. doi: 10.1093/nar/gky427

Weiller, F., Gerber, L., Trygg, J., Fangel, J. U., Willats, W. G. T., Driouich, A., et al. (2020). Overexpression of VviPGIP1 and NtCAD14 in tobacco screened using glycan microarrays reveals cell wall reorganisation in the absence of fungal infection. *Vaccines (Basel).* 8, 388. doi: 10.3390/vaccines8030388

Wu, Z., Han, S., Zhou, H., Tuang, Z. K., Wang, Y., Jin, Y., et al. (2019). Cold stress activates disease resistance in *Arabidopsis thaliana* through a salicylic acid dependent pathway. *Plant Cell Environ.* 42, 2645–2663. doi: 10.1111/pce.13579

Youn, B., Camacho, R., Moinuddin, S. G., Lee, C., Davin, L. B., Lewis, N. G., et al. (2006). Crystal structures and catalytic mechanism of the *Arabidopsis* cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4. *Org. Biomol. Chem.* 4, 1687–1697. doi: 10.1039/b601672c

Yu, X., Zhang, W., Zhang, Y., Zhang, X., Lang, D., and Zhang, X. (2019). The roles of methyl jasmonate to stress in plants. *Funct. Plant Biol.* 46, 197–212. doi: 10.1071/FP18106

Yusuf, C. Y. L., Nabilah, N. S., Taufik, N. A. A. M., Seman, I. A., and Abdullah, M. P. (2022). Genome-wide analysis of the CAD gene family reveals two *bona fide CAD* genes in oil palm. *3 Biotech.* 12, 149. doi: 10.1007/s13205-022-03208-0

Zhang, K., Qian, Q., Huang, Z., Wang, Y., Li, M., Hong, L., et al. (2006). GOLD HULL AND INTERNODE2 encodes a primarily multifunctional cinnamyl-alcohol dehydrogenase in rice. *Plant Physiol.* 140, 972–983. doi: 10.1104/pp.105.073007

Zhang, Y., Yan, H., Li, Y., Xiong, Y., Niu, M., Zhang, X., et al. (2006). Molecular cloning and functional analysis of 1-deoxy-d-xylulose 5-phosphate reductoisomerase from santalum album. *Genes*, 12(5), 626. doi: 10.3390/genes12050626

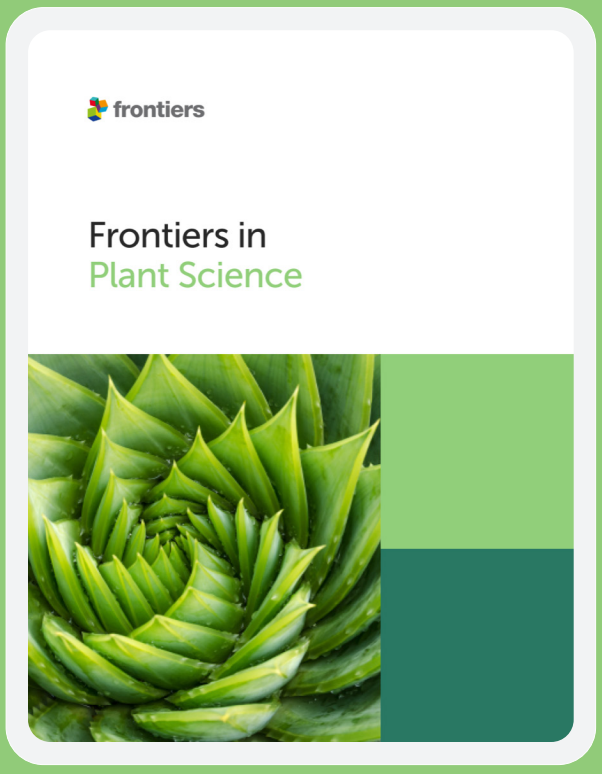
Frontiers in
Plant Science

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