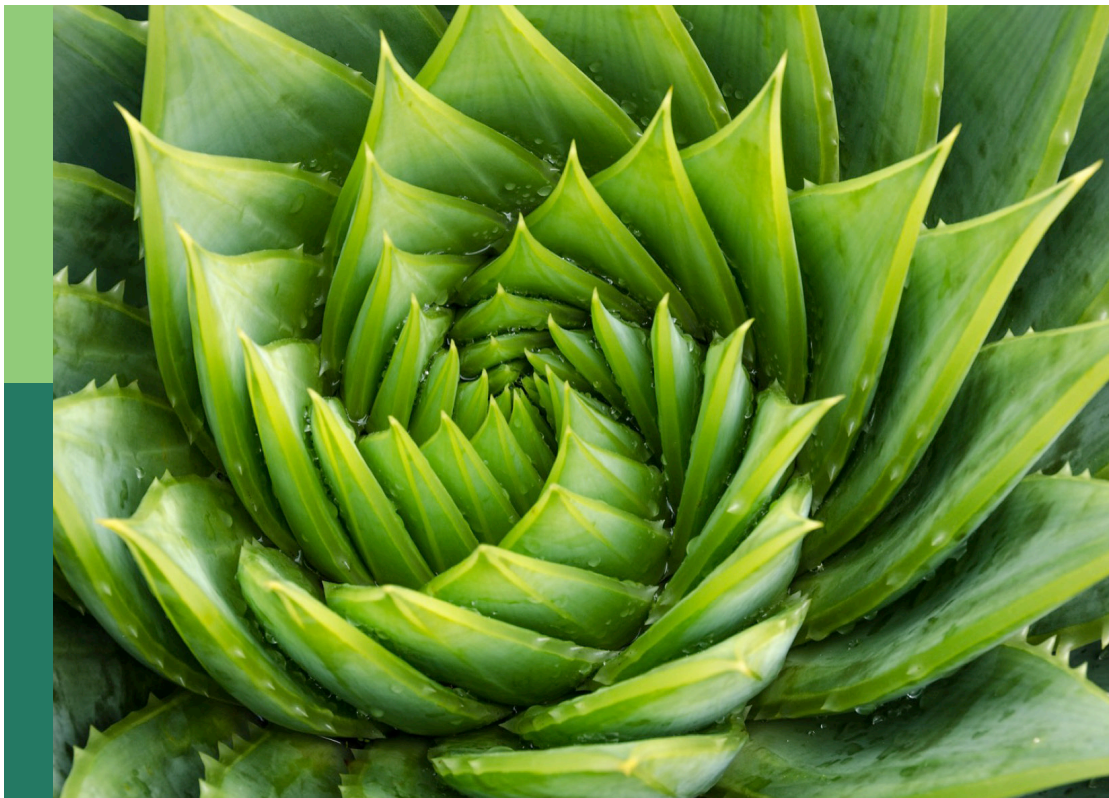
frontiers | Research Topics

Forage crop improvement for improved livestock production and nutrition

Edited by
Chris S. Jones, Jiyu Zhang, Jorge Fernando Pereira and Russell Jessup

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-83251-095-7
DOI 10.3389/978-2-83251-095-7

About Frontiers

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

Frontiers journal series

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

Dedication to quality

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

What are Frontiers Research Topics?

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

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

Forage crop improvement for improved livestock production and nutrition

Topic editors

Chris S. Jones — International Livestock Research Institute (Ethiopia), Ethiopia
Jiyu Zhang — Lanzhou University, China
Jorge Fernando Pereira — Embrapa Gado de Leite, Brazil
Russell Jessup — Texas A&M University, United States

Citation

Jones, C. S., Zhang, J., Pereira, J. F., Jessup, R., eds. (2023). *Forage crop improvement for improved livestock production and nutrition*.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-095-7

Table of contents

05 Editorial: Forage crop improvement for improved livestock production and nutrition
Jorge F. Pereira, Jiyu Zhang, Russell Jessup and Chris S. Jones

08 Triticale Improvement: Mining of Genes Related to Yellow Rust Resistance in Triticale Based on Transcriptome Sequencing
Fangyuan Zhao, Kuiju Niu, Xinhui Tian and Wenhua Du

23 Identification of Flowering Regulatory Networks and Hub Genes Expressed in the Leaves of *Elymus sibiricus* L. Using Comparative Transcriptome Analysis
Yuying Zheng, Na Wang, Zongyu Zhang, Wenhui Liu and Wengang Xie

42 RAD-Seq-Based High-Density Linkage Maps Construction and Quantitative Trait Loci Mapping of Flowering Time Trait in Alfalfa (*Medicago sativa* L.)
Xueqian Jiang, Tianhui Yang, Fan Zhang, Xijiang Yang, Changfu Yang, Fei He, Ruicai Long, Ting Gao, Yiwei Jiang, Qingchuan Yang, Zhen Wang and Junmei Kang

56 Full-Length Transcriptomics Reveals Complex Molecular Mechanism of Salt Tolerance in *Bromus inermis* L.
Qian Li, Jiaxing Song, Yi Zhou, Yingxia Chen, Lei Zhang, Yongzhen Pang and Bo Zhang

71 Transcriptional Analyses of Genes Related to Fodder Qualities in Giant Leucaena Under Different Stress Environments
Ahmed M. Bageel, Aaron Kam and Dulal Borthakur

85 Comparative Transcriptome and Proteome Analysis Provides New Insights Into the Mechanism of Protein Synthesis in Kenaf (*Hibiscus cannabinus* L.) Leaves
Chao Zhang, Yong Deng, Gaoyang Zhang, Jianjun Li, Aiping Xiao, Lining Zhao, Anguo Chen, Huijuan Tang, Li Chang, Gen Pan, Yingbao Wu, Jiangjiang Zhang, Cuiping Zhang, Ziggiju Mesenbet Birhanie, Hui Li, Juan Wu, Dawei Yang, Defang Li and Siqi Huang

101 High-Altitude Genetic Selection and Genome-Wide Association Analysis of Yield-Related Traits in *Elymus sibiricus* L. Using SLAF Sequencing
Zongyu Zhang, Yuying Zheng, Junchao Zhang, Na Wang, Yanrong Wang, Wenhui Liu, Shiqie Bai and Wengang Xie

119 A Genome-Wide Association Study Coupled With a Transcriptomic Analysis Reveals the Genetic Loci and Candidate Genes Governing the Flowering Time in Alfalfa (*Medicago sativa* L.)
Fei He, Fan Zhang, Xueqian Jiang, Ruicai Long, Zhen Wang, Yishi Chen, Mingna Li, Ting Gao, Tianhui Yang, Chuan Wang, Junmei Kang, Lin Chen and Qingchuan Yang

132 *Brittle Culm 15* mutation alters carbohydrate composition, degradation and methanogenesis of rice straw during *in vitro* ruminal fermentation
Siyu Yi, Xiumin Zhang, Jianjun Zhang, Zhiyuan Ma, Rong Wang, Duanqin Wu, Zhongshan Wei, Zhiliang Tan, Baocai Zhang and Min Wang

144 Analysis of *Elymus nutans* seed coat development elucidates the genetic basis of metabolome and transcriptome underlying seed coat permeability characteristics
Jing Zhou, Yan Li, Xun Wang, Yijia Liu, Rakefet David-Schwartz, Mira Weissberg, Shuiling Qiu, Zhenfei Guo and Fulin Yang

162 MtPT5 phosphate transporter is involved in leaf growth and phosphate accumulation of *Medicago truncatula*
Xue Wang, Chunxue Wei, Fei He and Qingchuan Yang

171 Zinc toxicity response in *Ceratoides arborescens* and identification of *CaMTP*, a novel zinc transporter
Xingyue Li, Lin Zhang, Haiyan Ren, Xiaoyu Wang and Fugui Mi

183 SgNramp1, a plasma membrane-localized transporter, involves in manganese uptake in *Stylosanthes guianensis*
Xiaoyan Zou, Rui Huang, Linjie Wang, Guihua Wang, Ye Miao, Idupulapati Rao, Guodao Liu and Zhijian Chen

197 Identification of QTL and candidate genes associated with biomass yield and Feed Quality in response to water deficit in alfalfa (*Medicago sativa* L.) using linkage mapping and RNA-Seq
Xueqian Jiang, Andong Yu, Fan Zhang, Tianhui Yang, Chuan Wang, Ting Gao, Qingchuan Yang, Long-Xi Yu, Zhen Wang and Junmei Kang

211 Multi-omics analyses reveal new insights into nutritional quality changes of alfalfa leaves during the flowering period
Yinghao Liu, Wenqiang Fan, Qiming Cheng, Lianyi Zhang, Ting Cai, Quan Shi, Zuo Wang, Chun Chang, Qiang Yin, Xiaowei Jiang and Ke Jin

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 05 December 2022
DOI 10.3389/fpls.2022.1082873

OPEN ACCESS

EDITED AND REVIEWED BY
Leo Marcelis,
Wageningen University and Research, Netherlands

*CORRESPONDENCE
Chris S. Jones
C.S.Jones@cgiar.org

SPECIALTY SECTION
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

RECEIVED 28 October 2022
ACCEPTED 23 November 2022
PUBLISHED 05 December 2022

CITATION
Pereira JF, Zhang J, Jessup R and Jones CS (2022) Editorial: Forage crop improvement for improved livestock production and nutrition. *Front. Plant Sci.* 13:1082873. doi: 10.3389/fpls.2022.1082873

Editorial: Forage crop improvement for improved livestock production and nutrition

Jorge F. Pereira[1], Jiyu Zhang[2], Russell Jessup[3] and Chris S. Jones[4]*

[1]Embrapa Gado de Leite, Juiz de Fora, Brazil, [2]State Key Laboratory of Herbage Improvement and Grassland Agro-Ecosystems, College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, China, [3]Department of Soil and Crop Science, Texas A&M University, College Station, TX, United States, [4]Feed and Forage Development, International Livestock Research Institute, Nairobi, Kenya

KEYWORDS

animal nutrition, forage quality, genome-wide association studies (GWAS), marker assisted selection, plant breeding, omics

Editorial on the Research Topic:
Forage crop improvement for improved livestock production and nutrition

We have a major challenge to sustainably grow the production of animal protein in order to feed an additional nearly two billion people by 2050. In addition, in some regions of the world, population is also predicted to be wealthier, which positively impacts consumption of animal protein. Increasing animal production is a multifaceted challenge and one component that needs to be tackled is the production and quality of the forages that they feed on. Improved forages can increase animal production per unit head and may also allow higher grazing intensity (more animals per hectare), leading to lower pressure on grazing land which is the most predominant way of feeding livestock. However, even in locations where animal feeding is based on cut and carry and indoor systems, forages are an important resource.

In order to improve forages for better livestock production, plants need to be bred by combining desired characteristics in one genotype or family. However, forage breeding faces many challenges including variable ploidy levels, different reproductive modes, and the need to evaluate traits that are laborious to assess. Besides, breeding efforts are distributed across hundreds of forage species that need to be adapted to a vast range of systems and environments worldwide.

Conventional forage breeding is a time-consuming process (that needs up to 10 years to develop a cultivar) and is mostly based on traits that are often poorly understood. Here, molecular sciences and biotechnology offer great possibilities to develop better breeding pipelines to pave the way for accelerated and more efficient forage breeding.

This Research Topic *"Forage Crop Improvement for Improved Livestock Production and Nutrition"* focused on studies leveraging molecular technologies to capture the opportunities from forage genetic diversity to support the accelerated development of new cultivars. The Research Topic contains 15 original research articles based on 11 legume and grass forage species, where several molecular approaches (genome-wide association studies - GWAS, transcriptomic, proteomic, metabolomic, linkage, and QTL mapping) were used to characterise traits including flowering time, heavy metal stress, nutritional quality, phosphate accumulation, rust resistance, salt tolerance, seed coat permeability, and yield.

The most common forage species studied in this Research Topic is alfalfa (*Medicago sativa*), a high-protein legume forage that is widely used especially in temperate regions. Flowering time in alfalfa, which effects biomass yield and quality, is affected by genetic and environmental factors. Thus, the identification of a marker (gene, protein, or metabolite) associated with flowering time could have important applications in alfalfa breeding. Two papers in this Research Topic report the identification of candidate genes associated with flowering time in alfalfa. Jiang et al. identified 16 and 22 QTL for early and late flowering traits, respectively, and gene expression analysis revealed thousands of up- and down-regulated transcripts. The QTL and gene expression data were integrated, resulting in the detection of seven flowering time gene candidates. He et al. found that geographical origin and breeding status strongly influenced alfalfa flowering time. The researchers combined GWAS and transcriptomic analyses to identify 38 candidate genes and two SNPs, located upstream of one candidate gene, were significantly associated with flowering time. The trait was also studied by Liu et al., but with the intention of understanding the molecular mechanisms underlying variation in nutritional quality. These authors used a multi-omics approach to evaluate chlorophyll, amino acid, and flavonoid content in alfalfa leaves at budding and full-bloom stages. As alfalfa matures the content of chlorophyll, amino acids, and flavonoids in leaves decreases. This was correlated with the up- or down-regulation of specific genes, which may guide future endeavours to obtain high-quality alfalfa. Several alfalfa quality traits, alongside biomass yield and plant height under water-limited environments, were studied by Jiang et al. By combining linkage mapping and RNA-seq analysis, they detected 22 common differentially expressed genes located in nine QTL regions that explained >10% of the phenotypic variation. This knowledge may guide new studies about the effect of drought on alfalfa nutritional quality.

After alfalfa, the forage species with most studies in the Research Topic is *Elymus sibiricus*. Another species from this same genus, *Elymus nutans*, was also studied. Zheng et al. reported the identification of regulatory networks and hub genes related to flowering time in *E. sibiricus*. After identifying accessions varying in booting, heading, and flowering times, transcriptome analyses revealed 72 candidate genes with four of them significantly upregulated in late-flowering accessions. Drought and salt stress appear to activate the flowering regulation pathway, and a SNP found in one of seven hub genes was detected and associated with late flowering. Zhang et al. provided relevant genetic information for the gigantic genome (6.86 Gbp) of *E. sibiricus*. They were interested in laying the foundation for molecular mechanisms of the species adaptation to high altitudes as well as of yield- and seed-related traits. Their results revealed genes associated with low oxygen and strong ultraviolet radiation (important factors in higher altitudes) and a total of 1,825 significant loci associated with 12 agronomic traits. For *Elymus nutans*, Zhou et al. were interested in elucidating the mechanism of seed coat permeability, which controls exchange of substances (water, gas, and nutrients) between the seed embryo and the outside environment, and can affect seed dormancy. The authors identified thiamine and salicylic acid as the key metabolites affecting seed coat permeability and expression of the *PR1* and *PAL* genes correlated with these two metabolites. These metabolites and genes may be important targets for better seed production of *E. nutans*.

The transport of phosphate, zinc and manganese were investigated in three papers of this Research Topic. Wang et al. were interested in investigating the role of the MtPT5 transporter in leaf growth and inorganic phosphate (Pi) accumulation of *Medicago truncatula*. By using expression analysis, complementation of *Arabidopsis* mutants, and overexpressing *MtPT5* in *M. truncatula*, the authors were able to demonstrate that MtPT5 plays important roles in vegetative growth and Pi nutrition. Li et al. focused on the impact of zinc (Zn) and the identification of Zn transporters in *Ceratoides arborescens*, a forage that can be grown in arid and semiarid regions. The *CaMTP* gene was detected in the root transcriptome of *C. arborescens* and complementation of *CaMTP* in a yeast mutant conferred its ability to grow under high Zn conditions. The identification of *CaMTP* significantly contributes to elucidating the molecular mechanisms underlying Zn toxicity in *C. arborescens*. Metal stress, caused by manganese (Mn), in the forage legume *Stylosanthes guianensis*, was investigated by Zou et al., who employed gene expression analyses to identify the *SgNramp1* gene. Higher gene expression was correlated with enhanced Mn uptake in *S. guianensis* and heterologous expression of *SgNramp1* complemented the phenotype of a Mn uptake-defective yeast. Taken together, these results pave the way for better understanding of the response to Mn stress in forages.

The other five papers in this Research Topic report on investigations of the response to biotic and abiotic stresses as well as molecular mechanisms involved in quality traits and better degradation during rumen fermentation. Zhao et al. used transcriptome analyses to elucidate the resistance mechanisms and candidate genes associated with yellow rust resistance in triticale. The results revealed a range of molecular defence

strategies in triticale when infected by yellow rust, and this list of candidate genes represents a resource for yellow rust resistance breeding. Transcriptome analysis was also employed by Li et al. to understand the molecular response of *Bromus inermis* to salt stress. Thousands of differentially expressed genes were obtained from leaves and roots and, after enrichment analyses, a set of key genes involved in salinity adaptation was detected. This list of genes now needs to be carefully studied in order to further understand the salinity adaptation mechanism in *B. inermis*. For fodder quality, Bageel et al. were interested in understanding the environmental conditions promoting or inhibiting mimosine and tannin biosynthesis, two secondary metabolites of giant leucaena (*Leucaena leucocephala* subsp. *glabrata*) that are undesirable in higher concentrations. Different environmental conditions (salinity, pH and nitrogen availability) had varying effects on mimosine and tannin biosynthesis. Using transcriptome analyses, the authors were able to identify two genes for mimosine metabolism and five for tannin biosynthesis. In general terms, mimosine content and the expression of mimosine synthase in the foliage were correlated and the increase in tannin production seemed to be affected by the expression of three tannin biosynthesis genes. These results may encourage the design of molecular approaches that inhibit the expression of these genes in order to reduce mimosine and tannins in giant leucaena foliage. The quality of forages not only involves the reduction of undesirable metabolites but also the increase of desirable ones. Zhang et al. studied crude protein content in order to understand the expression of this quality trait in kenaf (*Hibiscus cannabinus*) leaves. Proteomic and transcriptomic approaches revealed four genes associated with high crude protein content in kenaf leaves. These genes, and the proteins that they encode, need to be functionally characterized to help breeding high crude protein kenaf varieties. Better quality forages may also be obtained by evaluating their ruminal fermentation process or even by selecting materials that generate lower methane production. Yi et al. evaluated a rice *BC15* gene mutant, which influences cell wall composition. The authors observed that the *BC15* mutant had enhanced degradation at the early stage of rumen fermentation, which led to decreased methane production. On the other hand, the *BC15* mutant showed lower fibre degradation and cellulolytic fungi population indicating that its straw is more resistant for microbial degradation. It remains to be seen if other forage species with the *BC* (*Brittle Culm*) mutation may have similar characteristics.

This Research Topic demonstrates how the use of molecular sciences and biotechnology can improve the understanding of molecular mechanisms underlying important traits in forage species. The use of this information can be used to accelerate forage breeding in order to sustainably increase the production of animal protein and improve food security. We welcome everyone to explore the 15 research papers for deeper analysis of their methodologies and findings.

Author contributions

CJ and JP drafted the manuscript. JZ and RJ checked the manuscript and suggested modifications. All authors contributed to the Editorial and approved the submitted version.

Conflict of interest

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

Publisher's note

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 09 May 2022
doi: 10.3389/fpls.2022.883147

Triticale Improvement: Mining of Genes Related to Yellow Rust Resistance in Triticale Based on Transcriptome Sequencing

*Fangyuan Zhao, Kuiju Niu, Xinhui Tian and Wenhua Du**

College of Grassland Science, Key Laboratory of Grassland Ecosystem (Ministry of Education), Pratacultural Engineering Laboratory of Gansu Province, Sino-U.S. Centers for Grazing Land Ecosystem Sustainability, Gansu Agricultural University, Lanzhou, China

Yellow (stripe) rust caused by *Puccinia striiformis* f. sp. *tritici* (*Pst*) is a major destructive fungal disease of small grain cereals, leading to large yield losses. The breeding of resistant varieties is an effective, sustainable way to control yellow rust. Elucidation of resistance mechanisms against yellow rust and identification of candidate genes associated with rust resistance are thus crucial. In this study, seedlings of two *Triticosecale* Wittmack cultivars, highly resistant Gannong No. 2 and susceptible Shida No. 1, were inoculated with *Pst* race CYR34. Transcriptome sequencing (RNA-seq) was then used to investigate their transcriptional responses against pathogen infection before and after the appearance of symptoms—10 and 20 days after inoculation, respectively. According to the RNA-seq data, the number of upregulated and downregulated differentially expressed genes (DEGs) in the resistant cultivar was greater than in the susceptible cultivar. A total of 2,560 DEGs commonly expressed in the two cultivars on two sampling dates were subjected to pathway analysis, which revealed that most DEGs were closely associated with defense and metabolic activities. Transcription factor enrichment analysis indicated that the expressions of NAC, WRKY, and FAR1 families were also significantly changed. Further in-depth analysis of resistance genes revealed that almost all serine/threonine-protein kinases were upregulated in the resistant cultivar. Other genes related to disease resistance, such as those encoding disease-resistance- and pathogenesis-related proteins were differentially regulated in the two cultivars. Our findings can serve as a resource for gene discovery and facilitate elucidation of the complex defense mechanisms involved in triticale resistance to *Pst*.

Keywords: triticale, transcriptome, yellow rust, CYR34, resistance

OPEN ACCESS

Edited by:
Jorge Fernando Pereira, Embrapa Gado de Leite, Brazil

Reviewed by:
Jaspal Kaur, Punjab Agricultural University, India
Qiang Li, Northwest A & F University, China

****Correspondence:***
Wenhua Du
duwenhua1012@163.com

Specialty section:
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

Received: *24 February 2022*
Accepted: *19 April 2022*
Published: *09 May 2022*

Citation:
Zhao F, Niu K, Tian X and Du W (2022) Triticale Improvement: Mining of Genes Related to Yellow Rust Resistance in Triticale Based on Transcriptome Sequencing. Front. Plant Sci. 13:883147. doi: 10.3389/fpls.2022.883147

INTRODUCTION

Triticale (×*Triticosecale* Wittmack) is a synthetic hybrid derived from a cross of wheat (*Triticum* sp.) and rye (*Secale* sp.). This hybrid is high yielding and has favorable agronomic characteristics compared with wheat and rye, such as excellent performance in poor and acidic soils or under cold environments (González et al., 2005). Triticale has been developed into a

multi-purpose grain-forage species and can be used as a winter cover crop prior to planting of grain crops (Baron et al., 2015). Various diseases and insect pests lead to huge losses in the yield and quality of triticale, among which, yellow rust caused by *Puccinia striiformis* f. sp. *tritici* (*Pst*) has caused serious damage in cool climates (Gyawali et al., 2017). The pathogen prefers humid, cool habitats (3–15°C) and can overwinter on autumn-sown plants (Sodkiewicz et al., 2009). Because chemical and agronomic control methods are not effective when climatic conditions are favorable for the pathogen, the use of resistant cultivars is the most efficient, economically viable, environmentally friendly means of controlling the disease (Ren et al., 2015).

To be pathogenic, an organism must pierce through the cell wall and often the host plasma membrane as well (Zhang et al., 2014). Plants have evolved defense mechanisms to protect themselves from pathogen infection (Jones and Takemoto, 2004). In particular, plants typically use basal and resistance (*R*) gene-mediated defense mechanisms to mount a defensive response to pathogen attack that delays or arrests potential pathogen growth through pathogen-associated molecular pattern-triggered immunity (PTI) and effector-triggered immunity (ETI) systems (Jones and Dangl, 2006). *R* genes are the most effective weapons against pathogen invasion because they can specifically recognize the corresponding pathogen effectors or associated protein (s) that activate plant immune responses at the infection site. Although *R* genes are diverse, most contain conserved motifs, such as nucleotide-blinding site (NBS), leucine-rich repeat (LRR), toll protein and mammalian interleukin-1 receptor (TIR), coiled-coil (CC) or leucine zipper (LZ) structure, transmembrane domain (TM), and protein kinase domain (PK) motifs (Liu et al., 2007).

When plants are infected with various pathogens, the concentrations of endogenous hormones playing crucial roles in regulating plant immune responses, such as salicylic acid (SA), jasmonic acid (JA), and ethylene (ET), are altered (Bari and Jones, 2009). SA is generally involved in the activation of defense against biotrophic pathogens, whereas JA and ET are usually associated with defense against necrotrophic ones (Glazebrook, 2005). Other plant hormones previously implicated in plant growth and development, such as abscisic acid (ABA), auxins (indole-3-acetic acid, IAA), and brassinosteroids (BRs), also play vital roles in the regulation of the immune signaling network and the molding of plant–pathogen interactions, which results in a complex orchestration of plant immunity (Robert-Seilaniantz et al., 2011; Pieterse et al., 2012). Transcription factors (TFs), which control transcriptional regulation—an important method of intracellular regulation—are also involved in plant defense against pathogens. Multiple members of many TF families, such as AP2/ERF, NAC, MYB, and WRKY, participate in pathogen response (Ramamoorthy et al., 2008).

High-throughput sequencing is an efficient and economical molecular analysis strategy. RNA sequencing (RNA-Seq) has been widely applied to investigate plant–pathogen interactions at the transcriptional level in crop and forage plants, including wheat, rice (*Oryza sativa* L.), and alfalfa (*Medicago sativa* L.). This approach has also been used to identify metabolic pathways related to disease resistance, such as phenylpropanoid biosynthesis, plant–pathogen interaction, and plant hormone signal transduction pathways (Yuan et al., 2019). Dobon et al. (2016) performed a RNA-seq time-course analysis of susceptible and resistant wheat hosts infected with *Pst*, which has provided a framework to help clarify how *Pst* causes yellow rust in wheat. Recent studies have examined morphological variation and biochemical and physiological components of triticale. The functions of disease-resistance genes associated with powdery mildew, fusarium head blight, and slow rust have also been reported (Audenaert et al., 2014). The transcriptional responses associated with yellow rust resistance in triticale, a host of *Pst*, have not been reported, and the mechanism of rust resistance in this species is still unclear.

In this study, we used RNA-seq to investigate transcriptional regulation in yellow-rust resistant and susceptible triticale cultivars on two different dates following *Pst* infection. We also analyzed the key defensive metabolic pathways of triticale during *Pst* infection. Our identification of responsive genes in the two triticale cultivars after inoculation has helped elucidate molecular mechanisms underlying the response of triticale to yellow rust caused by *Pst* and has provided valuable genetic information for further functional genomic studies of triticale resistance to yellow rust.

MATERIALS AND METHODS

Plant Materials and Yellow Rust Pathogen Races

Triticale cultivars Gannong No. 2 (henceforth, Gannong 2) and Shida No. 1 (henceforth, Shida 1) were provided by the College of Grassland Science, Gansu Agricultural University, China. The yellow rust *Pst* race CYR34 was supplied by the Institute of Plant Protection, Gansu Academy of Agricultural Science, China. Gannong 2 and Shida 1 infected by *Pst* race CYR34, representing incompatible and compatible reactions, respectively.

Plant Cultivation, Inoculation Conditions, and Infection Assessment

Triticale plants were cultivated in pots (14 cm × 14 cm × 12 cm; 10 plants/pot) filled with 1:1 (v/v) field soil and vermiculite, which was supplemented before sowing with diammonium phosphate (0.45 g per pot). Plump, healthy triticale seeds of consistent maturity were sown in the filled pots, which were irrigated at 2-day intervals. Plants were cultured at 27°C/24°C (day/night) under a 16/8-h photoperiod (light/dark) and a relative humidity of 65%. Inoculation with *Pst* race CYR34, the most widespread race in China, was carried out 15 days later using the standard inoculation procedure detailed in Bozkurt et al. (2007). Briefly, plants during inoculation were incubated at 10°C under extremely high humidity for 24 h in the dark and then transferred to normal growth conditions (16-h light at 15°C–17°C and 8-h dark at 14°C–16°C). Fifteen days after inoculation, plant infection types (ITs) were assessed

on a scale of 0–4 (Roelfs et al., 1992), where 0, 0; 1, 2, 3, and 4 indicate immunity, near immunity, moderate resistance, high resistance, moderate susceptibility, and high susceptibility to yellow rust, respectively (Huang et al., 2014). Leaves of Gannong 2 and Shida 1 seedlings inoculated with CYR34 with different phenotypes were collected both prior to before and after the sporulation (susceptible cultivar Shida 1), namely, 10 and 20 days after inoculation, respectively. Samples collected from Gannong 2 on days 10 and 20 were designated as R10 and R20, respectively, with the corresponding samples from Shida 1 labeled as S10 and S20. All fresh samples were immediately frozen in liquid nitrogen and stored at −80°C until RNA extraction.

RNA Extraction, cDNA Library Construction, and Illumina Sequencing

Total RNA was extracted from leaf samples using an RNA simple Total RNA kit (Tiangen, Beijing, China). The extracted RNA was quantified on a NanoDrop ND-1000 instrument (Thermo Scientific), and RNA integrity was determined by 1% agarose gel electrophoresis. Construction of RNA-Seq and cDNA libraries was carried out with 1 μg RNA per sample using a NEBNext UltraTM RNA Library Prep Kit for Illumina (NEB, United States). Sequencing of the eight generated libraries (R10, S10, R20, and S20, with two biological replicates per sample) was performed on an Illumina HiSeq 2000 system (Illumina, San Diego, CA, United States) by Biomarker Technologies Co. (Beijing, China).

Sequence Read Mapping, Transcript Assembly, and Functional Annotation

Clean reads were obtained by removing adapter and poly-N-containing reads and low-quality sequences from the raw data. Q30 values, GC contents, and sequence duplication levels of the filtered data were then calculated. The clean reads were subsequently mapped to the wheat reference genome.[1] Reads with no more than one mismatch were further analyzed and annotated against the reference genome using Hisat2 (Kim et al., 2015). The mapped reads were assembled with StringTie software (Pertea et al., 2015) and used for mining new transcripts and genes of triticale. The assembled genes were aligned against several protein databases, including Nr, Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG), COG/KOG, Gene Ontology (GO), Pfam and eggNOG, using BLASTx (*E*-value $<1\times10^{-5}$; Altschul et al., 1997).

Differential Expression and Gene Ontology/Kyoto Encyclopedia of Genes and Genomes Enrichment Analyses

Differentially expressed genes (DEGs) were determined according to their expression abundances in different treatment groups, with new transcript and gene expressions calculated by the expected number of fragments per kilobase of transcript sequence per million base pairs sequenced (FPKM) method (Xue et al., 2010). GO enrichment analysis of DEGs was carried out using the GOseq R package based on Wallenius' noncentral hypergeometric distribution (Young et al., 2010). KEGG pathway enrichment analysis was performed using KOBAS software with a corrected $p<0.05$ (Mao et al., 2005). A comparative heatmap analysis of DEGs in two triticale cultivars (Gannong 2 and Shida 1) at 10 vs. 20 days after *Pst* infection was also carried out using the Biomarker Cloud platform.[2]

qRT-PCR Validation of DEGs

To experimentally evaluate the RNA-Seq results, we conducted a qRT-PCR analysis. Total RNA was extracted from leaf samples and reverse transcribed into cDNA using a PrimeScript RT reagent Kit with gDNA Eraser (Takara, Japan). qRT-PCR amplifications were conducted using SYBR Green Real-Time PCR Master Mix (Takara) on a LightCycler96 thermal cycler (Roche, Switzerland). Sixteen DEGs were randomly selected from kinase-related genes, disease-resistance protein genes and transcription factor that were upregulated in the resistant cultivar both at 10 and 20 days after inoculation, which were amplified with specific primers designed using Primer 5.0 software (**Table 1**), with three biological replicates performed per sample, and each with three technical repeats. The β-actin gene was used as an internal standard. Relative transcript levels of the selected genes were calculated by the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

RESULTS

Evaluation of Resistance

Fifteen days after inoculation, Gannong 2 and Shida 1 were photographed and assessed for disease severity (**Figure 1**). Compared with non-inoculated control seedlings, Gannong 2 seedlings infected with CYR34 exhibited a few necrotic stripes, but no sporulation, and were recorded as highly resistant, with an IT score of 1. In contrast, Shida 1 developed heavy sporulation after inoculation; it was highly susceptible to yellow rust race CYR34, with an IT score of 4. It was consistent with the identification results of these two cultivars at seedling stage by the Institute of Plant Protection, Gansu Academy of Agricultural Science, China. In addition, Gannong 2 was a resistant cultivar and Shida 1 was a susceptible cultivar at the adult plant stage, according to the identification by the institute.

Transcriptome Sequencing Analysis and Functional Annotation

Filtering of raw reads to remove adaptor-containing and low-quality reads yielded 59.30 Gb clean reads for subsequent analyses. An average of 5.75 Gb clean reads were generated per sample, Q30 and GC values ranged from 94.27%–95.77% and 53.41%–56.96%, respectively, and r^2 values between replicates varied between 0.974 and 0.990, these numbers,

[1]http://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Annotations/v1.1/

[2]http://www.biocloud.net//

TABLE 1 | Primers used for validation of RNA-seq values by qRT-PCR.

Gene ID	Forward primer	Reverse primer
TraesCS2D02G462000	CTCGCTGACCGGAGACATAC	TTCGTCAGGTTGCCGATCTC
TraesCS2A02G574800	CCACCTTTGGGAGGCTCAAT	GTTGCTGGCAAGGTTCAGTG
TraesCS5A02G487000	TGCTTGGATGGTCACGTCTC	TGCGTTCTCCTTGACAGCAT
TraesCS5A02G445700	GCATATGCAAGAACAGGCCG	TACCAACAATCGCCTTGGCT
TraesCS5B02G063600	CCCCTGCGATTATGCGTTTG	GGCAGAGACTGTTGGAGCTT
TraesCS7A02G425900	ACTACCAGACTGGGCAGCTA	CTTCAAGAACTTGTGCCGCC
TraesCS7D02G131800	AGTTGGAGATCTTGGGCGTG	AGACCATCAGTGAGTGCAGC
TraesCS7A02G069600	CCAATCTGACACCCAGTCCC	GTTGCTGCTCATTGGTGCAA
TraesCS4A02G482900	GTGTGGGCTAGCAGAATGGT	TCTGGCACTTCTTGACGACC
TraesCS1B02G284600	TCCCTTGCCTACCTGCTACT	CGATGCAGTAGGTCGATGCT
TraesCS4B02G387600	CGAAGGTTGGTGTGGCAAAG	TTGGTCACTGTTGTGCAGGT
TraesCS7D02G031000	TCCCTGCGCTCTCTTAGTCT	CCACTCGGGCATCTTTCCTT
TraesCS4B02G329100	GACGCTCGTCTTCTACCAGG	CCATCGCAAATGCTCTCTGC
TraesCS7B02G472600	TACCTCAAGCACCACGTCAC	GCACGAGGAAGTACCACTCC
TraesCS5A02G225600	CCCAGAGCCTACTTCCGATG	GATGGAGATGGAGCATGGCA
TraesCS5A02G259000	TTCCTGGAGTCTGCCACAAG	CCTGATGGTGTGCAGCGTA
β-Actin	GGAAAAGTGCAGAGAGACACG	TACAGTGTCTGGATCGGTGGT

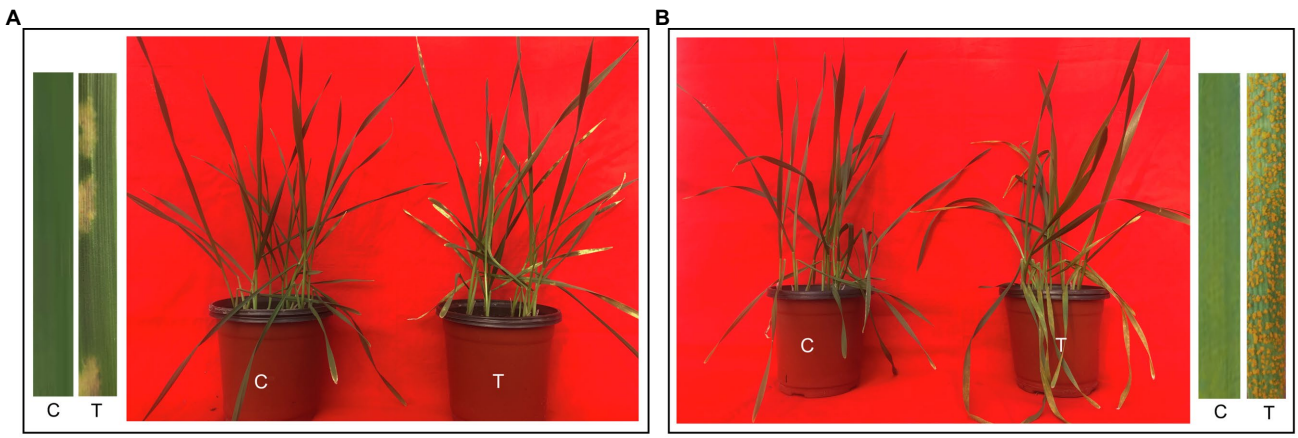


FIGURE 1 | Phenotypes of triticale cultivars Gannong 2 and Shida 1 at 15 days after inoculation with *Pst* race CYR34. **(A,B)** Responses of untreated and treated seedlings of Gannong 2 **(A)** and Shida 1 **(B)**. C, untreated control; T, treatment.

which reflect both a high correlation and slight variability among biological replicates, demonstrate the reliability of the RNA-seq data. On average, 65.35% of total reads were mapped to the wheat reference genome. A total of 33.20% (38,849), 72.44% (84,761), 30.28% (35,438), 45.80% (53,593), 75.22% (88,018), 66.63% (77,969), 88.11% (103,104), and 99.81% (116,796) of transcripts were successfully annotated against COG, GO, KEGG, KOG, Pfam, Swiss-Prot, and Nr databases, respectively, indicating that most transcripts had significant matches to genes of known function. The most abundant transcripts (79,090) were those longer than 1,000 bp (**Table 2**). Further investigation of BLAST results against the Nr database revealed that triticale transcripts were significantly similar to *Aegilops tauschii* proteins (65.85%), followed by *Triticum urartu* (14.36%) and *Triticum aestivum* (7.87%). In addition, other transcripts were similar to those of proteins of gramineous species, including *Hordeum vulgare* (4.13%) and *Brachypodium distachyon* (2.51%; **Figure 2**).

Identification of DEGs

DEGs were identified by comparing expression levels in plants of susceptible Shida 1 and resistant Gannong 2 at 10 and 20 days after inoculation. In each comparison, a FDR threshold of 0.05 and a $\log_2$ (fold change) ≥ 1 were used as criteria for determination of DEGs. The transcript abundance of each gene was further calculated in terms of FPKM and analyzed with Cufflinks (Trapnell et al., 2010). After inoculation with CYR34, 3,335 and 1,578 genes were, respectively, upregulated and downregulated in S10-*vs*-S20, whereas 3,814 were upregulated and 2,581 were downregulated in R10-*vs*-R20 (**Figure 3A**). It shows that the number of upregulated and downregulated DEGs in the resistant cultivar was greater than that of the susceptible cultivar. Compared to susceptible cultivar, a total of 5,869 DEGs (2,655 upregulated and 3,214 downregulated) were identified in the resistant cultivar at 10 days after inoculation (S10-*vs*-R10), and 5,685 DEGs (2,561 upregulated and 3,124 downregulated) were uncovered at 20 days after inoculation (S20-*vs*-R20; **Figure 3A**). It shows that there

TABLE 2 | Summary of the gene functional annotation and the length distribution of the transcripts.

Database	Number of annotated genes	The length distribution of the transcripts		
		<300	300–1,000	≥1,000
COG	38,849 (33.20%)	431	5,504	32,914
GO	84,761 (72.44%)	2,733	20,501	61,527
KEGG	35,438 (30.28%)	1,037	7,806	26,595
KOG	53,593 (45.80%)	863	10,320	42,410
Pfam	88,018 (75.22%)	1,687	19,334	66,997
Swiss-Prot	77,969 (66.63%)	1867	16,702	59,400
eggNOG	103,104 (88.11%)	2,897	25,499	74,708
Nr	116,796 (99.81%)	5,039	32,724	79,033
All	117,016	5,057	32,869	79,090

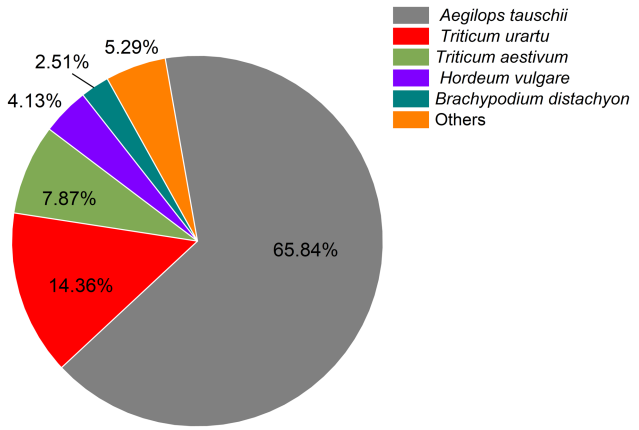


FIGURE 2 | Distribution of matching homologous species based on Nr annotation.

was no significant difference between these DEGs that upregulated and downregulated at 10 days and 20 days, respectively. According to a Venn diagram analysis, 2,560 DEGs were common to both S10-*vs*-R10 and S20-*vs*-R20 (**Figure 3B**).

GO and KEGG Enrichment Analysis of DEGs After Inoculation With CYR34

The above 2,560 DEGs were subjected to GO and KEGG pathway enrichment analyses. The GO enrichment analysis divided the DEGs into three main categories and 47 sub-categories, namely, 19, 15, and 13 sub-categories related to biological processes, cellular components, and molecular functions, respectively. Most enriched DEGs were classified into the biological process category and were mainly involved in functions associated with metabolic process (GO: 0008152), cellular processes (GO: 0009987), single-organism process (GO: 0044699), biological regulation (GO: 0065007), cell (GO: 0005623), binding (GO: 0005488), and catalytic activity (GO: 0003824; **Figure 4**). These results suggest that the majority of DEGs are related to multiple metabolic processes.

The KEGG enrichment analysis revealed that several KEGG pathways were significantly enriched after *Pst* infection, including phenylalanine metabolism (ko00360), plant–pathogen interaction (ko04626), plant hormone signal transduction (ko04075), phenylpropanoid biosynthesis (ko00940), biosynthesis of amino acids (ko01230), and carbon metabolism (ko01200) and so on. A total of 898 DEGs were associated with 117 KEGG pathways. The top 20 KEGG pathways based on the number of enriched DEGs are shown in **Table 3**. Taken together, these results suggest that a range of molecular defense strategies have evolved in triticale. As indicated by these results, the process is complex and likely regulated by the interaction of multiple pathways, such as various hormone signaling and plant–pathogen interaction pathways. We therefore investigated the DEGs involved in these two pathways.

DEGs Related to Phytohormone Signaling

To identify DEGs associated with hormonal response in triticale leaves infected with *Pst*, we analyzed the expressions of genes involved in hormone signal transduction pathways (**Figure 5**). Several genes designated as AUX-responsive, including auxin/indole-3-acetic acid (AUX/IAA), auxin-responsive Gretchen Hagen3 (GH3), and auxin response factor (ARF), were differentially expressed between resistant and susceptible cultivars (**Figure 5A**). Several DEGs involved in SA biosynthesis and signaling were also observed. Three DEGs were annotated as pathogenesis-related protein (PR-1) genes; one was upregulated in the resistant cultivar compared with the susceptible cultivar, and the other two were downregulated (**Figure 5B**). Two brassinosteroid insensitive 2 (BIN2) genes, involved in BR signaling, were significantly upregulated in the resistant cultivar relative to the susceptible one (**Figure 5C**). The only identified DEG encoding JAZ protein was downregulated in the resistant cultivar compared with the susceptible one (**Figure 5D**).

DEGs Related to Plant–Fungus Interaction

One DEG encoding a calcium-dependent protein kinase (CDPK) involved in induction of hypersensitive response and cell wall reinforcement was significantly upregulated in the susceptible cultivar compared with the resistant one. Furthermore, numerous DEGs encoding chitinases (CHI) and glucanase-related genes were upregulated in the susceptible cultivar relative to the resistant one, and the expressions of these genes gradually increased in both cultivars during *Pst* infection. Genes encoding Crassulacean acid metabolism/calmodulin-like proteins (CaM/CMLs) were differentially expressed between the two cultivars, with two DEGs upregulated in the resistant cultivar, and three upregulated in the susceptible cultivar. The expressions of these five genes gradually increased in both cultivars during *Pst* infection. Defense-related genes were also induced; in particular, one and two PR1 genes were upregulated in resistant and susceptible cultivars, respectively (**Figure 6**).

TFs Differentially Expressed in Response to *Pst*

Given that TFs can regulate gene expression by binding to target gene promoters (Singh et al., 2002), an analysis of TFs might provide a basis for understanding the regulatory network operating in triticale seedlings in response to *Pst*. Seventy one of the 2,560 DEGs under *Pst* infection in this study were

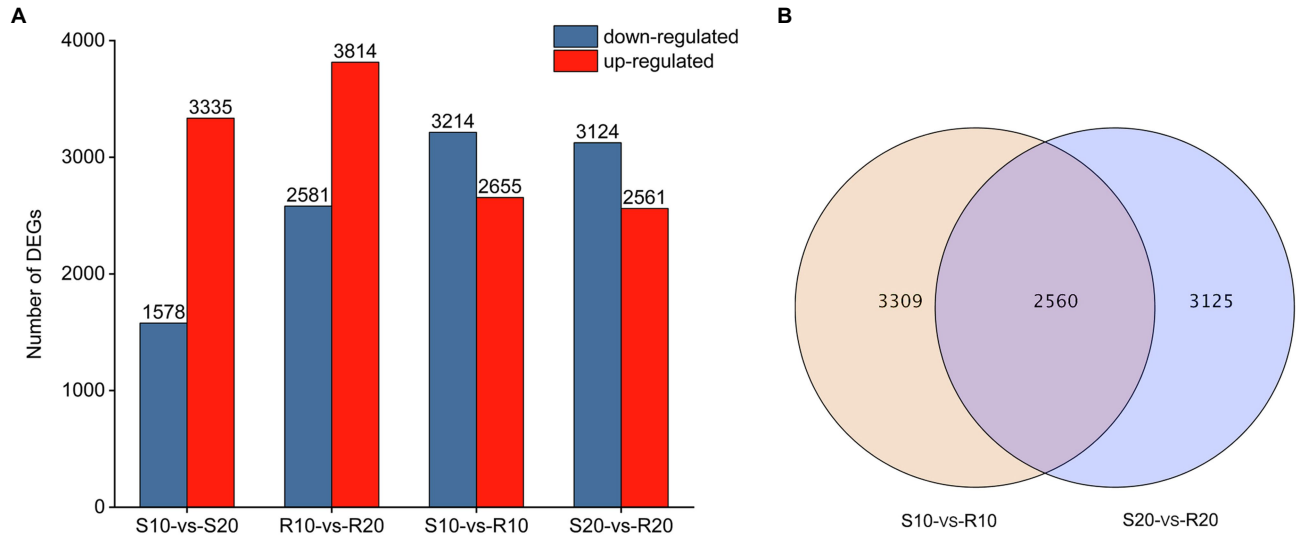


FIGURE 3 | Genes differentially expressed in two triticale cultivars at different treatment time points. **(A)** Number of differentially expressed genes (DEGs) per cultivar at each time point. **(B)** Venn diagram depicting the number of DEGs and their overlap at each time point.

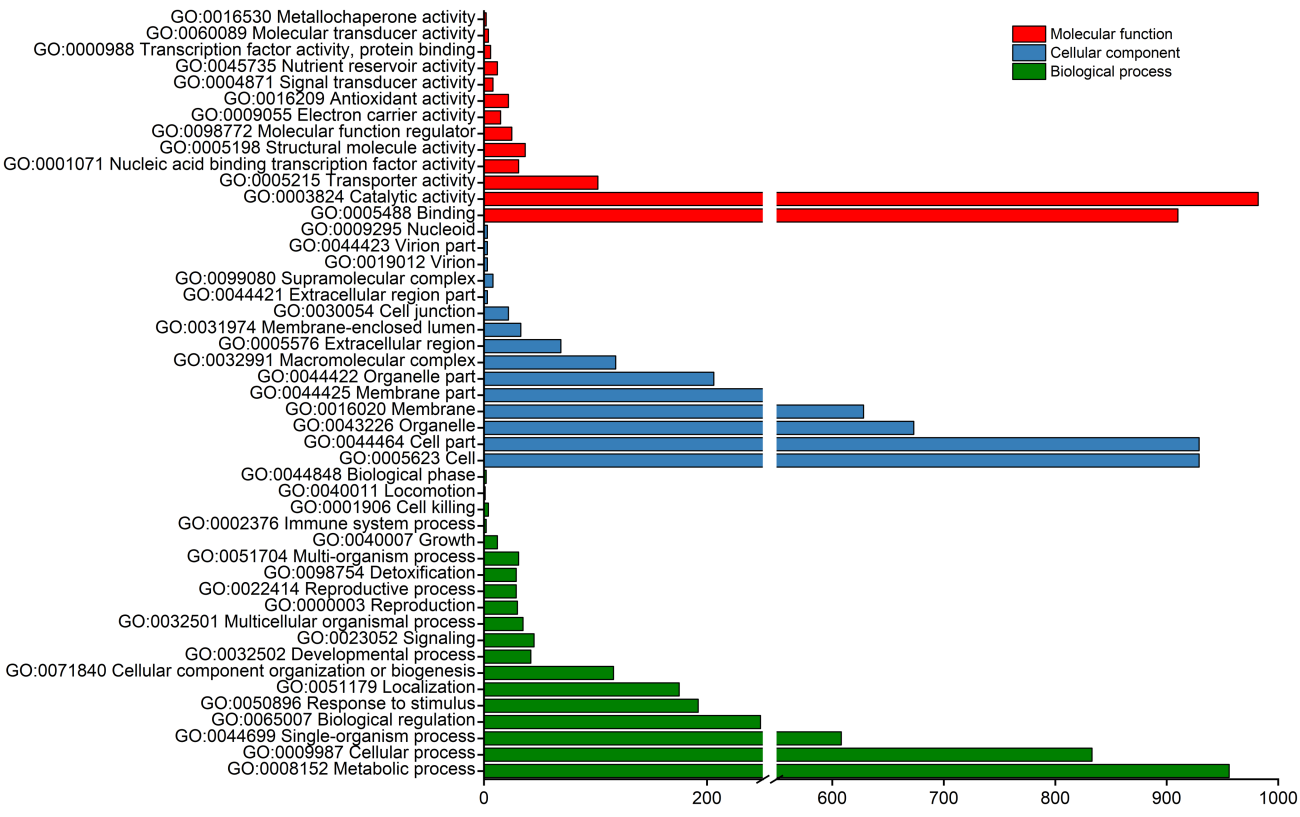


FIGURE 4 | GO enrichment analysis of DEGs. The main GO terms assigned to "biological process," "cellular component," and "molecular function" categories are shown.

identified as TF genes. Among them, 28 were upregulated in the resistant cultivar both at 10 and 20 days after inoculation, whereas 39 were downregulated. All the differentially expressed TF genes encoded members of 23 TF families (**Figure 7**), mainly NAC (16), WRKY (8), FAR1 (8), AP2/ERF-ERF (5), and MADS-MIKC (4) families.

TABLE 3 | Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways significantly enriched in DEGs.

KEGG pathway	Number of genes	*p*	ID
Phenylpropanoid biosynthesis	49	0.0003	ko00940
Biosynthesis of amino acids	39	7.41E-05	ko01230
Carbon metabolism	34	0.0101	ko01200
Amino sugar and nucleotide sugar metabolism	29	3.59E-05	ko00520
Starch and sucrose metabolism	25	0.2021	ko00500
Protein processing in endoplasmic reticulum	23	0.4620	ko04141
Ribosome	23	0.9793	ko03010
Phenylalanine metabolism	21	1.01E-08	ko00360
Plant-pathogen interaction	20	0.4177	ko04626
Alpha-linolenic acid metabolism	18	2.56E-05	ko00592
Cysteine and methionine metabolism	18	0.0054	ko00270
Glutathione metabolism	18	0.0677	ko00480
Glyoxylate and dicarboxylate metabolism	17	0.0003	ko00630
Pyrimidine metabolism	17	0.1279	ko00240
Glycine, serine and threonine metabolism	16	0.0007	ko00260
Galactose metabolism	16	0.0100	ko00052
Glycolysis/gluconeogenesis	16	0.1512	ko00010
RNA transport	16	0.4133	ko03013
Plant hormone signal transduction	16	0.9727	ko04075
Pyruvate metabolism	15	0.0075	ko00620

Characteristics of DEGs Associated With *Pst* Resistance

Numerous DEGs, the *R* gene such as those encoding disease-resistance proteins (RGAs, RPP13, and RPM1), and other pathogenesis-related proteins (PR1.1 and PR1-like) may involved in disease-resistance processes, were identified. Moreover, 31 DEGs encoding disease-resistance kinases were also identified, these included genes for serine/threonine-protein kinase (7), wall-associated receptor kinase (15), G-type lectin S-receptor-like serine/threonine-protein kinase (7), and calcium-dependent protein kinase (2). The expression levels of most kinase genes exhibited no significant differences at 10 vs. 20 days after inoculation. The exceptions were TraesCS5B02G063600, TraesCS7A02G425900, TraesCS7B02G326555, TraesCS7D02G418200, and TraesCS5A02G445700, which were weakly expressed before sporulation (10 days after inoculation) but significantly upregulated afterwards in both cultivars (**Figure 8**). Compared with their expressions in the susceptible cultivar, some of the disease-resistance kinase genes were upregulated (16) in the resistant cultivar, while others were downregulated (15), the genes upregulated and downregulated in the two cultivars were almost the same at 10 and 20 days after inoculation (**Figure 8**).

qRT-PCR Analysis of DEGs Associated With *Pst* Resistance

In this study, 16 DEGs (including 10 kinase-related genes, three disease-resistance protein genes and three TFs) were randomly selected for qRT-PCR analysis and were confirmed to be upregulated in the resistant cultivar both at 10 and 20 days after inoculation (**Figure 9**). Although some differences in the range of expression changes were observed, the expression trends of these genes based on fold changes in relative gene expression as measured by qRT-PCR were basically consistent with their corresponding transcript abundance changes detected by RNA-Seq; this indicates that the expressions of the 16 genes were reliable in the present study (**Figure 9**). All of these genes were more highly expressed in the resistant cultivar than in the susceptible cultivar under *Pst* infection. The four most significantly expressed genes (TraesCS2D02G462000, TraesCS7A02G425900, TraesCS4B02G387600, and TraesCS7D02G031,000, corresponding to *FLS2*, *WAK5*, *LecRK-IX.2*, and *RPM1*, respectively) may play an active role during PTI or ETI.

DISCUSSION

Yellow (stripe) rust is one of the most destructive diseases of plants, causing severe losses in small grain cereals, including triticale (Losert et al., 2017). The highly aggressive race CYR34 currently dominates the yellow rust population in wheat and triticale in China (Huang et al., 2019), therefore, clarifying the resistance level of triticale to CYR34 and screening the yellow rust resistance-related genes is of great significance in triticale yellow rust-resistance breeding. RNA-seq has provided a new avenue to study the responses of plants to pathogen. Until the present study, however, RNA-seq had not been exploited to investigate yellow rust resistance genes in triticale, and the mechanism of rust resistance in triticale was still unknown. We therefore used RNA-seq to analyze the expression of *Pst*-related genes in resistant and susceptible triticale cultivars at later stages (before and after sporulation of susceptible cultivar) to identify changes in gene expression that can be linked to key aspects of the infection process.

After inoculation with *Pst*, symptoms of chlorosis are generally observed 6–8 days later, whereas sporulation commences approximately 12–14 days post-inoculation under favorable conditions (Chen et al., 2014). In the present study, Shida 1 began to sporulate 13–15 days after inoculation, and the area of the spores on the leaves gradually increased, compared to Shida 1, Gannong 2 had only a few necrotic stripes, but no sporulation. Therefore, Shida 1 was identified as a highly

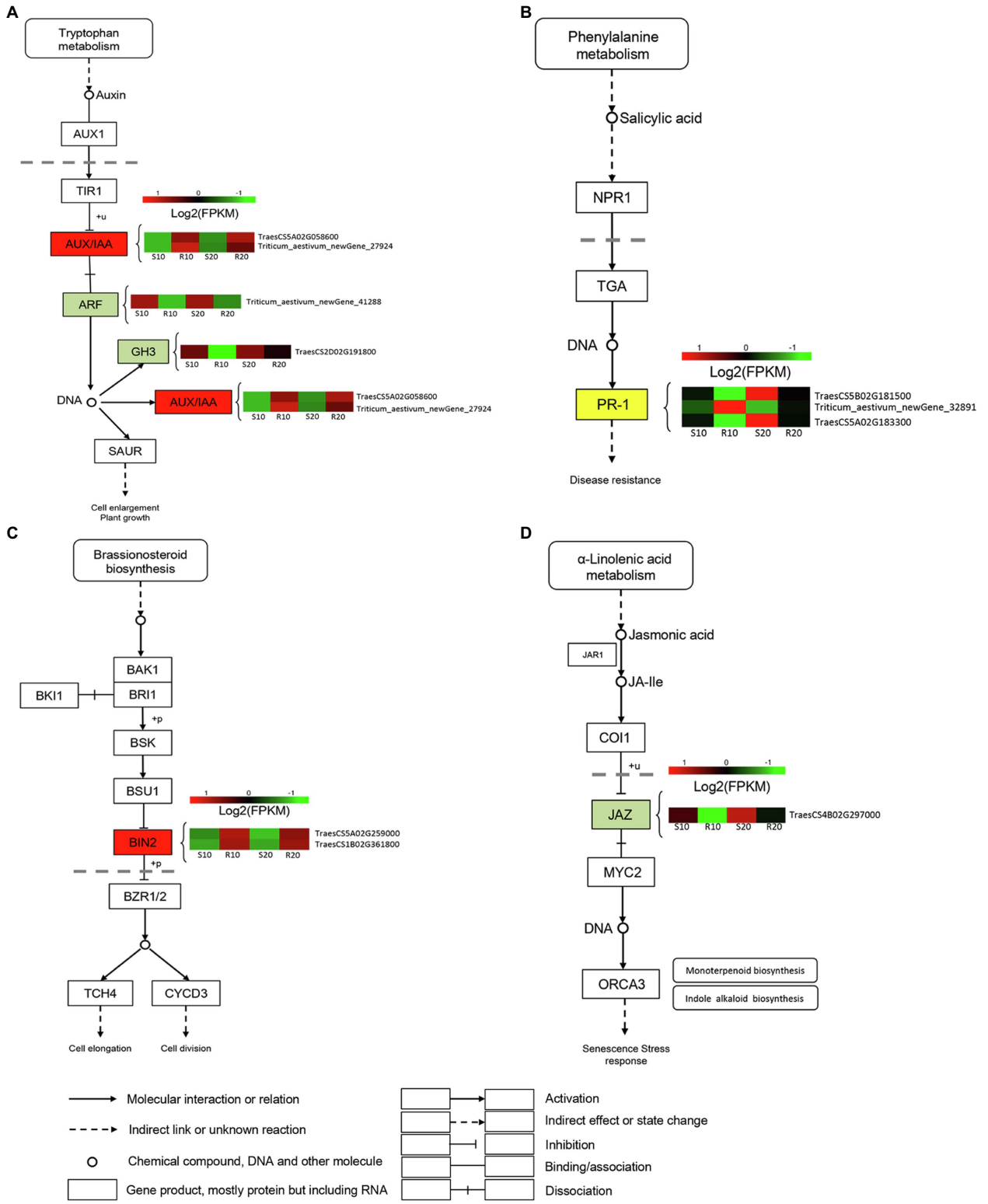


FIGURE 5 | DEGs related to phytohormone signaling transduction pathway. **(A)** Auxin (AUX); **(B)** salicylic acid (SA); **(C)** brassinosteroid (BR); and **(D)** jasmonic acid (JA). Expression values are presented as $\log_2$ FPKM values (red for upregulated and green for downregulated), and the vertical columns, from left to right, represent S10, R10, S20, and R20. The color of the background indicates the pattern of gene expression: red for upregulation, green for downregulation, and yellow for both upregulation and downregulation.

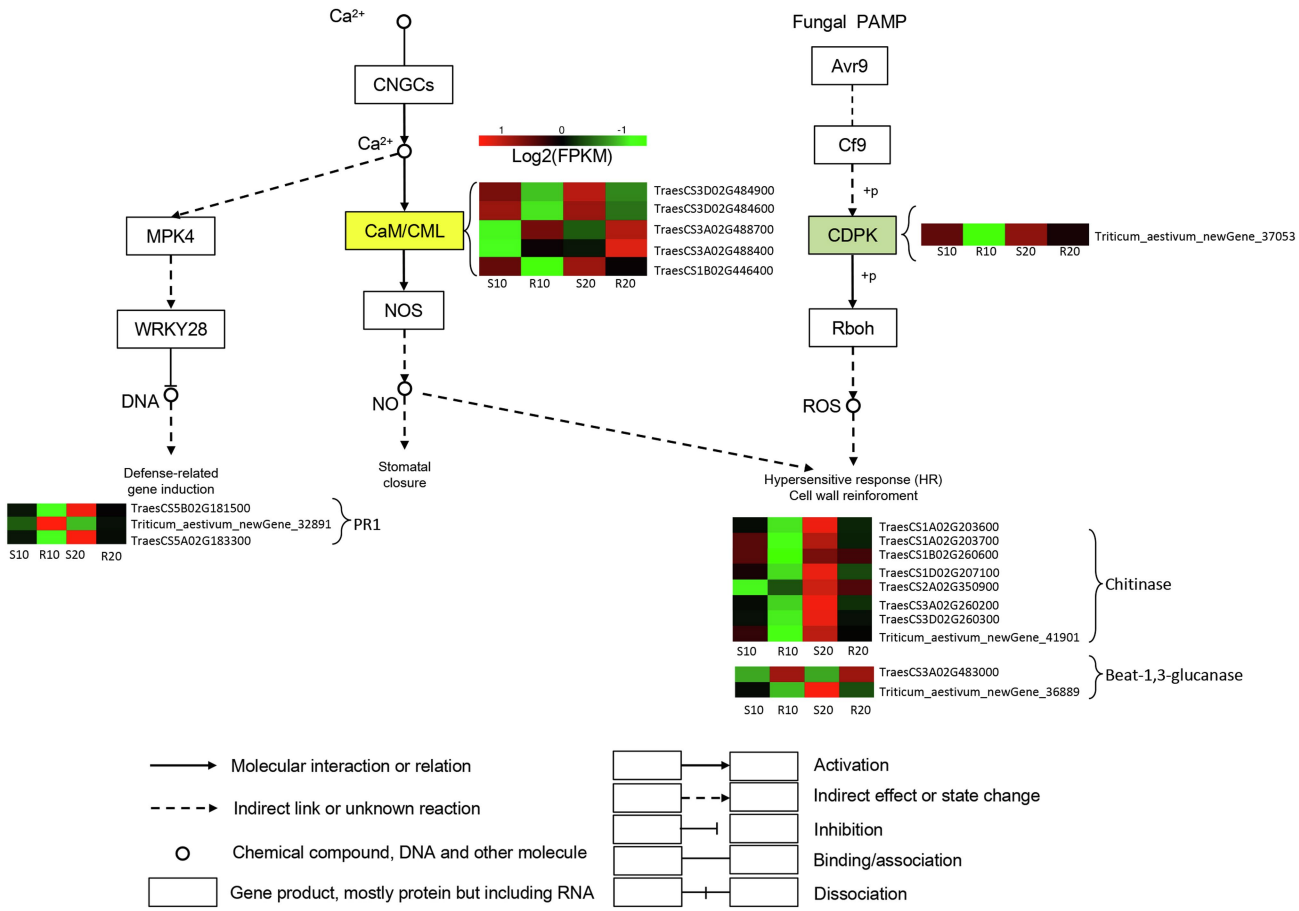


FIGURE 6 | DEGs related to triticale–*Pst* interaction pathway. Expression values are presented as $\log_2$ FPKM values (red for upregulated and green for downregulated), and the vertical columns, from left to right, represent S10, R10, S20, and R20. The color of the background indicates the pattern of gene expression: red for upregulation, green for downregulation, and yellow for both upregulation and downregulation.

susceptible cultivar, and Gannong 2 was a resistant cultivar, this is consistent with the symptoms described by Huang et al. (2014). Samples collected 10 and 20 days after inoculation were therefore used for an analysis of triticale defense mechanisms during infection. In the present study, DEGs in resistant and susceptible triticale were analyzed by RNA-Seq after inoculation with *Pst*. As a result, we identified 2,560 genes that were differentially expressed in the two cultivars during *Pst* infection. In a previous study, Yuan et al. (2019) found that more DEGs were induced in a *Fusarium graminearum*-resistant line than in a susceptible one. In our study, we found no significant difference in the number of DEGs before and after sporulation, but the number of DEGs in the resistant cultivar (Gannong 2) was greater than that of the susceptible cultivar (Shida 1). We thus hypothesize that a stronger defense response was activated in the resistant cultivar than in the susceptible one, similar to the results of Yuan et al. (2019). Our findings are also consistent with the results of Dobon et al. (2016), who showed that the *Pst* pathogen suppresses defense-related gene expression in a compatible interaction to enable successful colonization.

KEGG pathways enriched in DEGs were also analyzed. In a previous study, Hao et al. (2016) found that "phenylpropanoid biosynthesis" pathway were involved in the yellow rust defense response of wheat in the early stage. In our study, the most significant enrichment occurred in the "phenylpropanoid biosynthesis" pathway, and the number of DEGs was largest in this pathway, which indicated that this pathway was also involved in the response of triticale to *Pst* infection in a later stage. In addition, the number of DEGs associated with "plant–pathogen interaction" and "plant hormone signal transduction" was also higher in our study. Plant hormones such as SA, JA, and ET play major roles in the regulation of plant defense responses (Camehl et al., 2013). Auxin and other plant hormones have also been reported to be involved in the regulation of plant resistance to pathogens (Spoel and Dong, 2008). Dobon et al. (2016) have identified response pathways related to the wheat stress hormones SA, JA, ethylene, and ABA, whose balance is fine-tuned to regulate plant innate immunity. In the present study, we therefore focused on plant hormone signal transduction pathways (Ko04075). Annotation of DEGs classified 16 genes into plant hormone signal transduction

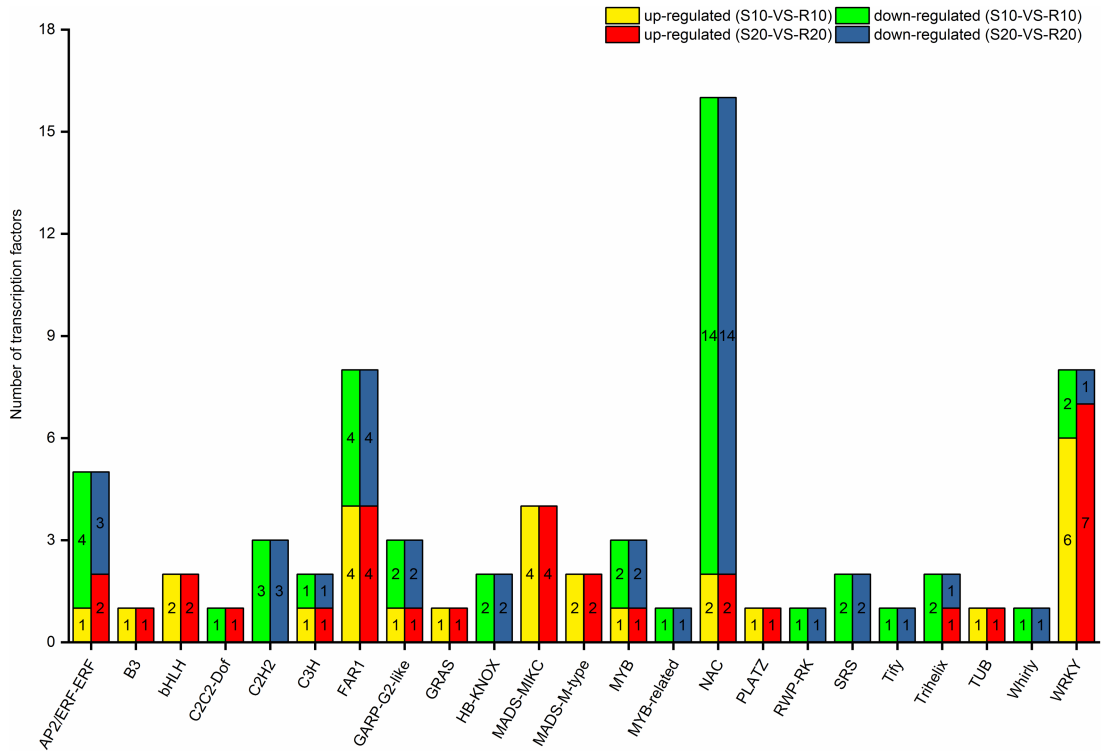


FIGURE 7 | Distribution of differentially expressed transcription factors in triticale leaves infected with *Pst*.

pathways, including those associated with auxin, BR, JA, and SA. AUX/IAA proteins are positive regulators of disease resistance, but auxin signaling is inactivated by AUX/IAA protein binding (Yang et al., 2014). After infection of the resistant cultivar (Gannong 2) by *Pst*, AUX/IAA gene expressions were upregulated, which may have suppressed the auxin signaling pathway to provide resistance to *Pst*. BRs promote plant growth and development and have a complex positive role in the plant immune defense response of Selaginella and Arabidopsis (Cheon et al., 2013), but they play a negative role in the interaction between rice and pathogens (He et al., 2017). De Vleesschauwer et al. (2012) have reported that BRs boost the root immunity of rice against infection by the pathogen *Pythium graminicola*. Brassinosteroid-INsensitive 2 (BIN2) is an Arabidopsis GSK3-like kinase that negatively regulates the intracellular signal transduction of BRs (Peng et al., 2010). In the present study, we found that the two DEGs encoding BIN2 were strongly upregulated in the resistant cultivar, possibly inhibiting BR signaling through over-expression of the *BIN2* gene. The role of BR in the interaction between triticale and *Pst* needs to be studied further. JAZ proteins act as negative regulators of JA signaling by suppressing MYC2 proteins (Yang et al., 2014), but JA is a positive regulator of immunity against necrotrophic pathogens (Tsuda and Somssich, 2015). In our study, we found that one *JAZ* gene was downregulated in the resistant cultivar (Gannong 2) but upregulated in the susceptible cultivar (Shida 1). We hypothesize that this downregulation is due to active suppression of *JAZ* gene expression by *Pst* in the resistant cultivar. NPR1 is a key regulator of systemic acquired resistance (SAR)-related PR gene expression, and NPR1-overexpressing plants have enhanced resistance to various pathogens without constitutively expressing the PR genes (Kinkema et al., 2000). No differential expression of NPR1 associated with SA signaling was detected in our study, and downstream PR-1 genes in resistant cultivar (Gannong 2) were both upregulated and downregulated. This result is in keeping with observations showing that defense responses to necrotrophic organisms are SA-independent (Kinkema et al., 2000).

A plant–pathogen interaction pathway (Ko04626) is also a component of the plant environmental adaptation and immune system (Yuan et al., 2019). In the present study, 20 DEGs were assigned to this plant–pathogen interaction pathway. Among them, one DEG encoding CDPK was identified and found to be significantly upregulated in the susceptible cultivar (Shida 1). We hypothesize that susceptible varieties can still perceive the fungus at later stages of infection, thereby triggering the signaling pathway required for ROS accumulation. ROS species accumulate during infection, resulting in hypersensitive response and cell wall reinforcement (Zhu et al., 2020). In plants, pathogen infection activates a series of defense responses, including the synthesis of PR proteins. β-1,3-glucanase is a type of PR protein that defends against fungal infection by hydrolyzing β-1,3-glucan, a fungal cell wall polymer, to inhibit fungal growth (Fujimori et al., 2016). Two DEGs encoding β-1,3-glucanase-related genes were identified in our study; one DEG (TraesCS3A02G483000) was significantly upregulated in

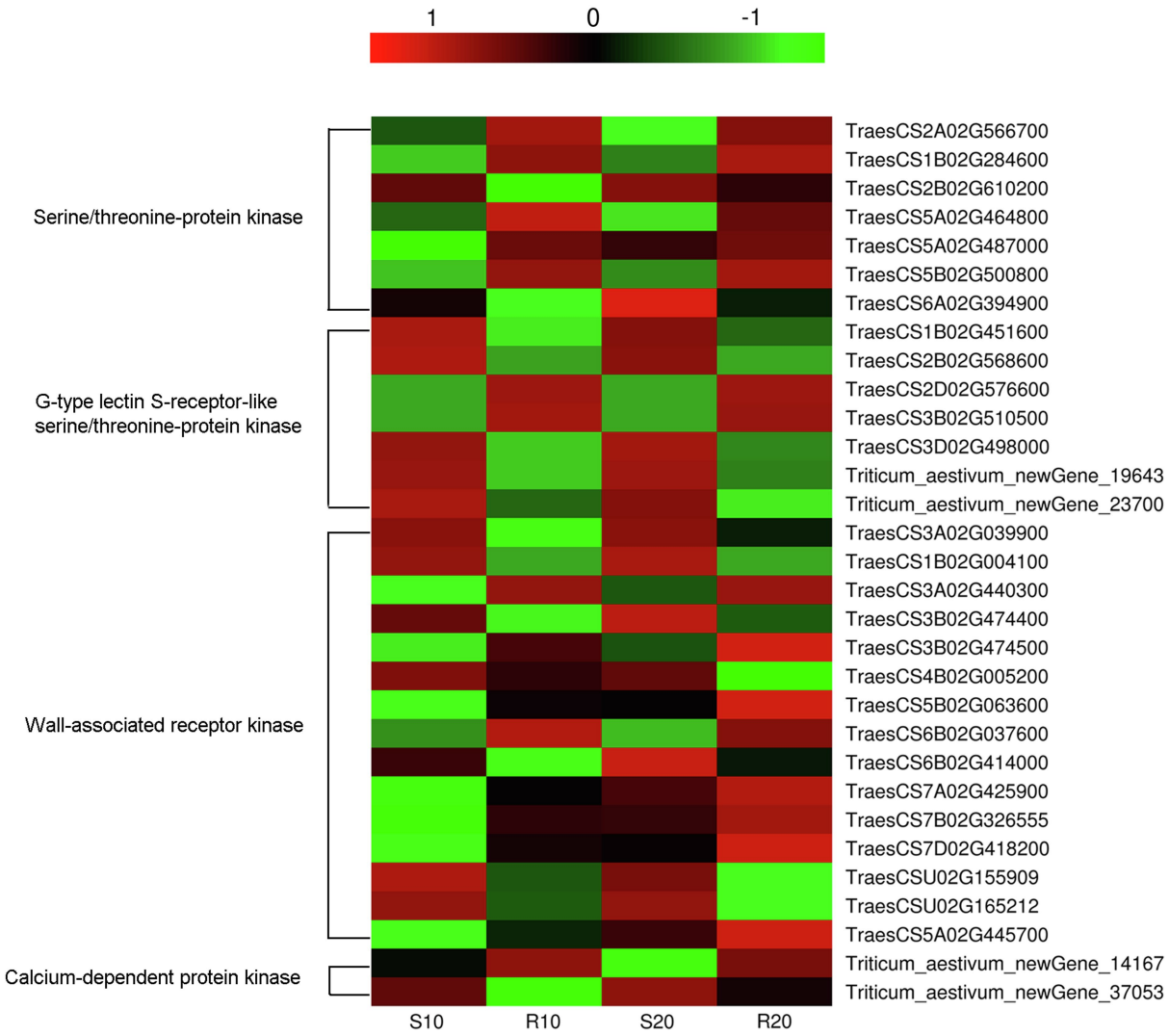


FIGURE 8 | Heatmap of transcript levels of disease-resistance kinases related to *Pst* resistance in leaves of resistant and susceptible triticale cultivars 10 and 20 days after *Pst* infection.

resistant cultivar (Gannong 2) and downregulated in susceptible cultivar (Shida 1), and the other one followed the opposite expression pattern. Several studies have shown that the chitinase activity of resistant varieties was higher than that of susceptible varieties at the early stage of pathogen infection (Zhao, 2004). In our study, DEGs related to chitinase were all downregulated in resistant cultivar Gannong 2 during *Pst* infection. We hypothesize that these chitinase-related DEGs may also have been expressed at a very early stage (germinating spores) in the resistant triticale cultivar. Previous studies have indicated that deregulation of CaM/CML genes or loss of CaM/CML function strongly affects immune responses (Cheval et al., 2013). In our study, three CaM/CMLs were markedly downregulated in resistant cultivar (Gannong 2) after *Pst* infection, whereas the other two were markedly upregulated. PR1 proteins are considered to be markers for an enhanced defensive state conferred by pathogen-induced SAR (Sekhon et al., 2006). Three DEGs were annotated as PR1 proteins in our study. The above DEGs likely participated in the defense response of the susceptible cultivar and the resistance response of the resistant cultivar, which suggests that triticale synthetically use multiple synthetic pathways to prevent fungal infection.

When plants are invaded by pathogens, some TFs can bind to specific *cis*-regulatory elements present in target gene promoters and activate downstream transduction pathways (Singh et al., 2002). TFs can therefore induce many downstream genes in plants to improve their resistance to disease (Tsuda and Somssich, 2015). TF families implicated in the regulation of plant defense mechanisms against pathogens mainly include WRKY, MYB, AP2/ERF, and bZIP (Alves et al., 2014). In the present study, members of NAC, WRKY, FAR1, AP2/ERF-ERF, GRAS, and MADS-MIKC TF families were differentially expressed in response to *Pst* infection. The largest number of differentially expressed TFs was NAC family members. Although the primary

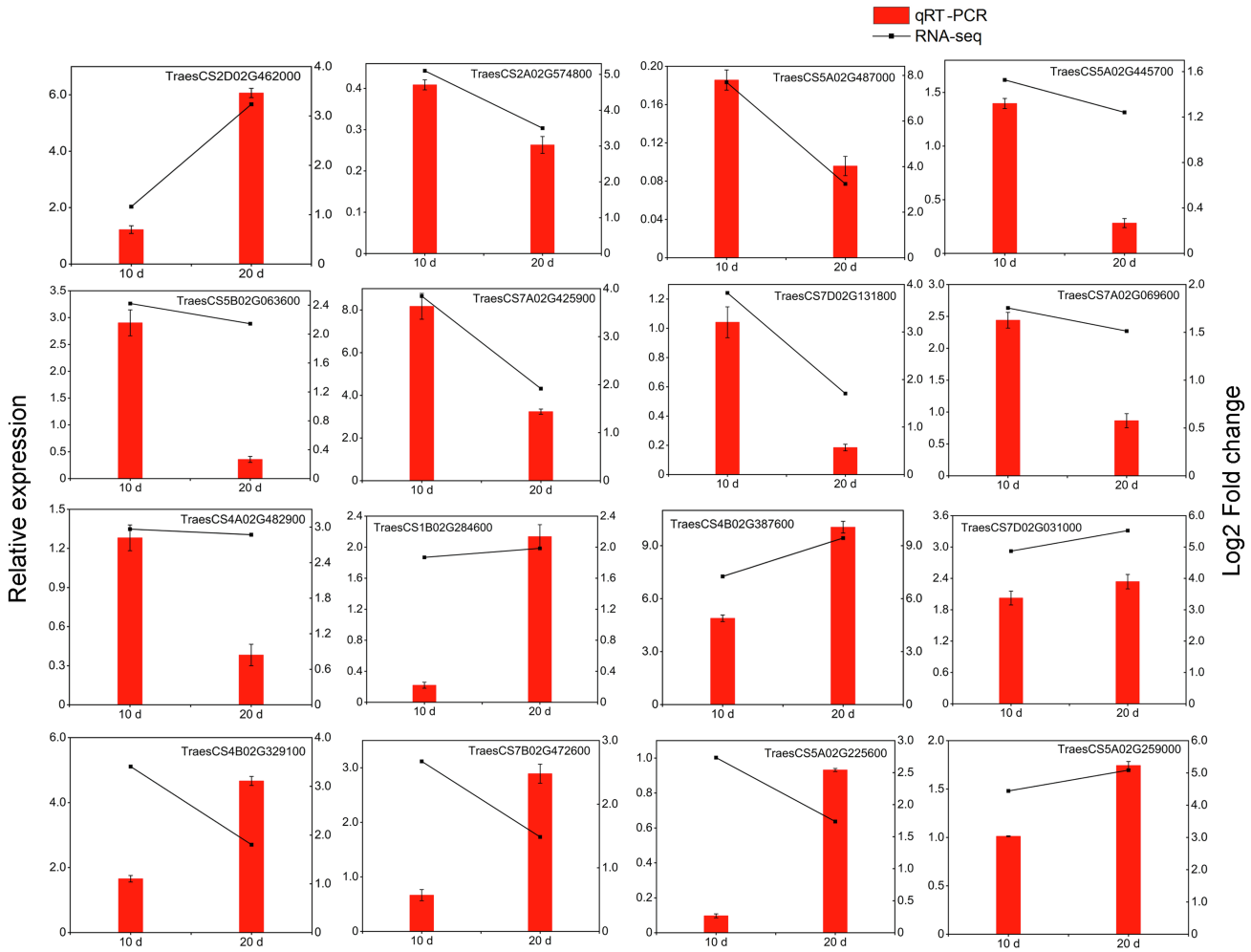


FIGURE 9 | qRT-PCR analysis of selected DEGs in triticale under the yellow rust infection. Red columns indicate relative expression levels determined by qRT-PCR (left *y*-axis), and dashed lines represent transcript level changes (log_2 fold change) based on FPKM values determined by RNA-Seq (right *y*-axis).

role of the NAC family is regulating responses to abiotic stresses such as drought and salinity, several NAC TFs in Arabidopsis (*At*ANAC019, *At*ANAC055 and *At*ANAC072) have also been identified as important immune components (Tsuda and Somssich, 2015). *Ta*NAC1 acts as a negative regulator of yellow rust resistance in wheat and is able to regulate JA and SA signal defense pathways (Wang et al., 2015). In our study, most of the NACs in the resistant cultivar (Gannong 2) were downregulated compared with the susceptible cultivar (Shida 1). The above-mentioned study demonstrated the importance of NAC TFs in plant disease resistance, a conclusion also confirmed by our observation of a large number of downregulated NAC genes in the resistant triticale cultivar after infection with *Pst*. WRKY is the most reported transcription factor associated with the disease resistance of plants, which plays an important regulatory role in the induction of defense responses to pathogens, and the major function of numerous distinct WRKY members in host immunity has been firmly established in Arabidopsis, barley, and rice (Pandey and Somssich, 2009). In addition, Cheng et al. (2015) have reported that the transcriptional regulation of *WRKY45*, *WRKY13*, and *WRKY42* is involved in the blast resistance of rice. In the present study, we identified eight WRKYs, six and seven of which were upregulated at 10 and 20 days after inoculation, respectively. Rice *OsWRKY45* has been reported to play a pivotal role in fungal disease resistance by mediating SA signaling (Shimono et al., 2012). In regards to the differentially expressed TFs in our study, some of these TFs may regulate the yellow rust resistance of triticale, which is consistent with the previous reports of positive and negative responses of TFs of wheat, rice, and other plants to fungal disease. The relationships of these TFs in signaling pathways of triticale against *Pst* are unknown, however, and their functions in triticale yellow rust resistance require further study.

The growth, development, and disease resistance-related functions of receptor-like protein kinases (RLKs), which are commonly found in plants, are a focus of increasing attention (Sun et al., 2004). As signaling molecule receptors, RLKs play

an important role in signal transduction. Most of these kinases exist at the nodes of various pathways, thereby forming a complex network with their upstream and downstream proteins (Li et al., 2008). The accumulation of intracellular signaling molecules induces specific phosphorylation and dephosphorylation cascades, with the ensuing metabolic changes leading to specific plant responses that include the activation of cadres of genes involved in particular defense responses (Gachomo et al., 2003). Some resistance genes are homologous to RLK genes in plants. Most RLKs involved in plant resistance belong to the LRR-RLK sub-family; examples include rice *Xa21* involved in bacterial pathogen resistance (Song et al., 1995) and Arabidopsis *FLS2* related to flagellin perception (Robatzek et al., 2006), which plays an important role in plant resistance response and pathogen recognition. Most RLKs in plants are serine/threonine kinases (Li et al., 2008). We also identified many serine/threonine-protein kinases in this study, most of which were found to be significantly upregulated in the resistant cultivar (Gannong 2) and downregulated in the susceptible cultivar (Shida 1), which may be why Gannong 2 can inhibit pathogen infection and exhibit high disease resistance. In addition, five kinase genes had significantly different expressions before (10 days) and after (20 days) sporulation, thus showing that they may have some functions in disease defense at a later stage. Furthermore, pathogens use effectors to interfere with the plant-based PTI defense system; on the other hand, plants have developed ETI responses based on effector recognition (Boller and Felix, 2009). Several disease-resistance proteins, such as RPM1, RPP13, and RGAs, were upregulated in resistant cultivar (Gannong 2) and may be involved in ETI. Plants maintain many resistance proteins (*R* genes) that directly or indirectly recognize effectors released from pathogens and initiate ETI (Kim et al., 2012).

Overall, this study represents the first step toward understanding of the molecular mechanisms involved in resistance to *Pst* in triticale. Our results provide a molecular basis for the elucidation of triticale–*Pst* interactions and can serve as a candidate gene resource for yellow rust resistance breeding in triticale.

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 at: NCBI—SRR18675448, SRR18675447, SRR18675446, SRR18675445, SRR18675444, SRR18675443, SRR18675442, SRR18675441.

AUTHOR CONTRIBUTIONS

FZ conducted the experiments, analyzed the data, and wrote the manuscript. KN participated in qRT-PCR experiment, provided technical support, and revised the manuscript critically. XT and WD provided the experimental materials. WD conceived the experiment and manuscript revision. All authors contributed to the article and approved the submitted version.

FUNDING

This study was financially supported by Science and Technology Program of Gansu Province (20YF8NA129 and 21ZD4NA012), Tibet Forage Industry Project (XZ202101ZD0003N), and the National Natural Science Foundation of China (31760702).

ACKNOWLEDGMENTS

The authors thank all the laboratory members for their useful suggestions, support and encouragement. The authors also appreciate Biomarker Technologies Co., Ltd. (Beijing, China) for the transcriptomics determination experiments.

REFERENCES

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Alves, M. S., Dadalto, S. P., Gonçalves, A. B., de Souza, G. B., Barros, V. A., and Fietto, L. G. (2014). Transcription factor functional protein-protein interactions in plant defense responses. *Proteomes* 2, 85–106. doi: 10.3390/proteomes2010085

Audenaert, K., Troch, V., Landschoot, S., and Haesaert, G. (2014). Biotic stresses in the anthropogenic hybrid triticale (×*Triticosecale* Wittmack): current knowledge and breeding challenges. *Eur. J. Plant Pathol.* 140, 615–630. doi: 10.1007/s10658-014-0498-2

Bari, R., and Jones, J. D. G. (2009). Role of plant hormones in plant defence responses. *Plant Mol. Biol.* 69, 473–488. doi: 10.1007/s11103-008-9435-0

Baron, V. S., Juskiw, P. E., and Aljarrah, M. (2015). "Triticale as a Forage," in *Triticale*. ed. F. Eudes (Canada: Springer, Cham Press), 189–212.

Boller, T., and Felix, G. (2009). A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu. Rev. Plant Biol.* 60, 379–406. doi: 10.1146/annurev.arplant.57.032905.105346

Bozkurt, O., Unver, T., and Akkaya, M. S. (2007). Genes associated with resistance to wheat yellow rust disease identified by differential display analysis. *Physiol Mol Plant P.* 71, 251–259. doi: 10.1016/j.pmpp.2008.03.002

Camehl, I., Sherameti, I., Seebald, E., Johnson, J. M., and Oelmüller, R. (2013). "Role of defense compounds in the beneficial interaction between *Arabidopsis thaliana* and *Piriformospora indica*," in *Piriformospora indica*. eds. A. Varma, G. Kost and R. Oelmüller (Springer, Berlin, Heidelberg Press), 239–250.

Chen, W. Q., Wellings, C., Chen, X. M., Kang, Z. S., and Liu, T. G. (2014). Wheat stripe (yellow) rust caused by *Puccinia striiformis* f. sp. tritici. *Mol. Plant Pathol.* 15, 433–446. doi: 10.1111/mpp.12116

Cheng, H. T., Liu, H. B., Deng, Y., Xiao, J. H., Li, X. H., and Wang, S. P. (2015). The *WRKY45-2 WRKY13 WRKY42* transcriptional regulatory cascade is required for rice resistance to fungal pathogen. *Plant Physiol.* 167, 1087–1099. doi: 10.1104/pp.114.256016

Cheon, J., Fujioka, S., Dilkes, B. P., and Choe, S. (2013). Brassinosteroids regulate plant growth through distinct signaling pathways in Selaginella and Arabidopsis. *PLoS One* 8:e81938. doi: 10.1371/journal.pone.0081938

Cheval, C., Aldon, D., Galaud, J. P., and Ranty, B. (2013). Calcium/calmodulin-mediated regulation of plant immunity. *Biochim. Biophys. Acta* 1833, 1766–1771. doi: 10.1016/j.bbamcr.2013.01.031

De Vleesschauwer, D., Van Buyten, E., Satoh, K., Balidion, J., Mauleon, R., Choi, I. R., et al. (2012). Brassinosteroids antagonize gibberellin- and salicylate-

mediated root immunity in rice. *Plant Physiol.* 158, 1833–1846. doi: 10.2307/41496323

Dobon, A., Bunting, D. C., Cabrera-Quio, L. E., Uauy, C., and Saunders, D. G. (2016). The host-pathogen interaction between wheat and yellow rust induces temporally coordinated waves of gene expression. *BMC Genomics* 17:380. doi: 10.1186/s12864-016-2684-4

Fujimori, N., Enoki, S., Suzuki, A., Naznin, H. A., Shimizu, M., and Suzuki, S. (2016). Grape apoplasmic β-1,3-glucanase confers fungal disease resistance in Arabidopsis. *Sci. Hortic.* 200, 105–110. doi: 10.1016/j.scienta.2016.01.008

Gachomo, E. W., Shonukan, O. O., and Kotchoni, S. O. (2003). The molecular initiation and subsequent acquisition of disease resistance in plants. *Afr. J. Biotechnol.* 2, 26–32. doi: 10.5897/AJB2003.000-1005

Glazebrook, J. (2005). Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.* 43, 205–227. doi: 10.1146/annurev.phyto.43.040204.135923

González, J. M., Muñiz, L. M., and Jouve, N. (2005). Mapping of QTLs for androgenetic response based on a molecular genetic map of × *Triticosecale* Wittmack. *Genome* 48, 999–1009. doi: 10.1139/g05-064

Gyawali, S., Verma, R. P. S., Kumar, S., Bhardwaj, S. C., Gangwar, O. P., Selvakumar, R., et al. (2017). Seedling and adult-plant stage resistance of a world collection of barley genotypes to stripe rust. *J. Phytopathol.* 166, 18–27. doi: 10.1111/jph.12655

Hao, Y. B., Wang, T., Wang, K., Wang, X. J., Fu, Y. P., Huang, L. L., et al. (2016). Transcriptome analysis provides insights into the mechanisms underlying wheat plant resistance to stripe rust at the adult plant stage. *PLoS One* 11:e0150717. doi: 10.1371/journal.pone.0150717

He, Y., Zhang, H., Sun, Z., Li, J., Hong, G., Zhu, Q., et al. (2017). Jasmonic acid-mediated defense suppresses brassinosteroid-mediated susceptibility to Rice black streaked dwarf virus infection in rice. *New Phytol.* 214, 388–399. doi: 10.1111/nph.14376

Huang, F., Li, W. Z., Chen, T. Q., Xiong, S. J., Wang, W., and Zhang, L. Y. (2014). Identification of the resistance to wheat stripe rust of Guizhou major wheat cultivars (lines). *Guizhou Agric Sci.* 42, 81–84+88. doi: 10.3969/j.issn.1001-3601.2014.02.021

Huang, L., Liu, T. G., Liu, B., Gao, L., Luo, P. G., and Chen, W. Q. (2019). Resistance evaluation of 197 Chinese wheat core germplasms to a new stripe rust race, CYR34. *Plant Prot.* 45, 148–154. doi: 10.16688/j.zwbh.2018067

Jones, J. D. G., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286

Jones, D. A., and Takemoto, D. (2004). Plant innate immunity—direct and indirect recognition of general and specific pathogen-associated molecules. *Curr. Opin. Immunol.* 16, 48–62. doi: 10.1016/j.coi.2003.11.016

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Kim, J., Lim, C. J., Lee, B. W., Choi, J. P., Oh, S. K., Ahmad, R., et al. (2012). A genome-wide comparison of NB-LRR type of resistance gene analogs (RGA) in the plant kingdom. *Mol. Cells* 33, 385–392. doi: 10.1007/s10059-012-0003-8

Kinkema, M., Fan, W. H., and Dong, X. N. (2000). Nuclear localization of NPR1 is required for activation of PR gene expression. *Plant Cell* 12, 2339–2350. doi: 10.2307/3871233

Li, L. Y., Jin, M. J., Liu, Q., and Liu, G. Z. (2008). Function and substrate identification of plant receptor-like kinases. *Chin. J. Biochem. Mol. Biol.* 24, 113–119. doi: 10.3969/j.issn.1007-7626.2008.02.006

Liu, J. L., Liu, X. L., Dai, L. Y., and Wang, G. L. (2007). Recent progress in elucidating the structure, function and evolution of disease resistance genes in plants. *J. Genet. Genomics* 34, 765–776. doi: 10.1016/S1673-8527(07)60087-3

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

Losert, D., Maurer, H. P., Leiser, W. L., and Würschum, T. (2017). Defeating the warrior: genetic architecture of triticale resistance against a novel aggressive yellow rust race. *Theor. Appl. Genet.* 130, 685–696. doi: 10.1007/s00122-016-2843-7

Mao, X. Z., Cai, T., Olyarchuk, J. G., and Wei, L. P. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21, 3787–3793. doi: 10.2307/1592215

Pandey, S. P., and Somssich, I. E. (2009). The role of WRKY transcription factors in plant immunity. *Plant Physiol.* 150, 1648–1655. doi: 10.1104/pp.109.138990

Peng, P., Zhao, J., Zhu, Y. Y., Asami, T., and Li, J. M. (2010). A direct docking mechanism for a plant GSK3-like kinase to phosphorylate its substrates. *J. Biol. Chem.* 285, 24646–24653. doi: 10.1074/jbc.M110.142547

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122

Pieterse, C. M. J., Van der Does, D., Zamioudis, C., Leon-Reyes, A., and Van Wees, S. C. M. (2012). Hormonal modulation of plant immunity. *Annu. Rev. Cell Dev. Biol.* 28, 489–521. doi: 10.1146/annurev-cellbio-092910-154055

Ramamoorthy, R., Jiang, S. Y., Kumar, N., Venkatesh, P. N., and Ramachandran, S. (2008). A comprehensive transcriptional profiling of the WRKY gene family in rice under various abiotic and phytohormone treatments. *Plant Cell Physiol.* 49, 865–879. doi: 10.1093/pcp/pcn061

Ren, Y., Li, S. R., Xia, X. C., Zhou, Q., He, Y. J., Wei, Y. M., et al. (2015). Molecular mapping of a recessive stripe rust resistance gene *yrMY37* in Chinese wheat cultivar Mianmai 37. *Mol. Breeding.* 35, 97. doi: 10.1016/S2095-3119(14)60781-4

Robatzek, S., Chinchilla, D., and Boller, T. (2006). Ligand-inducedendocytosis of the pattern recognition receptor *FLS2* in Arabidopsis. *Genes Dev.* 20, 537–542. doi: 10.1101/gad.366506

Robert-Seilaniantz, A., Grant, M., and Jones, J. D. G. (2011). Hormone crosstalk in plant disease and defense: more than just jasmonate-salicylate antagonism. *Annu. Rev. Phytopathol.* 49, 317–343. doi: 10.1146/annurev-phyto-073009-114447

Roelfs, A. P., Singh, R. P., and Saari, E. E. (1992). *Rust Disease of Wheat: Concepts and Methods of Disease Management*. Mexico: CIMMYT Press.

Sekhon, R. S., Kuldau, G., Mansfield, M., and Chopra, S. (2006). Characterization of Fusarium-induced expression of flavonoids and PR genes in maize. *Physiol. Mol. Plant P.* 69, 109–117. doi: 10.1016/j.pmpp.2007.02.004

Shimono, M., Koga, H., Akagi, A., Hayashi, N., Goto, S., Sawada, M., et al. (2012). Rice *WRKY45* plays important roles in fungal and bacterial disease resistance. *Mol. Plant Pathol.* 13, 83–94. doi: 10.1111/j.1364-3703.2011.00732.x

Singh, K. B., Foley, R. C., and Oate-Sánchez, L. (2002). Transcription factors in plant defense and stress responses. *Curr. Opin. Plant Biol.* 5, 430–436. doi: 10.1016/S1369-5266(02)00289-3

Sodkiewicz, W., Strzembicka, A., Sodkiewicz, T., and Majewska, M. (2009). Response to stripe rust (*Puccinia striiformis, Westend*. f. Sp. *tritici*) and its coincidence with leaf rust resistance in hexaploid introgressive triticale lines with Triticum monococcum genes. *J. Appl. Genet.* 50, 205–211. doi: 10.1007/BF03195674

Song, W. Y., Wang, G. L., Chen, L. L., Kim, H. S., Pi, L. Y., Hoisten, T., et al. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science* 270, 1804–1806. doi: 10.1126/science.270.5243.1804

Spoel, S. H., and Dong, X. N. (2008). Making sense of hormone crosstalk during plant immune responses. *Cell Host Microbe* 3, 348–351. doi: 10.1016/j.chom.2008.05.009

Sun, X. H., Lu, T. G., Jia, S. R., and Huang, D. F. (2004). Genome-wide bioinformatics analysis of the rice receptor-like kinase family. *Sci. Agric. Sin.* 37, 322–327. doi: 10.3321/j.issn:0578-1752.2004.03.002

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621

Tsuda, K., and Somssich, I. E. (2015). Transcriptional networks in plant immunity. *New Phytol.* 206, 932–947. doi: 10.1111/nph.13286

Wang, F. T., Lin, R. M., Feng, J., Chen, W. Q., Qiu, D. W., and Xu, S. C. (2015). TaNAC1 acts as a negative regulator of stripe rust resistance in wheat, enhances susceptibility to *Pseudomonas syringae*, and promotes lateral root development in transgenic *Arabidopsis thaliana*. *Front. Plant Sci.* 6:108. doi: 10.3389/fpls.2015.00108

Xue, J., Bao, Y. Y., Li, B. L., Cheng, Y. B., Peng, Z. Y., Liu, H., et al. (2010). Transcriptome analysis of the brown planthopper *Nilaparvata lugens*. *PLoS One* 5:e14233. doi: 10.1371/journal.pone.0014233

Yang, C., Li, L. Y., Feng, A. Q., Zhu, X. Y., and Li, J. X. (2014). Transcriptional profiling of the responses to infection by the false smut fungus *Ustilaginoidea virens* in resistant and susceptible rice varieties. *Can. J. Plant Pathol.* 36, 377–388. doi: 10.1080/07060661.2014.927925

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11:R14. doi: 10.1186/gb-2010-11-2-r14

Yuan, G. S., He, X. J., Li, H., Xiang, K., Liu, L., Zou, C. Y., et al. (2019). Transcriptomic responses in resistant and susceptible maize infected with *Fusarium graminearum*. *Crop J.* 8, 153–163. doi: 10.1016/j.cj.2019.05.008

Zhang, H., Yang, Y. Z., Wang, C. Y., Liu, M., Li, H., Fu, Y., et al. (2014). Large-scale transcriptome comparison reveals distinct gene activations in wheat responding to stripe rust and powdery mildew. *BMC Genomics* 15:898. doi: 10.1186/1471-2164-15-898

Zhao, H. (2004). Chitinase and plant protection. *J. Hebei Agric. Sci.* 8, 78–84. doi: 10.3969/j.issn.1088-1631.2004.02.019

Zhu, X. B., Ze, M., Chern, M., Chen, X. W., and Wang, J. (2020). Deciphering rice lesion mimic mutants to understand molecular network governing plant immunity and growth. *Rice Sci.* 27, 278–288. doi: 10.1016/j.rsci.2020.05.004

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 16 May 2022
doi: 10.3389/fpls.2022.877908

Identification of Flowering Regulatory Networks and Hub Genes Expressed in the Leaves of *Elymus sibiricus* L. Using Comparative Transcriptome Analysis

Yuying Zheng[1], *Na Wang*[1], *Zongyu Zhang*[1], *Wenhui Liu*[2] *and Wengang Xie*[1]*

[1] *The State Key Laboratory of Grassland Agro-ecosystems, Key Laboratory of Grassland Livestock Industry Innovation, Ministry of Agriculture and Rural Affairs, College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, China,* [2] *Key Laboratory of Superior Forage Germplasm in the Qinghai-Tibetan Plateau, Qinghai Academy of Animal Science and Veterinary Medicine, Xining, China*

OPEN ACCESS

Edited by:
Chris S. Jones, International Livestock Research Institute, Ethiopia

Reviewed by:
Linkai Huang, Sichuan Agricultural University, China
Zan Wang, Institute of Animal Sciences (CAAS), China

***Correspondence:**
Wengang Xie
xiewg@lzu.edu.cn

Specialty section:
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

Received: *17 February 2022*
Accepted: *19 April 2022*
Published: *16 May 2022*

Citation:
Zheng Y, Wang N, Zhang Z, Liu W and Xie W (2022) Identification of Flowering Regulatory Networks and Hub Genes Expressed in the Leaves of Elymus sibiricus L. Using Comparative Transcriptome Analysis. Front. Plant Sci. 13:877908. doi: 10.3389/fpls.2022.877908

Flowering is a significant stage from vegetative growth to reproductive growth in higher plants, which impacts the biomass and seed yield. To reveal the flowering time variations and identify the flowering regulatory networks and hub genes in *Elymus sibiricus*, we measured the booting, heading, and flowering times of 66 *E. sibiricus* accessions. The booting, heading, and flowering times varied from 136 to 188, 142 to 194, and 148 to 201 days, respectively. The difference in flowering time between the earliest- and the last-flowering accessions was 53 days. Furthermore, transcriptome analyses were performed at the three developmental stages of six accessions with contrasting flowering times. A total of 3,526 differentially expressed genes (DEGs) were predicted and 72 candidate genes were identified, including transcription factors, known flowering genes, and plant hormone-related genes. Among them, four candidate genes (*LATE*, *GA2OX6*, *FAR3*, and *MFT1)* were significantly upregulated in late-flowering accessions. *LIMYB*, *PEX19*, *GWD3*, *BOR7*, *PMEI28*, *LRR*, and *AIRP2* were identified as hub genes in the turquoise and blue modules which were related to the development time of flowering by weighted gene co-expression network analysis (WGCNA). A single-nucleotide polymorphism (SNP) of *LIMYB* found by multiple sequence alignment may cause late flowering. The expression pattern of flowering candidate genes was verified in eight flowering promoters (*CRY*, *COL*, *FPF1*, *Hd3*, *GID1*, *FLK*, *VIN3*, and *FPA*) and four flowering suppressors (*CCA1*, *ELF3*, *Ghd7*, and *COL4*) under drought and salt stress by qRT-PCR. The results suggested that drought and salt stress activated the flowering regulation pathways to some extent. The findings of the present study lay a foundation for the functional verification of flowering genes and breeding of new varieties of early- and late-flowering *E. sibiricus*.

Keywords: flowering, *E. sibiricus*, transcriptome analysis, candidate genes, expression pattern

INTRODUCTION

Over the past few decades, botanists have been studying the regulation of the timing of flowering (Kobayashi and Weigel, 2007). To reproduce successfully, plants must bloom at the right time (Zhao et al., 2020b). In the model plant *Arabidopsis thaliana*, several flowering regulatory pathways containing several genes have been reported for the molecular mechanism of flowering. The vernalization pathway regulates flowering through the occurrence of vernalization, mainly by repressing the expression of *FLOWERING LOCUS C* (*FLC*) (Srikanth and Schmid, 2011). The photoperiodic pathway regulates flowering through plants' sense of day and night length, and light signals are transformed into floral signals through *CONSTANS* (*CO*) (Andrés and Coupland, 2012). The gibberellin pathway is the pathway that induces flowering only through gibberellin signaling. It has been investigated in different plant species that both exogenous gibberellin and endogenous gibberellin can promote flowering (Osnato et al., 2012). The autonomic pathway is an endogenous regulator independent of other pathways and promotes flowering by suppressing the *FLC* gene expression, mainly involving *LUMINIDEPENDENS* (*LD*), *FLOWERING LOCUS D* (*FLD*), and *FLOWERING LOCUS KH DOMAIN* (*FLK*) (Michaels and Amasino, 2001). Some epigenetic factors are independent of photoperiod, gibberellin, and vernalization pathways called age pathways. The age pathway is mainly controlled by micro-RNA156 (miR156), and studies in *A. thaliana* have shown that nine miR156-targeted *SQUAMOSA PROMOTER BINDING protein-like* (*SPL*) genes are involved in flowering regulation (Zheng et al., 2019). In addition, stress-regulated flowering is an informal flowering regulatory pathway. Plants alter the flowering time to regulate their growth and developmental processes in response to changes in the external environmental conditions, such as drought and salt stress (Kazan and Lyons, 2016). In *A. thaliana*, integrator *FLOWERING LOCUS T* (*FT*), positive regulators *LEAFY* (*LFY*), and *SUPPRESSOR OF OVEREXPRESSION OF CO1* (*SOC1*) were induced to accelerate flowering and seed maturation under drought stress (Verslues and Juenger, 2011; Su et al., 2013). Photoperiod flowering time gene *GIGANTEA* (*GI*) is also a key regulator of drought escape response (Riboni et al., 2014). In wheat (*Triticum aestivum*) and barley (*Hordeum vulgare*), the effects of drought stress on flowering time are species-dependent. The genotypes with a winter major flowering time gene displayed a delay in flowering under drought; however,; the genotypes with a spring major flowering time gene performed a significant acceleration in flowering under drought (Gol et al., 2017). High-salt stress substantially represses the *FT* gene by a membrane-associated NAC transcription factor NTL8, resulting in the delay of flowering in *A. thaliana* (Kim et al., 2007). In addition, photoperiod flowering time gene *GI* has been reported to be degraded by the proteasome causing the delay of flowering in *A. thaliana* under salt stress (Kazan and Lyons, 2016).

Although, the flowering regulatory pathways are relatively conservative, different plants have their specificity with respect to flowering times. For instance, two specific pathways mediated by *Early heading date 1* (*Ehd1*) and *Days to heading on Chromosome 2* (*DTH2*), respectively, are found in rice (*Oryza sativa*) but not present in *Arabidopsis* (Naranjo et al., 2014). *Ehd1* under influencing by multigene promotes the expression of the florigen genes *Heading date 3a* (*Hd3a*) and *RICE FLOWERING LOCUS T1* (*RFT1)* in both short-day and long-day conditions (Kong et al., 2016). Similarly, *DTH2* expression induces the expression of *Hd3a* and *RFT1* to promote heading under long-day conditions (Zhao et al., 2015). Further studies are needed to explore whether these pathways only exist in *O. sativa* or co-exist in other Gramineae members. Therefore, there is a need of hour to reveal the flowering mechanism of different plants for breeding and production application.

Elymus sibiricus is a perennial, self-pollinating, and heterotetraploid ($2n = 4x = 28$) forage grass belonging to the genus *Elymus* of Triticeae in Poaceae (Xie et al., 2015b). It is widely distributed in high altitude areas in western and northern Eurasia; thus, it has good cold and drought resistance, making it suitable for ecological restoration of artificial grassland establishment in Qinghai-Tibet Plateau, China (Xie et al., 2015a). In *E. sibiricus*, the time of flowering determines whether it is suitable for hay or seed production. Late-flowering materials were selected for optimizing biomass yield, and early-flowering materials were selected for increasing seed yield production (Wolabu et al., 2016). The flowering process of herbage is affected by both external factors (light, temperature, and exogenous hormones) and internal factors (endogenous hormones, physiological conditions, and genetic characteristics), among which genetic characteristics are one of the most crucial factors (Shim and Jang, 2020). Exploring the complex process of flowering requires a high degree of coordination among various pathways. Until now, the genome of *E. sibiricus* has been sequenced but the data are not yet available (Xiong et al., 2021).

Hence, the transcriptomic study is a key approach on gene function and structure research of eukaryotic plants without reference genomes. Transcriptome sequencing is a technical method to study the transcription process of biological tissues or specific cells of a species at the RNA level under specific conditions (Van Dijk et al., 2018). It can obtain the gene expression information of species as a whole and create conditions for the connection between genome and proteome. As an effective tool, transcriptome sequencing is gradually applied to study herbage flowering. A previous report identified 77 flowering-related genes by transcriptome sequencing of inflorescences at different stages in orchardgrass (*Dactylis glomerata*) (Feng et al., 2017). In crested wheatgrass (*Agropyron cristatum*), a total of 113 flowering time-related genes, 123 MADS-box genes, and 22 *COL* genes were identified by RNA-seq of three successive growth stages (Zeng et al., 2017). Recently, weighted gene co-expression network analysis (WGCNA) has often been applied to transcriptome sequencing datasets. It is an analysis method based on multiple samples to explore the association between modules and focus phenotypes and mines the key genes in the co-expression network (Childs et al., 2011). This analysis method provides system-level insight into high sensitivity to genes and greatly reduces information loss (Pei et al., 2017).

Therefore, high-throughput transcriptome sequencing and WGCNA were used in this study to jointly discover the specific flowering hub genes and deduce flowering regulatory networks in the leaf of early- and late-flowering accessions in *E. sibiricus*. The expression patterns of flowering candidate genes were analyzed under abiotic stresses. This study could provide new insights to explore the actual flowering regulation mechanisms and help to cultivate high-yield and high-quality varieties of *E. sibiricus*.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

A total of 66 *E. sibiricus* accessions originated from four countries, including China, Russia, Mongolia, and Kyrgyzstan (**Supplementary Table 1**), were used in this study. A total of ten healthy, plump, and uniform seeds were selected from each accession and placed evenly in 9-cm Petri dishes covered with a filter paper. They were incubated and germinated in an incubator at a constant temperature 25°C with 16:8-h light/dark regime. When the seeds grew to 8 weeks old, three healthy seedlings were selected and transplanted to the Yuzhong test field of Lanzhou University, Gansu, China (latitude 35°34'N, longitude 103°34'E, elevation 1,720 m) with plant spacing of 0.5 m and row spacing of 0.4 m. We recorded the time of leaf sheath wrapping the spikelet inflated after flag leaf grown out as booting time, the time of three spikelets tip breaking through leaf sheath per plant as the heading time, and the time of the inferior palea and the palea opened with anther and filaments emerging as the flowering time. The number of days was calculated from 1st January according to the method provided by Xie et al. (2012). A number of five other inflorescences-related traits, such as reproductive branches per plant (RBP), spikelet number of rachis (SNR), floret number per spikelet (FNS), 100-seed weight (SW1), and breaking tensile strength (BTS), were also measured.

RNA Extraction, Library Construction, and RNA-Seq

According to the records of booting, heading, and flowering time, three early-flowering accessions (PI598781, PI655199, and LQ10) and three late-flowering accessions (PI531665, and PI531669, PI595169) were chosen for the RNA-seq. The fresh leaves from each accession were collected in triplicates at the three developmental stages as follows: booting stage, heading stage, and flowering stage. The collected samples were stored at −80°C. The three sampling stages included booting from 15 May, heading from 21 May, and flowering from 27 May for early-flowering accessions, and booting from 5 June, heading from 10 June, and flowering from 21 June for late-flowering accessions.

Total RNAs were extracted from *E. sibiricus* samples using a UNIQ-10 Column TRIzol Total RNA Extraction Kit (Sangon Biotech, Shanghai). The purity, concentration, and integrity of RNA samples were detected by Nanodrop, Qubit 2.0, and Agilent 2100 methods, respectively. The qualified samples were used for library construction. A total of eighteen sequencing libraries were generated using Hieff NGS® MaxUp II Dual-mode mRNA Library Prep Kit for Illumina® MaxUp II (Yeasen Biotechnology, Shanghai). Library construction, RNA-seq, assembly, and quality analysis were conducted by Breeding Biotechnologies Company (Shaanxi, China) on Illumina HiSeq 2500 sequencing platform.

Sequence Data Processing, *De novo* Assembly, and Annotation

Raw data were obtained after transcriptome sequencing of the library constructed by *E. sibiricus*. After filtering raw data, the joint sequences and primer sequences in the raw data were removed, empty sequences and low-quality sequences were screened out, and clean data were obtained for assembly. Unigenes were obtained by *de novo* transcriptome assembly by Trinity-v2.4.0.

The obtained functional gene sequences were annotated by Nr (NCBI non-redundant protein sequence), Gene Ontology (GO), Clusters of Orthologous Groups (COG), Karyotic Orthologous Groups (KOG), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Swiss-Prot (annotated protein sequence) (Altschul et al., 1997). HMMER-v3.2.1 software was used for comparison with the Pfam (Protein Family) database (Eddy, 1998).

Construction of Weighted Gene Co-Expression Networks and Identification of Modules

All DEGs were used for the construction of weighted gene co-expression networks. Filtering differentially expressed genes to remove low-expression genes, DEGs with FPKM > 0 in at least nine samples were retained. A total of 2,354 DEGs were obtained for WGCNA. The trait datasets mainly included the development time (DT), reproductive branches per plant (RBP), spikelet number of rachis (SNR), floret number per spikelet (FNS), 100-seed weight (SW1), and breaking tensile strength (BTS). The WGCNA assumed that the genetic network followed scale-free networks. The similarity matrix of gene co-expression and gene network formed the adjacency function, the similarity matrix was transformed into the adjacency matrix, and then, the expression correlation coefficient was calculated to construct the gene's hierarchical clustering tree. Modules were divided according to the clustering relationship between genes, and modules with similar expression patterns were merged. Finally, the correlation between phenotypic traits of modules and samples was analyzed, and the hub genes in the network were found by the R language package (Langfelder and Horvath, 2008).

Validation Analysis of Transcriptome Data by qRT-PCR

The reliability of transcriptome data was verified by quantitative real-time PCR (qRT-PCR) experiments. The complementary DNA (cDNA) reverse transcription of total RNA was performed using PrimeScript™ RT reagent Kit with gDNA Eraser (TaKaRa, Dalian, China). A total of twelve flowering-related candidate genes were selected randomly, and *Protein Phosphatase 2A* (*PP2A*) was used as a reference gene to evaluate the expression stability of the candidate genes (Zhang et al., 2019). The primers of a reference gene and candidate genes were designed by PrimerQuest Tool (http://sg.idtdna.com/Primerquest/Home/Index) (**Supplementary Table 2**).

The qRT-PCR was performed using an SYBR Premix Ex Taq^{TM} II kit (TaKaRa, Dalian, China) on a Roche LightCycler480 quantitative PCR instrument. The final reaction volume was 20 μl, and each reaction mixture contained abm ®EvaGreen qPCR MasterMix-no dye 10 μl, forward and reverse primers 0.8 μl, cDNA 1 μl, and ddH_2O 7.4 μl. The following reaction procedures were used the enzyme activation at 95°C for 10 min, denaturation at 94°C for 15 s, annealing at 60°C for 1 min, denaturation and annealing for 40 cycles, and then the melting curve analyzed. The candidate gene expression level was calculated by the $2^{-\Delta\Delta Ct}$ method.

Gene Expression Analysis of Flowering Candidate Genes

In this study, the expression of flowering candidate genes in early- and late-flowering under different concentrations of salt stress and drought stress was investigated by a greenhouse pot experiment. The expression pattern of flowering candidate genes was studied by qRT-PCR technique. We combined qRT-PCR to study the expression patterns of candidate genes and elucidate the flowering molecular mechanism of *E. sibiricus*. Two accessions PI598781 and PI531665 were subjected to salt stress and drought stress at the heading stage. Under salt stress, 150, 200, and 250 mmol/L NaCl solutions (prepared with distilled water as mother liquor) were used to irrigate *E. sibiricus*. In the meanwhile, 10, 15, and 20% PEG6000 (polyethylene glycol) were used to irrigate *E. sibiricus* to simulate drought stress. After treatment for 6 h, the leaf samples were collected, stored at −80°C, and used to analyze the expression patterns of flowering candidate genes under different stress treatments. RNA extraction, cDNA preparation, and qRT-PCR analysis were carried out as presented above. Primer information on reference genes and candidate genes is shown in **Supplementary Table 2**.

RESULTS

Flowering Time Variations of 66 *E. sibiricus* Accessions

The booting, heading, and flowering times of 66 *E. sibiricus* accessions were measured (**Figure 1**). The booting time varied from 136 to 188 days, the heading time varied from 142 to 194 days, and flowering time varied from 148 to 201 days. The cluster heatmap showed that 66 *E. sibiricus* accessions were divided into two groups, one group included 43 early- and mid-flowering accessions and another group included 19 late-flowering accessions (**Figure 2**). The booting, heading, and flowering times for the earliest-flowering accession was 136, 142, and 148 days, respectively. For the last-flowering accession, the booting, heading, and flowering times were 188, 194, and 201 days, respectively. The difference in flowering times between the earliest-flowering accession and the last-flowering accession was 53 days. The variation coefficients of booting time, heading time, and flowering time across all accessions were 7.43, 6.98, and 6.92%, respectively. Hence, we selected three early-flowering accessions (PI598781, PI655189, and LQ10) and three late-flowering accessions (PI531665, PI531669, and PI595166) for further transcriptome analysis at booting, heading, and flowering stages (**Table 1**).

Sequencing, Assembly, and Functional Annotation

A total of 18 samples (L1B, L1H, L1F, L2B, L2H, L2F, L3B, L3H, L3F, E1B, E1H, E1F, E2B, E2H, E2F, E3B, E3H, E3F) were employed to construct the cDNA library. After trimming, we got 119.63Gb databases (**Supplementary Table 3**). Pearson's correlation coefficient between the three biological replicates of different tissues varied from 0.69 to 0.98, indicating the high quality of the replicates (**Supplementary Figure 1**). *De novo* assembly of high-quality sequences through Trinity software produced more than 180,000 unigenes. The number of unigenes with the length of 200 to 300 bp was the highest (57,094). The number of unigenes decreased as the length of the sequence increased and increased again when the length of the sequence was >3,000 (**Supplementary Figure 2**). A total of 89.90% unigenes were successfully annotated in COG, GO, KEGG, KOG, Pfam, Swiss-Prot, TrEMBL, Nr, and Nt databases (**Supplementary Table 4**). Currently, homologous species matched results in Nr database showed that *Aegilops tauschii* (23,554, 29%) was the best match for *E. sibiricus*, followed by *H. vulgare* (15,885, 20%) (**Supplementary Figure 3**).

Identification of Differentially Expressed Genes and KEGG Enrichment Analysis

The differentially expressed genes (DEGs) were filtered with expression levels FDR < 0.01, $\log_2$-fold change ≥2. Based on this criterion, we identified 3,526 DEGs at booting, heading, and flowering stages, of which 936 (576 upregulated, 360 downregulated), 1,162 (584 upregulated, 578 downregulated), and 1,428 (555 upregulated, 873 downregulated) were predicted from "EB vs. LB," "EH vs. LH," "EF vs. LF," respectively (**Figure 3A**). Among them, 363 DEGs were found to be common among three DEG sets. On the other hand, 323, 396, and 784 DEGs were found to be specific in "EB vs. LB," "EH vs. LH," and "EF vs. LF," respectively (**Figure 3B**).

The biological functions and metabolic pathways in the three sets of DEGs were inquired based on the KEGG database. In EB vs. LB, 63 unigenes were annotated in 43 metabolic pathways, of which the "Glycolysis/Gluconeogenesis" was the most highly represented pathway. A total of 52 unigenes were annotated in 32 metabolic pathways in EH vs. LH. Among them, "flavonoid biosynthesis" was the most enriched pathway. In EF vs. LF, a total of 109 unigenes were enriched in 49 metabolic pathways, and the "starch and sucrose metabolism" was the highest enrich pathway, followed by the "galactose metabolism" pathway (**Supplementary Figure 4**).

Lots of DEGs involved in glycolysis or gluconeogenesis, RNA transport, flavonoid biosynthesis, peroxisome, purine metabolism, pyrimidine metabolism, starch and sucrose metabolism, galactose metabolism, and plant hormone signal pathways were recognized as the candidate genes for flowering regulation. At the booting stage, DEGs involved in glycolysis or gluconeogenesis, RNA transport, and flavonoid biosynthesis

FIGURE 1 | The image of *E. sibiricus* morphology. **(A)** Early-flowering accession. **(B)** Late-flowering accession. **(C)** The morphology of booting time, heading time, and flowering time in *E. sibiricus*. **(A)** and **(B)** were taken at 167 days according to the calculation method in this study.

were found to be highly expressed in the early-flowering accession, except for an alcohol dehydrogenase-like 3 gene (*TRINITY_DN33709_c5_g1*) and a glyceraldehyde-3-phosphate dehydrogenase gene *GAPCP1* (*TRINITY_DN41564_c1_g1*) (**Figure 4A**). At the heading stage, DEGs involved in RNA transport and peroxisome were upregulated in the early-flowering accession (**Figure 4B**). At the flowering stage, DEGs involved in starch and sucrose metabolism, galactose metabolic, and plant hormone signal pathways were highly expressed in late-flowering accessions relative to early-flowering accessions at the flowering stage, except for a glycosyl transferase family gene, *TRINITY_DN41331_c0_g1* (**Figure 4C**).

Identification of Flowering Candidate Genes and Expression Patterns

Based on the gene function annotation, 156 DEGs involved in the vernalization pathway (6), photoperiod pathway

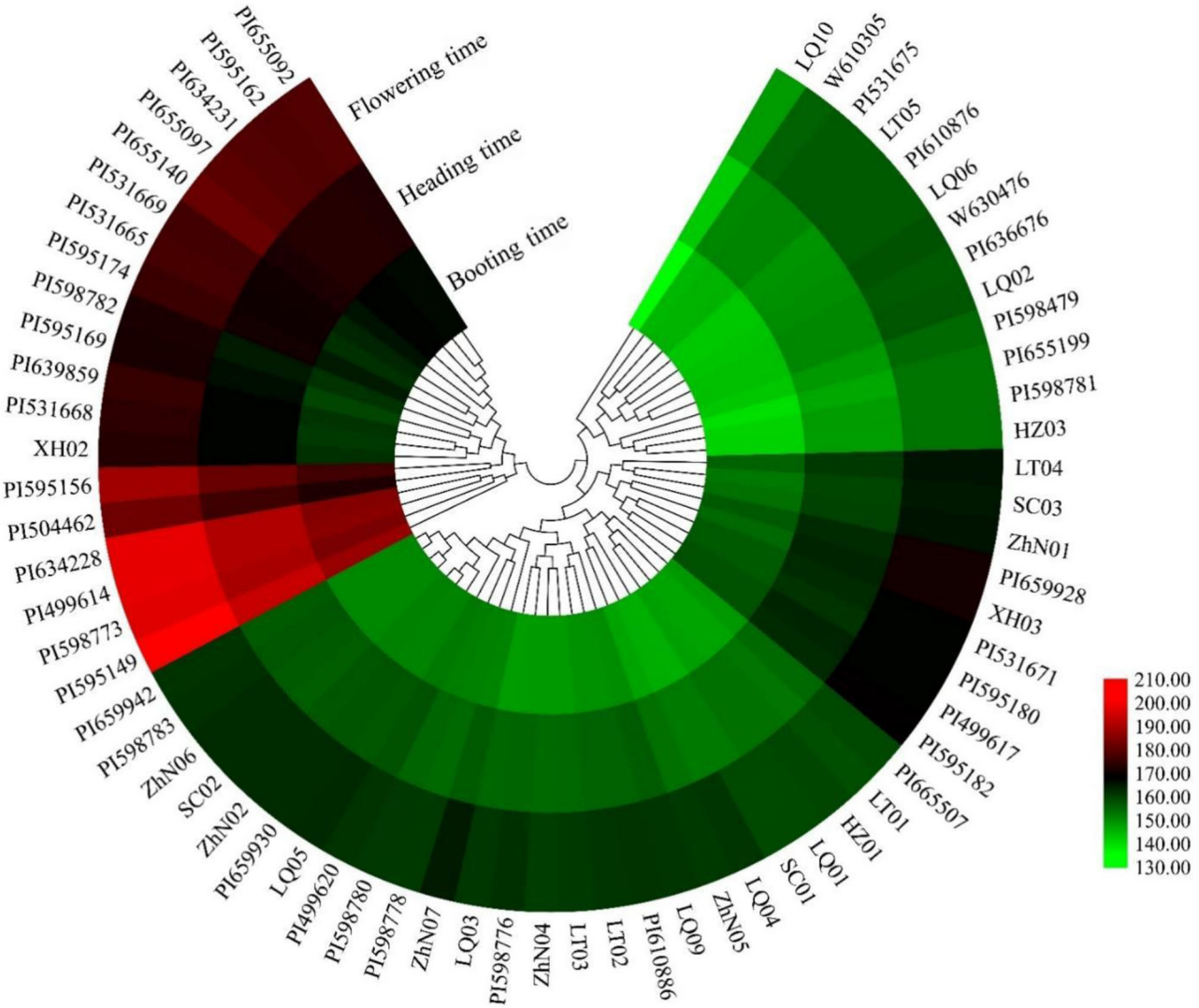


FIGURE 2 | The heatmap of the booting time, heading time, and flowering time of 66 *E. sibiricus* accessions.

TABLE 1 | The information of six *E. sibiricus* accessions for transcriptome sequencing.

Sample name	Accession	Origin	Booting time (d)	Heading time (d)	Flowering time (d)
E1	PI598781	Baikal, Buryat, Russia	141	148	153
E2	PI655199	Hongyuan, Sichuan, China	140	147	153
E3	LQ10	Luqu, Gansu, China	136	142	148
L1	PI531665	Beijing, China	162	170	177
L2	PI531669	Xining, Qinghai, China	163	170	177
L3	PI595169	Xinjiang, China	159	166	171

(82), circadian clock (12), gibberellin pathway (17), autonomic pathway (16), age pathway (5), and central integrator (18) were identified as flowering candidate genes (**Supplementary Table 5**). Among them, seventy-two candidate genes were divided into three parts, namely, transcription factors (27), known flowering genes (14), and

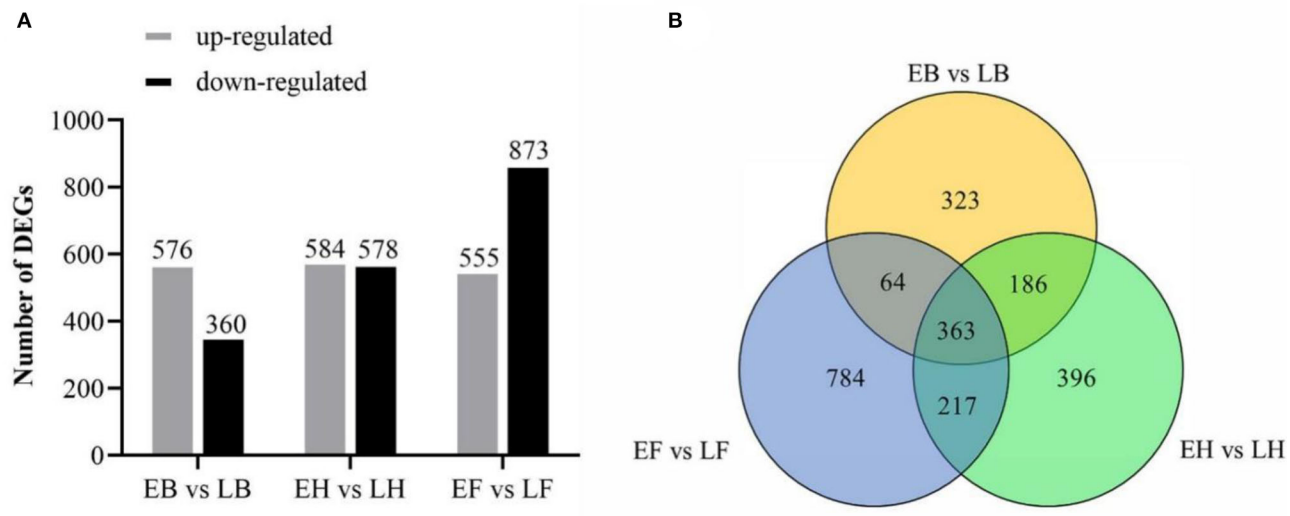


FIGURE 3 | The number of DEGs of two opposing flowering accessions. **(A)** The column chart of DEGs number in three stages. **(B)** Venn diagram represents the number of overlapping DEGs.

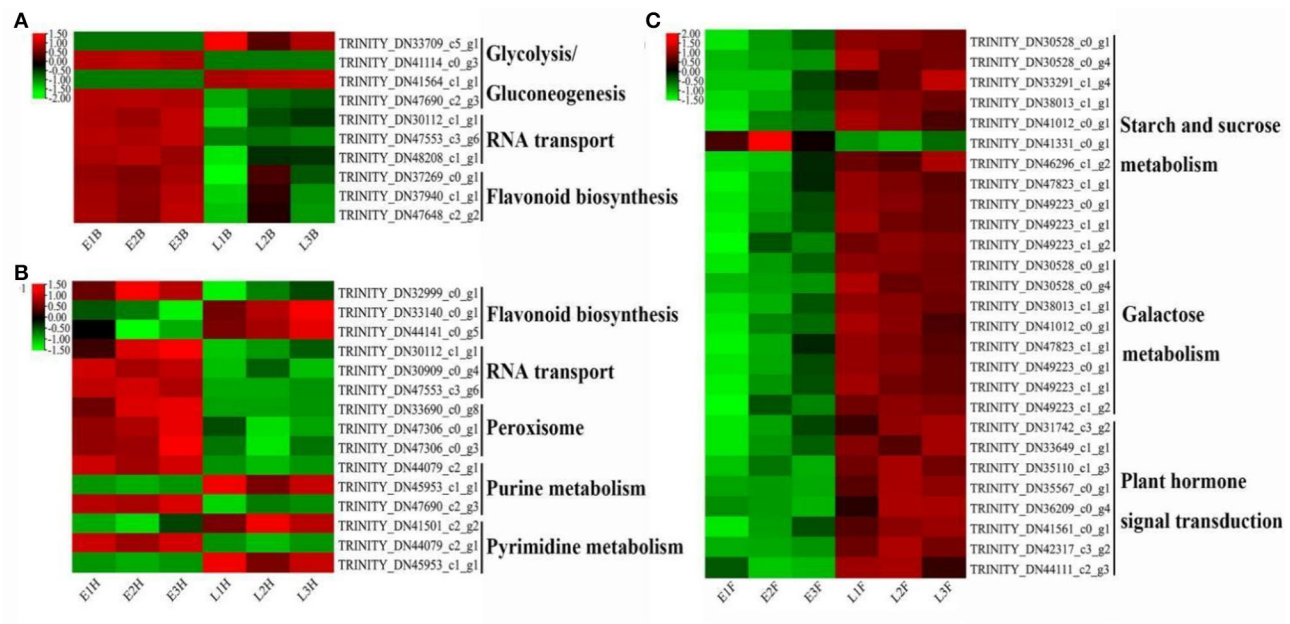


FIGURE 4 | The heatmap of DEGs derived from KEGG pathways with two opposing flowering accessions. **(A)** EB vs. LB. **(B)** EH vs. LH. **(C)** EF vs. LF.

plant hormone signal transduction (31). All transcription factors were identified from nine transcription factor families. The bZIP and MYB families have the most numbers of candidate genes, and the remaining 16 transcription factors were from the AP2/ERF (2), bHLH (3), Homebox (3), WRKY (2), HSF (2), MADS-box (1), and NAC family (3). Candidate genes identified in plant hormone signal transduction were involved in gibberellin, auxin, abscisic acid, jasmonic acid, and salicylic acid. A total of six genes involved in jasmonic acid, *TRINITY_DN39355_c3_g1* (**Figure 5A**), *TRINITY_DN47809_c0_g2* (**Figure 5A**), *TRINITY_DN50314_c1_g2* (**Figure 5B**), *TRINITY_DN34038_c2_g1* (**Figure 5C**), *TRINITY_DN30282_c0_g2* (**Figures 5B,C**), and *TRINITY_DN47564_c2_g1* (**Figures 5B,C**), were

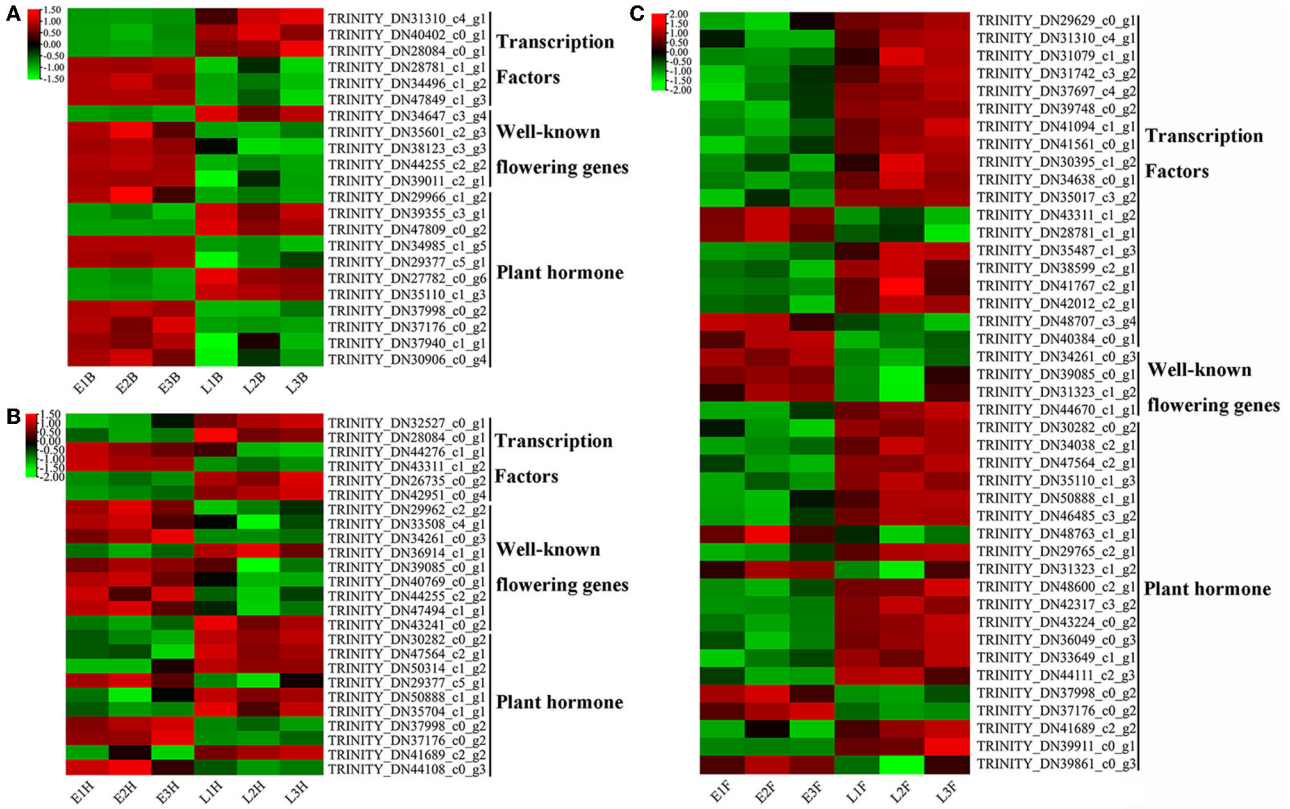


FIGURE 5 | The heatmap of candidate genes expression in *E. sibiricus*. **(A)** EB vs. LB. **(B)** EH vs. LH. **(C)** EF vs. LF.

upregulated in late-flowering accessions; a total of two genes involved in salicylic acid, *TRINITY_DN30906_c0_g4* (**Figure 5A**), and *TRINITY_DN44108_c0_g3* (**Figure 5B**) were downregulated in late-flowering accessions. The homologous genes of known flowering genes were mostly upregulated in early-flowering accessions except for four genes. At the booting stage, a transcriptional repressor *LATE*, *TRINITY_DN34647_c3_g4*, was significantly upregulated in late-flowering accessions (**Figure 5A**). At the heading stage, two genes were significantly upregulated in late-flowering accessions which are gibberellin degradation enzyme gene *GA2OX6* (*TRINITY_DN43241_c0_g2*) and far-red elongated hypocotyl gene *FAR3* (*TRINITY_DN36914_c1_g1*) (**Figure 5B**). At the flowering stage, a flowering regulation repressor *MFT1*, *TRINITY_DN44670_c1_g1*, was significantly upregulated in late-flowering accessions (**Figure 5C**). These four candidate genes affected the formation of late-flowering accessions in *E. sibiricus*.

Based on these candidate genes captured in multiple different flowering-related pathways and known regulatory networks in Arabidopsis and rice, we constructed a putative flowering regulation network in *E. sibiricus* (**Figure 6**).

Construction of Weighted Gene Co-Expression Networks and Identification of Modules

After screening the low expressed genes, 2,354 DEGs were retained for the WGCNA. A total of five distinct modules were identified *via* the block-wise module function, in which each module was labeled with different colors. Based on the approximate scale-free topology criterion, we have chosen the soft threshold power 24 when the correlation coefficient squared was 0.8 to define the adjacency matrix (**Figure 7A**). The corresponding relationship between the soft threshold and the mean value of gene linkage coefficient met the construction requirements (**Figure 7B**).

According to the correlation coefficient between DEGs, a gene clustering tree was constructed, with modules of gray (60), turquoise (1,046), blue (982), brown (200), and yellow (66) (**Figure 8A**). Combined with **Figure 8B**, there was a significant difference among the five modules, especially the blue module (0.89), which was strongly related to the development time of flowering (DT); the turquoise module (−0.85) was negatively related to the DT trait. In addition, it could be seen that the blue and turquoise modules were still closely related to DT

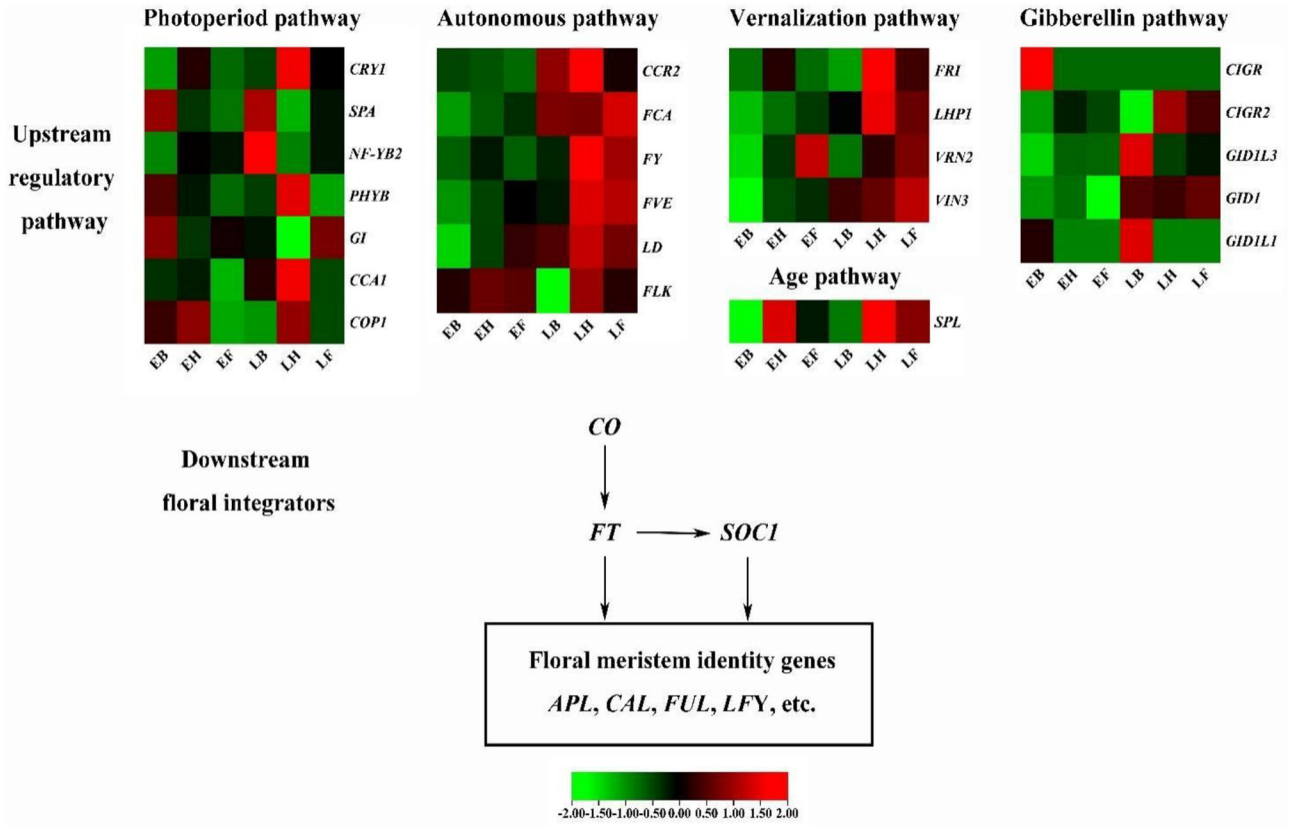


FIGURE 6 | The putative flowering regulation network in *E. sibiricus*. Lines with arrows indicate activation.

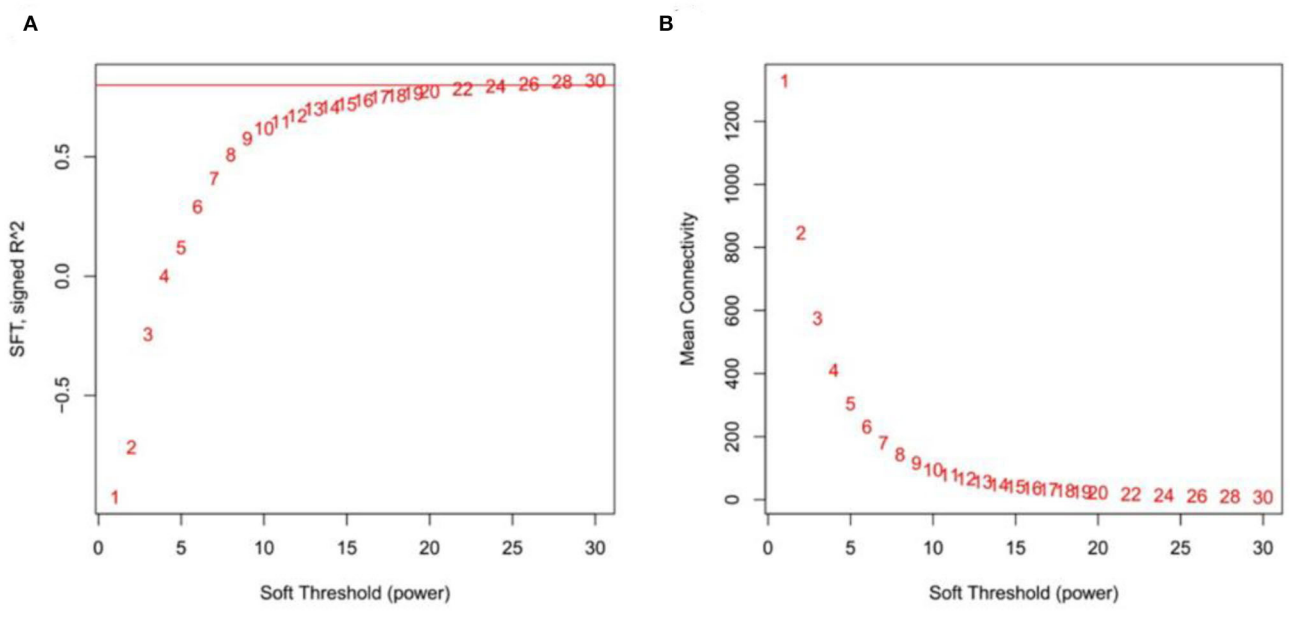


FIGURE 7 | Determination soft threshold. **(A)** Scale independence. **(B)** Mean connectivity.

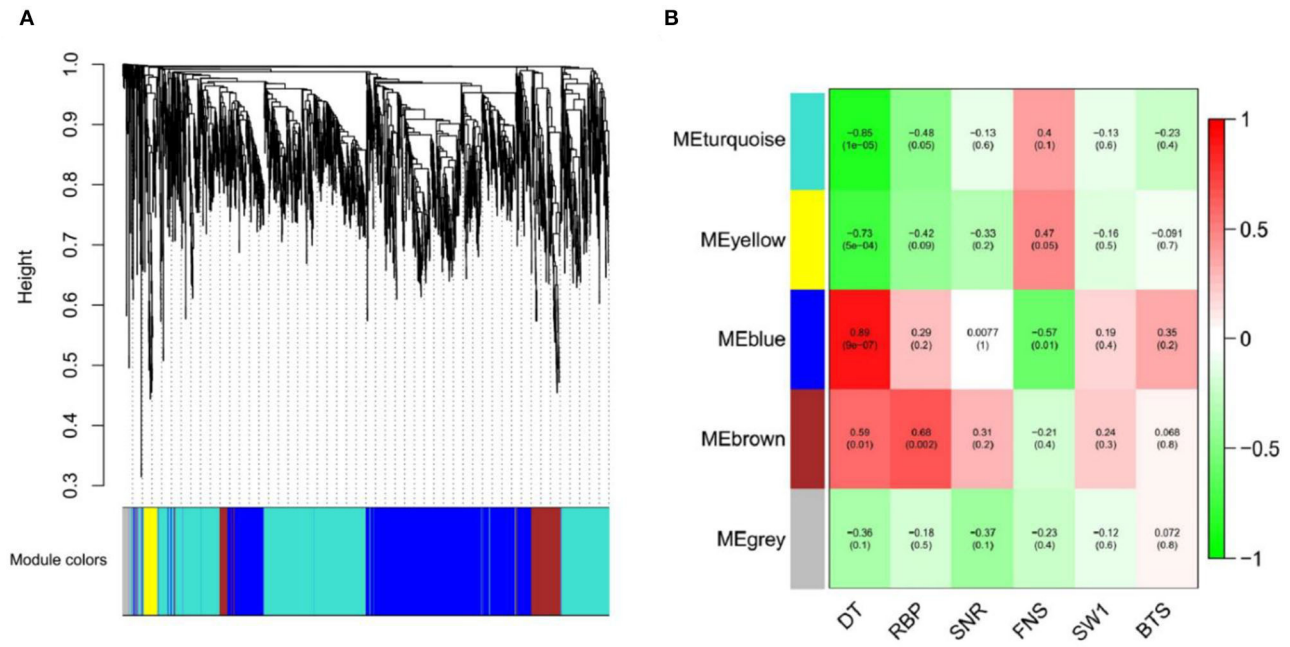


FIGURE 8 | Gene cluster tree and module diagram, correlation heatmap between modules and traits. **(A)** Cluster Dendrogram. **(B)** Module–trait relationships.

trait by the significance analysis between DT trait and DEGs (**Supplementary Figure 5**).

Therefore, we visualized the co-expression network between the blue module (**Figure 9A**) and the turquoise module (**Figure 9B**). A total of seven hub genes were identified in the two modules. The hub genes *L10-interacting MYB domain-containing protein* (*LIMYB*), *Peroxisome biogenesis protein 19* (*PEX19*), and *Glucan-water dikinase 3* (*GWD3*) were identified in the blue module; the hub genes *Boron transporter 7* (*BOR7*), *Pectin methylesterase inhibitor28* (*PMEI28*), *Leucine-rich repeat* (*LRR*), and E3 ubiquitin-protein ligase gene *AIRP2* were identified in the turquoise module (**Table 2**). The seven hub genes with high connectivity in the modules were more compatible with flowering development time, laying a foundation for the subsequent research on flowering genes of *E. sibiricus*.

Allelic Variation Analysis of Hub Gene *LIMYB*

To reveal the potential allelic variation of hub gene *LIMYB* for flowering regulation, 30 *E. sibiricus* accessions were selected, including 10 early-flowering accessions, 10 mid-flowering accessions, and 10 late-flowering accessions, for allelic variation analysis. The result of multiple sequence alignment showed that a single-nucleotide polymorphism (SNP) was found in the 320-bp ORF region of the hub gene *LIMYB*. At 320 bp, 10 early-flowering accessions were completely consistent with the original sequence, whereas 10 mid-flowering accessions and 10 late-flowering accessions mutated from G (guanine) to A (adenine) (**Figure 10A**). The base mutation at 320 bp could have resulted in the amino acid mutation of the mid- and late-flowering accessions from R(arginine) to K(lysine) (**Figure 10B**).

Validation of the Expression of Flowering-Related Genes

To confirm the authenticity of transcriptome sequencing data, twelve flowering candidate genes were selected for qRT-PCR verification, including *LF*, *CIGR*, *CCR2*, *CRY*, *COL*, *FPF1*, *FT*, *HD3*, *FLT*, *FLK*, *GID1*, and *COL4*. In E3 and L3 materials, *LF*, *CCR2*, *FPF1*, and *FLT* were all downregulated at heading and flowering time. The gene *COL* was upregulated in three stages, whereas *Hd3* was downregulated in three stages. The expression trends of the 12 flowering candidate genes in the early- and late-flowering accessions of *E. sibiricus* were similar to the transcriptome sequencing results (**Figure 11**). Linear regression analysis showed a positive correlation between the transcriptome sequencing data and the qRT-PCR results ($r = 0.704$, $p < 0.05$).

Expression Analysis of Flowering-Related Genes Under Salt Stress

A total of twelve flowering candidate genes, including *CCA1*, *Ghd7*, *FPF1*, *FPA*, *Hd3*, *VIN3*, *ELF3*, *CRY*, *COL*, *FLK*, *GID1*, and *COL4,* were selected for expression analysis of flowering-related genes under salt stress (**Figure 12**). In early-flowering accessions, the genes *Ghd7*, *FPF1*, *FPA*, *Hd3*, *VIN3*, *GID1*, and *COL4* were significantly upregulated under 150, 200, and 250 mmol/L NaCl treatments. The expression of genes *CCA1*, *ELF3*, *CRY*, and *COL* was not significant compared with the control under salt stress; however, the expression of genes *CRY* and *COL* was an

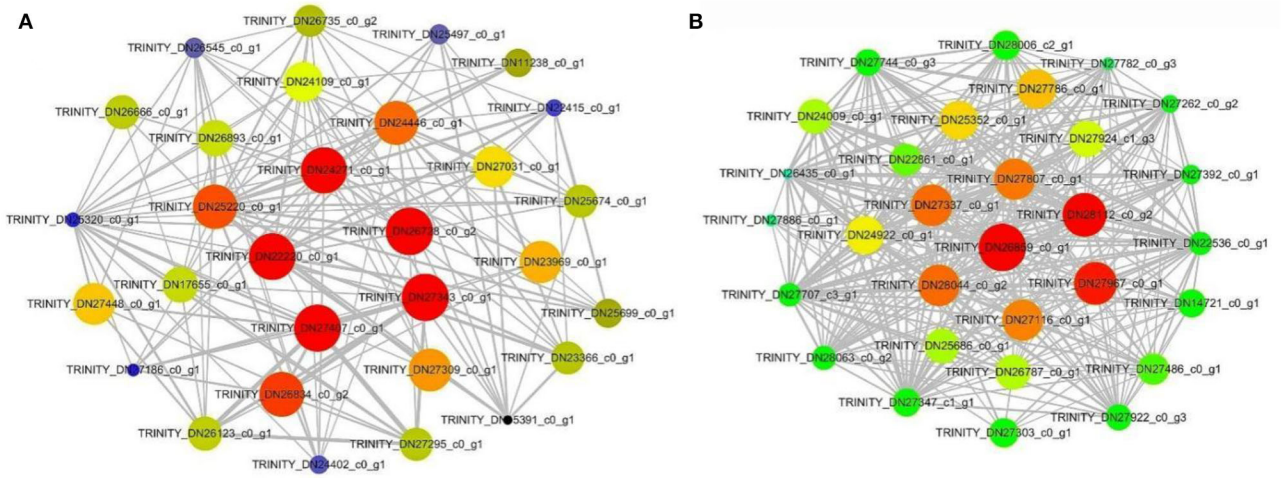


FIGURE 9 | Blue module and turquoise module co-expression network visualization. **(A)** Blue. **(B)** Turquoise.

TABLE 2 | Hub genes.

Number	Gene ID	Gene name	Module	Annotation
1	TRINITY_DN26735_c0_g2	*LIMYB*	blue	MYB transcription factor
2	TRINITY_DN25320_c0_g1	*PEX19*	blue	Peroxisome biogenesis protein
3	TRINITY_DN26834_c0_g2	*GWD3*	blue	Pyruvate phosphate dikinase
4	TRINITY_DN27392_c0_g1	*BOR7*	turquoise	Boron transporter
5	TRINITY_DN27886_c0_g1	*PMEI28*	turquoise	Pectinesterase inhibitor
6	TRINITY_DN25352_c0_g1	*LRR*	turquoise	Leucine-rich repeat receptor-like protein kinase
7	TRINITY_DN27337_c0_g1	*AIRP2*	turquoise	E3 ubiquitin-protein ligase

upward trend under the three levels. Moreover, the gene *FLK* was only upregulated in 150 and 200 mmol/L NaCl treatments. In the late-flowering accessions, the genes *COL*, *FLK*, and *GID1* were significantly upregulated under 150, 200, and 250 mmol/L NaCl treatments. The *COL4* gene was the only one that was not significantly expressed under the three treatments. Most of the candidate genes, including *CCA1*, *Ghd7*, *FPF1*, *FPA*, *Hd3*, *VIN3*, and *CRY,* were just significantly upregulated under 150 mmol/L NaCl treatment, and their expression was downregulated when the salt stress increased. On the contrary, the expression of *ELF3* was significantly upregulated under 200 and 250 mmol/L NaCl treatments.

Under salt stress, the expression of most of these candidate genes was upregulated. The high concentration of salt stress stimulated the upregulation of candidate genes was confirmed in early-flowering accessions, whereas the high concentration of salt stress may block or suppress certain flowering pathways in late-flowering accessions.

Expression Analysis of Flowering-Related Genes Under Drought Stress

Likewise, the expression analysis of twelve flowering candidate genes was detected under drought stress (**Figure 13**). In the early-flowering accessions, the expression of genes *Ghd7*, *FPF1*, *FPA*, *Hd3*, *ELF3*, *FLK*, *GID1*, *COL*, and *COL4* was significantly upregulated under 10% PEG6000 treatment. The expression quantities of these genes were low when the stress was strong, except that *COL* was also significantly upregulated under 15% PEG6000 treatment. The genes *CCA1*, *VIN3*, and *CRY* were not significantly expressed compared with the control. In the late-flowering accessions, the genes *FPF1*, *FPA*, *VIN3*, *ELF3*, *CRY*, and *GID1* were significantly upregulated under 10% PEG6000 treatment. Under 20% PEG6000 treatment, the flowering promoter *FLK* and the gibberellin receptor *GID1* were significantly upregulated, and the flowering suppressor *Ghd7* was downregulated. The result indicated that high-level stress-triggered changes in the expression of these genes led to drought escape. Interestingly, the gene *COL4* was downregulated under three levels of stress, suggesting that *COL4* may be only involved in early-flowering regulation. In addition, the genes *CCA1*, *Hd3*, and *COL* were not significantly expressed under drought stress.

A certain degree of drought stress could promote the response of the external environmental stimuli and endogenous signals to influence the flowering process. In the late-flowering accessions, the drought escape behavior of plants was more obvious.

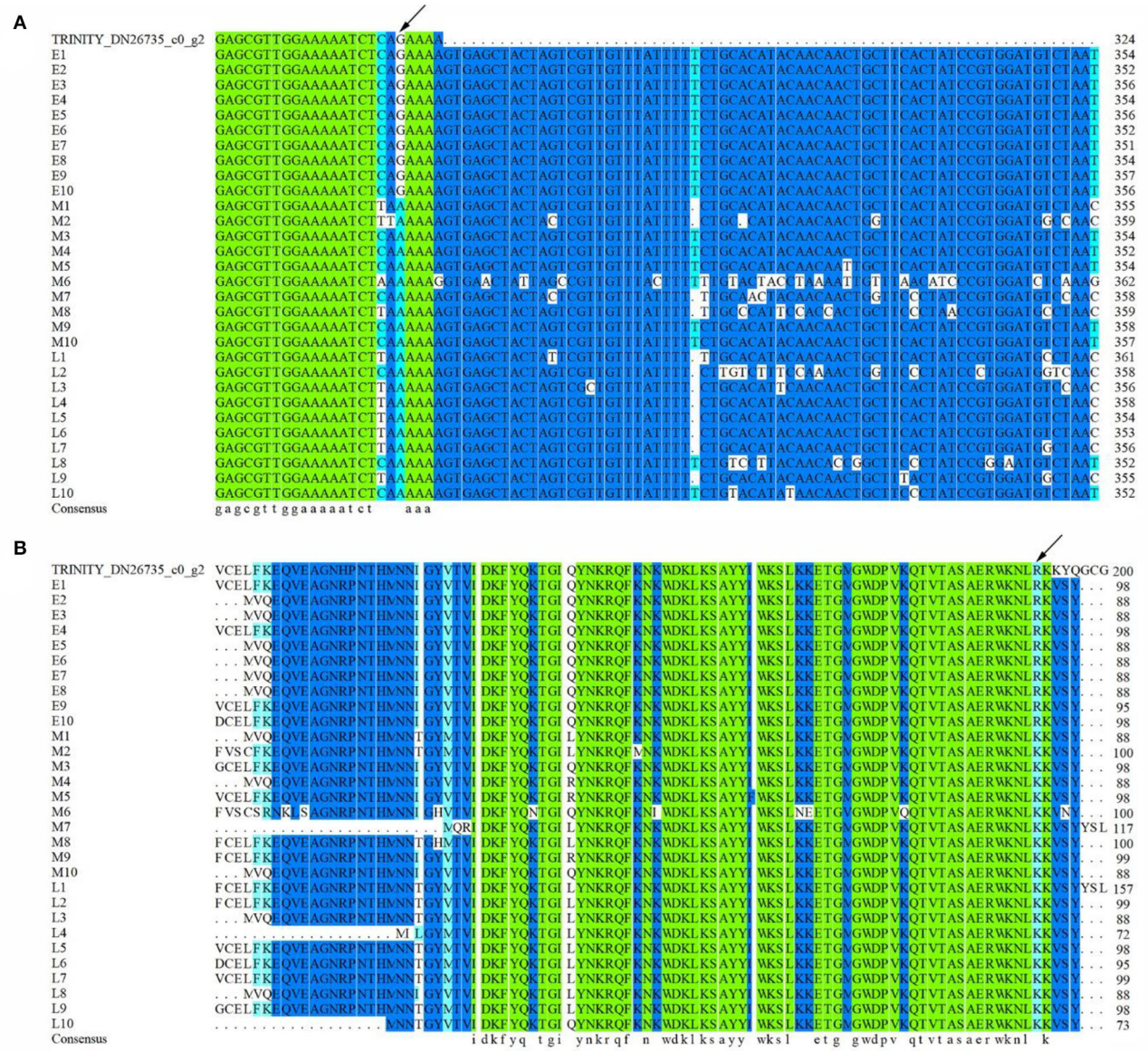


FIGURE 10 | Multiple sequence alignment of hub gene *LIMYB*. **(A)** DNA sequences. **(B)** Amino acid sequences.

DISCUSSION

Variation of Flowering Time in *E. sibiricus*

In this study, the booting stage, heading stage, and flowering stage of 66 *E. sibiricus* accessions from different regions showed a great variation. These germplasms with different flowering times are beneficial for the breeding of early- and late-maturing varieties of *E. sibiricus*. Flowering is often affected by altitude, latitude, temperature, rainfall, and other factors (Linder, 2020). In general, high-altitude and high-latitude populations bloom earlier as compared to lower altitudes and latitudes. However, we found the correlation between the flowering time of 66 *E. sibiricus* accessions and altitude or latitude was weak. The weak correlation may be the result of environmental diversity in China and the different correlations between flowering time and environmental factors in different environments (Cremer et al., 1998; Huang et al., 2012). In future experiments, we can collect more information on environmental factors at the collection site to clarify the flowering mechanism of *E. sibiricus*.

Multiple Gene Regulation Contributes to Flowering Variation in *E. sibiricus*

The flowering process of *E. sibiricus* is not controlled by a single gene, but by multiple flowering regulatory pathways. A large number of genes were differentially expressed in the flowering

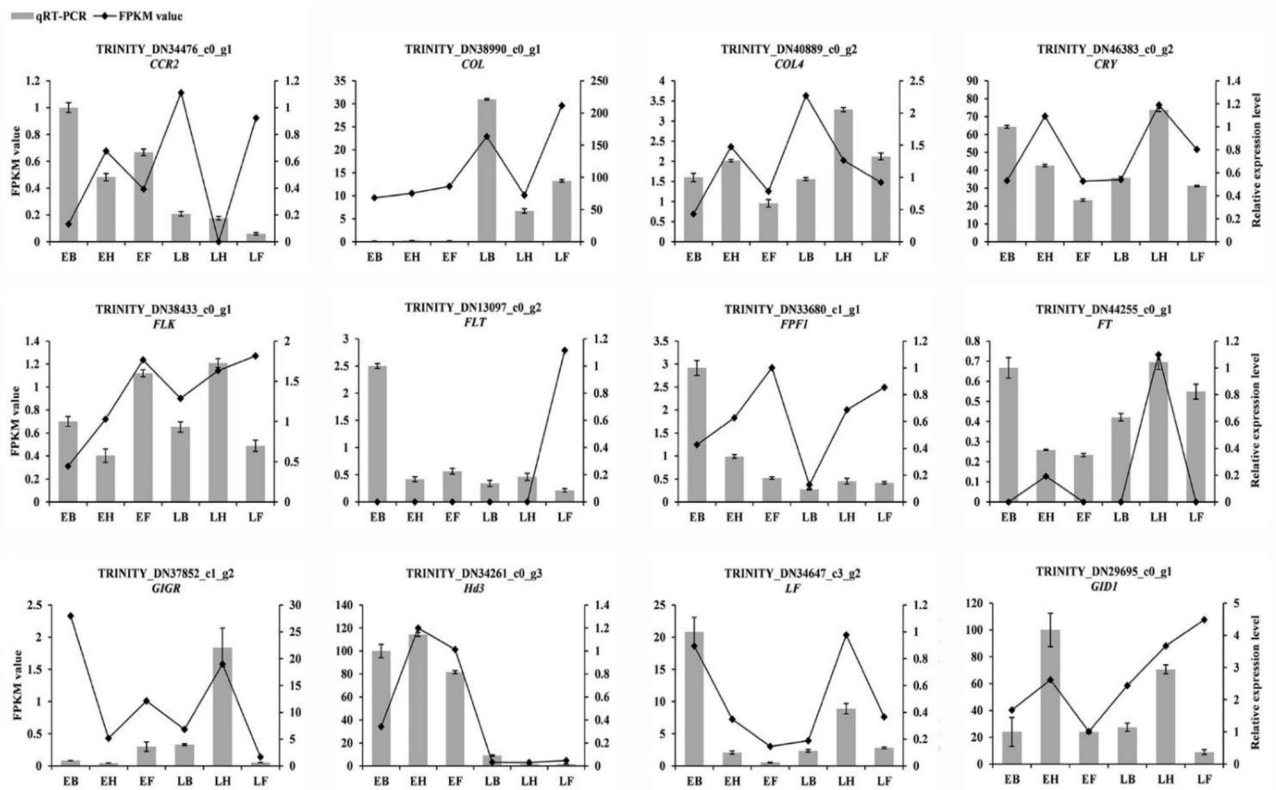


FIGURE 11 | Validation of the expression by qRT-PCR. Bar charts indicate values of qRT-PCR. Line plots indicate values of FPKM value.

stage and heading stage between two opposite accessions which could contribute to the difference in flowering in *E. sibiricus*. In EH vs. LH, "flavonoid biosynthesis" was the most enriched pathway. *Dihydroflavonol-4-reductase* (*DFR*) was upregulated in the pathway that was a key enzyme regulating anthocyanin and proanthocyanin synthesis in the flavonoid biosynthesis pathway (Chen et al., 2020). In EF vs. LF, "starch and sucrose metabolism" was the mainly enriched pathway. The two pathways were also reported in the integrative analysis of flower bud differentiation, flower bud elongation, and floral anthesis in Loquat (*Eriobotrya japonica*) (Jing et al., 2020). Based on these KEGG pathways, not only known flowering genes but also transcription factors, and plant hormone signal transduction, were selected as the flowering candidate genes of *E. sibiricus*. Most genes from MYB, bHLH, AP2, WRKY, and MADS families have positive regulation on flower transformation (Matías-Hernández et al., 2016). *WRKY75*, a WRKY DNA-binding protein, positively regulates flowering by acting on *FT*, and thus, overexpression of *WRKY75* can promote flowering in Arabidopsis (Zhang et al., 2018b). *REM16* is one of the members of AP2/B3-like transcription factor family that acts upstream of *SOC1* and *FT* in the flowering pathway and delays the flowering when it silences (Yu et al., 2020). Previous studies showed that transcription factors of NAC, bZIP, and Homebox families accelerated the flowering and enhanced plant tolerance to drought stress (Jakoby et al., 2002; Minh-Thu et al., 2018; Zhang et al., 2018a). Heat shock transcription factors (HSFs) are involved in flowering development and heat stress response by activating heat shock proteins (Liang et al., 2021). It is concluded that some transcription factors are not only involved in flowering development but also respond to abiotic stress, which is beneficial to the cultivation of early- and late-flowering-resistant varieties.

Plant hormone signaling coordinates flowering in higher plants. We found candidate genes related to auxin, gibberellins (GAs), abscisic acid (ABA), salicylic acid (SA), and jasmonic acid (JA) involved in *E. sibiricus* flowering process. Auxin, GAs, and ABA are the traditional phytohormones. The most common auxin is indole-3-acetic acid (IAA). IAA and GA regulate DELLA protein which inhibits flowering, such as *AUXIN RESPONSE FACTOR* (*ARF*) and *GA-insensitive* (*GAI*) (Yamaguchi et al., 2013). The orthologs of *ARF* were also found in our study. Abscisic acid receptors *PYL4*, *PYL5*, and *PYL9* were differentially expressed in the early- and late-flowering accessions of *E. sibiricus* in our study. A recent study showed that *ABSCISIC ACID-INSENSITIVE MUTANT* 5 (*ABI5*) modulated the *PYL*-mediated ABA responses to promote flowering by activating *FT* (Zhao et al., 2020a). SA and JA are the plant hormones that are attaining deep concerns in recent years because they are actively involved in responses to stress conditions in plants. *Salicylic acid-binding protein 2* (*SABP2*) found in this study was only reported to be related to stress resistance (Haq et al., 2020), whereas *JASMONATE-RESISTANT1* (*JAR1*) was involved in the photoperiod pathway (Chen et al., 2018). Previous

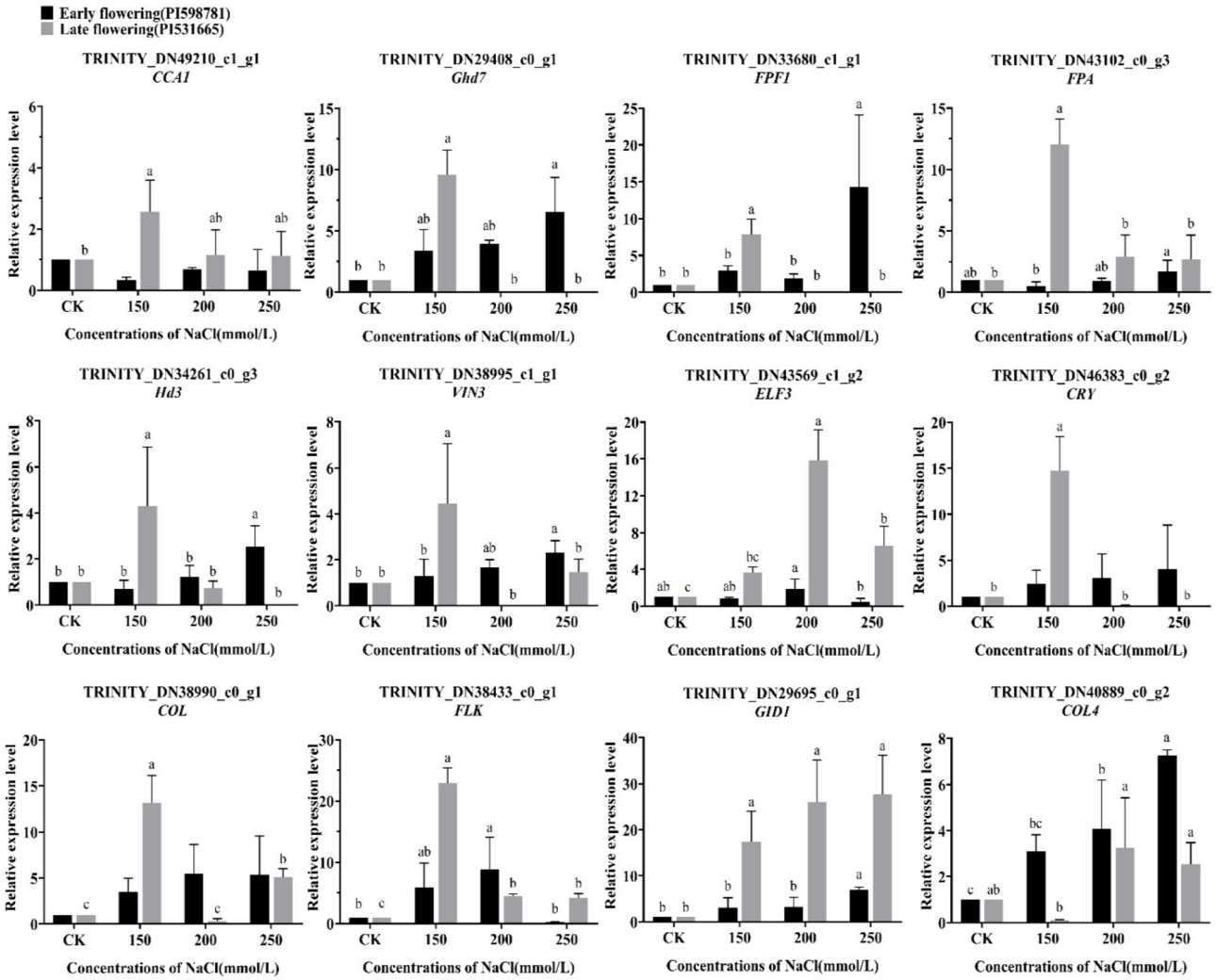


FIGURE 12 | The expression pattern of 12 flowering candidate genes in the early- and late-flowering accessions under salt stress by qRT-PCR analysis.

studies have revealed that SA inhibited flowering through *SUMO E3 ligase* (*SIZ1*) which facilitates protein conjugation with a small ubiquitin-like modifier of *FLD*, and the action of *CORONATINE INSENSITIVE 1* (*COI1*) can be inhibited by JA (Campos-Rivero et al., 2017). The hormone regulation of flowering time in *E. sibiricus* should be further explored. The putative flowering regulation network constructed in the study has a reference value for flowering regulation of *E. sibiricus*.

Flowering Hub Genes Selected by WGCNA

A total of seven hub genes were found in the blue module and the turquoise module that could be the key genes in future research on *E. sibiricus* flowering regulation. The genes *LIMYB*, *PEX19*, and *GWD3* may be the flowering promoters of *E. sibiricus*. *LIMYB* is an MYB domain protein that interacts with L10 and belongs to the MYB transcription factor family, which is widely involved in plant growth and development and metabolic regulation (Zorzatto et al., 2015). Many transcription factors from the MYB family are involved in flowering regulation, such as *EARLY FLOWERING MYB PROTEIN* (*EFM*), which directly inhibited *FT* expression in leaf vasculature and delayed flowering time in *A. thaliana*, so it can be speculated that *LIMYB* gene may be involved in the regulation of flowering time of *E. sibiricus* (Yan et al., 2014). The peroxisome is essential for normal development in plants. In Arabidopsis, *PEX19* plays an important role in the biosynthesis and transport of peroxisome membrane proteins, and peroxisome division and inheritance (McDonnell et al., 2016). It is closely related to cell senescence and may be involved in flowering regulation through the age pathway. Glucan water dikinase (GWD) enzymes are the catalytic enzymes in plant starch phosphorylation and play key roles in starch metabolism; among them, *GWD3* is one of the two clades of the GWD isoforms (Mdodana et al., 2019). It could affect grain filling, starch accumulation, and starch biosynthesis of *E. sibiricus* after flowering.

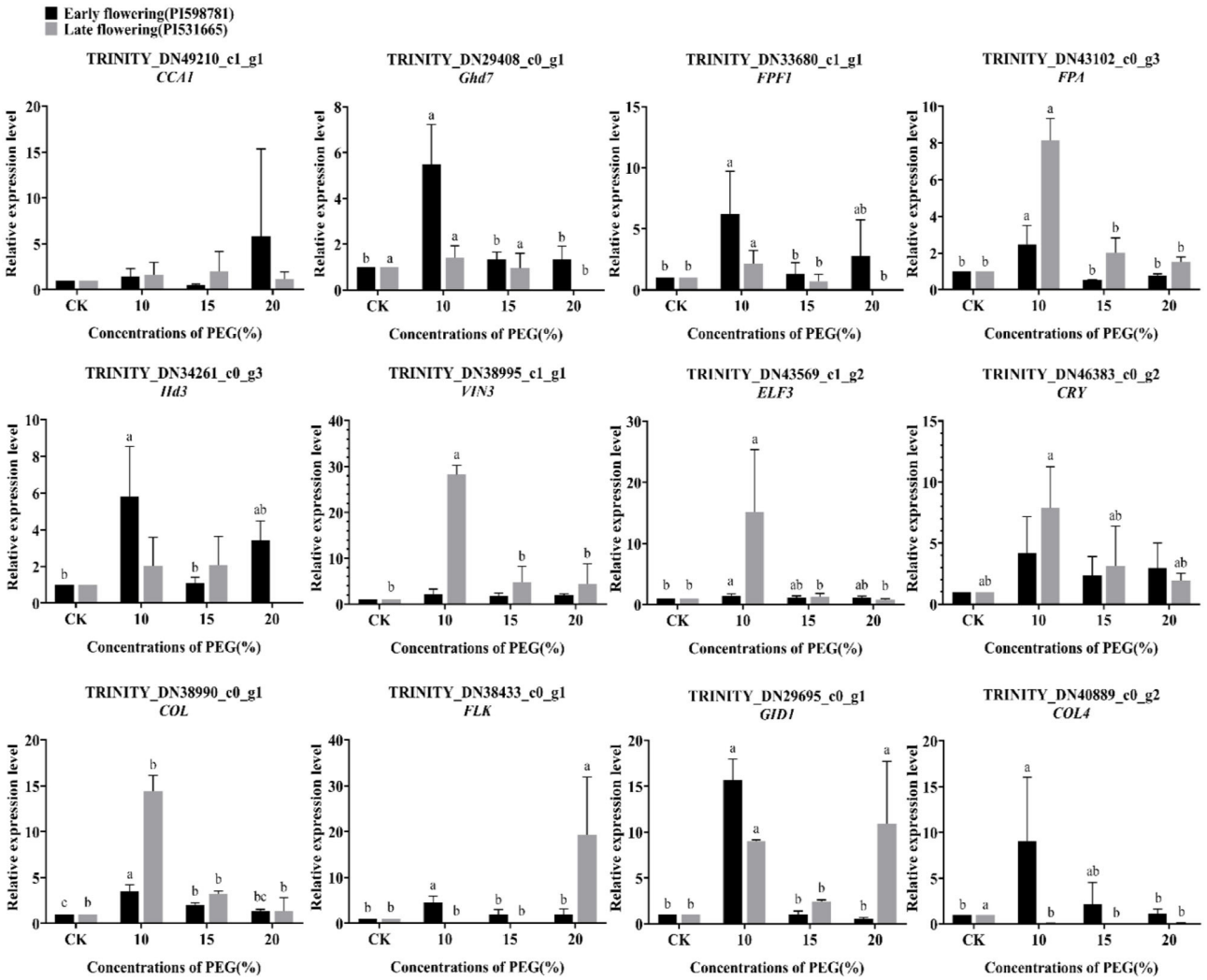


FIGURE 13 | The expression pattern of 12 flowering candidate genes in the early- and late-flowering accessions under drought stress by qRT-PCR analysis.

The genes *BOR7, PMEI28, LRR, and AIRP2* may be the flowering suppressors of *E. sibiricus*. *BOR7* is a boron transporter that carries an indispensable trace element during the flowering of plants. In *O. sativa*, the boron transporter, *OsBOR4*, is involved in normal pollen germination and tube elongation (Tanaka et al., 2013). It can also be verified whether *BOR7* affects pollen activity of *E. sibiricus* or not. *PMEI28* belongs to the carbohydrate esterase family. Pectin methylesterase can cut the ester bond between galacturonic acid and methyl, so the change in methyl esterification level plays a vital role in the growth and development of plants. Research has shown that overexpression of *OsPMEI28* in rice induced dwarf phenotypes and reduced stem diameter (Nguyen et al., 2017). *LRR* is the largest known subfamily of receptor-like protein kinases, which involves the whole life of plants, such as regulating plant growth and participating in stress response and defense response (Chakraborty et al., 2019). Similarly, the wheat *LRK10* gene involved in drought resistance was found to be related to the early flowering of alfalfa by RNA-seq (Ma et al., 2021). E3 ubiquitin ligase *AIRP2* could take part in the photoperiod pathway. Various components involved in the regulation of flowering photoperiod are the target proteins of E3 ubiquitin ligase, which can achieve ubiquitin degradation under the mediation of E3 ubiquitin ligase, resulting in photoperiod signal transformation and influencing the flowering of plants (Piñeiro and Jarillo, 2013). Moreover, *AtAIRP2* positively regulates abscisic acid (ABA) response in *A. thaliana* (Oh et al., 2017).

Allelic Variation Analysis of Hub Gene *LIMYB*

The *L10-INTERACTING MYB DOMAIN-CONTAINING PROTEIN* (*LIMYB*) is a newly identified MYB family protein. The gene interaction with *NIK1* can completely downregulate the genes involved in the translation mechanism, and overexpression

of *LIMYB* inhibits ribosomal protein genes at the transcriptional level, leading to inhibition of protein synthesis. At present, few works of literature have reported the role of *LIMYB* in plants. In this study, a single-nucleotide polymorphism (SNP) was found in the 320-bp ORF region of the hub gene *LIMYB* by allelic variation analysis which was not found in dbSNP (https://www.ncbi.nlm.nih.gov/SNP/). However, it is unclear that what types of biological functions are carried out through this SNP locus in *E.sibiricus*. There are two classical methods for SNP function verification. One is gene knockout which may not obtain phenotypes (Song et al., 2018; Wang et al., 2021), and the other is to predict changes in functional pathways through bioinformatic analysis and then conduct experimental verification (Ramsey et al., 2020). High cost and long test period are the difficulties of SNP function verification. Genome-wide association studies (GWAS) have become one of the most important methods for the identification of target quantitative trait loci (Paudel et al., 2021). Based on the GWAS results in soybean (*Glycine max*), two significant and sixteen non-significant SNP markers mapped to 18 candidate genes were involved in six major flowering pathways (Kim et al., 2020). In *O. sativa*, 309 SNPs related to flowering traits were identified by GWAS (Liu et al., 2021). In our subsequent experiment, GWAS will be conducted to verify whether the same SNP can be found in *LIMYB*. If so, functional validation will be conducted. Recently, reports have identified SNP characteristics that predict validation success in GWAS hits (Liu et al., 2021). These characteristics can be used to select SNPs for validation and downstream functional studies to provide the reference for our study.

Expression Analysis of Flowering Candidate Genes Under Abiotic Stress

Many types of abiotic stresses usually induce flowering to ensure the survival of future generations, including light intensity, ultraviolet light, high temperature, low temperature, malnutrition, nitrogen deficiency, drought, salt, hypoxia, and so on (Wada and Takeno, 2010). Nevertheless, different stress factors can also inhibit or delay flowering in many plant species (Takeno, 2016). In this study, the expression pattern of flowering candidate genes between early- and late-flowering accessions in *E. sibiricus* was verified by qRT-PCR, containing eight accelerated flowering genes (*CRY*, *COL*, *FPF1*, *Hd3*, *GID1*, *FLK*, *VIN3*, and *FPA*) (Kania et al., 1997; Schomburg et al., 2001; Monna et al., 2002; Griffiths et al., 2003; Hirano et al., 2008; Liu et al., 2008; Kim et al., 2010; Rodríguez-Cazorla et al., 2015) and four delayed flowering genes (*CCA1*, *ELF3*, *Ghd7*, and *COL4*) (Xue et al., 2008; Lu et al., 2012; Steinbach, 2019; Zhao et al., 2021).

Under salt stress, the majority of candidate genes were upregulated in early-flowering accessions that resulted in acceleration or delay in flowering. In late-flowering accessions, all the genes promoting flowering were upregulated, and all the genes inhibiting flowering except *ELF3* were only expressed at 150 mmol/L NaCl treatment. The studies of salt stress originated from salt accumulation caused by irrigation. Previous report showed that the salt overly sensitive pathway and photoperiod pathway co-regulated flowering time and stress tolerance (Park et al., 2016). Both acceleration and delay in flowering are the defensive behaviors of plants under salt stress to ensure reproduction. But, a long delay in flowering under high concentration stress may lead to an insufficient reproductive time of late-flowering accessions and thus affect seed setting. Reports indicated that the expression of *ELF3* not only delayed flowering but also enhanced salt tolerance of soybean and Arabidopsis (Sakuraba et al., 2017; Cheng et al., 2020). The study also confirmed the function of the *ELF3* gene in *E. sibiricus*. Under drought stress, many plants speed up the flowering process in the absence of water to shorten their life cycle and this process is known as drought-escape (DE) response (Kooyers, 2015). In this study, the majority of candidate genes were upregulated under low concentration PEG6000 in both early- and late-flowering accessions. However, at 20% PEG6000 treatment, all genes that inhibit flowering were downregulated, showing a DE response. Furthermore, some plants have other strategies, such as *RICE CENTRORADIALIS 1* (*RCN1*), which delays the flowering of rice under drought stress (Wang et al., 2020). The flowering promoter *COL4* was found always downregulated under any concentration of drought in late-flowering accessions as presented in this study.

Overall, stress activated the flowering regulation pathway to some extent, and flowering candidate genes of *E. sibiricus* were expressed under drought and salt stress. Under high salt and drought stress, plants preferentially accelerate their life processes and produce seeds before death. The network between flowering genes and abiotic stress needs further investigations in *E. sibiricus* for the development of stress-tolerant varieties with high seed yield and biomass.

CONCLUSION

In this study, we revealed flowering time variations of 66 *E. sibiricus* accessions collected from different countries and identified flowering regulatory networks and hub genes expressed in the leaves by comparative transcriptome analysis. A total of 3,526 DEGs were predicted and 72 candidate genes were identified, of which *LATE*, *GA2OX6*, *FAR3*, and *MFT1* were recognized as late-flowering genes. The WGCNA analysis revealed 7 hub genes controlling flowering in *E. sibiricus*. The result of multiple sequence alignment indicated that single-nucleotide polymorphism (SNP) of hub gene *LIMYB* may cause late flowering. The validation of the expression pattern of flowering candidate genes suggested that stress may activate the flowering regulation pathway to some extent. The research on flowering regulatory networks and these potential candidate genes will be valuable resources for further flowering functional research in *E. sibiricus*.

DATA AVAILABILITY STATEMENT

The *E. sibiricus* transcriptome sequencing data have been deposited in National Center for Biotechnology Information (NCBI), under accession number PRJNA665941.

AUTHOR CONTRIBUTIONS

WX conceived and designed the experiments and acquired financial support for the project leading to this publication. NW performed field management and sampling. YZ and NW performed the qRT-PCR and multiple sequence alignment experiments. YZ, NW, ZZ, and WL analyzed the data. YZ performed data visualization and wrote the initial draft. WX and YZ revised the manuscript. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was funded by the Gansu Provincial Science and Technology Major Projects (19ZD2NA002), Chinese National Natural Science Foundation (31971751), the State's Key Project of Research and Development Plan (2019YFC0507702), the open projects of Key Laboratory of Superior Forage Germplasm in the Qinghai-Tibetan Plateau (2020-ZJ-Y03), and the Fundamental Research Fund for the Central Universities (lzujbky-2021-ct21).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.877908/full#supplementary-material

Supplementary Figure 1 | The clustering heat map of samples for transcriptome analysis.

Supplementary Figure 2 | The unigenes distribution of *E. siburicus*.

Supplementary Figure 3 | The top 10 species distribution of transcripts that were annotated on the basis of homology.

Supplementary Figure 4 | KEGG classification map **(A–C)** and enrich scatter plot of KEGG pathway **(D–F)** of DEGs. **(A,D)** EB vs. LB. **(B,E)** EH vs. LH. **(C,F)** EF vs. LF.

Supplementary Figure 5 | The bar chart of correlation between DT trait and module genes.

Supplementary Table 1 | The origin of 66 *E. siburicus* accessions.

Supplementary Table 2 | The primers information of *E. siburicus*.

Supplementary Table 3 | Sequencing data evaluation statistics in *E. sibiricus*.

Supplementary Table 4 | The statistics of unigenes annotation in *E. sibiricus*.

Supplementary Table 5 | The information of flowering pathway genes.

REFERENCES

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Andrés, F., and Coupland, G. (2012). The genetic basis of flowering responses to seasonal cues. *Nat. Rev. Genet.* 13, 627–639. doi: 10.1038/nrg3291

Campos-Rivero, G., Osorio-Montalvo, P., Sánchez-Borges, R., Us-Camas, R., Duarte-Aké, F., and De-La-Peña, C. (2017). Plant hormone signaling in flowering: an epigenetic point of view. *J. Plant Physiol.* 214, 16–27. doi: 10.1016/j.jplph.2017.03.018

Chakraborty, S., Nguyen, B., Wasti, S. D., and Xu, G. (2019). Plant leucine-rich repeat receptor kinase (lrr-rk): structure, ligand perception, and activation mechanism. *Molecules* 24, 3081. doi: 10.3390/molecules24173081

Chen, H. J., Fu, T. Y., Yang, S. L., and Hsieh, H. L. (2018). FIN219/JAR1 and cryptochrome1 antagonize each other to modulate photomorphogenesis under blue light in *Arabidopsis*. *PLoS Genet.* 14, e1007248. doi: 10.1371/journal.pgen.1007248

Chen, X., Liu, W., Huang, X., Fu, H., Wang, Q., Wang, Y., et al. (2020). Arg-type dihydroflavonol 4-reductase genes from the fern *Dryopteris erythrosora* play important roles in the biosynthesis of anthocyanins. *PLoS ONE* 15, e0232090. doi: 10.1371/journal.pone.0232090

Cheng, Q., Gan, Z., Wang, Y., Lu, S., Hou, Z., Li, H., et al. (2020). The soybean gene *j* contributes to salt stress tolerance by up-regulating salt-responsive genes. *Front. Plant Sci.* 11, 272. doi: 10.3389/fpls.2020.00272

Childs, K. L., Davidson, R. M., and Buell, C. R. (2011). Gene coexpression network analysis as a source of functional annotation for rice genes. *PLoS ONE* 6, e22196. doi: 10.1371/journal.pone.0022196

Cremer, F., Havelange, A., Saedler, H., and Huijser, P. (1998). Environmental control of flowering time in *Antirrhinum majus*. *Physiol. Plant.* 104, 345–350. doi: 10.1034/j.1399-3054.1998.1040308.x

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics* 14, 755–763. doi: 10.1093/bioinformatics/14.9.755

Feng, G., Huang, L., Li, J., Wang, J., Xu, L., Pan, L., et al. (2017). Comprehensive transcriptome analysis reveals distinct regulatory programs during vernalization and floral bud development of orchardgrass (*Dactylis glomerata* L.). *BMC Plant Biol.* 17, 216. doi: 10.1186/s12870-017-1170-8

Gol, L., Tomé, F., and Von Korff, M. (2017). Floral transitions in wheat and barley: interactions between photoperiod, abiotic stresses, and nutrient status. *J. Exp. Bot.* 68, 1399–1410. doi: 10.1093/jxb/erx055

Griffiths, S., Dunford, R. P., Coupland, G., and Laurie, D. A. (2003). The evolution of CONSTANS-like gene families in barley, rice, and *Arabidopsis*. *Plant Physiol.* 131, 1855–1867. doi: 10.1104/pp.102.016188

Haq, M. I., Thakuri, B. K. C., Hobbs, T., Davenport, M. L., and Kumar, D. (2020). Tobacco SABP2-interacting protein SIP428 is a SIR2 type deacetylase. *Plant Physiol. Biochem.* 152, 72–80. doi: 10.1016/j.plaphy.2020.04.034

Hirano, K., Ueguchi-Tanaka, M., and Matsuoka, M. (2008). GID1-mediated gibberellin signaling in plants. *Trends Plant Sci.* 13, 192–199. doi: 10.1016/j.tplants.2008.02.005

Huang, H. R., Yan, P. C., Lascoux, M., and Ge, X. J. (2012). Flowering time and transcriptome variation in *Capsella bursa-pastoris* (*Brassicaceae*). *New Phytol.* 194, 676–689. doi: 10.1111/j.1469-8137.2012.04101.x

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., et al. (2002). bZIP transcription factors in *Arabidopsis*. *Trends Plant Sci.* 7, 106–111. doi: 10.1016/S1360-1385(01)02223-3

Jing, D., Chen, W., Hu, R., Zhang, Y., Xia, Y., Wang, S., et al. (2020). An integrative analysis of transcriptome, proteome and hormones reveals key differentially expressed genes and metabolic pathways involved in flower development in loquat. *Int. J. Mol. Sci.* 21, 5107. doi: 10.3390/ijms21145107

Kania, T., Russenberger, D., Peng, S., Apel, K., and Melzer, S. (1997). *FPF1* promotes flowering in *Arabidopsis*. *Plant Cell* 9, 1327–1338. doi: 10.1105/tpc.9.8.1327

Kazan, K., and Lyons, R. (2016). The link between flowering time and stress tolerance. *J. Exp. Bot.* 67, 47–60. doi: 10.1093/jxb/erv441

Kim, D. H., Zografos, B. R., and Sung, S. (2010). Mechanisms underlying vernalization-mediated *VIN3* induction in *Arabidopsis*. *Plant Signal. Behav.* 5, 1457–1459. doi: 10.4161/psb.5.11.13465

Kim, K. H., Kim, J. Y., Lim, W. J., Jeong, S., Lee, H. Y., Cho, Y., et al. (2020). Genome-wide association and epistatic interactions of flowering time in soybean cultivar. *PLoS ONE* 15, e0228114. doi: 10.1371/journal.pone.0228114

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

Kobayashi, Y., and Weigel, D. (2007). Move on up, it's time for change–mobile signals controlling photoperiod-dependent flowering. *Genes Dev.* 21, 2371–2384. doi: 10.1101/gad.1589007

Kong, D. Y., Chen, S. J., Zhou, L. G., Gao, H., Luo, L. J., and Liu, Z. C. (2016). Research progress of photoperiod regulated genes on flowering time in rice. *Yi Chuan* 38, 532–542. doi: 10.16288/j.yczz.15-478

Kooyers, N. J. (2015). The evolution of drought escape and avoidance in natural herbaceous populations. *Plant Sci.* 234, 155–162. doi: 10.1016/j.plantsci.2015.02.012

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

Liang, Y., Wang, J., Zheng, J., Gong, Z., Li, Z., Ai, X., et al. (2021). Genome-wide comparative analysis of heat shock transcription factors provides novel insights for evolutionary history and expression characterization in cotton diploid and tetraploid genomes. *Front. Genet.* 12, 658847. doi: 10.3389/fgene.2021.658847

Linder, H. P. (2020). The evolution of flowering phenology: an example from the wind-pollinated African restionaceae. *Ann. Bot.* 126, 1141–1153. doi: 10.1093/aob/mcaa129

Liu, C., Tu, Y., Liao, S., Fu, X., Lian, X., He, Y., et al. (2021). Genome-wide association study of flowering time reveals complex genetic heterogeneity and epistatic interactions in rice. *Gene* 770, 145353. doi: 10.1016/j.gene.2020.145353

Liu, H., Yu, X., Li, K., Klejnot, J., Yang, H., Lisiero, D., et al. (2008). Photoexcited *CRY2* interacts with *CIB1* to regulate transcription and floral initiation in *Arabidopsis*. *Science* 322, 1535–1539. doi: 10.1126/science.1163927

Lu, S. X., Webb, C. J., Knowles, S. M., Kim, S. H., Wang, Z., and Tobin, E. M. (2012). *CCA1* and *ELF3* Interact in the control of hypocotyl length and flowering time in *Arabidopsis*. *Plant Physiol.* 158, 1079–1088. doi: 10.1104/pp.111.189670

Ma, D., Liu, B., Ge, L., Weng, Y., Cao, X., Liu, F., et al. (2021). Identification and characterization of regulatory pathways involved in early flowering in the new leaves of alfalfa (*Medicago sativa* L.) by transcriptome analysis. *BMC Plant Biol.* 21, 8. doi: 10.1186/s12870-020-02775-9

Matías-Hernández, L., Aguilar-Jaramillo, A. E., Cigliano, R. A., Sanseverino, W., and Pelaz, S. (2016). Flowering and trichome development share hormonal and transcription factor regulation. *J. Exp. Bot.* 67, 1209–1219. doi: 10.1093/jxb/erv534

McDonnell, M. M., Burkhart, S. E., Stoddard, J. M., Wright, Z. J., Strader, L. C., and Bartel, B. (2016). The early-acting peroxin *PEX19* is redundantly encoded, farnesylated, and essential for viability in *Arabidopsis thaliana*. *PLoS ONE* 11, e0148335. doi: 10.1371/journal.pone.0148335

Mdodana, N. T., Jewell, J. F., Phiri, E. E., Smith, M. L., Oberlander, K., Mahmoodi, S., et al. (2019). Mutations in glucan, water dikinase affect starch degradation and gametophore development in the moss *Physcomitrella patens*. *Sci. Rep.* 9, 15114. doi: 10.1038/s41598-019-51632-9

Michaels, S. D., and Amasino, R. M. (2001). Loss of *FLOWERING LOCUS C* activity eliminates the late-flowering phenotype of *FRIGIDA* and autonomous pathway mutations but not responsiveness to vernalization. *Plant Cell* 13, 935–941. doi: 10.1105/tpc.13.4.935

Minh-Thu, P. T., Kim, J. S., Chae, S., Jun, K. M., Lee, G. S., Kim, D. E., et al. (2018). A WUSCHEL homeobox transcription factor, OsWOX13, enhances drought tolerance and triggers early flowering in Rice. *Mol. Cells* 41, 781–798. doi: 10.14348/molcells.2018.0203

Monna, L., Lin, X., Kojima, S., Sasaki, T., and Yano, M. (2002). Genetic dissection of a genomic region for a quantitative trait locus, *Hd3*, into two loci, *Hd3a* and *Hd3b*, controlling heading date in rice. *Theor. Appl. Genet.* 104, 772–778. doi: 10.1007/s00122-001-0813-0

Naranjo, L., Talón, M., and Domingo, C. (2014). Diversity of floral regulatory genes of japonica rice cultivated at northern latitudes. *BMC Genomics* 15, 101. doi: 10.1186/1471-2164-15-101

Nguyen, H. P., Jeong, H. Y., Jeon, S. H., Kim, D., and Lee, C. (2017). Rice pectin methylesterase inhibitor28 (*OsPMEI28*) encodes a functional *PMEI* and its overexpression results in a dwarf phenotype through increased pectin methylesterification levels. *J. Plant Physiol.* 208, 17–25. doi: 10.1016/j.jplph.2016.11.006

Oh, T. R., Kim, J. H., Cho, S. K., Ryu, M. Y., Yang, S. W., and Kim, W. T. (2017). *AtAIRP2 E3 ligase* affects ABA and high-salinity responses by stimulating its *ATP1*/*SDIRIP1* substrate turnover. *Plant Physiol.* 174, 2515–2531. doi: 10.1104/pp.17.00467

Osnato, M., Castillejo, C., Matías-Hernández, L., and Pelaz, S. (2012). *TEMPRANILLO* genes link photoperiod and gibberellin pathways to control flowering in *Arabidopsis*. *Nat. Commun.* 3, 808. doi: 10.1038/ncomms1810

Park, H. J., Kim, W. Y., and Yun, D. J. (2016). A new insight of salt stress signaling in plant. *Mol. Cells* 39, 447–459. doi: 10.14348/molcells.2016.0083

Paudel, D., Dareus, R., Rosenwald, J., Muñoz-Amatriaín, M., and Rios, E. F. (2021). Genome-wide association study reveals candidate genes for flowering time in Cowpea [*Vigna unguiculata* (L.) Walp.]. *Front. Genet.* 12, 667038. doi: 10.3389/fgene.2021.667038

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

Piñeiro, M., and Jarillo, J. A. (2013). Ubiquitination in the control of photoperiodic flowering. *Plant Sci.* 198, 98–109. doi: 10.1016/j.plantsci.2012.10.005

Ramsey, S. A., Liu, Z., Yao, Y., and Weeder, B. (2020). Combining eQTL and SNP annotation data to identify functional noncoding SNPs in GWAS trait-associated regions. *Methods Mol. Biol.* 2082, 73–86. doi: 10.1007/978-1-0716-0026-9_6

Riboni, M., Robustelli Test, A., Galbiati, M., Tonelli, C., and Conti, L. (2014). Environmental stress and flowering time: the photoperiodic connection. *Plant Signal. Behav.* 9, e29036. doi: 10.4161/psb.29036

Rodríguez-Cazorla, E., Ripoll, J. J., Andújar, A., Bailey, L. J., Martínez-Laborda, A., Yanofsky, M. F., et al. (2015). K-homology nuclear ribonucleoproteins regulate floral organ identity and determinacy in arabidopsis. *PLoS Genet.* 11, e1004983. doi: 10.1371/journal.pgen.1004983

Sakuraba, Y., Bülbül, S., Piao, W., Choi, G., and Paek, N. C. (2017). Arabidopsis *EARLY FLOWERING3* increases salt tolerance by suppressing salt stress response pathways. *Plant J.* 92, 1106–1120. doi: 10.1111/tpj.13747

Schomburg, F. M., Patton, D. A., Meinke, D. W., and Amasino, R. M. (2001). *FPA*, a gene involved in floral induction in *Arabidopsis*, encodes a protein containing RNA-recognition motifs. *Plant Cell* 13, 1427–1436. doi: 10.1105/TPC.010017

Shim, J. S., and Jang, G. (2020). Environmental signal-dependent regulation of flowering time in rice. *Int. J. Mol. Sci.* 21, 6155. doi: 10.3390/ijms21176155

Song, Y., Zhang, Y., Chen, M., Deng, J., Sui, T., Lai, L., et al. (2018). Functional validation of the albinism-associated tyrosinase T373K SNP by CRISPR/Cas9-mediated homology-directed repair (HDR) in rabbits. *EBioMedicine* 36, 517–525. doi: 10.1016/j.ebiom.2018.09.041

Srikanth, A., and Schmid, M. (2011). Regulation of flowering time: all roads lead to Rome. *Cell. Mol. Life Sci.* 68, 2013–2037. doi: 10.1007/s00018-011-0673-y

Steinbach, Y. (2019). The *Arabidopsis thaliana CONSTANS-LIKE 4* (*COL4*) - A modulator of flowering time. *Front. Plant Sci.* 10, 651. doi: 10.3389/fpls.2019.00651

Su, Z., Ma, X., Guo, H., Sukiran, N. L., Guo, B., Assmann, S. M., et al. (2013). Flower development under drought stress: morphological and transcriptomic analyses reveal acute responses and long-term acclimation in *Arabidopsis*. *Plant Cell* 25, 3785–3807. doi: 10.1105/tpc.113.115428

Takeno, K. (2016). Stress-induced flowering: the third category of flowering response. *J. Exp. Bot.* 67, 4925–4934. doi: 10.1093/jxb/erw272

Tanaka, N., Uraguchi, S., Saito, A., Kajikawa, M., Kasai, K., Sato, Y., et al. (2013). Roles of pollen-specific boron efflux transporter, *OsBOR4*, in the rice fertilization process. *Plant Cell Physiol.* 54, 2011–2019. doi: 10.1093/pcp/pct136

Van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018). The third revolution in sequencing technology. *Trends Genet.* 34, 666–681. doi: 10.1016/j.tig.2018.05.008

Verslues, P. E., and Juenger, T. E. (2011). Drought, metabolites, and Arabidopsis natural variation: a promising combination for understanding adaptation to water-limited environments. *Curr. Opin. Plant Biol.* 14, 240–245. doi: 10.1016/j.pbi.2011.04.006

Wada, K. C., and Takeno, K. (2010). Stress-induced flowering. *Plant Signal. Behav.* 5, 944–947. doi: 10.4161/psb.5.8.11826

Wang, X., Hayes, J. E., Xu, X., Gao, X., Mehta, D., Lilja, H. G., et al. (2021). Validation of prostate cancer risk variants rs10993994 and rs7098889 by CRISPR/Cas9 mediated genome editing. *Gene* 768, 145265. doi: 10.1016/j.gene.2020.145265

Wang, Y., Lu, Y., Guo, Z., Ding, Y., and Ding, C. (2020). *RICE CENTRORADIALIS 1*, a *TFL1*-like gene, responses to drought stress and regulates rice flowering transition. *Rice* 13, 70. doi: 10.1186/s12284-020-00430-3

Wolabu, T. W., Zhang, F., Niu, L., Kalve, S., Bhatnagar-Mathur, P., Muszynski, M. G., et al. (2016). Three *FLOWERING LOCUS T*-like genes function as potential florigens and mediate photoperiod response in sorghum. *New Phytol* 210, 946–959. doi: 10.1111/nph.13834

Xie, W., Zhang, J., Zhao, X., Zhang, J., and Wang, Y. (2015a). Siberian wild rye (*Elymus sibiricus* L.): Genetic diversity of germplasm determined using DNA fingerprinting and SCoT markers. *Biochem. Syst. Ecol.* 60, 186–192. doi: 10.1016/j.bse.2015.04.021

Xie, W., Zhao, X., Zhang, J., Wang, Y., and Liu, W. (2015b). Assessment of genetic diversity of Siberian wild rye (*Elymus sibiricus* L.) germplasms with variation of seed shattering and implication for future genetic improvement. *Biochem. Syst. Ecol.* 58, 211–218. doi: 10.1016/j.bse.2014.12.006

Xie, W. G., Robins, J. G., and Bushman, B. S. (2012). A genetic linkage map of tetraploid orchardgrass (*Dactylis glomerata* L.) and quantitative trait loci for heading date. *Genome* 55, 360–369. doi: 10.1139/g2012-026

Xiong, Y., Lei, X., Bai, S., Xiong, Y., Liu, W., Wu, W., et al. (2021). Genomic survey sequencing, development and characterization of single- and multi-locus genomic SSR markers of *Elymus sibiricus* L. *BMC Plant Biol.* 21, 3. doi: 10.1186/s12870-020-02770-0

Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L., et al. (2008). Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nat. Genet.* 40, 761–767. doi: 10.1038/ng.143

Yamaguchi, N., Wu, M. F., Winter, C. M., Berns, M. C., Nole-Wilson, S., Yamaguchi, A., et al. (2013). A molecular framework for auxin-mediated initiation of flower primordia. *Dev. Cell* 24, 271–282. doi: 10.1016/j.devcel.2012.12.017

Yan, Y., Shen, L., Chen, Y., Bao, S., Thong, Z., and Yu, H. (2014). A MYB-domain protein EFM mediates flowering responses to environmental cues in *Arabidopsis*. *Dev. Cell* 30, 437–448. doi: 10.1016/j.devcel.2014.07.004

Yu, Y., Qiao, L., Chen, J., Rong, Y., Zhao, Y., Cui, X., et al. (2020). Arabidopsis *REM16* acts as a B3 domain transcription factor to promote flowering time via directly binding to the promoters of *SOC1* and *FT*. *Plant J.* 103, 1386–1398. doi: 10.1111/tpj.14807

Zeng, F., Biligetu, B., Coulman, B., Schellenberg, M. P., and Fu, Y. B. (2017). RNA-Seq analysis of gene expression for floral development in crested wheatgrass (*Agropyron cristatum* L.). *PLoS ONE* 12, e0177417. doi: 10.1371/journal.pone.0177417

Zhang, H., Cui, X., Guo, Y., Luo, C., and Zhang, L. (2018a). Picea wilsonii transcription factor NAC2 enhanced plant tolerance to abiotic stress and participated in *RFCP1*-regulated flowering time. *Plant Mol. Biol.* 98, 471–493. doi: 10.1007/s11103-018-0792-z

Zhang, J., Xie, W., Yu, X., Zhang, Z., Zhao, Y., Wang, N., et al. (2019). Selection of suitable reference genes for RT-qPCR gene expression analysis in siberian wild rye (*Elymus sibiricus*) under different experimental conditions. *Genes* 10, 451. doi: 10.3390/genes10060451

Zhang, L., Chen, L., and Yu, D. (2018b). Transcription factor WRKY75 interacts with DELLA proteins to affect flowering. *Plant Physiol.* 176, 790–803. doi: 10.1104/pp.17.00657

Zhao, H., Nie, K., Zhou, H., Yan, X., Zhan, Q., Zheng, Y., et al. (2020a). *ABI5* modulates seed germination via feedback regulation of the expression of the *PYR/PYL/RCAR* ABA receptor genes. *New Phytol.* 228, 596–608. doi: 10.1111/nph.16713

Zhao, H., Xu, D., Tian, T., Kong, F., Lin, K., Gan, S., et al. (2021). Molecular and functional dissection of *EARLY-FLOWERING 3* (*ELF3*) and *ELF4* in *Arabidopsis*. *Plant Sci.* 303, 110786. doi: 10.1016/j.plantsci.2020.110786

Zhao, J., Chen, H., Ren, D., Tang, H., Qiu, R., Feng, J., et al. (2015). Genetic interactions between diverged alleles of *Early heading date 1* (*Ehd1*) and *Heading date 3a* (*Hd3a*)/ *RICE FLOWERING LOCUS T1* (*RFT1*) control differential heading and contribute to regional adaptation in rice (*Oryza sativa*). *New Phytol.* 208, 936–948. doi: 10.1111/nph.13503

Zhao, L., Richards, S., Turck, F., and Kollmann, M. (2020b). Information integration and decision making in flowering time control. *PLoS ONE* 15, e0239417. doi: 10.1371/journal.pone.0239417

Zheng, C., Ye, M., Sang, M., and Wu, R. (2019). A regulatory network for miR156-SPL module in *Arabidopsis thaliana*. *Int. J. Mol. Sci.* 20, 6166. doi: 10.3390/ijms20246166

Zorzatto, C., Machado, J. P., Lopes, K. V., Nascimento, K. J., Pereira, W. A., Brustolini, O. J., et al. (2015). NIK1-mediated translation suppression functions as a plant antiviral immunity mechanism. *Nature* 520, 679–682. doi: 10.1038/nature14171

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 26 May 2022
doi: 10.3389/fpls.2022.899681

RAD-Seq-Based High-Density Linkage Maps Construction and Quantitative Trait Loci Mapping of Flowering Time Trait in Alfalfa (*Medicago sativa* L.)

Xueqian Jiang[1], Tianhui Yang[2], Fan Zhang[1], Xijiang Yang[1], Changfu Yang[1], Fei He[1], Ruicai Long[1], Ting Gao[2], Yiwei Jiang[3], Qingchuan Yang[1], Zhen Wang[1] and Junmei Kang[1]**

[1] Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China, [2] Institute of Animal Science, Ningxia Academy of Agricultural and Forestry Sciences, Yinchuan, China, [3] Department of Agronomy, Purdue University, West Lafayette, IN, United States

OPEN ACCESS

Edited by:
Jiyu Zhang,
Lanzhou University, China

Reviewed by:
Muhammet Şakiroğlu,
Adana Alparslan Turkes Science and Technology University, Turkey
Zefeng Yang,
Yangzhou University, China

***Correspondence:**
Zhen Wang
wangzhen@caas.cn
Junmei Kang
kangjunmei@caas.cn

Specialty section:
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

Received: *19 March 2022*
Accepted: *25 April 2022*
Published: *26 May 2022*

Citation:
Jiang X, Yang T, Zhang F, Yang X, Yang C, He F, Long R, Gao T, Jiang Y, Yang Q, Wang Z and Kang J (2022) RAD-Seq-Based High-Density Linkage Maps Construction and Quantitative Trait Loci Mapping of Flowering Time Trait in Alfalfa (Medicago sativa L.). Front. Plant Sci. 13:899681. doi: 10.3389/fpls.2022.899681

Alfalfa (*Medicago sativa* L.) is a perennial forage crop known as the "Queen of Forages." To dissect the genetic mechanism of flowering time (FT) in alfalfa, high−density linkage maps were constructed for both parents of an F1 mapping population derived from a cross between Cangzhou (P1) and ZhongmuNO.1 (P2), consisting of 150 progenies. The FT showed a transgressive segregation pattern in the mapping population. A total of 13,773 single-nucleotide polymorphism markers was obtained by using restriction-site associated DNA sequencing and distributed on 64 linkage groups, with a total length of 3,780.49 and 4,113.45 cM and an average marker interval of 0.58 and 0.59 cM for P1 and P2 parent, respectively. Quantitative trait loci (QTL) analyses were performed using the least square means of each year as well as the best linear unbiased prediction values across 4 years. Sixteen QTLs for FT were detected for P1 and 22 QTLs for P2, accounting for 1.40–16.04% of FT variation. RNA-Seq analysis at three flowering stages identified 5,039, 7,058, and 7,996 genes that were differentially expressed between two parents, respectively. Based on QTL mapping, DEGs analysis, and functional annotation, seven candidate genes associated with flowering time were finally detected. This study discovered QTLs and candidate genes for alfalfa FT, making it a useful resource for breeding studies on this essential crop.

Keywords: alfalfa, flowering time, genetic map, QTL, RAD-seq, RNA-seq

INTRODUCTION

Alfalfa (*Medicago sativa* L.) is an important forage species with high yields, quality, and adaptability. This species has a wide distribution all over the world (Radovic et al., 2009). China is one of the largest producers of alfalfa with a long history of its cultivation and utilization for livestock. However, the increasing demands for livestock production pose a challenge for the alfalfa industry (Bai et al., 2018), mainly due to a lack of alfalfa cultivars with a wide range of adaptation. This obstacle has been severely limiting the expansion of alfalfa cultivation

and production in China. Until 2017, approximately 71% of the alfalfa was grown in the four provinces of northwestern China (Xinjiang, Inner Mongolia, Gansu, and Shaanxi). Therefore, it is crucial to develop alfalfa cultivars with improved yield and adaptability to meet the growing needs of alfalfa production.

Flowering time (FT), a major fitness component, is a key trait for the seasonal and geographical adaptation of flowering plants (Kooyers, 2015). FT is a signal that allows plants to change from vegetative growth to reproductive growth (Srikanth and Schmid, 2011), which plays an important role in alfalfa biomass yield. Premature flowering often leads to a shorter vegetative growth stage that results in reduced biomass yield, whereas delaying the flowering could increase plant yield and biomass of perennial crops, including alfalfa. Given the importance of FT, it has been suggested that the development of the novel varieties with modified FT may result in higher biomass production due to a better adaptation of plants to local environments and climate conditions (Schiffers et al., 2013).

Flowering time is a complex quantitative trait determined by genotype, environment, and the interactions of genotype and environment (Koornneef et al., 1998; Upadhyaya et al., 2015). Previous research showed that shifting FT-related genes enables plants to gain a wider adaptability (Diaz et al., 2012; Weller et al., 2019; Lu et al., 2020). For instance, *Tof11* and *Tof12* delayed FT under long photoperiods and improved plant adaptation to higher latitudes during soybean domestication (Lu et al., 2020). In addition, the mechanism of flowering time has been well studied in model plants like Arabidopsis and in crop species such as rice (*Oryza sativa* L.) (Hori et al., 2016; Kinoshita and Richter, 2020). Numerous candidate genes have been identified and their intricate regulatory networks are being deciphered, including *FLOWERING LOCUS T* T (*FT*), *CONSTANS* (*CO*), and FLOWERING LOCUS C (*FLC*)/mads affecting FLOWERING (*MAF*) (Song et al., 2013; Kinoshita and Richter, 2020). *FT*-like genes are involved in photoperiod sensing, and homologues of *FT* have also been identified to control flowering time in several legume species. For example, five homologues of *FT* were characterized in *Medicago truncatula*, two of which were flowering inductors (Hecht et al., 2011; Laurie et al., 2011). Five homologues were detected in alfalfa, but only *MsFTa1* was found to be a flowering inducer (Lorenzo et al., 2020). Nevertheless, no functional homologues of *CO* were identified in *Medicago truncatula* (Wong et al., 2014), suggesting that the *CO*-like genes were not conserved in legumes, including alfalfa.

Recently, RNA-sequencing (RNA-seq) has been widely applied in various species due to the rapid advancement of the technology and decreased price. However, through RNA-seq, a large quantity of differentially expressed genes (DEGs) are usually identified. Therefore, the strategy of QTL mapping combined with RNA-seq is considered to be a promising method to rapidly identify potential candidate genes. For example, integration of RNA-seq and QTLs mapping have been carried out to identify candidate genes associated with cadmium tolerance in barley (Hordeum vulgare) (Derakhshani et al., 2020), genes for pod number variation in rapeseed (*Brassica napus* L.) (Ye et al., 2017), and high-temperature stress-responsive genes in tomato (*Solanum lycopersicum* L.) (Wen et al., 2019).

As alfalfa is a long-day plant, its flowering is affected by both photoperiod and temperature (Major et al., 1991; Lorenzo et al., 2020). Although some FT-related QTLs have been investigated in alfalfa (Adhikari et al., 2019; Zhang et al., 2020), the genetic basis of flowering time remains largely unexplored. In previous studies, due to the lack of an alfalfa reference genome, it was difficult to construct high−density linkage maps and compare QTLs in different mapping populations. All of these studies were made feasible by the availability of a high-quality, chromosome-level alfalfa genome (Chen H. et al., 2020). RAD-seq is an economical and useful method of generating SNP markers across the whole genome (Huang et al., 2017; Lee et al., 2019). Therefore, the objectives of this study were to construct linkage maps using genome-wide SNP markers based on RAD-seq, detect the QTLs related to FT, and identify candidate genes by integrating the QTLs and RNA-seq.

MATERIALS AND METHODS

Plant Materials and Phenotyping

An F_1 population consisting of 150 progenies was developed by crossing a tetraploid early flowering and low forage yield alfalfa landrace (Cangzhou, CF000735, paternal parent, P1) with a late flowering and high forage yield cultivar (Zhongmu NO.1, CF0032020, maternal parent, P2). The seeds of the F_1 population were planted and the clones of progenies and two parents were initially propagated in a greenhouse in 2012 in the Chinese Academy of Agricultural Sciences Research Station in Langfang, Hebei Province (39.59°N, 116.59°E). In 2013, the cloned plants were transferred to the field in Langfang, Hebei Province (Zhang et al., 2019), and propagated again in 2016 to ensure genetic uniformity of all plant materials. The field experiment was a randomized complete block design with three replications, where one clone for each progeny and parents was planted in per replication.

In this study, FT was defined as the date of the first flower appeared in individual plants. It was recorded daily from the date when the first flower appeared in a plant until all plants flowered. We collected FT data across 4 years from 2017 to 2020. FT dates were expressed as photothermal units (PTU) using the method described by Grabowski et al. (2017). The PTU was estimated using the following formula;

$$dPTU = ((minT + maxT)/2 - -10)^{*}dayL/24$$

where dPTU = daily PTU; minT = the minimum recorded temperature; maxT = the maximum recorded temperature; dayL = the number of hours between sunrise and sunset. PTU starts accumulating after 5 consecutive days with (minT + maxT) > 10°C. The information of minT and maxT was obtained from the climatological station,[1] and dayL data was collected from the website.[2]

[1] http://tianqi.2345.com/wea_history/60198.htm

[2] https://richurimo.51240.com/beijing__richurimo/

We estimated the PTU values using the least-squares mean (LS mean), performing for each year separately with PROC GLM (v. 9.4; SAS Institute, Cary, NC, United States). The best linear unbiased prediction (BLUP) values were calculated using the "lme4" package in R (Bates et al., 2014). The R package lme4 was also used to estimate the broad-sense heritability (H^2) according to the following formula:

$$H^2 \quad \frac{\sigma_g^2}{\sigma_g^2 + (\sigma_r^2 + \sigma_\varepsilon^2)/r}$$

where H^2 = broad-sense heritability; $\sigma^2{}_g$ = genotypic variance; $\sigma^2{}_r$ = block variance; $\sigma^2{}_\varepsilon$ = residual error variance; r, the numbers of blocks. The LS means of each year along with BLUP values were used as the phenotype data for QTL analysis.

DNA Extraction and Sequencing

The information on DNA isolation and RAD library construction has been described in detail in our previous study (Zhang et al., 2019). Briefly, young leaf tissues from the parents and F_1 progenies were collected and frozen in liquid nitrogen. DNA was extracted using the CWBIO Plant Genomic DNA Kit (CoWin Biosciences, Beijing, China) according to the manufacturer's protocol. DNA concentrations were measured using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, United States). RAD library preparation and sequencing were performed at ORI-GENE (Beijing, China), using a Hi-seq X Ten (Illumina). The RAD sequences were deposited in the NCBI Sequence Read Archive (PRJNA503672).

Reference-Based Single-Nucleotide Polymorphism Calling and Genetic Linkage Maps Construction

Sequencing data was filtered using trimmomatic software with default parameters (Bolger et al., 2014). Paired-end sequencing reads were mapped to the alfalfa reference genome (Chen H. et al., 2020) with BWA-MEM (Li, 2013). Samtools was used to translate SAM files to BAM files and then to sort BAM files using default parameters (Li et al., 2009). GATK4.0 was used to mark duplicate reads and detect single-nucleotide polymorphism (SNP) (McKenna et al., 2010). Furthermore, SNP data was filtered using Vcftools (Danecek et al., 2011) with the missing rate less than 50%, minor allele frequency greater than 0.05 and mean read depth greater than 5. Single-dose allele (SDA) markers with a ratio of less than 2:1 and missing data rates of $\leq$ 20% were retained for constructing genetic linkage maps (Li et al., 2014). The regression method of JoinMap 4.0[3] was used to order the markers within the chromosomes (Stam, 1993). Kosambi map function was used to translate recombination frequency into map distances.

Quantitative Trait Loci Analysis for Flowering Time

Combining the phenotypic data and the genetic linkage maps, QTLs for FT were identified in the F_1 population using the Inclusive Composite Interval Mapping with an additive effect (ICIM-ADD) in the QTL IciMapping software[4] (Meng et al., 2015). The mapping parameter of each step for ICIM-ADD was set at 1.0 cM, and the LOD threshold was set at 3.0. When the genetic locations of the QTLs (at a 99% significance level) overlapped in different years, they were defined as a single QTL. The QTLs detected for each parental map were drawn on linkage maps using MapChart 2.3 (Voorrips, 2002). QTLs were named as: *qFT* + linkage group no., or named as *qFT* + linkage group no. + an ordered number designating one of multiple QTLs in a single linkage group. For example, *qFT4.1-2* indicates the second QTL in the Chr4.1 linkage group.

RNA-Sequencing Analysis

The flower samples were collected at three flowering stages (budding stage, BS; initial flowering stage, IFS and full flowering stage, FFS) from two parental alfalfa plants and immediately frozen in liquid nitrogen and stored at −80°C. Altogether, 12 flower samples were collected for all two replicates. Total RNA was isolated using the TRIZOL reagent (Invitrogen, CA, United States) according to the manufacturer's protocol. High-quality total RNA samples were sent to the Novogene (Beijing, China) and sequenced with an Illumina HiSeq 2000 platform. The obtained RNA-Seq data were submitted to the Sequence Read Archive of the National Center for Biotechnology Information (NCBI) with the accession number PRJNA817856.[5]

Estimation of RNA-Seq data quality was performed by FastQC v0.11.9,[6] and low-quality sequences were filtered by fastp v0.12.4 (Chen et al., 2018). All filtered clean reads were mapped to the XJDY reference genome using hisat2 v4.8.2 (Kim et al., 2019), and the alignments were sorted using SAMTools (Li et al., 2009). Read counts for each sample were generated using featureCounts v2.0.1 (Liao et al., 2014). DEGs with an adjusted p-value $\leq$ 0.0001 and expression level of $|\log2 \text{ fold change}| \geq 3$ was performed with the R package DESeq2 (Love et al., 2014).

Comparison With Previously Reported Quantitative Trait Loci and Potential Candidate Gene Identification

The sequences of flanking markers for each previously reported QTL were used for local blast (E-value of $1e^{-5}$) against the alfalfa reference genome (cultivar XinJiangDaYe) (Chen H. et al., 2020), which was downloaded online.[7] TBtools software (Chen C. et al., 2020) was used to complete the following analysis. First, we identified the physical positions based on the flanking marker sequences of the QTLs with the function of –BLAST–BLAST GUI Wrapper–Several Sequences to a Big File. Then we extracted genes on the alfalfa genome based on the overlapped physical positions in different genetic backgrounds (function: –Sequence Toolkit–GFF3/GTF Manipulate–GXF Regoin Overlap). DEGs

[3]www.kyazma.nl

[4]www.isbreeding.net

[5]https://www.ncbi.nlm.nih.gov/bioproject/PRJNA817856

[6]https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

[7]https://figshare.com/projects/whole_genome_sequencing_and_assembly_of_Medicago_sativa/66380

within QTL intervals were annotated based on BLSAT search in NCBI,[8] Ensembl,[9] and eggNOG-mapper.[10]

Quantitative Real-Time PCR

The RNA-seq data were validated using quantitative real-time PCR (qRT-PCR). On a 7500 Real-Time PCR System (Applied Biosystems, CA, United States), qRT-PCR was implemented in triplicate for each sample of two parents at three flowering stages using the SYBR Premix Ex Taq (Takara, Japan). The relative gene expression level was calculated by the $2^{-\Delta\Delta Ct}$ method.

[8]https://www.ncbi.nlm.nih.gov/

[9]https://ensembl.gramene.org/

[10]http://eggnog-mapper.embl.de/

RESULTS

Phenotypic Data Analysis

Phenotypic data of the F_1 population and parents were collected in Langfang, Hebei Province for four consecutive

TABLE 1 | Statistical analysis of LS means of photothermal units values in parents and F1 population across 4 years.

Year	Parents		F_1 population				H^2
	P1	P2	Range	Mean	CV (%)	Se	
2017	15.44	242.94***	150.77∼293.40	192.01	14.47	2.13	0.74
2018	191.65	256.78**	185.23∼288.11	221.06	11.28	1.61	0.71
2019	49.56	213.72**	135.46∼233.17	171.52	15.15	1.64	0.68
2020	153.06	223.16*	143.15∼271.69	184.61	16.00	2.18	0.78

*CV, Coefficient of variance; se, Standard error; H2, broad-sense heritability. Asterisks indicate significant differences in FT between parents (t-test, *P < 0.05; **P < 0.01; ***P < 0.001).*

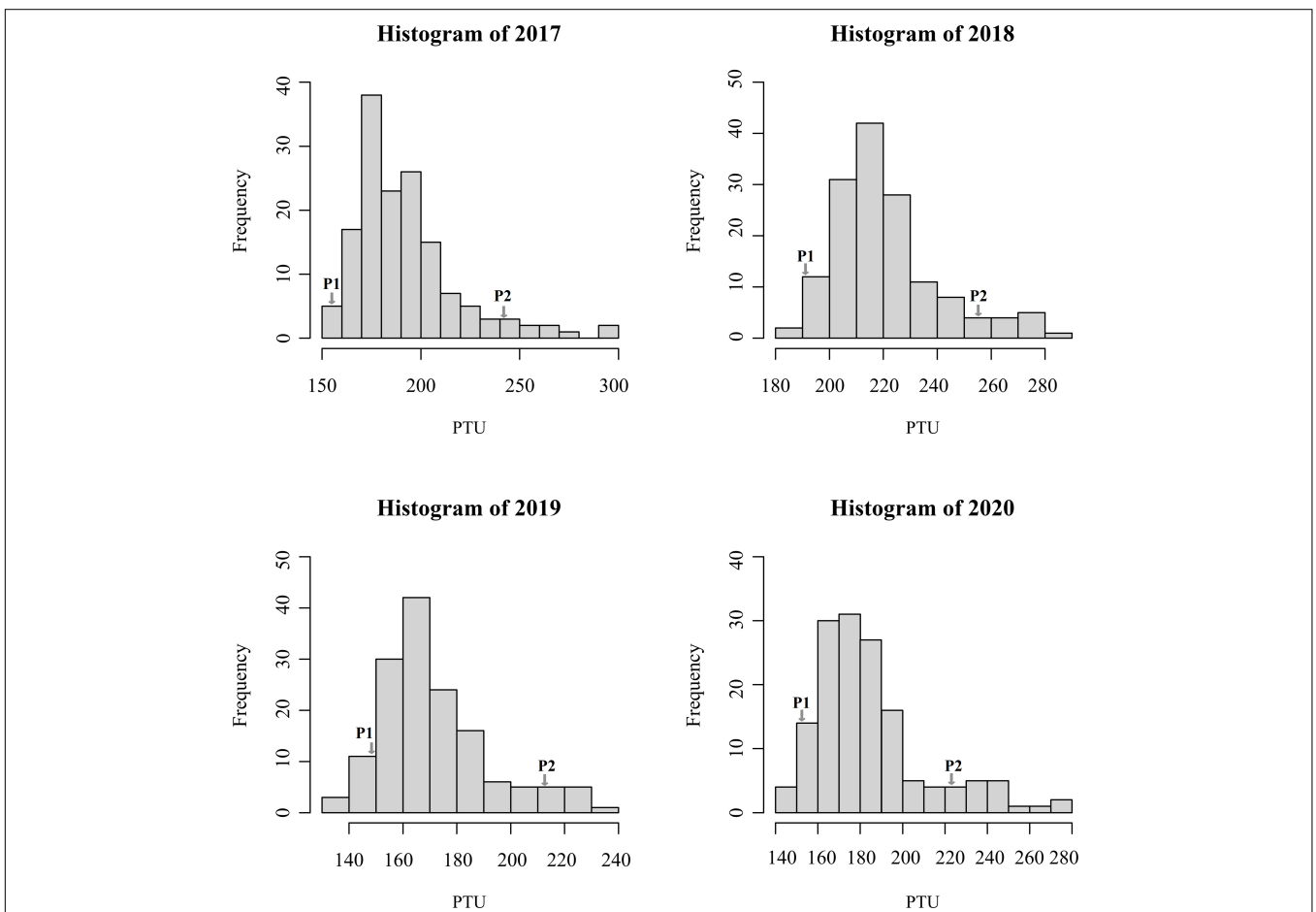


FIGURE 1 | Frequency distributions of photothermal unit values in the F_1 population at Langfang, Hebei Province from the year 2017 to 2020.

years from 2017 to 2020. Our previous work showed that P1 and P2 differed in yield and leaf size (He et al., 2019; Zhang et al., 2019). In addition, FT appeared to differ considerably (**Table 1**), suggesting that these two parents represent substantial variations among alfalfa varieties. Phenotypic variations in FT among the F_1 population ranged from 11.28 to 16.00% (**Table 1**), and transgressive segregation indicated the FT trait was quantitatively inherited. The frequency distribution of FT revealed a nearly normal distribution for FT trait in all 4 years (**Figure 1**). In our previous study, the same mapping population was used to identify QTLs associated with biomass yield (Zhang et al., 2019). A weak negative correlation was observed between FT and yield in the mapping population with correlation coefficients ranging from −0.05 to −0.47 (**Figure 2**), indicating that the genotypes with earlier flowering time (shorter vegetative growth time) tended to have lower yields. The result was similar to that of a previous study on alfalfa (Adhikari et al., 2019). The H^2 of FT ranged from 0.68 to 0.78 (**Table 1**). The relatively high heritability suggested that genotypic variance accounted for a high proportion of total phenotype variance for FT in this mapping population.

Genetic Linkage Map Construction

Each F_1 progeny was subjected to RAD-seq to detect genome-wide SNPs. A total of 161,169 raw SNPs were detected by aligning the sequences to the XinJiangDaYe reference genome (Chen H. et al., 2020). SNPs were filtered with mean read depth < 5, missing rate > 20%, heterozygosity ≥ 2:1. After filtering, 7,252 and 7,404 high-quality SDA markers were obtained for P1 and P2 parents, respectively; which were then used to construct high-density linkage maps. The total genetic lengths were 3780.49 cM for P1 and 4113.45 cM for P2 maps (**Supplementary Table 1**). The P1 linkage maps consisted of 6,844 SDA markers with an average genetic distance of 0.58 cM and a map length from 84.30 (chr3.3) to 145.22 cM (chr6.3) (**Supplementary Tables 1–3** and **Figure 3**). In P1, the lowest marker density was 0.85 on chr2.2 and the highest was 0.31 on chr6.2 (**Supplementary Table 1** and **Figure 3**). The P2 linkage maps contained 6,929 SDA markers with an average genetic distance of 0.59 cM, and a map length from 88.19 (chr7.2) to 160.14 cM (chr5.3) (**Supplementary Table 1** and **Figure 3**). Among the 32 linkage groups of P2, the lowest marker density was 0.89 on chr2.3 and the highest was 0.36 on chr6.2 (**Supplementary Table 1** and **Figure 3**).

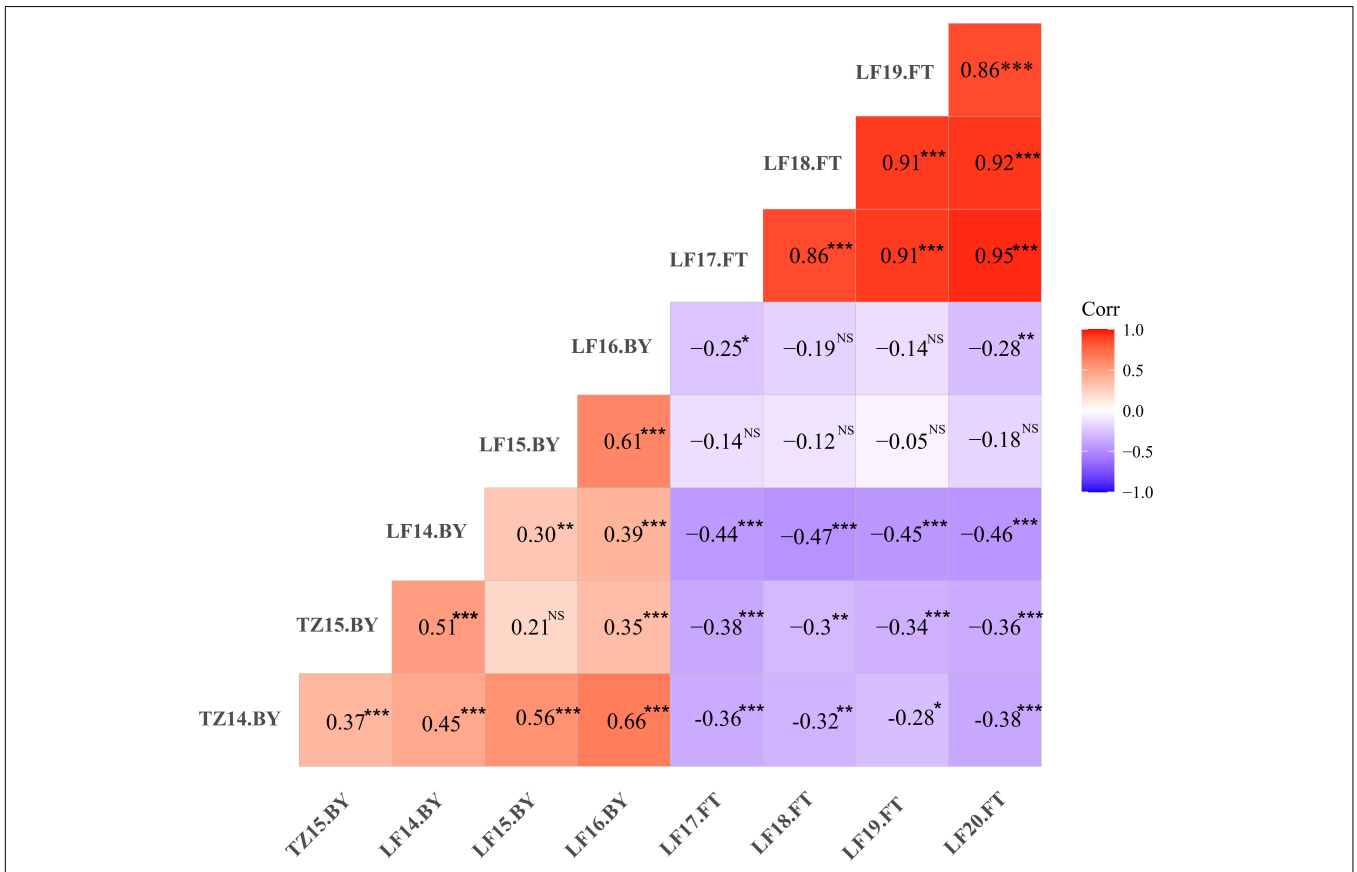


FIGURE 2 | Phenotypic correlations of flowering time (FT) and biomass yield (BY) among different environments in the F_1 population. Red and blue colors represent positive and negative correlations (*r*) between traits, respectively. TZ; Tongzhou; LF, Langfang, e.g., TZ14.BY indicates biomass yield data collected at Tongzhou in 2014. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, NS, non-significant. The results of yield were published in our previous study Zhang et al. (2019).

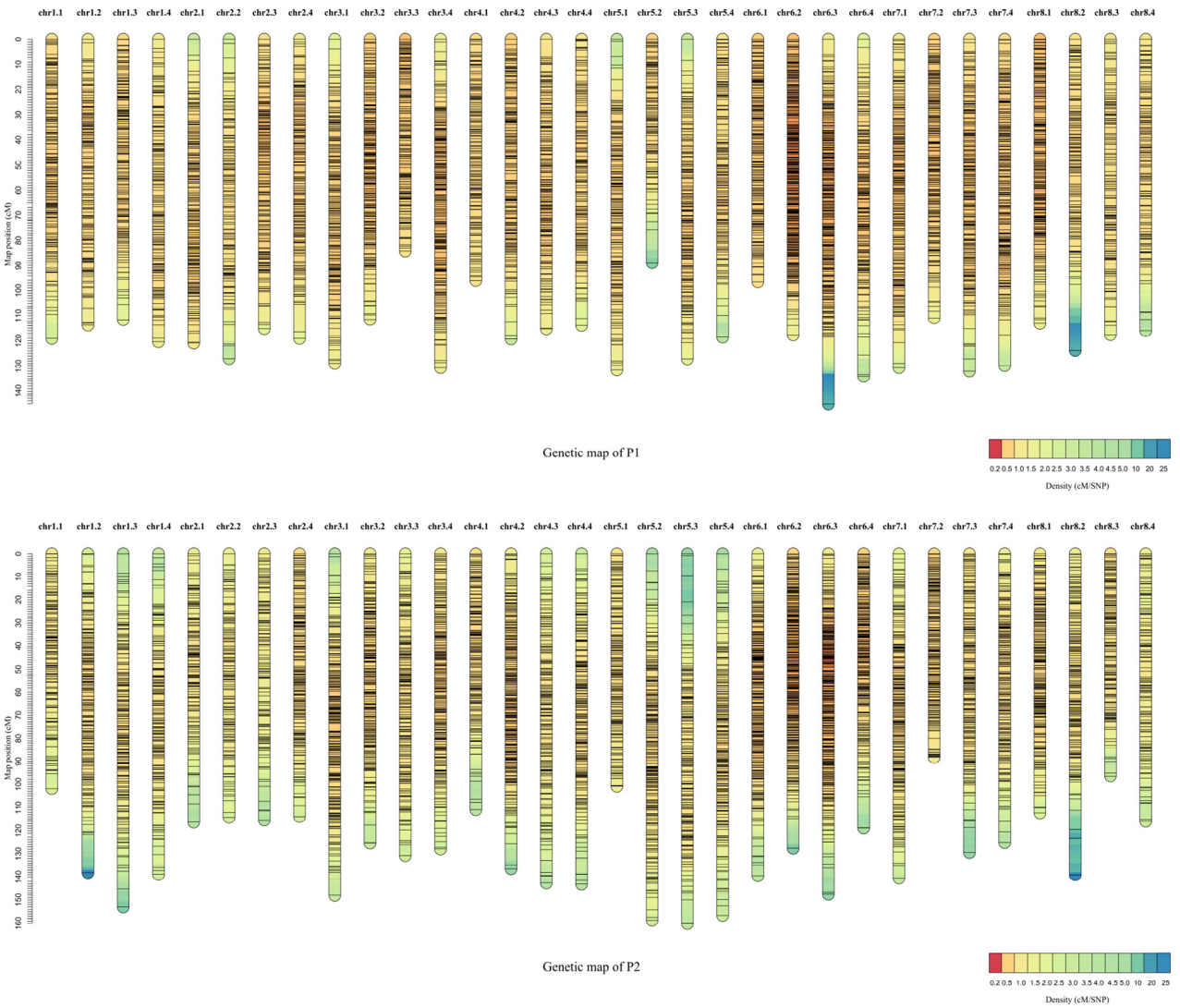


FIGURE 3 | Genetic linkage maps for the F_1 population. Distribution of 6,844 and 6,929 single-dose allele (SDA) markers in the parent (P1) and parent (P2) genetic maps, respectively. *X*-axis is the chromosome number and *Y*-axis describes the CentiMorgan distance. Detailed information on the maps is listed in **Supplementary Table 1**.

Quantitative Trait Loci Mapping for Flowering Time

Using the ICIM-ADD mapping method, the LS means of each year and BLUP values across 4 years were used for QTL mapping. A total of 16 QTLs for FT were detected for P1 (**Table 2** and **Supplementary Figure 1**). The percentage of phenotypic variance explained (PVE) by each QTL was from 4.13 to 16.04%, with the highest PVE observed on *qFT7.4-3*. Twenty-two QTLs for FT were detected for P2 parent with PVE values ranging from 1.40 to 13.78% and the highest PVE was observed on *qFT2.3-2* (**Table 2** and **Supplementary Figure 2**).

Five QTLs (*qFT4.1-1*, *qFT5.2-1*, *qFT6.3-1*, *qFT7.4-1*, and *qFT7.4-2*) in P1 parent and four QTLs (*qFT1.2-2*, *qFT2.3-2*, *qFT7.4-4,* and *qFT8.3*) in P2 parent were independently co-detected in multi-environments (**Table 2**). In P1, *qFT6.3-1* was identified in three grown environments (2017, 2019, and 2020) on chr6.3 at a position of 95.5–96.5 cM with an average PVE value of 12.03% (**Table 2**). Three QTLs (*qFT7.4-1*, *qFT7.4-2,* and *qFT7.4-4*) were identified on chr7.4 across multiple environments. Of them, *qFT7.4-4* on P2 was detected in all 4 years and BLUP values with an average LOD value of 5.85 and a PVE value of 8.87 (**Table 2**).

Novel Environment-Stable Quantitative Trait Locis Controlling Flowering Time

Four intervals were novel signals as they were detected in more than one environment and did not overlap with previously known FT-related QTLs. All of them were identified in the P1 parent, including *qFT4.1-1*, *qFT5.2-1*, *qFT6.3-1*, and *qFT7.4-1* (**Table 2**). The average PVE value of these four loci were

9.56, 6.32, 12.03, and 8.31%. Among them, *qFT6.3-1* was identified in three environments (2017, 2019, and 2020) with the highest average PVE value of 12.03% and an average LOD value of 8.64.

Consistent Quantitative Trait Locis Controlling Flowering Time

Four novel QTLs for FT such as *qFT4.1-1, qFT5.2-1, qFT6.3-1*, and *qFT7.4-1* were detected in this study, enriching our current knowledge of genetic control of FT in alfalfa. We also detected seven QTLs that overlapped with known QTLs. These QTLs, two from P1 and five from P2, may be considered as consistent QTLs (**Table 2**). Specifically, *qFT1.1-2* located at 45.2–68.9 Mb on chr1.1 was in the same region as *qflower-5* and *Tof-d1* (Adhikari et al., 2019; Zhang et al., 2020), while *qFT1.2-1* and *qFT1.2-2* fell into a known region of 55.0–77.5 Mb with *qflower-20*, *Tof-d6*, *Tof-d10,* and *Tof-d2* on chr1.2 (Adhikari et al., 2019; Zhang et al., 2020). On chr2.3, *qFT2.3-2* was overlapped with *qflower-6* (Zhang et al., 2020). The physical position of *qFT7.4-4* was relatively large (13.0 ∼ 51.2 Mb), and overlapped with *qFT7.4-2*, *Tof-d5*, *Tof-d4,* and *Tof-n6* (Adhikari et al., 2019). *qFT8.3* was located at 66.9–82.1 Mb on chr8.3, overlapped with *qflower-28* (Zhang et al., 2020). The results demonstrated that these QTLs could be robust signals for genetic control of FT in alfalfa. Other QTLs were detected in only one environment with a lower LOD value, suggesting that these QTLs may be environment-specific.

Integration of Differentially Expressed Genes With Quantitative Trait Locis

A larger number of significant DEGs were found at three flowering stages between two parents (P2 vs. P1). At BS, between P1 and P2, 3,329 and 1,710 DEGs were up- and downregulated, respectively (**Figure 4A**). At IFS, between P1 and P2, 5,075 and 1,983 DEGs were up- and down-regulated, respectively (**Figure 4A**). Between P1 and P2 at FFS, a total of 5,691 and 2,305 DEGs were up- and down-regulated, respectively (**Figure 4A**). These DEGs in three alfalfa flowering stages were then categorized into the biological process (BP), cellular component (CC), and molecular function (MF) categories using GO enrichment analysis (**Supplementary Table 5**). Specifically, in the BP terms, DEGs were mainly enriched in "cell morphogenesis involved in differentiation," "pollen tube growth," "developmental growth involved in morphogenesis," "anther morphogenesis," and "stamen morphogenesis." Also, these DEGs had a role in "response to low light intensity stimulus," "regulation of vernalization response," and "response to blue light" among others (**Supplementary Table 5** and **Figure 5**). The results suggested that these DEGs may play a role in the regulation of floral organ development and flowering time.

A comparison of the results from QTL mapping and RNA-seq analysis was employed to reveal candidate genes regulating the FT of alfalfa. A total of 230 common DEGs at three flowering stages were found in the QTL intervals (**Supplementary Table 6** and **Figure 4B**). Of which, 26 DEGs were annotated as candidate genes associated with flowering time in plants (**Figure 4C**). Interestingly, seven candidates were co-localized within consistent QTLs or novel environment-stable QTLs. Among them, six DEGs resided in three consistent QTL regions of *qFT1.1-2* (*MS.gene89579* and *MS.gene004686*), *qFT1.2-1* and *qFT1.2-2* (*MS.gene32142* and *MS.gene40752*), *qFT7.4-2* and *qFT7.4-4* (*MS.gene072377* and *MS.gene024098*). The last one, *MS.gene059196*, was found in a novel environment-stable QTL region of *qFT4.1-1*. The relative gene expression levels of the seven candidates were validated by qRT-PCR. The results showed that RNA-seq and qRT-PCR results were highly consistent and the seven candidate genes were differently expressed in two parents at three flowering stages (**Figure 6**).

DISCUSSION

Large phenotypic variations of FT were observed in the F_1 mapping population. The range of FT was comparable to a previous study identified in alfalfa population (Zhang et al., 2020). Alfalfa genotypes with late flowering showed an enhanced biomass yield (Adhikari et al., 2019). Our results also demonstrated that biomass yield was negatively correlated with FT in the F_1 mapping population. Large phenotypic variations and the high heritability of FT provide an important basis for analyzing QTL mapping in alfalfa.

Although FT-related QTLs have been reported in previous studies (Adhikari et al., 2019; Zhang et al., 2020), there were some limitations due to the mapping population and the lack of a reference genome. For example, FT-related QTLs were identified in a pseudo-testcross F_1 population derived from cultivars 3010 and CW 1010, but the two parents did not show significant differences in FT traits. Moreover, the flowering date was recorded as the day of the year (e.g., January 1st = 1) and was only recorded every 3 days (Adhikari et al., 2019). In addition, the genetic linkage maps were previously constructed by SNP markers generated through genotype−by−sequencing, not RAD-seq (Adhikari et al., 2019; Zhang et al., 2020). This could result in some differences in marker density and quality for creating genetic linkage maps. In this study, the two parental materials greatly varied for FT, and genome-wide SDA markers were detected with XinJiangDaYe reference genome. Therefore, our results on the constructed genetic maps and identified QTLs would add additional values to a better understanding of the genetic control of FT in alfalfa.

High-Density Genetic Linkage Maps

The creation of a high-density and quality linkage map is an important prerequisite for QTL mapping and map-based cloning of genes. Alfalfa is autotetraploid and highly heterozygous, which has impeded the construction of a high-density linkage map (Li et al., 2014). An early study reported that a tetraploid alfalfa linkage map had seven linkage groups with 443 cM because of low marker density (Brouwer and Osborn, 1999). Powerful new sequencing techniques can generate a large number of SNP markers for linkage map construction (Andolfatto et al., 2011; Yang et al., 2012). Li et al. (2014) first used SDA markers

TABLE 2 | QTLs detected for least-square means (LS means) and the best linear unbiased prediction (BLUP) values of flowering time (FT) using ICIM-ADD in the F1 population.

Parent	QTL	Year	Chr.	Location (cM)	Left marker	Right marker	LOD	PVE (%)	Add
	qFT1.2-1	2020	1.2	81.5∼82.5	chr1.2__77498697	chr1.2__55012028	4.06	6.33	15.34
	qFT1.4	2017	1.4	33.5∼34.5	chr1.4__9568849	chr1.4__8751633	3.59	4.56	−12.30
	qFT2.3-1	BLUP	2.3	78.5∼79.5	chr2.3__58327605	chr2.3__51410345	4.84	6.10	−9.99
	qFT3.3-1	BLUP	3.3	50.5∼51.5	chr3.3__62421973	chr3.3__69012912	3.83	4.13	−8.38
	qFT4.1-1	2017, BLUP	4.1	63.5∼64.5	chr4.1__64642892	chr4.1__77321616	4.06, 5.82	6.68, 12.43	−15.04, −14.43
	qFT4.1-2	2017	4.1	92.5∼93.5	chr4.1__75331077	chr4.1__75711577	4.43	5.43	−13.60
	qFT4.2	BLUP	4.2	9.5∼10.5	chr4.2__34610643	chr4.2__11026558	3.06	4.46	8.57
P1	***qFT5.2-1***	2017, 2020, BLUP	5.2	4.5∼5.5	chr5.2__45638433	chr5.2__44056158	6.71, 5.48, 5.32	6.99, 5.83, 6.13	−15.25, −14.74, −10.04
	qFT5.2-2	2018	5.2	34.5∼35.5	chr5.2__57773538	chr5.2__56102093	4.83	6.23	11.26
	qFT6.1	BLUP	6.1	76.5∼77.5	chr6.1__44126752	chr6.1__76961406	5.73	8.14	11.71
	qFT6.3-1	2017, 2019, 2020	6.3	95.5∼96.5	chr6.3__52410664	chr6.3__61869379	12.03, 8.97, 4.91	14.37, 15.39, 6.32	21.86, 17.05, 15.31
	qFT7.3-1	2017	7.3	11.5∼13.5	chr7.3__8960937	chr7.3__12357404	3.28	4.28	11.91
	qFT7.3-2	2019	7.3	64.5∼65.5	chr7.3__63489699	chr7.3__50361610	4.51	7.2	11.65
	qFT7.4-1	2017, 2020, BLUP	7.4	62.5∼63.5	chr7.4__62411379	chr7.4__61234796	4.91, 3.82, 5.85	7.63, 6.52, 10.78	16.06, 15.64, 13.41
	qFT7.4-2	2018, 2020	7.4	17.5∼18.5	chr7.4__35788679	chr7.4__40755488	5.01, 8.18	8.02, 11.09	−12.91, −20.55
	qFT7.4-3	2018	7.4	93.5∼94.5	chr7.4__87919782	chr7.4__85044135	7.27	16.04	−18.10
	qFT1.1-1	2017	1.1	68.5∼69.5	chr1.1__75472691	chr1.1__70841814	3.22	3.54	−10.00
	qFT1.1-2	2019	1.1	50.5∼51.5	chr1.1__45209340	chr1.1__56558159	4.40	2.25	6.80
	qFT1.1-3	BLUP	1.1	9.5∼10.5	chr1.1__22559325	chr1.1__16294404	4.22	6.54	11.18
	qFT1.2-2	2017, 2019	1.2	91.5∼92.5	chr1.2__71070950	chr1.2__55221961	5.18, 18.81	5.75, 9.08	−13.21, −13.37
	qFT1.2-3	2020	1.2	104.5∼105.5	chr1.2__71071980	chr1.2__81663663	3.98	9.79	−17.22

(Continued)

TABLE 2 | (Continued)

Parent	QTL	Year	Chr.	Location (cM)	Left marker	Right marker	LOD	PVE (%)	Add
	qFT1.3	2019	1.3	101.5∼102.5	chr1.3__64152948	chr1.3__65763782	11.64	9.47	−13.28
	qFT2.3-2	2017, 2019, 2020, BLUP	2.3	78.5∼79.5	chr2.3__59083503	chr2.3__59862317	8.66, 16.26, 3.89, 6.46	13.56, 9.44, 9.94, 13.78	−20.02, −13.42, −17.72, −16.07
P2	*qFT2.3-3*	2018	2.3	64.5∼65.5	chr2.3__46634301	chr2.3__56553524	4.12	7.45	−13.13
	qFT3.2-1	2019	3.2	23.5∼24.5	chr3.2__5624964	chr3.2__29030021	4.36	4.80	9.46
	qFT3.2-2	2019	3.2	104.5∼105.5	chr3.2__90944356	chr3.2__78402762	8.63	3.8	−8.79
	qFT3.2-3	BLUP	3.2	53.5∼54.5	chr3.2__41429169	chr3.2__37443429	3.64	8.61	−12.56
	qFT3.3-2	2019	3.3	15.5∼16.5	chr3.3__13703202	chr3.3__12599738	4.46	2.48	−6.80
	qFT4.1-3	2019	4.1	64.5∼65.5	chr4.1__63974266	chr4.1__80043069	11.04	5.04	−9.83
	qFT4.3	2017	4.3	97.5∼98.5	chr4.3__49979397	chr4.3__58477555	3.66	4.66	−11.46
	qFT5.1	2019	5.1	9.5∼10.5	chr5.1__20167931	chr5.1__18706007	5.87	2.97	7.53
	qFT5.2-3	2017	5.2	142.5∼144.5	chr5.2__82957406	chr5.2__82957429	6.96	7.71	14.73
	qFT5.2-4	2019	5.2	141.5∼142.5	chr5.2__80311479	chr5.2__82957410	6.10	2.66	6.99
	qFT6.3-2	2019	6.3	28.5∼29.5	chr6.3__5222282	chr6.3__5981497	12.64	5.72	−10.45
	qFT6.3-3	2019	6.3	33.5∼34.5	chr6.3__5307300	chr6.3__6965213	3.81	2.45	6.71
	qFT7.3-3	2019	7.3	59.5∼60.5	chr7.3__38899574	chr7.3__63369976	5.59	6.83	−11.60
	qFT7.4-4	2017, 2018, 2019, 2020, BLUP	7.4	39.5∼40.5	chr7.4__13015502	chr7.4__51199977	5.05, 5.32, 11.26, 3.45, 4.15	7.42, 10.99, 6.97, 9.37, 9.61	14.46, 15.74, 11.35, 16.80, 13.10
	qFT8.3	2017, 2019, 2020, BLUP	8.3	87.5∼88.5	chr8.3__75627434	chr8.3__81295867	6.89, 3.68, 3.48, 3.62	7.83, 1.40, 6.41, 5.76	−14.93, −5.10, −13.94, −10.17

Sixteen and Twenty-one QTLs for FT were mapped on P1 and P2 linkage maps, respectively. Consistent QTLs were underlined; Novel QTLs were bold. Left marker, the marker on the left of the LOD peak; right marker, the marker on the right of the LOD peak; LG, linkage group; interval (cM), 1-LOD support interval; LOD, the logarithm of the odds; PVE, the percentage of the phenotypic variation explained by QTL; Add, the additive effects of the QTL.

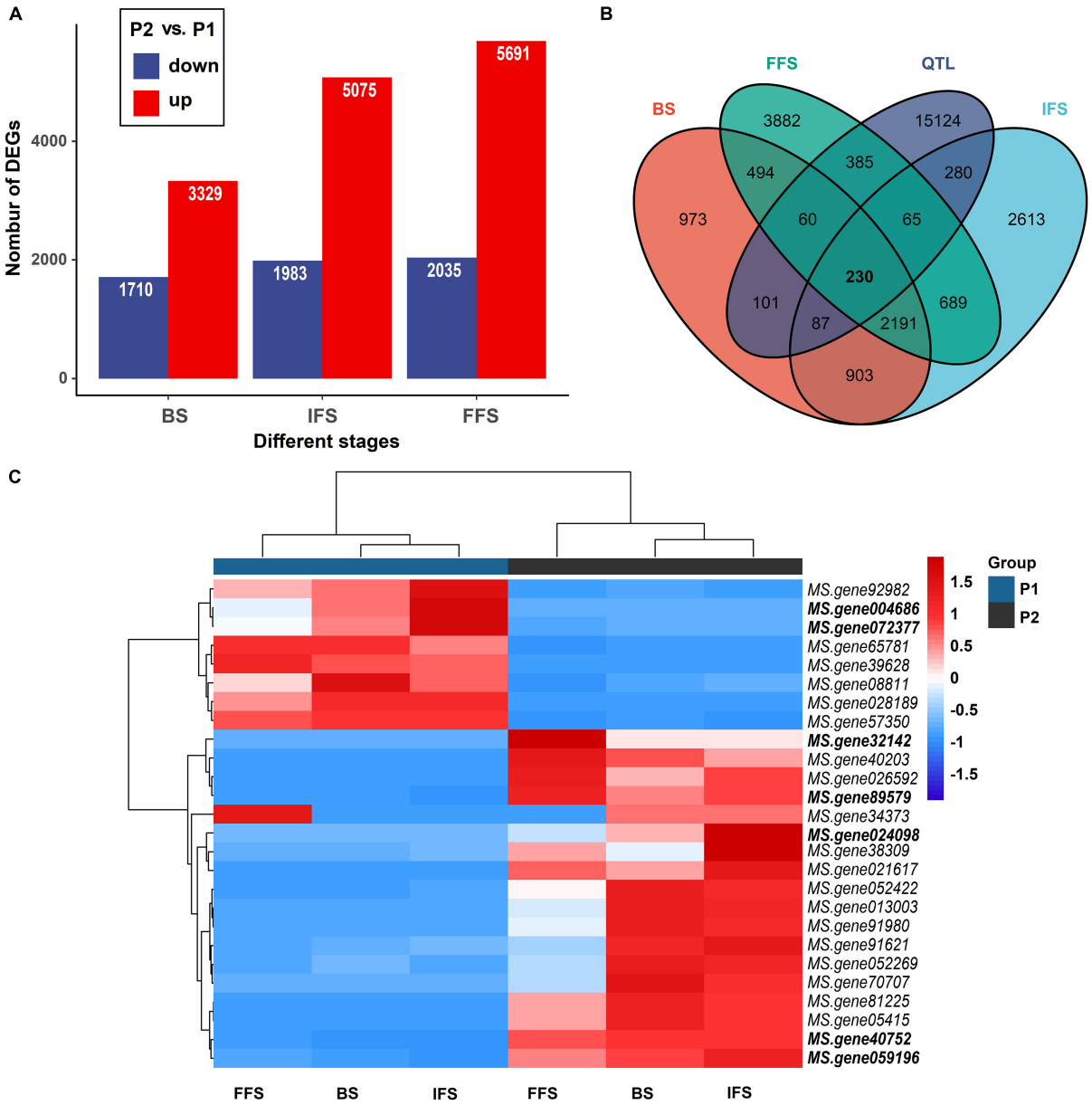


FIGURE 4 | Differentially expressed genes (DEGs) within QTL intervals. **(A)** Number of downregulated genes (blue) and upregulated genes (red) between two parents at three flowering stages. BS, budding stage, IFS, initial flowering stage and FFS, full flowering stage. **(B)** The number of DEGs within QTL intervals. **(C)** Heatmap clustering the DEGs associated with flowering time by their expression abundance. DEGs within the consistent QTL intervals or new environment-stable QTLs were bold.

to construct high-density linkage maps for two alfalfa parents to identify QTLs associated with important agronomic traits. Their genetic maps with 64 linkage groups contained 3,591 SDA markers, with an average density of one marker per 1.5 and 1.0 cM for the maternal and paternal parents, respectively (Li et al., 2014). However, the deficiency of genomic information has been the major limitation for genetic map construction. The first available alfalfa genome sequence makes it possible to construct high-density genetic maps (Chen H. et al., 2020). In this study, the maps contained 13,773 SDA markers on 64 linkage groups across both parents, with an average marker distance of 0.58 and 0.59 cM for the P1 and P2 maps, respectively. These maps will be useful for future QTL mapping studies and the application of marker-assisted selection (MAS).

Population size and quality of linkage map greatly influence the power of QTL detection. Adhikari et al. (2019) performed QTL mapping for the flowering time by using a population of 181 F1 progenies in four environments. The heritability of flowering phenotypes in three of these environments (0.69, 0.74, and 0.75) was similar to our study, while only three QTLs individually

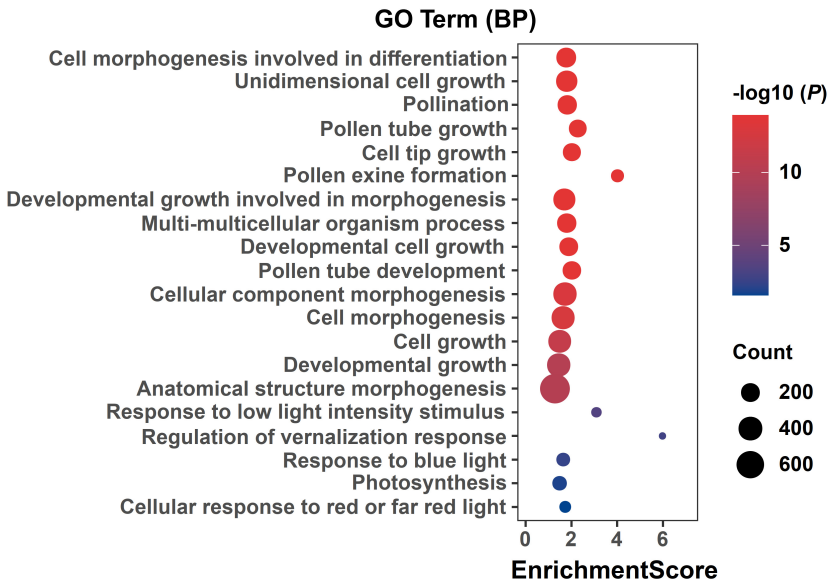


FIGURE 5 | GO enrichment of DEGs in three alfalfa flowering stages. The ordinate represents different GO terms (biological progress) and the abscissa represents enrichment Score. Circle size represents the gene number while circle color represents the value of −log10 (*P*). The *P*-value was corrected by Benjamini and Hochberg (BH) method.

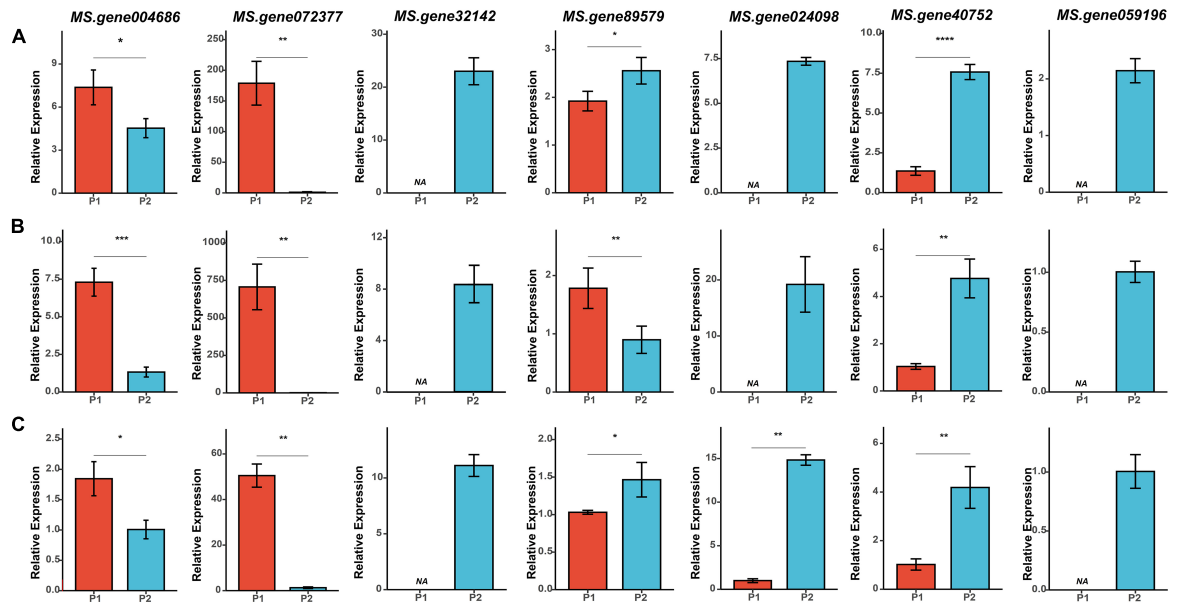


FIGURE 6 | qRT-PCR analysis of the seven candidate genes at three flowering stages in two parents. **(A–C)** Relative expression of the seven candidate genes at BS (budding stage), IFS (initial flowering stage), and FFS (full flowering stage), respectively. Asterisk symbols indicate significant differences between two parents (*t*-test, $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$, $^{****}P < 0.0001$).

co-localized in two environments. In a population of 392 F1 progenies, Zhang et al. (2020) identified three environment-stable QTLs, with one found in three environments and two found in two environments. In this study, we identified eight environment-stable QTLs, suggesting that our population size was big enough for QTL identification. Moreover, the high-density linkage map constructed in this study had an increased power of QTL detection compared with maps developed in previous studies (Adhikari et al., 2019; Zhang et al., 2020). Nonetheless, in our study, a major limitation remains due to the only use of SDA markers. Many polymorphic SNP markers not classified as SDA were not used for the genetic map construction, resulting in the loss of genetic information and the reduced power of QTL detection (Li et al., 2014). As a result, despite the high

heritability of flowering time, only one environment-stable QTL was discovered across all four environments in our study.

Analysis of Candidate Genes for Flowering Time

Previous studies into FT mainly focused on QTL mapping in alfalfa (in genomics level) (Adhikari et al., 2019; Zhang et al., 2020). Although integration of QTL mapping and RNA-seq has been applied to identify candidate genes in rice, maize (*Zea mays* L.), wheat (*Triticum aestivum* L.), and other crops (Kuang et al., 2020; Lei et al., 2020; Han et al., 2022), but such approaches have not been used for discovering alfalfa FT genes. In this study, flower samples at three flowering stages from two parents of the F_1 mapping population were characterized using the RNA-seq technology, and the identified DEGs were combined with QTLs to narrow the candidates. Finally, based on DEG annotations within QTL intervals, seven potential genes associated with flowering time were detected.

Two candidates, *MS.gene89579* and *MS.gene004686*, located within qFT1.1-2. *MS.gene89579*, was annotated as a casein kinase I-like protein, which was previously reported to function in regulating flowering time through gibberellin signaling (Hori et al., 2013; Kang et al., 2020). Another one, *MS.gene004686*, was annotated as a basic helix-loop-helix (BHLH) transcription factor. Its homologous gene in Arabidopsis has been reported to control flowering time in response to blue light (Liu et al., 2013). In the interval of *qFT1.2-1* and *qFT1.2-2*, two candidates (*MS.gene32142* and *MS.gene40752*) associated with flowering time were annotated. *MS.gene32142* homolog was ethylene receptor ETR2, which can delay flowering and cause starch accumulation in stems (Wuriyanghan et al., 2009). *MS.gene40752* homolog was glycogen synthase kinase 3, which played an important role in preventing precocious flowering through phosphorylating CO (Chen Y. et al., 2020). In the physical interval of two consistent QTLs (*qFT7.4-2* and *qFT7.4-4*), *MS.gene072377* and *MS.gene024098* were identified as candidate genes. *MS.gene072377*, annotated as leucine-rich repeat protein FLOR 1, promoted flowering transition in long days (Torti et al., 2012). *MS.gene024098* was a member of UDP-glucosyltransferase family, and its homologous gene can regulate flowering time via *flowering locus C* (Wang et al., 2012). In the physical interval of qFT4.1-1, *MS.gene059196* was annotated as a pleiotropic regulatory locus. Its mutants (prl1) caused an early flowering phenotype in Arabidopsis (Weihmann et al., 2012).

In plants, the *FLOWERING LOCUS T* (FT) family of genes is known to be associated with photoperiod sensing, which is involved in the initiation of flowering. By functional annotation, two homologues of *FT*, *MS.gene51911* and *MS.gene87044*, were identified in a consistent QTL interval (*qFT7.4-2* and *qFT7.4-4*). They were annotated as *FTc* and *FTb*, respectively. Interestingly, *MsFTb1* could partially complement the phenotype of late flowering mutants (*ft-10*), while *MsFTc* had no effect on the flowering time in alfalfa (Lorenzo et al., 2020). The above-mentioned genes in the QTL region were not among the DEGs detected by RNA-Seq analysis, potentially due to the specific expression of FT genes in leaves (Huang et al., 2005).

CONCLUSION

Plant flowering is regulated by genetic and environmental signals. The work presented here revealed the genetic mechanisms controlling flowering time in the alfalfa F_1 population through QTL mapping. High-density genetic maps were created and 38 QTLs for FT were detected, including seven consistent QTLs and four novel environment-stable QTLs. Seven candidate genes involved in flowering regulation were annotated by QTL mapping and DEGs analysis. The functions of these candidate genes could be further verified in alfalfa. The loci generated by the current QTL mapping study are valuable resources for enhancing breeding programs aimed at improving flowering time and yield in alfalfa.

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/bioproject/PRJNA817856.

AUTHOR CONTRIBUTIONS

QY, JK, and ZW conceived, designed the experiments, and developed the mapping population. XJ, FH, XY, and CY collected phenotypic data. XJ, FZ, TY, and TG performed data analysis. XJ and ZW wrote the manuscript. YJ and JK revised and finalized the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (32071868), the earmarked fund for the China Agriculture Research System (CARS-35-04), the Key Research Project of Ningxia Province for Alfalfa Breeding Program (2019NYY203), and the Agricultural Science and Technology Innovation Program (ASTIP-IAS14).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.899681/full#supplementary-material

Supplementary Figure 1 | Distribution of QTLs for flowering time (FT) using LS means of each environment and BLUP values in P1 parent.

Supplementary Figure 2 | Distribution of QTLs for flowering time (FT) detected using LS means of each environment and BLUP values in P2 parent.

Supplementary Table 5 | GO enrichment analysis of the DEGs.

REFERENCES

Adhikari, L., Makaju, S. O., and Missaoui, A. M. (2019). QTL mapping of flowering time and biomass yield in tetraploid alfalfa (Medicago sativa L.). *BMC Plant Biol.* 19:359.

Andolfatto, P., Davison, D., Erezyilmaz, D., Hu, T. T., Mast, J., Sunayama-Morita, T., et al. (2011). Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res.* 21, 610–617. doi: 10.1101/gr.115402.110

Bai, Z., Ma, W., Ma, L., Velthof, G. L., Wei, Z., Havlík, P., et al. (2018). China's livestock transition: Driving forces, impacts, and consequences. *Sci. Adv.* 4:eaar8534. doi: 10.1126/sciadv.aar8534

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv* 1406, 133–199. doi: 10.1007/0-387-22747-4_4

Bolger, A. M., Marc, L., and Bjoern, U. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Brouwer, D. J., and Osborn, T. C. (1999). A molecular marker linkage map of tetraploid alfalfa (Medicago sativa L.). *Theor. Appl. Genet.* 99, 1194–1200. doi: 10.1007/s001220051324

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

Chen, H., Zeng, Y., Yang, Y., Huang, L., Tang, B., Zhang, H., et al. (2020). Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. *Nat. Commun.* 11:2494. doi: 10.1038/s41467-020-16338-x

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

Chen, Y., Song, S., Gan, Y., Jiang, L., Yu, H., and Shen, L. (2020). SHAGGY-like kinase 12 regulates flowering through mediating CONSTANS stability in Arabidopsis. *Sci. Adv.* 6:eaaw0413. doi: 10.1126/sciadv.aaw0413

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

Derakhshani, B., Jafary, H., Maleki Zanjani, B., Hasanpur, K., Mishina, K., Tanaka, T., et al. (2020). Combined QTL mapping and RNA-Seq profiling reveals candidate genes associated with cadmium tolerance in barley. *PLoS One* 15:e0230820. doi: 10.1371/journal.pone.0230820

Diaz, A., Zikhali, M., Turner, A. S., Isaac, P., and Laurie, D. A. (2012). Copy number variation affecting the Photoperiod-B1 and Vernalization-A1 genes is associated with altered flowering time in wheat (Triticum aestivum). *PLoS One* 7:e33234. doi: 10.1371/journal.pone.0033234

Grabowski, P. P., Evans, J., Daum, C., Deshpande, S., Barry, K. W., Kennedy, M., et al. (2017). Genome-wide associations with flowering time in switchgrass using exome-capture sequencing data. *New Phytol.* 213, 154–169. doi: 10.1111/nph.14101

Han, Q., Zhu, Q., Shen, Y., Lee, M., Lübberstedt, T., and Zhao, G. (2022). QTL mapping low-temperature germination ability in the maize IBM Syn10 DH Population. *Plants* 11:214. doi: 10.3390/plants11020214

He, F., Kang, J., Zhang, F., Long, R., Yu, L.-X., Wang, Z., et al. (2019). Genetic mapping of leaf-related traits in autotetraploid alfalfa (Medicago sativa L.). *Mol. Breed.* 39:147.

Hecht, V., Laurie, R. E., VanderSchoor, J. K., Ridge, S., Knowles, C. L., Liew, L. C., et al. (2011). The Pea GIGAS Gene is a flowering locus t homolog necessary for graft-transmissible specification of flowering but not for responsiveness to photoperiod. *Plant Cell* 23, 147–161. doi: 10.1105/tpc.110.081042

Hori, K., Matsubara, K., and Yano, M. (2016). Genetic control of flowering time in rice: integration of Mendelian genetics and genomics. *Theor. Appl. Genet.* 129, 2241–2252. doi: 10.1007/s00122-016-2773-4

Hori, K., Ogiso-Tanaka, E., Matsubara, K., Yamanouchi, U., Ebana, K., and Yano, M. (2013). Hd16, a gene for casein kinase I, is involved in the control of rice flowering time by modulating the day-length response. *Plant J.* 76, 36–46. doi: 10.1111/tpj.12268

Huang, L., Yang, Y., Zhang, F., and Cao, J. (2017). A genome-wide SNP-based genetic map and QTL mapping for agronomic traits in Chinese cabbage. *Sci. Rep.* 7:46305. doi: 10.1038/srep46305

Huang, T., Böhlenius, H., Eriksson, S., Parcy, F., and Nilsson, O. J. S. (2005). The mRNA of the Arabidopsis gene FT moves from leaf to shoot apex and induces flowering. *Science* 309, 1694–1696. doi: 10.1126/science.1117768

Kang, J., Cui, H., Jia, S., Liu, W., Yu, R., Wu, Z., et al. (2020). Arabidopsis thaliana MLK3, a plant-specific casein kinase 1, negatively regulates flowering and phosphorylates histone H3 in Vitro. *Genes* 11:345. doi: 10.3390/genes11030345

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

Kinoshita, A., and Richter, R. (2020). Genetic and molecular basis of floral induction in Arabidopsis thaliana. *J. Exp. Bot.* 71, 2490–2504. doi: 10.1093/jxb/eraa057

Koornneef, M., Alonso-Blanco, C., Peeters, A. J., and Soppe, W. (1998). Genetic Control of Flowering Time in Arabidopsis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 345–370.

Kooyers, N. J. (2015). The evolution of drought escape and avoidance in natural herbaceous populations. *Plant Sci.* 234, 155–162. doi: 10.1016/j.plantsci.2015.02.012

Kuang, C. H., Zhao, X. F., Yang, K., Zhang, Z. P., Ding, L., Pu, Z. E., et al. (2020). Mapping and characterization of major QTL for spike traits in common wheat. *Physiol. Mol. Biol. Plants* 26, 1295–1307. doi: 10.1007/s12298-020-00823-0

Laurie, R. E., Diwadkar, P., Jaudal, M., Zhang, L., Hecht, V. R., Wen, J., et al. (2011). The Medicago FLOWERING LOCUS T Homolog, MtFTa1, Is aKey RegulatorofFlowering Time. *Plant Physiol.* 156, 2207–2224. doi: 10.1104/pp.111.180182

Lee, B.-Y., Kim, M.-S., Choi, B.-S., Nagano, A. J., Wu, R. S. S., Takehana, Y., et al. (2019). Construction of High-Resolution RAD-Seq based linkage map, anchoring reference genome, and QTL mapping of the sex chromosome in the marine medaka oryzias melastigma. *G3* 9, 3537–3545. doi: 10.1534/g3.119.400708

Lei, L., Zheng, H., Bi, Y., Yang, L., Liu, H., Wang, J., et al. (2020). Identification of a major qtl and candidate gene analysis of salt tolerance at the bud burst stage in rice (Oryza sativa L.) using QTL-Seq and RNA-Seq. *Rice* 13, 1–14. doi: 10.1186/s12284-020-00416-1

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 1303:3997.

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

Li, X., Wei, Y., Acharya, A., Jiang, Q., Kang, J., and Brummer, E. C. (2014). A Saturated Genetic linkage map of autotetraploid alfalfa (Medicago sativa L.) developed using genotyping-by-sequencing is highly syntenous with the medicago truncatula genome. *G3 Genesgenet.* 4, 1971–1979. doi: 10.1534/g3.114.012245

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, Y., Li, X., Li, K., Liu, H., Lin, C., and Quail, P. H. (2013). Multiple bHLH proteins form heterodimers to mediate CRY2-dependent regulation of flowering-time in arabidopsis. *PLoS Genet.* 9:e1003861. doi: 10.1371/journal.pgen.1003861

Lorenzo, C. D., Garcia-Gagliardi, P., Antonietti, M. S., Sanchez-Lamas, M., Mancini, E., Dezar, C. A., et al. (2020). Improvement of alfalfa forage quality and management through the down-regulation of MsFTa1. *Plant Biotechnol. J.* 18, 944–954. doi: 10.1111/pbi.13258

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, 1–21. doi: 10.1186/s13059-014-0550-8

Lu, S., Dong, L., Fang, C., Liu, S., Kong, L., Cheng, Q., et al. (2020). Stepwise selection on homeologous PRR genes controlling flowering and maturity during soybean domestication. *Nat. Genet.* 52, 428–436. doi: 10.1038/s41588-020-0604-7

Major, D. J., Hanna, M. R., and Beasley, B. W. (1991). Photoperiod response characteristics of alfalfa (Medicago sativa L.) cultivars. *Can. J. Plant Sci.* 71, 87–93. doi: 10.4141/cjps91-010

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for

analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

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

Radovic, J., Sokolovic, D., and Markovic. (2009). Alfalfa-most important perennial forage legume in animal husbandry. *Biotechnol. Anim. Husb.* 25, 465–475. doi: 10.2298/bah0906465r

Schiffers, K., Bourne, E. C., Lavergne, S., Thuiller, W., and Travis, J. M. (2013). Limited evolutionary rescue of locally adapted populations facing climate change. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 368:20120083. doi: 10.1098/rstb.2012.0083

Song, Y. H., Ito, S., and Imaizumi, T. (2013). Flowering time regulation: photoperiod- and temperature-sensing in leaves. *Trends Plant Sci.* 18, 575–583. doi: 10.1016/j.tplants.2013.05.003

Srikanth, A., and Schmid, M. (2011). Regulation of flowering time: all roads lead to Rome. *Cell. Mol. Life Sci.* 68, 2013–2037. doi: 10.1007/s00018-011-0673-y

Stam, P. (1993). Construction of integrated genetic linkage maps by means of a new computer package: join Map. *Plant J.* 3, 739–744. doi: 10.1111/j.1365-313x.1993.00739.x

Torti, S., Fornara, F., Vincent, C., Andrés, F., Nordström, K., Göbel, U., et al. (2012). Analysis of the Arabidopsis shoot meristem transcriptome during floral transition identifies distinct regulatory patterns and a leucine-rich repeat protein that promotes flowering. *Plant Cell* 24, 444–462. doi: 10.1105/tpc.111.092791

Upadhyaya, H. D., Bajaj, D., Das, S., Saxena, M. S., Badoni, S., Kumar, V., et al. (2015). A genome-scale integrated approach aids in genetic dissection of complex flowering time trait in chickpea. *Plant Mol. Biol.* 89, 403–420. doi: 10.1007/s11103-015-0377-z

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

Wang, B., Jin, S. H., Hu, H. Q., Sun, Y. G., Wang, Y. W., Han, P., et al. (2012). UGT87A2, an Arabidopsis glycosyltransferase, regulates flowering time via FLOWERING LOCUS C. *New Phytol.* 194, 666–675. doi: 10.1111/j.1469-8137.2012.04107.x

Weihmann, T., Palma, K., Nitta, Y., and Li, X. (2012). Pleiotropic regulatory locus 2 exhibits unequal genetic redundancy with its homolog PRL1. *Plant Cell Physiol.* 53, 1617–1626. doi: 10.1093/pcp/pcs103

Weller, J. L., Vander Schoor, J. K., Perez-Wright, E. C., Hecht, V., Gonzalez, A. M., Capel, C., et al. (2019). Parallel origins of photoperiod adaptation following dual domestications of common bean. *J. Exp. Bot.* 70, 1209–1219. doi: 10.1093/jxb/ery455

Wen, J., Jiang, F., Weng, Y., Sun, M., Shi, X., Zhou, Y., et al. (2019). Identification of heat-tolerance QTLs and high-temperature stress-responsive genes through conventional QTL mapping. QTL-seq and RNA-seq in tomato. *BMC Plant Biol.* 19:398. doi: 10.1186/s12870-019-2008-3

Wong, A., Hecht, V. F., Picard, K., Diwadkar, P., Laurie, R. E., Wen, J., et al. (2014). Isolation and functional analysis of CONSTANS-LIKE genes suggests that a central role for CONSTANS in flowering time control is not evolutionarily conserved in Medicago truncatula. *Front. Plant Sci.* 5:486. doi: 10.3389/fpls.2014.00486

Wuriyanghan, H., Zhang, B., Cao, W. H., Ma, B., Lei, G., Liu, Y. F., et al. (2009). The ethylene receptor ETR2 delays floral transition and affects starch accumulation in rice. *Plant Cell* 21, 1473–1494. doi: 10.1105/tpc.108.065391

Yang, H., Tao, Y., Zheng, Z., Li, C., Sweetingham, M. W., and Howieson, J. G. (2012). Application of next-generation sequencing for rapid marker development in molecular plant breeding: a case study on anthracnose disease resistance in Lupinus angustifolius L. *BMC Genom.* 13:318. doi: 10.1186/1471-2164-13-318

Ye, J., Yang, Y., Chen, B., Shi, J., Luo, M., Zhan, J., et al. (2017). An integrated analysis of QTL mapping and RNA sequencing provides further insights and promising candidates for pod number variation in rapeseed (Brassica napus L.). *BMC Genom.* 18:71. doi: 10.1186/s12864-016-3402-y

Zhang, F., Kang, J., Long, R., Yu, L. X., Sun, Y., Wang, Z., et al. (2020). Construction of high-density genetic linkage map and mapping quantitative trait loci (QTL) for flowering time in autotetraploid alfalfa (Medicago sativa L.) using genotyping by sequencing. *Plant Genome* 13:45. doi: 10.1002/tpg2.20045

Zhang, F., Kang, J., Long, R., Yu, L. X., Wang, Z., Zhao, Z., et al. (2019). High-density linkage map construction and mapping QTL for yield and yield components in autotetraploid alfalfa using RAD-seq. *BMC Plant Biol.* 19:165. doi: 10.1186/s12870-019-1770-6

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 09 June 2022
doi: 10.3389/fpls.2022.917338

Full-Length Transcriptomics Reveals Complex Molecular Mechanism of Salt Tolerance in *Bromus inermis* L.

Qian Li[1,2,3†], *Jiaxing Song*[4†], *Yi Zhou*[5], *Yingxia Chen*[1,2], *Lei Zhang*[1,2], *Yongzhen Pang*[3*] and *Bo Zhang*[1,2*]

[1] Key Laboratory of Grassland Resources and Ecology of Western Arid Region, Ministry of Education, College of Grassland Science, Xinjiang Agricultural University, Urumqi, China, [2] Key Laboratory of Grassland Resources and Ecology of Xinjiang, College of Grassland Science, Xinjiang Agricultural University, Urumqi, China, [3] Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China, [4] College of Grassland Agriculture, Northwest A&F University, Yangling, China, [5] School of Agriculture Food and Wine, The University of Adelaide, Urrbrae, SA, Australia

Bromus inermis L. (commonly known as smooth bromegrass) is a grass species with high nutritional value, great palatability, cold tolerance, and grazing resistance, which has been widely cultivated for pasture and sand fixation in northern and northwestern China. Salt stress is a main environmental factor limiting growth and production of smooth bromegrass. In this study, we performed PacBio Iso-Seq to construct the first full-length transcriptome database for smooth bromegrass under 300 mM NaCl treatment at different time points. Third-generation full-length transcriptome sequencing yielded 19.67 G polymerase read bases, which were assembled into 355,836 full-length transcripts with an average length of 2,542 bp. A total of 116,578 differentially expressed genes were obtained by comparing the results of third-generation sequencing and second-generation sequencing. GO and KEGG enrichment analyses revealed that multiple pathways were differently activated in leaves and roots. In particular, a number of genes participating in the molecular network of plant signal perception, signal transduction, transcription regulation, antioxidant defense, and ion regulation were affected by NaCl treatment. In particular, the CBL-CIPK, MAPK, ABA signaling network, and SOS core regulatory pathways of Ca^{2+} signal transduction were activated to regulate salt stress response. In addition, the expression patterns of 10 salt-responsive genes were validated by quantitative real-time PCR, which were consistent with those detected by RNA-Seq. Our results reveal the molecular regulation of smooth bromegrass in response to salt stress, which are important for further investigation of critical salt responsive genes and molecular breeding of salt-tolerant smooth bromegrass.

Keywords: smooth bromegrass, full-length transcriptomics, salt tolerance, ion signal transduction, SOS regulatory pathways

OPEN ACCESS

Edited by:
Jiyu Zhang,
Lanzhou University, China

Reviewed by:
Jianxiu Liu,
Institute of Botany, Jiangsu Province (CAS), China
Marcelo Nogueira Do Amaral,
Federal University of Pelotas, Brazil

***Correspondence:**
Bo Zhang
xjauzb@126.com
Yongzhen Pang
pangyongzhen@caas.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

Received: *11 April 2022*
Accepted: *09 May 2022*
Published: *09 June 2022*

Citation:
Li Q, Song J, Zhou Y, Chen Y, Zhang L, Pang Y and Zhang B (2022) Full-Length Transcriptomics Reveals Complex Molecular Mechanism of Salt Tolerance in Bromus inermis L. Front. Plant Sci. 13:917338. doi: 10.3389/fpls.2022.917338

INTRODUCTION

The salt content of the soil surface is high in arid and semi-arid areas due to low precipitation, high evaporation, and inadequate water resource management (Rengasamy, 2010). Soil salinity exists in almost every irrigated area in the world, as well as in non-irrigated farmland and pasture (Ivushkin et al., 2019; Kearl et al., 2019). High content of salt in the soil causes dehydration of plant roots,

imbalance of intracellular osmotic regulation, and osmotic stress. In addition, salt stress affects biological functions in plant cells, for example, inhibiting its ability to uptake water and nutrients, accelerating chlorophyll degradation, inhibiting photosynthesis, and blocking various metabolic reactions (Karan and Subudhi, 2012; Deinlein et al., 2014; Al-Farsi et al., 2020). The most direct damage is high osmotic pressure resulting from low water use efficiency and ionic toxicity (mainly Na^+) (Chinnusamy et al., 2006). Plants generally respond to salt stress through various mechanisms, including osmotic adjustment, ion transmembrane transport, ion compartmentalization, and active oxygen scavenging (ROS) (Liu et al., 2021).

In response to high salt stress, plants have evolved a number of signal transduction pathways involving regulatory genes and proteins for stress perception, stress signaling, and downstream metabolites (Bhattarai et al., 2020). Under salt stress, various stress signal receptors located on the cell membrane can quickly perceive changes in the external environment, and second messengers such as Ca^{2+}, ROS, and phytohormone are rapidly generated in the cytoplasm to gradually decode and amplify salt stress signals (Fahad et al., 2014; Steinhorst and Kudla, 2019). These signals regulate downstream TFs through a cascade reaction to further alter transcript levels of many TF genes, such as *AP2/ERF*, *bZIP*, *bHLH*, *WRKY*, *MYB*, *NAC*, and *DREB* (Zhao et al., 2020). Thereafter, the expression of various osmotic stress-responsive genes (e.g., *P5CSs* and *P5CRs*) and ionic stress-responsive genes (e.g., *HKTs*, *AKTs*, and *NSCCs*) is eventually affected and ultimately contributes to plant salt tolerance (Zhu, 2016). Several signaling pathways involved in salt stress response have been revealed, including calcineurin B-like protein (CBL)–interacting protein kinase (CBL-CIPK), mitogen-activated protein kinase (MAPK), SOS, plant hormone (e.g., ABA, Eth, BR, and GAs), and calcium-dependent protein kinase (CDPK) pathways (Zhang et al., 2020; Zhao et al., 2020).

Smooth bromegrass is a perennial fine pasture of the grass family, which is widely used as hay, silage, and pasture for ruminants and dairy production (Ferdinandez and Coulman, 2001; Smart et al., 2006). It is mainly cultivated in northern China. However, due to climatic factors and poor agricultural management, soil salinization in these regions is becoming worse, which limits the production and utilization of smooth bromegrass (Wang and Yan, 2013 Wang et al., 2021). Therefore, it is important to explore molecular mechanisms of smooth bromegrass under salt stress and identify valuable resources for further breeding programs.

Currently, next-generation sequencing (NGS) provides an accurate and comprehensive analysis of differentially expressed genes and has made significant progress in understanding plant responses to drought and salt stress. NGS based on Illumina sequencing technology has been applied in *Spartina pectinata* (Robben and Gonzalez, 2016), *Hordeum vulgare* (Zhang et al., 2021), *Spartina alterniflora* (Hana, 2015), and *Sorghum bicolor* (Cui et al., 2018) to explore the molecular response mechanism under salt stress. Meanwhile, the single-molecule real-time (SMRT) sequencing provides third-generation sequencing (TGS) technology, which outperforms NGS technology in terms of read length (Rhoads and Au, 2015). In comparison to standard NGS technology, PacBio RSII TGS employs SMRT isoform sequencing (Iso-Seq), which provides longer read length, uniform coverage, and high accuracy (Dong et al., 2015). Therefore, SMRT sequencing technology has been used in *Zea mays* (Wang et al., 2016a), *Sorghum bicolor* (Abdel-Ghany et al., 2016), *Medicago sativa* (Luo et al., 2019), and *Arabidopsis pumila* (Yang et al., 2018) to explore how plants respond to stress.

In this study, we combined NGS and SMRT sequencing to generate a full-length transcriptome in smooth bromegrass. We analyzed the expression profile of differentially expressed genes in leaves and roots of smooth bromegrass under salt stress. In addition, we also identified a set of key genes that are likely involved in salinity adaptation in smooth bromegrass. These sequencing data will facilitate future research on gene function characterization and salinity adaptation mechanism in smooth bromegrass.

MATERIALS AND METHODS

Plant Material and Growth

Bromus inermis cv. Wusu No. 1 is a variety that is widely cultivated at Qitai Grassland Station in Xinjiang. It possesses characteristics of high yield, superior drought tolerance, cold resistance, rapid regeneration, early greening, and disease resistance. The full-grained seeds of Wusu No. 1 smooth bromegrass were sterilized and placed into a 26°C/16°C (light/dark, 8h/16h) light incubator for germination. At the two-leaf stage, the seedlings were transplanted into soil containing vermiculite and perlite at a ratio of 3:1 and grew in a greenhouse. The plants were irrigated with a fresh 1/2 Hoagland solution every 3 days. When the seedlings grew to the 4-leaf stage, they were selected and carefully transplanted into a triangular flask containing a 1/2 Hoagland solution to recover for 3 days. These seedlings were then treated with a 1/2 Hoagland solution containing 300 mM NaCl, and leaves and roots from 10 seedlings (with a triplicate) were sampled at 0 h, 12 h, 24 h, 36 h, and 48 h, frozen in liquid nitrogen, and stored at −80 °C.

Analysis on Physiological Index

Chlorophyll was measured with a SPAD-502 chlorophyll meter, and the SPAD value was recorded with 10 repeats. A handy PEA portable plant fluorometer was used to determine the maximum photochemical efficiency. A DDS-11A conductivity meter was utilized to measure relative conductivity, and the relative water content method was used to measure fresh weight (FW) immediately after harvest. The plants were placed in deionized water at 4°C for 8 h. The weight (TW) saturated with water absorption was measured, dried at 105°C for 10 min, and moved to 80°C for 24 h. The dry weight (DW) of the plants was measured, and the RWC was calculated using the following formula: RWC (%) = (FW-DW)/(TW-DW) × 100. The parameters of MDA, proline, SOD, and GSH content were determined by using kits developed by Nanjing Jiancheng Bioengineering Research Institute Co., Ltd. Measurement methods were referred to the instructions of the plant malondialdehyde (MDA) assay kit (A003-3-1), proline assay kit (A107-1-1), superoxide dismutase (SOD) assay kit

(A001-3-2), and reduced glutathione (GSH) assay kit (A006-2-1), respectively, as previously reported (Wang et al., 2016b, 2019; Meng et al., 2020).

Sequencing, Assembly, and Annotation of SMRT Library

RNAs were extracted by using an EasyPure Plant RNA Kit (No. ER301), and a Nanodrop 2000 spectrophotometer was used to detect the purity of RNA. A SMARTer PCR cDNA Synthesis Kit was employed for reverse transcription. The PacBio Sequel platform (Pacific Biosciences: http://www.pacb.com) was used for sequencing, and the PacBio official software package SMRTlink was used to process the original offline data. The subread sequence and the circular consensus sequence (CCS) were produced by correcting the subreads, and the sequences were then classified as full-length or non-full-length based on the presence of 5'-end primer, 3'-end primer, or poly-A tail. The full-length sequences were then clustered using ICE to get the cluster consensus sequence. In order to measure the correlation and similarity among samples, the PCA and Pearson correlation coefficient (R^2) were analyzed by using the R package.

We used LoRDEC software (Leena and Eric, 2014) to analyze and correct the second-generation data and third-generation PacBio data. The corrected transcripts were sequenced and clustered by CD-HIT software (Fu et al., 2012), and redundant and similar sequences were removed. CDS predictive analysis was performed using ANGEL software (Shimizu et al., 2006), and the fault-tolerant mode was adopted by default. We compared all predicted protein coding sequences with protein and nucleotide databases using BLASTX with the following databases: NR, Nt, Pfam, KOG/COG, SWISS-PROT, KEGG, and GO. The cutoff E-value for NCBI non-redundant protein (NR) and NCBI non-redundant nucleotide (NT) was $\leq 10^{-5}$, and that for the protein family was E-value $\leq 10^{-5}$. The cutoff E-values were all $\leq 10^{-5}$ for KOG (http://www.ncbi.nlm.nih.gov/COG/), SWISS-PROT (Amos and Rolf, 2000), and KEGG analyses (http://www.genome.jp/kegg/) (Minoru et al., 2004). The setting parameters for Gene Ontology were referred to Ashburner et al. (http://www.geneontology.org/) (Ashburner et al., 2000), and GO enrichment analysis was performed by using agriGO v2.0 (p-value $\leq$ 0.05) (Tian et al., 2017). All differentially expressed genes were analyzed by using iTAK for online plant transcription factor prediction (http://itak.feilab.net/cgi-bin/itak/index.cgi) (Yi et al., 2016).

Illumina Sequencing and Sequence Assembly

Magnetic beads with oligo (dT) were used to enrich mRNA, and six-base random primers (random hexamers) were employed to synthesize cDNA. The double-stranded cDNAs were purified with AMPure XP beads, and PCR amplification was performed to construct the cDNA library. An Agilent 2,100 Bioanalyzer was used to detect the insert size of the library. After qualification, the Illumina second-generation high-throughput sequencing platform was used, and raw reads were filtered to obtain clean reads. Trinity (Grabherr et al., 2011) was used to assemble the transcripts of clean reads, and software RSEM (Dewey and Li, 2011) was used to calculate the gene expression using the FPKM method (Trapnell et al., 2010). Based on the average FPKM values under each treatment in roots and leaves, both the absolute values of the $|\log_2$ (fold change)$| \geq 1$ and the adjusted P-value < 0.05 were used as thresholds to identify DEGs.

RT-qPCR Analysis

In order to verify the accuracy of the sequencing results, 10 differential genes were selected for RT-qPCR analysis. The TransScript One-Step gDNA Removal and cDNA Synthesis Super-Mix Reverse Transcription Kit were used for reverse transcription. Total RNAs extracted from leaves and roots from the control and salt treatment groups were used to synthesize cDNA. Primer Premier 5 software was used to design primers (**Supplementary Table S1**) (Regina et al., 2010; Bahrini et al., 2011). Real-time PCR was performed using the Applied Biosystems 7500/7500 fast real-time PCR system (ABI, Foster City, California, USA) and the SYBR Green PCR Master Mix system (Takara). We referred to the Trans-Start Top Green qPCR Super-Mix manual for RT-qPCR with triplicates. The $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001) was used to quantitatively calculate the relative expression level of candidate genes.

Statistical Analysis

Statistical data analysis was performed using Excel 2016 software and the SPSS software package (ver. 22.0; SPSS Inc., Chicago, IL, USA). MEV 4.9 software (https://sourceforge.net/pro-jects/mev-tm4/files/mev-tm4/) was used for cluster analysis and expression pattern analysis.

RESULTS

Effects of Salt Treatment on Physiological Changes of Smooth Bromegrass

Under the treatment of 300 mM NaCl, seedlings of smooth bromegrass gradually withered from 12 h to 48 h (**Supplementary Figure S1**). Compared with the control group (CK), the relative water content gradually decreased under NaCl treatment, while the relative conductivity gradually increased and reached the peak at 48 h (**Figures 1A,B**). No significant difference in the chlorophyll SPAD value was found for the treatment group when compared with the control group, but the photochemical efficiency was reduced by 10.98% after 48-h treatment compared with the CK (**Figures 1C,D**). Proline and Na^+ contents increased during treatment time (**Figures 1E,F**), while the K^+ content increased slightly at 12 h and decreased gradually from 36 h to 48 h (**Figure 1G**).

The malondialdehyde content can be used as an indicator of membrane lipid peroxidation. The malondialdehyde level in leaves of smooth bromegrass was considerably elevated by 37.19% at 48 h in the NaCl treatment group compared to the control (**Figure 1H**). In order to investigate the antioxidant defense system of smooth bromegrass cells under NaCl treatment, activities of key enzymes such as superoxide dismutase (SOD) and reduced glutathione (GSH) were tested in both control and treatment groups. It showed that SOD activity

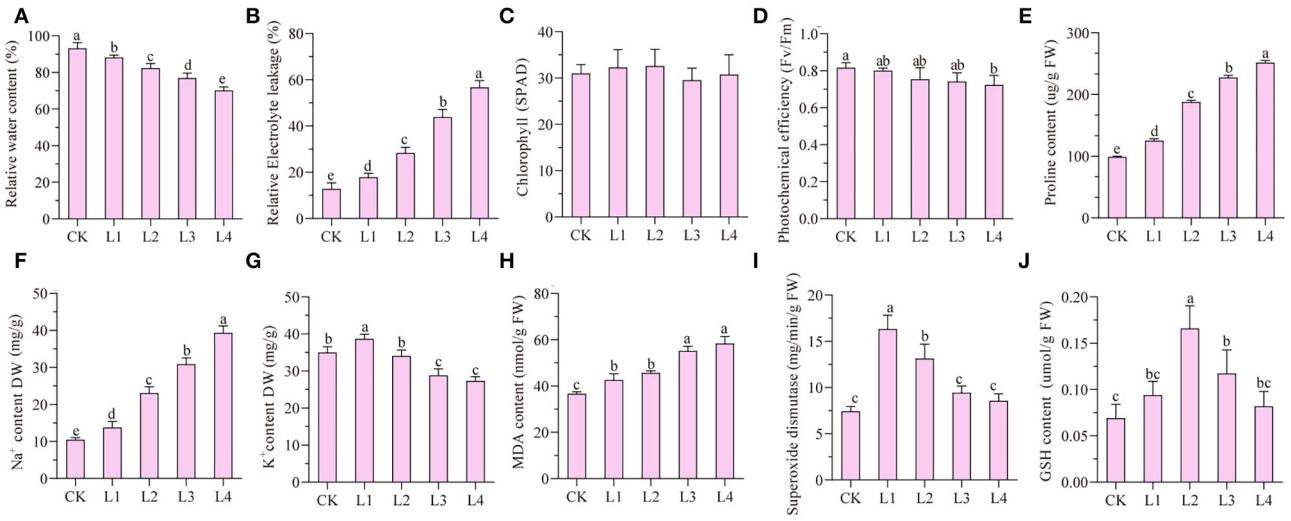


FIGURE 1 | Analyses of dynamic physiological effects under 300 mM NaCl stress in the leaves. CK: untreated control; L1: salt stress treatment for 12h; L2: 24 h; L3: 36 h; L4: 48 h. **(A)** rRelative water content; **(B)** relative electrolyte leakage; **(C)** chlorophyll content (SPAD); **(D)** photochemical efficiency; **(E)** proline content; **(F)** Na^+ content; **(G)** K^+ content; **(H)** MDA content; **(I)** superoxide dismutase; **(J)** GSH content.

reached its peak after a 12-h treatment and started to decrease from 24 h to 48 h (**Figure 1I**). Unlike SOD, the GSH content increased considerably to a maximum at 24 h and then decreased from 36 h to 48 h (**Figure 1J**).

Statistical Analysis of Transcriptome Sequencing Data

The leaves and roots of smooth bromegrass treated with 300 mM NaCl at five time points were separately sampled and subjected to NGS sequencing. In total, about 226.11 G clean reads were generated by Illumina sequencing (**Supplementary Table S2**). The PacBio RS II platform was utilized to create the Iso-Seq library and single-molecule sequencing. In total, 19.67 G polymerase read bases and 516,677 polymerase reads were obtained. It showed that the final insert subread is 19.06 G, and the number of final insert subreads is 8,354,897 (**Table 1**).

In total, 369,646 full-length (FL) reads, 41,968 non-full-length (NFL) reads, and 355,836 full-length non-chimeric (FLNC) reads (**Supplementary Table S3**) were obtained. To eliminate redundancy, we used the ICE algorithm to cluster FLNC sequences with the same transcript, and a total of 202,837 consensus numbers were obtained. The length of the polished consensus sequence for each sample and the length distribution map are shown in **Supplementary Figures S2A,B**. In order to test the repeatability between samples, both principal component analysis and Pearson correlation were performed for a total of 30 samples. The principal component analysis showed the samples for L0 and R0 from leaves and roots were grouped separately, which was far away from the samples treated with salt (**Supplementary Figure S3**). Correlation analysis showed the Pearson correlation coefficient between samples of the same treatment was greater than 0.8 (**Supplementary Figure S4**),

TABLE 1 | Full-length transcriptome sequencing results.

Library	Bases(G)	Number	Mean length	N50
Polymerase	19.67	516677	38,079	64,895
Subreads	19.06	8354897	2,282	2,730

indicating that sampling was relatively accurate with small error, and the data analyzed later are also relatively reliable. Furthermore, the consensus sequences obtained from the subreads yielded 413,135 circular consensus sequences (CCS), and the average number of CCS was 17. FLNC reads accounts for 86.13% of the CCS (**Supplementary Table S5**).

We used LoRDEC software (Leena and Eric, 2014) to correct and analyze the Illumina data and SMRT PacBio data, and CD-HIT software (Fu et al., 2012) to remove redundant and similar sequences, and finally, we obtained 202,837 transcripts (**Table 2**). Among all these transcripts, only 1.66% of them were >500 bp, 37.77% of them were 2-3kb, and 31.57% of them were longer than 3kb (**Table 2**). Our results confirmed that SMRT sequencing provides a large number of full-length and high-quality transcripts and that using Illumina data to correct low-quality SMRT reads improved the accuracy of long reads of PacBio (**Figure 2A**). The average length of genes detected by SMRT is more than that detected by Illumina. Finally, 116,578 unigenes were obtained after the removal of redundant sequences, and 1.11% of the retrieved unigenes were >500 bp, whereas 39.10% of them were longer than 3 kb (**Table 2**, **Figure 2B**).

Analysis of Gene Function Annotation

For more accurate gene analysis, we used ANGEL software (Shimizu et al., 2006) to predict CDS regions and obtained

the distribution map of CDS (**Supplementary Figure S3A**). Unigenes were used for gene function annotation, with the following databases: Nr, Nt, Pfam, KOG/COG, SWISS-PROT, KEGG, and GO. In total, 66,847 (57.34%) unigenes had their annotation based on the GO database, and 104,784 (89.88%) and 107,022 (91.80%) unigenes based on KEGG and Nr databases, respectively (**Figure 2C**; **Supplementary Table S4**). Based on the plant species in the Nr database, a large number of unigenes were annotated to *Hordeum vulgare*, indicating that smooth bromegrass is closely related to *Hordeum vulgare* (**Supplementary Figure S3B**). Five major databases (viz., Nr, GO, KEGG, KOG, and Nt) jointly annotated 45,200 shared unigenes, and 523 were solely found in the NR database, 486 in the GO database, 32 in KEGG, 41 in KOG, and 3,679 in NT (**Figure 2D**).

TABLE 2 | Statistical table of length frequency distribution before and after transcript de-redundancy.

Transcripts length interval	<500bp	500-1kbp	1k-2kbp	2k-3kbp	>3kbp	Total
Number of transcripts	3,362	11,784	59,214	64,439	64,038	202,837
Number of Genes	1,295	6,639	25,736	37,338	45,570	116,578

Analysis of Differentially Expressed Genes

Compared with the control, 1,727 unigenes were specifically expressed in leaves under NaCl treatment at 12 h (L1), 7,523 at 24 h (L2), 1,458 at 36 h (L3), and 1,068 at 48 h (L4), and 4,753 of them were shared by these four comparison pairs (**Figure 3A**). Among differentially expressed genes, 7,954, 12,247, 6,283, and

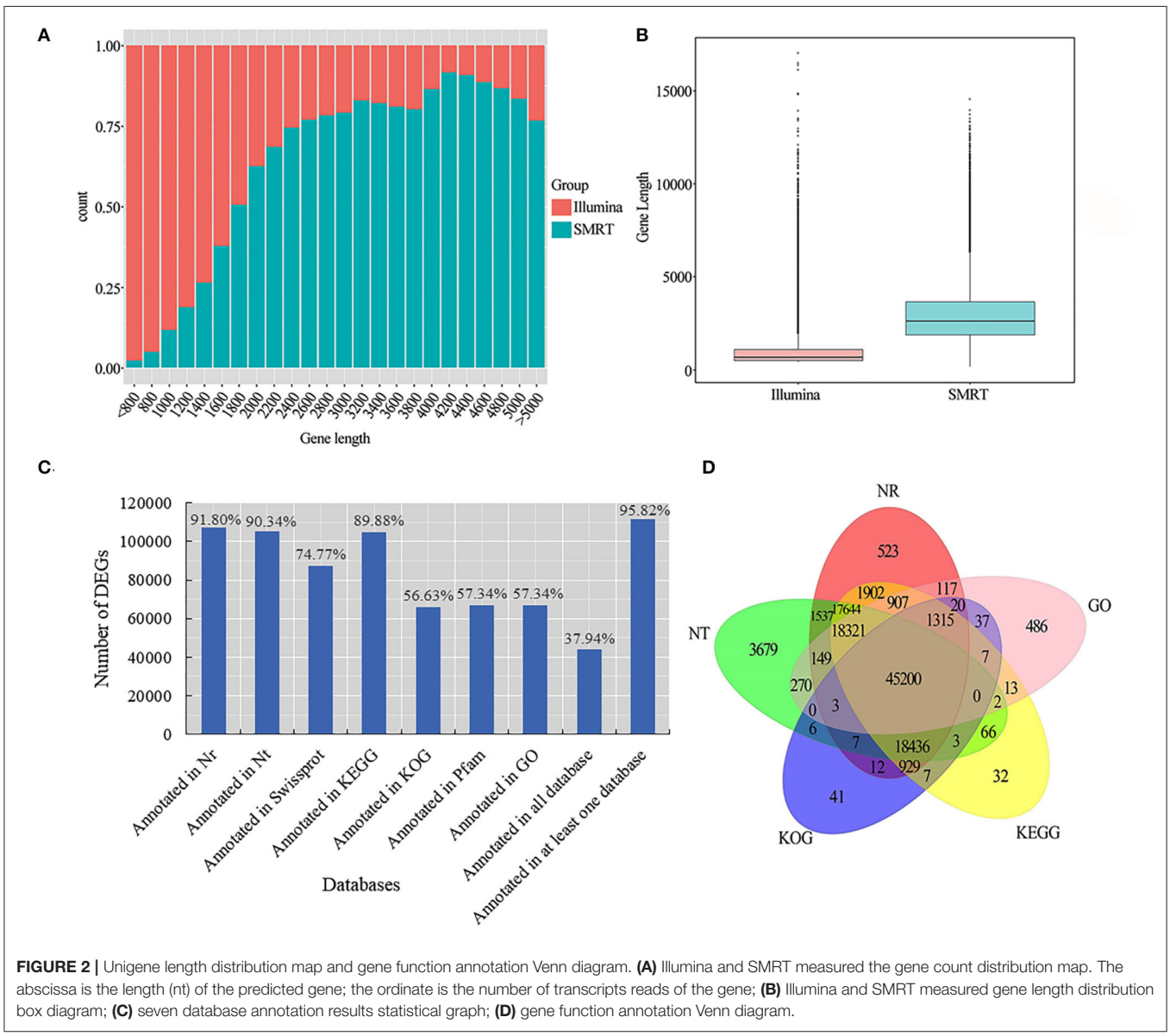


FIGURE 2 | Unigene length distribution map and gene function annotation Venn diagram. **(A)** Illumina and SMRT measured the gene count distribution map. The abscissa is the length (nt) of the predicted gene; the ordinate is the number of transcripts reads of the gene; **(B)** Illumina and SMRT measured gene length distribution box diagram; **(C)** seven database annotation results statistical graph; **(D)** gene function annotation Venn diagram.

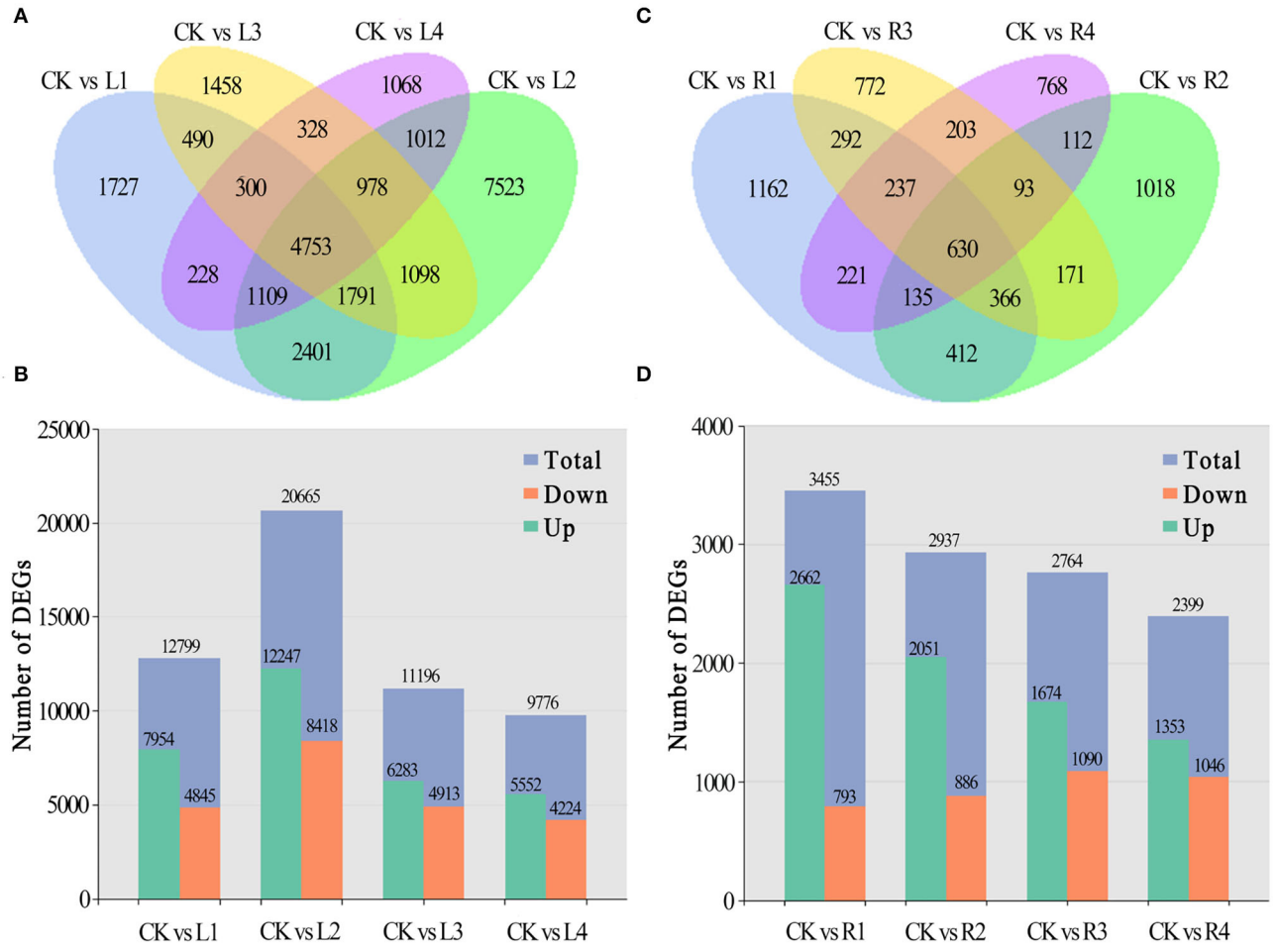


FIGURE 3 | Distribution of DEGs in different periods. **(A,B)** Numbers of DEGs in leaves and roots of smooth bromegrass under different stress times compared with the control; **(C,D)** Summary of the numbers of up- and downregulated DEGs in leaves and roots of smooth bromegrass.

5,552 unigenes were upregulated, and 4,845, 8,418, 4,913, and 4,224 unigenes were downregulated for the four comparison pairs, respectively (**Figure 3B**).

In roots, 1,162 unigenes were specifically expressed at the R1 stage under NaCl treatment in comparison with the control group, 1,018 at the R2 stage, 772 at the R3 stage, and 768 at the R4 stage, and a total of 630 unigenes were shared by the four comparison pairs (**Figure 3C**). Meanwhile, 3,455, 2,937, 2,764, and 2,399 unigenes were found to be differentially expressed in roots of smooth bromegrass at four different periods in comparison with the control group. Among them, 2,662, 2,051, 1,674, and 1,353 unigenes were upregulated, and 793, 886, 1,090, and 1,046 unigenes were downregulated at the four stages, respectively (**Figure 3D**).

Analysis of GO Annotation

The GO database was used for the annotation of DEGs between the control and treatment groups. The heat map for GO of DEGs was generated based on the calculated -$\log_2$(*p*-value), and a total of 17 biological processes were identified (**Supplementary Table S5**). Among them, the carbohydrate metabolism process, cellular amine metabolism process, and response to oxidative stress process showed higher expression levels in each stage under salt stress (**Figure 4A**). These data were also analyzed by using a direct acyclic graph (DAG) tree, resulting in the identification of response to oxidative stress, carbohydrate metabolic process, cellular amine metabolic process, and other important processes (**Figure 4B**).

Furthermore, we found seven processes related to cell components (**Supplementary Table S5**), which were expressed differently in both leaves and roots. The DAG tree analysis also revealed important processes of the polyphosphate kinase complex and membrane (**Figure 4C**). By analyzing these data, we also found 18 processes related to molecular functions (**Supplementary Table S5**), and these processes varied in different stages under NaCl treatment. In particular, the catalytic activity process was significantly expressed in leaves of smooth bromegrass at all treatment time points (**Figure 4A**). The oxidoreductase activity process exhibited relatively high expression in various processing stages in both leaves and

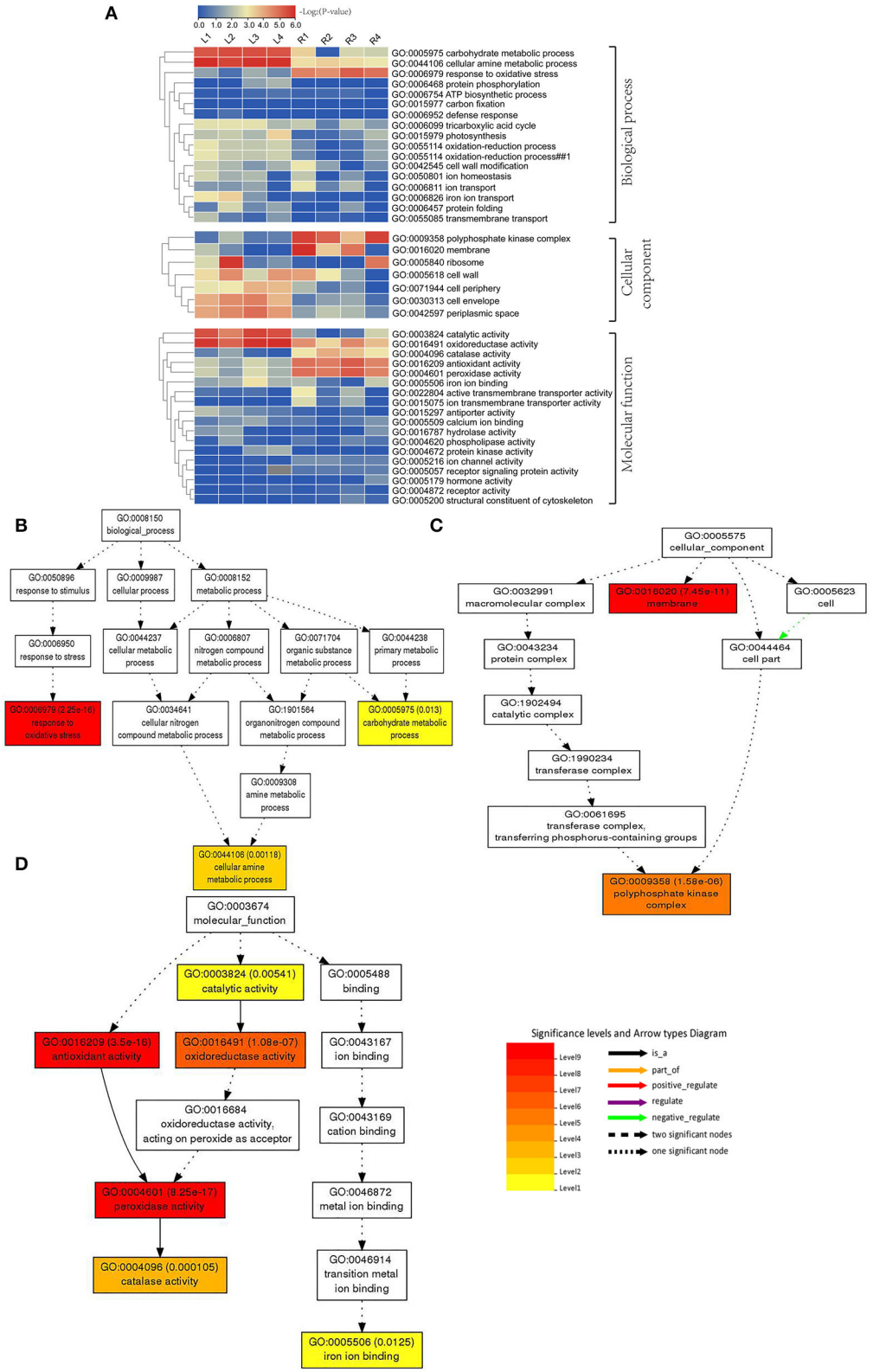


FIGURE 4 | GO enrichment results of DEGs after NaCl treatment. **(A)** GO annotation results of DEGs; **(B–D)** GO term for DEGs that respond to salt stress. Each box represents a GO term; the depth of the color represents the degree of enrichment, and the darker the color, the higher the degree of enrichment.

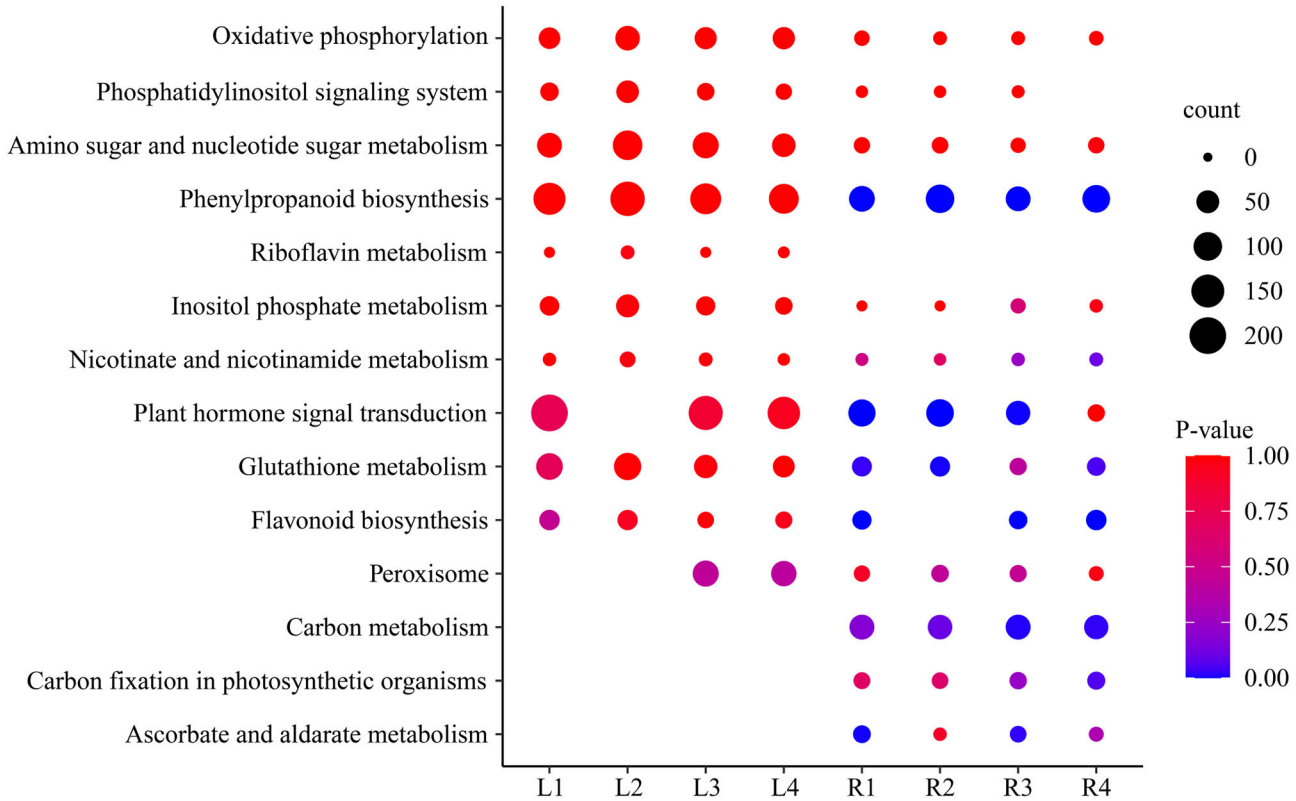


FIGURE 5 | Scatter plot of the enriched KEGG pathway of DEGs under NaCl stress.

roots. In the process of peroxidase activity, catalase activity and antioxidant activity showed higher expression in roots. Moreover, several important molecular functional processes were also detected by using DAG analysis, such as catalytic activity, antioxidant activity, oxidoreductase activity, peroxidase activity, catalase activity, and iron ion binding (**Figure 4D**).

Analysis of KEGG Pathway

Analysis of the KEGG metabolic pathway was performed with DEGs between the control group and the four treatment groups in both leaves and roots, and 14 important pathways were significantly enriched (**Supplementary Table S6**), including ascorbate and aldarate metabolism, oxidative phosphorylation, glutathione metabolism, plant hormone signal transduction, phenylpropanoid biosynthesis, and starch and sucrose metabolism (**Figure 5**). These pathways were affected in different time periods under salt stress. The phenylpropanoid biosynthesis pathway, oxidative phosphorylation, phosphatidylinositol signaling system, and sugar metabolism were highly expressed, with more genes in leaves than in roots. Interestingly, the pathways of riboflavin metabolism only occurred in leaves but not in roots, while the carbon metabolism pathway, carbon fixation in photosynthetic organisms, and ascorbate and aldarate metabolism only occurred in roots but not in leaves, indicating they are tissue-specific pathways in smooth bromegrass (**Figure 5**).

Analyses of Gene Expression Levels Involved in Specific Pathways

In this study, we also analyzed ABA signaling pathway genes and identified 56 genes encoding signal receptor sensors related to NaCl stress through DEGs annotation analysis (**Supplementary Figure S6A** and **Supplementary Table S7**). In addition, 3 *RLKs*, 8 *CDPKs*, 20 *CIPKs*, 15 *CMLs*, and 10 *MAPKs* in the signal sensor pathway were also identified (**Supplementary Figure S6B** and **Supplementary Table S8**).

Meanwhile, we also found that most of the DEGs were associated with antioxidant enzymes (*SODs*, *PODs*, *GPXs*, *CATs*, and *APXs*), non-enzymatic antioxidants (*GSTs*, *GSHs*, and *GSRs*), and proline synthase (*P5CSs*, and *P5CRs*), and their corresponding coding genes were all regulated by NaCl stress at various levels (**Figure 6A**; **Supplementary Table S9**). We found that the expression levels of 9 *AKTs*, 11 *AVPs*, 7 *CHXs*, 6 *CNGCs*, 3 *HKTs*, 3 *KATs*, and 6 *PMAs* genes were also affected under NaCl treatment in both leaves and roots (**Figure 6B**; **Supplementary Table S10**). Furthermore, we identified 12 *NHX*-, 10 SOS1-, 5 SOS2-, and 4 SOS3-related genes (**Figure 6C**, **Supplementary Table S11**), and these genes

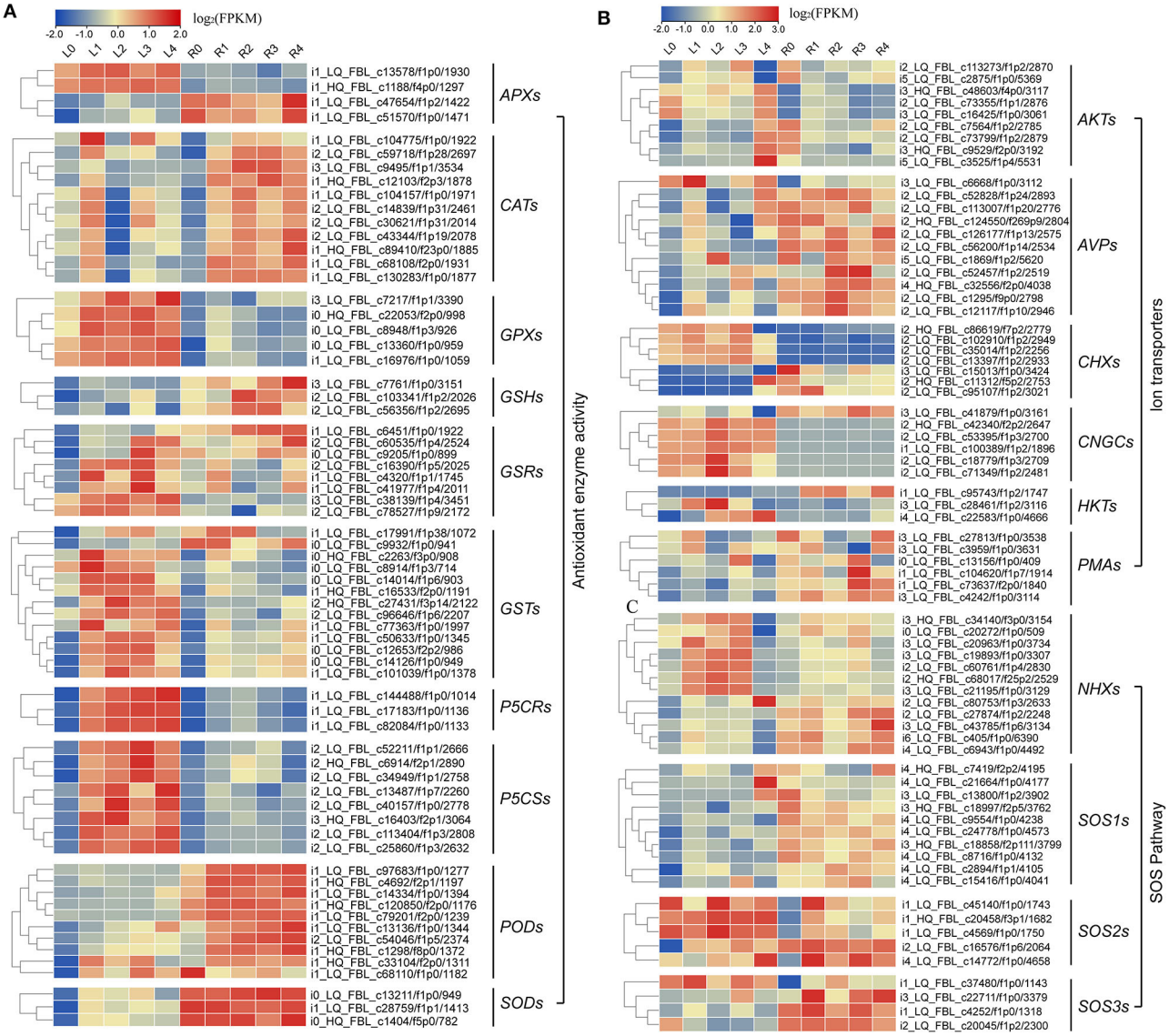


FIGURE 6 | Heat map plot of the expression levels of the DEGs involved in antioxidative, SOS, and ion transporter pathway under NaCl stress. **(A)** Heat map plot of the antioxidative defense-related DEGs by NaCl stress; **(B,C)** heat map plot of the DEGs of SOS pathway and ion transporters by NaCl stress.

were upregulated after salt stress, indicating that Na^+ regulatory mechanism in smooth bromegrass was activated.

Identification of Salt-Responsive Transcription Factor Genes

In this study, a total of 1,223 transcription factor genes were identified and classified into 53 families. Among them, the bZIP family is the largest family with 127 genes, followed by 96 *NAC*-, 76 *bHLH*-, 73 *AP2/ERF*-, and 70 *MYB*-related genes (**Figure 7**; **Supplementary Table S12**). In order to investigate the expression of these TF genes, they were all analyzed at all the treatment time points. It was found that the expression levels of many transcription factor genes were highly upregulated at different time points after salt stress, including several main TF family genes *AP2/ERF*, *bHLH*, *bZIP*, *GRAS*, *MYB*, and *WRKY* (**Supplementary Figure S7**). In particular, many of them showed distinct expression profiles between leaves and roots (**Supplementary Figure S7**). The dynamic changes in the expression level of these TF genes implied their important roles in regulating plant salt tolerance.

Validation of NaCl Stress-Related Genes by RT-qPCR

To validate the transcriptome data, 10 DEGs were randomly selected, and their expression levels were analyzed in the control and treatment groups by RT-qPCR (**Supplementary Figure S8A**), with *actin* gene as an internal reference. These genes included the putative peroxide gene,

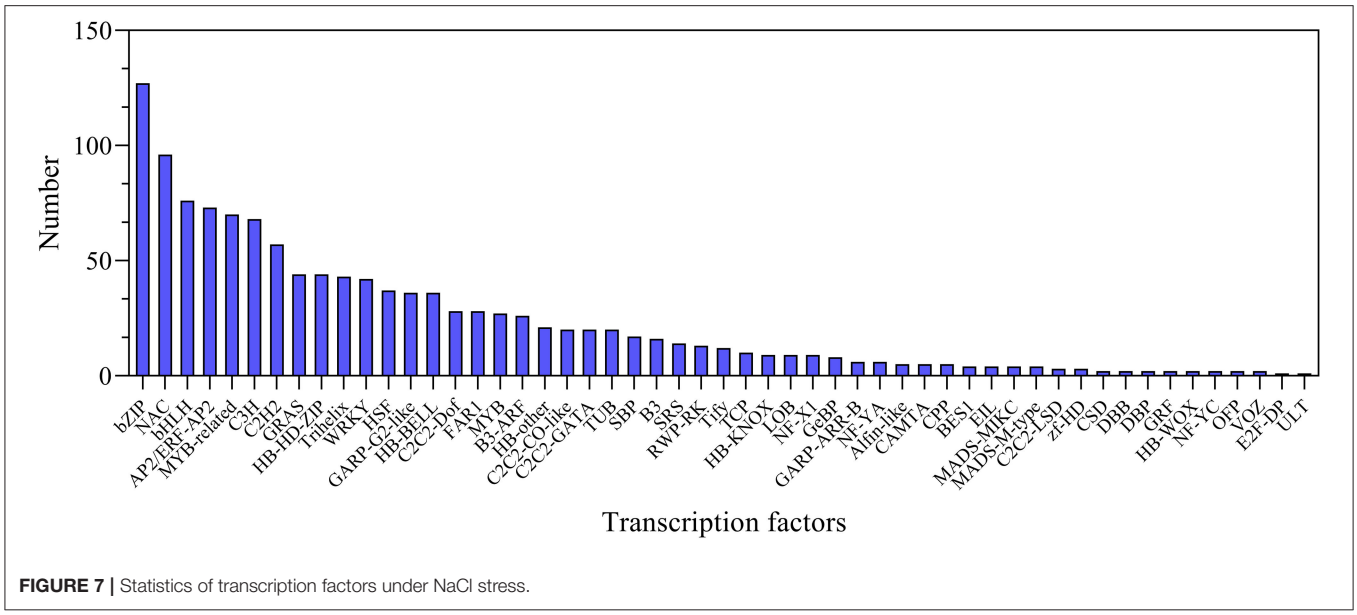


FIGURE 7 | Statistics of transcription factors under NaCl stress.

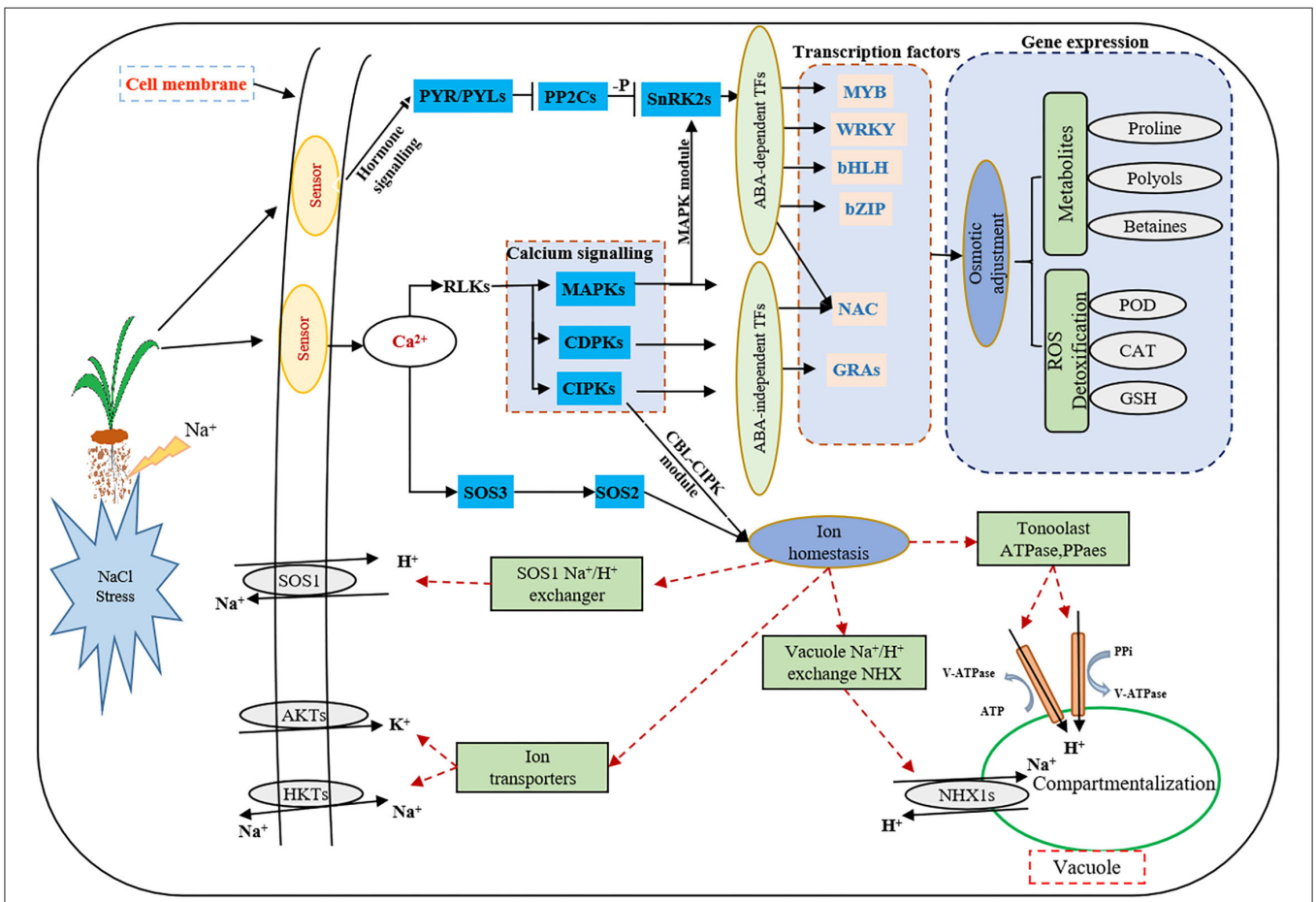


FIGURE 8 | Model describing the signal transduction pathways involved in the acquisition of salt tolerance in smooth bromegrass. Red dashed lines indicate proposed regulation, solid arrowheads indicate activation, and vertical solid lines indicate suppression.

MYB39, and CYP73A2 (**Supplementary Figure S8A**). By comparing and analyzing these results with the RNA-Seq data (**Supplementary Figure S8B**), it was found that the expression of these 10 selected genes showed similar expression profiles in the leaves and roots at four time points (**Supplementary Figure S8A**). Further analyses showed that RT-qPCR and RNA-Seq data were correlated linearly either in leaves or in roots at 12 h, 24 h, 36 h, and 48 h (**Supplementary Figure S8B–E**), indicating RNA-Seq data were reliable for further gene screening and expression level comparison.

DISCUSSION

At present, the third-generation sequencing technology (e.g., PacBio SMRT) has greatly facilitated the *de novo* assembly of transcriptomes of many higher organisms and has helped to overcome the problems of short splicing and incomplete information for the species without reference genomes (Cheng et al., 2017; Chao et al., 2019; Kan et al., 2020; Shen et al., 2021). Meanwhile, second-generation sequencing (e.g., Illumina platform) has been widely used to obtain more comprehensive annotation information of many plant species (Yang et al., 2018; Luo et al., 2019). In this study, for the first time, the Illumina platform was used to study the transcriptome changes of global genes in response to salt stress in the leaves and roots of smooth bromegrass, and the reference sequence was obtained from its full-length transcript database using PacBio SMRT sequencing technology. After screening, we acquired 19.67 G polymerase read bases (**Table 1**) and 116,578 distinct genes in 30 sample libraries (**Table 2**). Of these genes, 95.82% of DEGs were annotated in at least one database (**Figure 3C**), and five major databases (viz., Nr, GO, KEGG, KOG, and Nt) jointly annotated 45,200 unigenes (**Figure 3D**), which is higher than *Elymus sibiricus* (79.81%) (Zhou et al., 2016); however, it is lower than alfalfa (99%) (Luo et al., 2019). Compared with the control, 12,799, 20,665,11,196, and 9,776 DEGs were specifically expressed in leaves under NaCl treatment at 12 h, 24 h, 36 h, and 48 h, respectively (**Figure 3B**), and in roots, 3,455, 2,937, 2764, and 2,399 DEGs were identified in four different periods (**Figure 3D**). The candidate genes were analyzed in the control and treatment groups by RT-qPCR (**Supplementary Figure S8A**), which were consistent with the RNA-Seq data, implying our sequencing data were reliable (**Supplementary Figure S8**). Our study provided valuable and nearly complete sequence information on smooth bromegrass. Furthermore, we depicted a global molecular mechanism model of salt response in smooth bromegrass (**Figure 8**), with all these transcriptome information.

Under salt stress conditions, plants produced a large number of genes related to osmotic regulation, which regulate the accumulation of primary metabolites in plants such as proline, betaine, and sugar alcohol to maintain cell osmotic potential and improve their own salt tolerance (Xiong et al., 2017; Singh et al., 2018; Gao et al., 2019). Our studies showed that proline and soluble sugar contents increased in smooth bromegrass leaves under salt treatment (**Figure 1**), which was consistent with other studies in *Arabidopsis thaliana* (Yang et al., 2018), *Medicago sativa* (Luo et al., 2019), *Sorghum bicolor* (Cui et al., 2018), *Arundo donax* (Sicilia et al., 2020), and *Zea mays* (Kakumanu et al., 2012), indicating that plants share similar salt tolerance mechanisms. At the molecular level, the GO categories "response to oxidative stress," "ion transport," "catalytic activity," "antioxidant activity," and "hormone activity", and the KEGG pathway categories "plant hormone signal transduction," "phenylpropanoid biosynthesis," and "flavonoid biosynthesis" were significantly enriched under salt stress (**Figures 4**, **5**), which were consistent with rice (*Oryza sativa*) (Rabbani et al., 2003), maize (*Zea mays*) (Sun et al., 2015), and *Sophora alopecuroides* (Zhu et al., 2021). These results implied secondary metabolic processes and antioxidant systems are activated through secondary metabolites (phenylpropane and flavonoids), hormones, ion transport (Na^+ and K^+), antioxidant enzymes (SOD and GSH), and second messengers (such as ROS and Ca^{2+}) in response to salt stress. Consistently, our physiological work also revealed that SOD and GSH contents increased at the physiological level by NaCl stress (**Figures 1I,J**), and increased levels of Na^+ and K^+ contents were consistent with the ion transport pathway (**Figures 1F,G**). The physiological regulation mechanism of salt stress usually includes osmotic regulation, ion transmembrane transport, ion compartmentalization, and active oxygen scavenging. Plant hormones and transcription factors also respond to salt stress (Zhu, 2016; Geng et al., 2020; Luo et al., 2020), which enabled smooth bromegrass to respond to salt stress and further adapt to changes in the external environment by ROS accumulation.

Under NaCl stress, the osmotic stress caused by the increase in the Na^+ concentration in the rhizosphere of plants reduces the water potential. On the other hand, this ion stress induces a nutritional imbalance in plants (Huang et al., 2012; Al-Farsi et al., 2020). The active oxygen signal pathway is activated in response to osmotic stress, and excessive ROS causes serious damage to plant cell membranes, proteins, nucleic acids, etc. (Yun et al., 2011; Zhang et al., 2016). Plants produce redox-sensitive TFs and other molecular sensors to perceive stress signals, thus increasing the ROS concentration and activating ROS defense/metabolic pathways for ROS clearance (Chawla et al., 2013; Fidalgo et al., 2015; Mahmoud et al., 2020). Physiological studies on smooth bromegrass leaves revealed that NaCl produced oxidative damage, resulting in a substantial reduction in the chlorophyll content, an elevation in relative electrical conductivity, and a reduction in the relative water content (**Figure 1**). Smooth bromegrass activates the antioxidant defense system by adjusting the activity of antioxidants (POD, CAT, and SOD) and the content of osmotic regulators (PRO) (**Supplementary Table S9**). The same transcriptome data also revealed that NaCl stress activates the antioxidant defense system at the molecular level. When exposed to NaCl treatment GO-type oxidoreductase activity and KEGG pathways peroxisome and glutathione metabolism, as well as hormone signal transduction, were all enriched, as illustrated

in **Figures 4**, **5**. In addition, most of the DEGs associated with antioxidant enzymes (*SODs*, *PODs*, *GPXs*, *CATs*, and *APXs*), non-enzymatic antioxidants (*GSTs*, *GSHs*, and *GSRs*), and proline synthase (*P5CSs* and *P5CRs*) were all regulated by NaCl stress (**Supplementary Table S9**). These physiological indicators are consistent with the results of the transcriptome data, which strongly indicate the important role of these DEGs in responding to osmotic stress and therefore protecting smooth bromegrass from ROS damage.

Plants respond to external stress by activating signal pathways, including the ROS signal pathway, the MAPK cascade signal system, the ABA signal response pathway, and Ca^{2+}-dependent proteins (Wang et al., 2016b; Danquah et al., 2014; Steinhorst and Kudla, 2019). Osmotic stress signals are sensed by protein receptors on the cell membrane. These stress signal receptors can quickly sense changes in the external environment. Subsequently, many second messengers such as Ca^{2+}, inositol phosphate, ROS, and ABA signals will be rapidly produced in the cytoplasm to further decode and amplify the salt stress signal (Mehlmer et al., 2010; Zhu, 2016; Cao et al., 2020). The second messenger stimulates downstream signals through protein phosphorylation cascade changes, such as CDPKs, CBLs, CIPKs, CMLs, and MAPKs, linking the perception of external stimuli with cellular responses (Danquah et al., 2014; Bakshi et al., 2019). In this study, we identified a number of genes encoding signal receptor sensors related to NaCl stress through DEG annotation analysis (**Supplementary Figure S4B**), including *CDPKs*, *CIPKs*, *CMLs*, and *MAPKs* (**Supplementary Table S8**). Since roots are directly exposed to salt stress, the gene expression of smooth bromegrass exhibits distinct expression profiles in leaves and roots.

The characteristic of the CBL-CIPK regulatory network provides a great contribution to the efficient transduction of Ca^{2+} signals and also plays an important role in regulating the response of plants to salt stress (Thoday-Kennedy et al., 2015; Liu et al., 2021). The CBL-CIPK module can interact with many ion transporters and regulate their activity, especially Na^+/K^+ transporters, which are essential for ion homeostasis (Pandey et al., 2015; Fan et al., 2016). Studies have reported that cyclic nucleotide-gated channels (CNGCs) can mediate the flow of Ca^{2+} into the cytoplasm under stress and maintain ion balance in the cell (Kugler et al., 2009; Jin et al., 2015). In many cases, these ion transporters interact with the SOS signaling network in a synergistic or antagonistic manner to maintain ion "balance" in plants under salt stress conditions. In our study, the GO enrichment of DEGs revealed genes involved in ion homeostasis, iron ion transport, ion channel activity, ion transmembrane transporter activity, and iron ion binding, as well as the ion transporter (**Figure 4**). Meanwhile, genes involved in this signaling pathway were identified, including *AKTs*, *AVPs*, *HKTs*, and *PMAs* (**Figure 6B**; **Supplementary Table S10**), indicating that these DEGs may directly or indirectly promote the tolerance mechanism of smooth bromegrass to salt stress. The SOS signal pathway is an important signal transduction pathway for plant salt tolerance, and it plays an important role in regulating Na^+ and K^+ homeostasis and salt tolerance (Deinlein et al., 2014; Steinhorst and Kudla, 2019). In total, 31 genes related to NHXs and SOS pathways were identified (**Supplementary Table S11**), and they were largely upregulated by salt stress (**Figure 6B**), indicating that Na^+ regulatory mechanism was activated in smooth bromegrass, which was consistent with an increase in the Na^+ content (**Figure 1F**).

ABA is an important plant hormone that plays a key role in regulating plant growth and resisting adversity stress. The ABA signaling pathway is also activated in the early stage of osmotic stress (Raghavendra et al., 2010; Komatsu et al., 2020). Osmotic stress causes plants to accumulate a significant quantity of active oxygen, and NADPH oxidase in the ABA pathway has a certain mitigating effect on alleviating osmotic stress (Zhao et al., 2020). In this study, 36 ABA signaling pathway-related genes were analyzed, including 5 *PYR/PYLs*, 19 *PP2Cs*, and 12 *SnRK2s* (**Supplementary Figure S6A**; **Supplementary Table S7**), and the expression level of these genes increased significantly after salt stress treatment in smooth bromegrass.

TFs are key components in the osmotic stress signal pathway, and they are involved in the signal perception of salt stress and the expression of downstream key genes in response to salt stress. They participate in a variety of biological processes and play an important role in regulating plant growth and stress response (Hartmann et al., 2015; Yang et al., 2020). The transcriptome of smooth bromegrass revealed that a number of different transcription factor families were affected by salt stress, including *bZIP*, *NAC*, *bHLH*, *WRKY*, *AP2/ERF*, *MYB*-related, *C3H*, *C2H2*, *GRAS*, and *MYB* (**Figure 7**), which was similar to those major TF families as in wheat under salt stress (Malik et al., 2020). These transcription factor genes have also been found in response to salt stress of *Arabidopsis thaliana* (Yang et al., 2018), *Caragana korshinskii* (Li et al., 2016), and *Cynodon dactylon* (Hu et al., 2015). Transcription factors such as *AP2/ERF*, *bHLH*, *bZIP*, *C2H2*, *C3H*, *GRAS*, *NAC*, *MYB*-related, and *WRKY* showed different expression patterns in leaves and roots after NaCl stress (**Supplementary Figure S7**), indicating that NaCl had different effects on the complex transcriptional regulation of smooth bromegrass adaption to salt stress.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. All sequencing data of PacBio Iso-Seq and Illumina SGS RNA-Seq in this article can be obtained in NCBI SRA (http://www.ncbi.nlm.nih.gov/sra): SRR15634023 and SRR15633796-SRR15633815.

AUTHOR CONTRIBUTIONS

JS and QL designed this experiment, performed the experiments, and drafted the manuscript. JS, YC, and LZ analyzed experimental data and visualized it. BZ, YP, and YC revised the manuscript and directed the study. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was funded by the China Agriculture Research System of MOF and MARA (CARS-34).

ACKNOWLEDGMENTS

We sincerely thank Drs. Rongchen Yang and Jing Yang at Northwest A&F University for their help and suggestions for this experiment and manuscript. We also thank Drs. Shuzhen Zhang, Yanhui Zhang, Yuxiang Wang, Lili Gu, and Chenjian Li for their help and advice during the experiment.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.917338/full#supplementary-material

REFERENCES

Abdel-Ghany, S. E., Hamilton, M., Jacobi, J. L., Ngam, P., Devitt, N., Schilkey, F., et al. (2016). A survey of the sorghum transcriptome using single-molecule long reads. *Nat. Commun.* 7, 11706. doi: 10.1038/ncomms11706

Al-Farsi, S. M., Nawaz, A., Anees Ur, R., Nadaf, S. K., Al-Sadi, A. M., Siddique, K. H. M., et al. (2020). Effects, tolerance mechanisms and management of salt stress in lucerne (*Medicago sativa*). *Crop. Pasture. Sci.* 71, 411–428. doi: 10.1071/CP20033

Amos, B., and Rolf, A. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucl. Acids Res.* 28, 45–48. doi: 10.1093/nar/28.1.45

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Bahrini, I., Ogawa, T., Kobayashi, F., Kawahigashi, H., and Handa, H. (2011). Overexpression of the pathogen-inducible wheat TaWRKY45 gene confers disease resistance to multiple fungi in transgenic wheat plants. *Breed* 61, 319–326. doi: 10.1270/jsbbs.61.319

Bakshi, P., Handa, N., Gautam, V., Kaur, P., Sareen, S., Mir, B., et al. (2019). *Role and Regulation of Plant Hormones as a Signal Molecule in Response to Abiotic Stresses.* Cambridge: Woodhead Publ Ltd. p. 303–317. doi: 10.1016/B978-0-12-816451-8.00018-6

Bhattarai, S., Biswas, D., Fu, Y. B., and Biligetu, B. (2020). Morphological, physiological, and genetic responses to salt stress in alfalfa: a review. *Agronomy-Basel.* 10, 577. doi: 10.3390/agronomy10040577

Cao, Y., Zhang, M., Liang, X., Li, F., and Jiang, C. (2020). Natural variation of an EF-hand Ca^{2+}-binding-protein coding gene confers saline-alkaline tolerance in maize. *Nat. Commun.* 11, 186. doi: 10.1038/s41467-019-14027-y

Chao, Y. H., Yuan, J. B., Guo, T., Xu, L. X., Mu, Z. Y., and Han, L. B. (2019). Analysis of transcripts and splice isoforms in *Medicago sativa* L. by single-molecule long-read sequencing. *Plant Mol. Biol.* 99, 219–235. doi: 10.1007/s11103-018-0813-y

Chawla, S., Jain, S., and Jain, V. (2013). Salinity induced oxidative stress and antioxidant system in salt-tolerant and salt-sensitive cultivars of rice (*Oryza Sativa* L.). *J. Plant Biochem. Biot.* 22, 27–34. doi: 10.1007/s13562-012-0107-4

Cheng, B., Furtado, A., and Henry, R. J. (2017). Long-read sequencing of the coffee bean transcriptome reveals the diversity of full-length transcripts. *Gigascience.* 6, 13. doi: 10.1093/gigascience/gix086

Chinnusamy, V., Zhu, J., and Zhu, J. K. (2006). Salt stress signaling and mechanisms of plant salt tolerance. *Genet. Eng.* 27, 141–177. doi: 10.1007/0-387-25856-6_9

Cui, J., Ren, G., Qiao, H., Xiang, X., Huang, L., and Chang, J. (2018). Comparative transcriptome analysis of seedling stage of two *Sorghum* cultivars under salt stress. *J. Plant Growth Regul.* 37, 986–998. doi: 10.1007/s00344-018-9796-9

Danquah, A., Zelicourt, A. D., Colcombet, J., and Hirt, H. (2014). The role of ABA and MAPK signaling pathways in plant abiotic stress responses. *Biotechnol. Adv.* 32, 40–52. doi: 10.1016/j.biotechadv.2013.09.006

Deinlein, U., Stephan, A. B., Horie, T., Luo, W., and Schroeder, J. I. (2014). Plant salt-tolerance mechanisms. *Trends Plant Sci.* 19, 371–379. doi: 10.1016/j.tplants.2014.02.001

Dewey, C. N., and Li, B. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* 12, 323–323. doi: 10.1186/1471-2105-12-323

Dong, L., Liu, H., Zhang, J., Yang, S., Kong, G., Chu, J. S. C., et al. (2015). Single-molecule real-time transcript sequencing facilitates common wheat genome annotation and grain transcriptome research. *BMC Genom.* 16, 1039. doi: 10.1186/s12864-015-2257-y

Fahad, S., Hussain, S., Matloob, A., Khan, F. A., and Huang, J. (2014). Phytohormones and plant responses to salinity stress: a review. *Plant Growth Regul.* 75, 391–404. doi: 10.1007/s10725-014-0013-y

Fan, Z., Li, L., Jiao, Z., Chen, Y., and Zheng, J. (2016). Characterization of the Calcineurin B-Like (CBL) gene family in maize and functional analysis of *ZmCBL9* under abscisic acid and abiotic stress treatments. *Plant Sci.* 253, 118–129. doi: 10.1016/j.plantsci.2016.09.011

Ferdinandez, Y., and Coulman, B. E. (2001). Nutritive values of *Smooth Bromegrass*, Meadow Bromegrass, and Meadow × Smooth Bromegrass hybrids for different plant parts and growth stages. *Crop Sci.* 41, 473–478. doi: 10.2135/cropsci2001.412473x

Fidalgo, F., Santos, A., Santos, I., and Salema, R. (2015). Effects of long-term salt stress on antioxidant defence systems, leaf water relations and chloroplast ultrastructure of potato plants. *Ann. Appl. Biol.* 145, 185–192. doi: 10.1111/j.1744-7348.2004.tb00374.x

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565

Gao, Y. L., Cui, Y. J., Long, R. C., Sun, Y., Zhang, T. J., Yang, Q. C., et al. (2019). Salt-stress induced proteomic changes of two contrasting alfalfa cultivars during germination stage. *J. Sci. Food Agric.* 99, 1384–1396. doi: 10.1002/jsfa.9331

Geng, W., Li, Z., Hassan, M. J., and Peng, Y. (2020). Chitosan regulates metabolic balance, polyamine accumulation, and Na^+ transport contributing to salt tolerance in creeping bentgrass. *BMC Plant Biol.* 20, 506. doi: 10.1186/s12870-020-02720-w

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotech.* 29, 644–652. doi: 10.1038/nbt.1883

Hana, Z.andkarimi, Renesh, B.edre, Julio, S.olis, and Venkata, M.angu, Niranjan, and Baisakh (2015). Sequencing and expression analysis of salt-responsive miRNAs and target genes in the halophyte smooth cordgrass (*Spartina alternifolia* L.). *Mol. Biol. Rep.* 42, 1341–1350. doi: 10.1007/s11033-015-3880-z

Hartmann, L., Pedrotti, L., Weiste, C., Fekete, A., Schierstaedt, J., GöTtler, J., et al. (2015). Crosstalk between two bZIP signaling pathways orchestrates salt-induced metabolic reprogramming in *Arabidopsis* roots. *Plant Cell.* 27, 2244–2260. doi: 10.1105/tpc.15.00163

Hu, L. X., Li, H. Y., Chen, L., Lou, Y. H., and Amombo, E. (2015). RNA-seq for gene identification and transcript profiling in relation to root growth of bermudagrass (*Cynodon dactylon*) under salinity stress. *BMC Genomics.* 16, 1–12. doi: 10.1186/s12864-015-1799-3

Huang, G. T., Ma, S. L., Bai, L. P., Li, Z., Hui, M., Jia, P., et al. (2012). Signal transduction during cold, salt, and drought stresses in plants. *Mol. Biol. Rep.* 39, 969–987. doi: 10.1007/s11033-011-0823-1

Ivushkin, K., Bartholomeus, H., Bregt, A. K., Pulatov, A., and Sousa, L. D. (2019). Global mapping of soil salinity change. *Remote Sens. Environ.* 231, 111260. doi: 10.1016/j.rse.2019.111260

Jin, Y., Jing, W., Zhang, Q., and Zhang, W. (2015). Cyclic nucleotide gated channel 10 negatively regulates salt tolerance by mediating Na^+ transport in *Arabidopsis*. *J. Plant Res.* 128, 211–220. doi: 10.1007/s10265-014-0679-2

Kakumanu, A., Ambavaram, M. M. R., Klumas, C., Krishnan, A., Batlang, U., Myers, E., et al. (2012). Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. *Plant Physiol.* 160, 846–867. doi: 10.1104/pp.112.200444

Kan, L., Liao, Q. C., Su, Z. Y., Tan, Y. S., Wang, S. Y., and Zhang, L. (2020). Single-Molecule Real-Time sequencing of the *Madhuca pasquieri* (Dubard) Lam. transcriptome reveals the diversity of full-length transcripts. *Forests* 11, 866. doi: 10.3390/f11080866

Karan, R., and Subudhi, P. K. (2012). *Approaches to Increasing Salt Tolerance in Crop Plants.* New York, NY: Springer. p. 63–88. doi: 10.1007/978-1-4614-0634-1_4

Kearl, J., Mcnary, C., Lowman, J. S., Mei, C., and Nielsen, B. L. (2019). Salt-tolerant halophyte rhizosphere bacteria stimulate growth of alfalfa in salty soil. *Front. Microbiol.* 10, 1849. doi: 10.3389/fmicb.2019.01849

Komatsu, K., Takezawa, D., and Sakata, Y. (2020). Decoding ABA and osmostress signalling in plants from an evolutionary point of view. *Plant. Cell Environ.* 43, 2894–2911. doi: 10.1111/pce.13869

Kugler, A., K?Hler, B., Palme, K., Wolff, P., and Dietrich, P. (2009). Salt-dependent regulation of a CNG channel subfamily in *Arabidopsis*. *BMC Plant Biol.* 9, 140. doi: 10.1186/1471-2229-9-140

Leena, S., and Eric, R. (2014). LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30, 3506–3514. doi: 10.1093/bioinformatics/btu538

Li, S., Fan, C., Li, Y., Zhang, J., Sun, J., Chen, Y., et al. (2016). Effects of drought and salt-stresses on gene expression in Caragana korshinskii seedlings revealed by RNA-Seq. *BMC Genom.* 17, 1–19. doi: 10.1186/s12864-016-2562-0

Liu, C., Mao, B., Yuan, D., Chu, C., and Duan, M. (2021). Salt tolerance in rice: physiological responses and molecular mechanisms. *The Crop J.* 10, 13–25 doi: 10.1016/j.cj.2021.02.010

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

Luo, D., Zhou, Q., Wu, Y., Chai, X., Liu, W., Wang, Y., et al. (2019). Full-length transcript sequencing and comparative transcriptomic analysis to evaluate the contribution of osmotic and ionic stress components towards salinity tolerance in the roots of cultivated alfalfa (*Medicago sativa* L.). *BMC Plant Biol.* 19, 32. doi: 10.1186/s12870-019-1630-4

Luo, H., Zhou, Z., Song, G., Yao, H., and Han, L. (2020). Antioxidant enzyme activity and microRNA are associated with growth of Poa pratensis callus under salt stress. *Plant Biotechnol. Rep.* 14, 429–438. doi: 10.1007/s11816-020-00620-x

Mahmoud, Z. S., El-Sayed, M. D., Ali, O., and Mostafa, M. R. (2020). Pumpkin seed protein hydrolysate treatment alleviates salt stress effects on *Phaseolus vulgaris* by elevating antioxidant capacity and recovering ion homeostasis. *Sci. Hortic.* 271, 109495, doi: 10.1016/j.scienta.2020.109495

Malik, A., Gul, A., Hanif, U., Kubra, G., Bibi, S., Ali, M., et al. (2020). "Salt responsive transcription factors in wheat", in *Climate Change and Food Security with Emphasis on Wheat*. Elsevier, Pakistan. 107–127. doi: 10.1016/B978-0-12-819527-7.00007-8

Mehlmer, N., Wurzinger, B., Stael, S., Hofmann-Rodrigues, D., Csaszar, E., Pfister, B., et al. (2010). The Ca^{2+}-dependent protein kinase CPK3 is required for MAPK-independent salt-stress acclimation in *Arabidopsis*. *Plant J.* 63, 484–498. doi: 10.1111/j.1365-313X.2010.04257.x

Meng, X., Liu, S., Dong, T., Xu, T., and Zhu, M. (2020). Comparative transcriptome and proteome analysis of salt-tolerant and salt-sensitive sweet potato and overexpression of *IbNAC7* confers salt tolerance in *Arabidopsis*. *Front. Plant Sci.* 11, 572540. doi: 10.3389/fpls.2020.572540

Minoru, K., Susumu, G., Shuichi, K., Yasushi, O., and Masahiro, H. (2004). The KEGG resource for deciphering the genome. *Nucl. Acids Res.* 32, 277–280. doi: 10.1093/nar/gkh063

Pandey, G. K., Kanwar, P., Singh, A., Steinhorst, L., Pandey, A., Yadav, A. K., et al. (2015). Calcineurin B-like protein-interacting protein kinase *CIPK21* regulates osmotic and salt stress responses in *Arabidopsis*. *Plant Physiol.* 169, 780–792. doi: 10.1104/pp.15.00623

Rabbani, M. A., Maruyama, K., Abe, H., Khan, M. A., Katsura, K., Ito, Y., et al. (2003). Monitoring expression profiles of rice genes under cold, drought, and high salinity stresses and abscisic acid application using cDNA microarray and RNA gel blot analyses. *Plant Physiol.* 133, 1755–67. doi: 10.1104/pp.103.025742

Raghavendra, A. S., Gonugunta, V. K., Christmann, A., and Grill, E. (2010). ABA perception and signalling. *Trends Plant Sci.* 15, 395–401. doi: 10.1016/j.tplants.2010.04.006

Regina, A., Kosar-Hashemi, B., Ling, S., Li, Z., Rahman, S. and Morell, M. (2010). Control of starch branching in barley defined through differential RNAi suppression of starch branching enzyme IIa and IIb. *J. Exper. Bota.* 61, 1469–1482. doi: 10.1093/jxb/erq011

Rengasamy, P. (2010). Soil processes affecting crop production in salt-affected soils. *Plant Biol.* 37, 613–620. doi: 10.1071/FP09249

Rhoads, A., and Au, K. F. (2015). Sequencing and its applications. *Genomics, Proteom. Bioinform.* 13, 278–289. doi: 10.1016/j.gpb.2015.08.002

Robben, M., and Gonzalez, J. (2016). RNA-Seq analysis of the salt tolerance response in prairie cordgrass. *The FASEB J.* 30, lb165–lb165. doi: 10.1096/fasebj.30.1_supplement.lb165

Shen, G. M., Ou, S. Y., He, C., Liu, J., and He, L. (2021). Full length sequencing reveals novel transcripts of detoxification genes along with related alternative splicing events and lncRNAs in *Phyllotreta striolata*. *PLoS ONE* 16, e0248749. doi: 10.1371/journal.pone.0248749

Shimizu, K., Adachi, J., and Muraoka, Y. (2006). ANGLE: a sequencing errors resistant program for predicting protein coding regions in unfinished cDNA. *J. Bioinform. Comput. Biol.* 4, 649–664. doi: 10.1142/S0219720006002260

Sicilia, A., Santoro, D. F., Testa, G., Cosentino, S. L., and Piero, A. (2020). Transcriptional response of giant reed (*Arundo donax* L.) low ecotype to long-term salt stress by unigene-based RNA-seq. *Phytochemist.* 177, 112436. doi: 10.1016/j.phytochem.2020.112436

Singh, V., Singh, A. P., Bhadoria, J., Giri, J., Singh, J., V., et al. (2018). Differential expression of salt-responsive genes to salinity stress in salt-tolerant and salt-sensitive rice (*Oryza sativa* L.) at seedling stage. *Protoplasma*. 255, 1667–1681. doi: 10.1007/s00709-018-1257-6

Smart, A. J., Schacht, W. H., Volesky, J. D., and Moser, L. E. (2006). Seasonal changes in dry matter partitioning, yield, and crude protein of intermediate wheatgrass and *Smooth Bromegrass*. *Agron. J.* 98, 986–991. doi: 10.2134/agronj2005.0233

Steinhorst, L., and Kudla, J. (2019). How plants perceive salt. *Nature* 572, 318–320. doi: 10.1038/d41586-019-02289-x

Sun, C., Gao, X., Fu, J., Zhou, J., and Wu, X. (2015). Metabolic response of maize (*Zea mays* L.) plants to combined drought and salt stress. *Plant Soil*. 388:99–117. doi: 10.1007/s11104-014-2309-0

Thoday-Kennedy, E. L., Jacobs, A. K., and Roy, S. J. (2015). The role of the CBL-CIPK calcium signalling network in regulating ion transport in response to abiotic stress. *Plant Growth Regul.* 76, 3–12. doi: 10.1007/s10725-015-0034-1

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucl. Acids Res.* 45, W122–W129. doi: 10.1093/nar/gkx382

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Baren, M., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621

Wang, B., Tseng, E., Regulski, M., Clark, T. A., Hon, T., Jiao, Y., et al. (2016a). Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat. Commun.* 7, 11708. doi: 10.1038/ncomms11708

Wang, X., Gao, F., Bing, J., Sun, W., Feng, X., Ma, X. F., et al. (2019). Overexpression of the *Jojoba Aquaporin* gene, *ScPIP1*, enhances drought and salt tolerance in transgenic *Arabidopsis*. *Int. J. Mol. Sci.* 20, 153. doi: 10.3390/ijms20010153

Wang, X., Li, X., Cai, D., Lou, J., and Liu, F. (2021). Salinification and salt transports under aeolian processes in potential desertification regions of China. *Sci. Total Environ.* 782, 146832. doi: 10.1016/j.scitotenv.2021.146832

Wang, X. L., Li, X. Zhang, S., Korpelainen, H., and Li, C. Y. (2016b). Physiological and transcriptional responses of two contrasting Populus clones to nitrogen stress. *Tree Physiol.* 36, 628–644. doi: 10.1093/treephys/tpw019

Wang, Y., and Yan, L. (2013). Land exploitation resulting in soil salinization in a desertoasis ecotone. *Catena* 100, 50–56. doi: 10.1016/j.catena.2012.08.005

Xiong, J. B., Sun, Y., Yang, Q. C., Tian, H., Zhang, H. S., Liu, Y., et al. (2017). Proteomic analysis of early salt stress responsive proteins in alfalfa roots and shoots. *Proteome Sci* .15, 1–19. doi: 10.1186/s12953-017-0127-z

Yang, L., Jin, Y., Huang, W., Sun, Q., Liu, F., and Huang, X. (2018). Full-length transcriptome sequences of ephemeral plant Arabidopsis pumila provides

insight into gene expression dynamics during continuous salt stress. *BMC Genomics* 19, 717–731. doi: 10.1186/s12864-018-5106-y

Yang, Y., Yu, T. F., Ma, J., Chen, J., and Xu, Z. S. (2020). The Soybean bZIP transcription factor gene *GmbZIP2* confers drought and salt resistances in transgenic plants. *Int. J. Mol. Sci.* 21, 670. doi: 10.3390/ijms21020670

Yi, Z., Chen, J., Sun, H., Rosli, H. G., Pombo, M. A., Zhang, P., et al. (2016). iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* 9, 1667–1670. doi: 10.1016/j.molp.2016.09.014

Yun, S. B., Oh, H., Rhee, S. G., and Yoo, Y. D. (2011). Regulation of reactive oxygen species generation in cell signaling. *Mol. Cells* 32, 491–509. doi: 10.1007/s10059-011-0276-3

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

Zhang, M., Smith, J. A. C., Harberd, N. P., and Jiang, C. F. (2016). The regulatory roles of ethylene and reactive oxygen species (ROS) in plant salt stress responses. *Plant Mol. Biol.* 91, 651–659. doi: 10.1007/s11103-016-0488-1

Zhang, X., Zhang, L., Chen, Y., Wang, S., Fang, Y., Zhang, X., et al. (2021). Genome-wide identification of the SOD gene family and expression analysis under drought and salt stress in barley. *Plant Growth Regul.* 94, 49–60. doi: 10.1007/s10725-021-00695-8

Zhao, C., Zhang, H., Song, C., Zhu, J. K., and Shabala, S. (2020). Mechanisms of plant responses and adaptation to soil Salinity. *Innovat.* 1, 100017. doi: 10.1016/j.xinn.2020.100017

Zhou, Q., Luo, D., Ma, L., Xie, W., Wang, Y., Wang, Y., et al. (2016). Development and cross-species transferability of EST-SSR markers in Siberian wildrye (*Elymus sibiricus* L.) using Illumina sequencing. *Sci. Rep.* 6, 20549. doi: 10.1038/srep20549

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

Zhu, Y., Wang, Q., Wang, Y., Xu, Y., Li, J., Zhao, S., et al. (2021). Combined transcriptomic and metabolomic analysis reveals the role of phenylpropanoid biosynthesis pathway in the salt tolerance process of *Sophora alopecuroides*. *Int. J. Molec. Sci.* 22, 2399. doi: 10.3390/ijms22052399

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 16 June 2022
doi: 10.3389/fpls.2022.885366

Transcriptional Analyses of Genes Related to Fodder Qualities in Giant Leucaena Under Different Stress Environments

*Ahmed M. Bageel, Aaron Kam and Dulal Borthakur**

Department of Molecular Biosciences and Bioengineering, University of Hawai'i at Mānoa, Honolulu, HI, United States

Leucaena leucocephala subsp. *glabrata* (giant leucaena) is a tree legume, whose foliage is used as a fodder for animals because of its high protein content. In spite of being a highly nutritious fodder, giant leucaena foliage has two undesirable secondary metabolites, mimosine and tannin. The amounts of mimosine and tannin in giant leucaena foliage are known to vary under different environmental conditions. Giant leucaena was grown under different salinity, pH and nitrogen availability conditions. It produced the highest amounts of mimosine at pH 6.0–7.0, whereas, variation in soil pH did not affect tannin concentrations. Salinity stress had negative effects on both mimosine and tannin concentrations, while nitrogen abundance promoted both mimosine and tannin production. Seven genes for mimosine and tannin metabolism were isolated from a transcriptome library of giant leucaena. These were mimosine synthase, mimosinase, chalcone synthase, flavanone 3β-hydroxylase, dihydroflavonol reductase, leucoanthocyanidin reductase, and anthocyanidin reductase. The highest level of mimosine synthase activity was observed in the absence of salt in the soils. Mimosine synthase activities had strong positive correlation with mimosine concentrations in the foliage (R^2 = 0.78) whereas mimosinase expression did not appear to have a direct relationship with salt concentrations. The expression of mimosine synthase was significantly higher in the leucaena foliage under nitrogen abundant condition than in nitrogen deficiency conditions, while mimosinase expression was significantly higher under nitrogen deficiency condition than in nitrogen abundance conditions. Mimosine concentrations in the foliage were positively correlated with the expression levels of mimosine synthase but not mimosinase. Similarly, the concentrations of tannin were positively correlated with expression levels of dihydroflavonol reductase, leucoanthocyanidin reductase, and anthocyanidin reductase. Understanding of the environmental conditions that promote or inhibit transcription of the genes for mimosine and tannin biosynthesis should help to design environmental conditions that inhibit transcription of these genes, resulting in reduced levels of these compounds in the leucaena foliage.

Keywords: giant leucaena, mimosine, tannin, environmental stress, gene expression, fodder quality

OPEN ACCESS

Edited by:
Jorge Fernando Pereira, Embrapa Gado de Leite, Brazil

Reviewed by:
Shijiang Cao, Fujian Agriculture and Forestry University, China
Zi Shi, Beijing Academy of Agriculture and Forestry Sciences, China

***Correspondence:**
Dulal Borthakur
dulal@hawaii.edu

Specialty section:
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

Received: *11 March 2022*
Accepted: *13 May 2022*
Published: *16 June 2022*

Citation:
Bageel AM, Kam A and Borthakur D (2022) Transcriptional Analyses of Genes Related to Fodder Qualities in Giant Leucaena Under Different Stress Environments. Front. Plant Sci. 13:885366. doi: 10.3389/fpls.2022.885366

INTRODUCTION

Giant leucaena is a tree legume, which is widely grown in many tropical and subtropical countries as a source for nutritious fodder. It can be managed as a dwarf bush by repeated harvest of its foliage by pruning (Ishihara et al., 2018). On the other hand, common leucaena grows as a bushy shrub even without pruning. Giant leucaena is ∼2.5 times more productive than common leucaena for fodder production (Bageel et al., 2020). Because of the high protein content (18–30%) in its foliage, it is often described as the "alfalfa of the tropics" (Bageel et al., 2020). Although leucaena is considered to be a nutritious fodder, it has two major secondary metabolites, mimosine and tannin, which reduce the fodder quality if they are present in high concentrations. These two metabolites vary in their concentrations depending on the environmental conditions that leucaena is exposed to (Bageel and Borthakur, 2022).

Mimosine is a non-protein amino acid present in all parts of the leucaena plant; young leaves, growing shoot tips, and seeds contain the highest concentration of mimosine (Xuan et al., 2006). The concentrations of mimosine in the foliage and other parts of the leucaena plants change in response to various environmental stresses; therefore, mimosine has been described as a stress-response molecule (Honda and Borthakur, 2021). Mimosine is also secreted in the root exudates to the rhizosphere, where it works as a phytosiderophore and facilitates metallic cation uptake from the soil (Honda and Borthakur, 2020). It also serves as an antioxidant that mitigates free radicals and limits oxidative damage (Honda and Borthakur, 2021). On the other hand, mimosine is considered a toxic compound that negatively affects animals when ingested in high amounts. According to Szyszka and Ter Meulen (1984), consuming low amounts of mimosine (<0.14 g/kg BW) does not cause harmful effects in animals. Therefore, it is desirable to reduce mimosine concentrations in leucaena through inter-varietal and inter-species crosses, and selection. Recently, two key genes that are responsible for synthesis and breakdown of mimosine have been identified; they are mimosine synthase and mimosinase, respectively (Negi et al., 2014; Harun-Ur-Rashid et al., 2018).

Tannins are phenolic compounds that play important roles in forage quality. High amounts of tannins have negative impacts on the foliage quality, it reduces ruminants' digestion (Reed, 1995). On the other hand, moderate amounts of tannins have positive impacts on foliage quality by disrupting protein foam in the rumen that prevents rumen pasture bloating (McMahon et al., 2000). In addition, moderate amounts of tannins reduce the number of internal parasites and increase the titer of small peptides and amino acids without compromising total digestion (Niezen et al., 1995). As a result, livestock can more efficiently produce meat, fur, and milk by consuming foliage with moderate concentrations of tannins compared with tannin-free foliage (Min et al., 1998).

Giant leucaena foliage is considered superior to common leucaena foliage for a number of reasons: (i) it remains in juvenile stage for a relatively longer period of time, making it more palatable and digestible; (ii) it produces much higher amounts of foliage; (iii) it produces much less pods and seeds; and (iv) it produces less amounts of tannins. Cultivars such as K636 and KX2 are known to have low amounts of tannins (Bageel et al., 2020). It was shown previously that foliage yield, and the concentrations of protein, mimosine and tannin change under different soil salinity and pH conditions. The goal of the present study was to determine the expression of a number of genes related to fodder quality under different salinity, pH conditions, and nitrogen availability. The leucaena genome has not been sequenced and genes for tannin biosynthesis have not been isolated from leucaena. Therefore, the coding sequences for these genes were isolated through transcriptome sequencing. To determine if the coding sequences for these genes in giant leucaena differ from those of common leucaena, these sequences were isolated from the transcriptome libraries of both giant and common leucaena and compared. For studying expression of genes related to fodder quality, seven genes were selected; two for mimosine metabolism (mimosine synthase and mimosinase), and five genes for tannin biosynthesis. Among these five genes, two genes were for general phenylpropanoid pathway and three genes were specific for tannin biosynthesis. The phenylpropanoid pathway genes were chalcone synthase (CHS), and flavanone 3β-hydroxylase (F3H) and genes for tannin biosynthesis were dihydroflavonol reductase (DFR), leucoanthocyanidin reductase (LAR), and anthocyanidin reductase (ANR) (**Figure 1**).

MATERIALS AND METHODS

Transcriptome Analysis

Plant Material

Seeds of giant (K636) and common leucaena were collected from Waimanalo Research Station, University of Hawaii. The seeds were scarified, germinated, and grown in pots as described earlier (Bageel and Borthakur, 2022). Three-months old seedlings (shoots and roots) were harvested, rinsed with water, and placed immediately in four different containers (giant shoots, giant roots, common shoots, and common roots) filled with liquid nitrogen.

RNA Extraction

Total RNA was extracted and purified using the reagents and protocol described by Ishihara et al. (2016). To remove contamination of phenolic compounds, the isolated total RNA samples were precipitated with glycogen and resuspended in RNAse-free dH_2O. To confirm the RNA's quality, the RNA yields were analyzed by checking the RNA Integrity Numbers (RIN) using Agilent Technologies 2100 Bioanalyzer, and RIN were 7.30, indicating acceptable level yield and purity (**Supplementary Figure 1**). The samples were then treated with TURBO DNA free Kit to remove possible DNA contamination according to manufacturer's instructions.

Poly-A RNA Sequencing

Cleaned total RNA extracted from the foliage and root samples were combined in two Eppendorf tubes, one for giant leucaena and the other for common leucaena. These two sample tubes were sent to the LC Sciences company for transcriptome

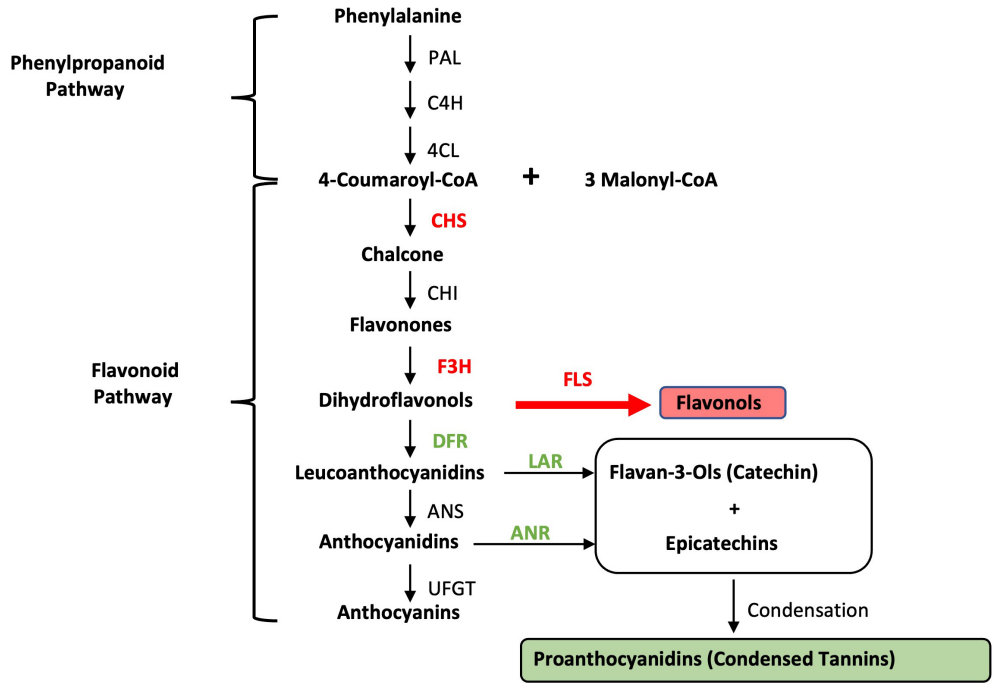


FIGURE 1 | Condensed tannin biosynthesis pathway. PAL, phenylalanine ammonia-lyase; C4H, cinnamic acid 4-hydroxylase; 4CL, 4-coumarate; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3β-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol reductase; LAR, leucoanthocyanidin reductase; ANS, anthocyanidin synthase; ANR, anthocyanidin reductase; UFGT, UDP-glucose flavonoid 3-*O* glucosyltransferase. The genes/enzymes used in this study are shown in bold.

sequencing where Illumina's TruSeq-stranded-mRNA sample preparation protocol was followed. LC Sciences purified mRNA from the samples by using oligo-(dT) magnetic beads which attach to poly(A)-tails of mRNAs. All mRNA were fragmented to small contigs by using divalent cation buffer in high temperature. Paired-end sequencing on the resulting fragments was performed using Illumina's NovaSeq 6000 sequencing system which produced the raw reads.

Transcriptome Assembly

The raw reads were cleaned using Cutadapt-3.4 to remove bases with a quality score less than 20, alongside contamination from the Illumina universal adapters used during sequencing. The removal of adapter contaminations and low-quality reads were confirmed by using FastQC-0.11.9 resulting in high quality reads. The high-quality reads were then *de novo* assembled into a transcriptome using Trinity-2.12.0 that contains predicted mRNA transcripts. Transcripts generated by Trinity were clustered into groups based on shared component sequences that are loosely considered to represent the same gene. The longest transcript in each of the gene clusters was designated as a unigene and was used as a representative for downstream gene-level analyses.

Functional Annotation

The primary unigene library was annotated by the Non-Redundant (nr), Swissprot, Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. The translated amino acid sequences for each unigene were compared against entries in the databases using DIAMOND accelerated BLASTx and alignments with an e-value $< 1e^{-5}$ were used as annotations.

However, these annotations are limited to the unigene cluster as a whole and do not provide the specific, transcript-level information that is needed to create qRT-PCR primers. To determine the appropriate coding nucleotide sequences from the transcripts, annotations from each database were searched for labels of the desired genes. Because names can vary between databases and entries, annotations containing aliases, alternate names, and protein isoforms of our desired genes were also considered. Unigenes with annotations of interest had their related transcripts extracted and compared to the nr and Swissprot databases using blastx. Transcripts were selected if either of their translated amino acid alignments covered a large portion of a desired protein in its alignment; had a high (>90%) similarity; and contained the protein's functionally important domains and residues. These transcripts had their coding nucleotides for the aligned areas used as representative sequences for creating qRT-PCR primers.

Biochemical Assays and Gene Expression

Plant Material

Approximately 100 giant leucaena seedlings were grown in the greenhouse in large pots containing potting mix planting soil

(Sam's Club) for 1 year, and only irrigated with tap water. Thus, the planting soil has lost most of its minerals, and that was confirmed by the EC probe where the dripped water from the soil pots showed close mineral concentration that was shown on sink water. These seedlings were used in the following three experiments.

(i) Twenty of the seedlings, which showed uniformity in size and health condition, were selected for the experiment. These plants were pruned to develop fresh foliage uniformly, and grouped into four groups representing four different salinity levels (Zero, 50, 100, and 150 mM NaCl). Every group was irrigated every other day with 1/18 strength of Hoagland solution adjusted to pH 7.0 and the salinity level, which the group is representing for 3 months. Then, 300 mg of young leaves were harvested and placed in liquid nitrogen immediately, and kept in −80°C for RNA extraction. The rest of the green foliage was harvested for mimosine and tannin quantifications.

(ii) Another group of twenty-five seedlings, which showed uniformity in size and health condition were selected for this experiment. These plants were pruned to develop fresh foliage uniformly, and grouped into five groups representing (pH 5.0, 6.0, 7.0, 8.0, and 9.0). Each group was irrigated every other day with 1/18 strength of Hoagland solution adjusted to the pH level which the group is representing for 3 months. Then, 300 mg of young leaves were harvested and placed in liquid nitrogen immediately, and kept in −80°C for RNA extraction. The rest of the green foliage was harvested for mimosine and tannin quantifications.

(iii) Ten seedlings that showed uniformity in size and health condition were selected for this experiment. These plants were pruned to develop fresh foliage uniformly, and grouped into two groups representing nitrogen abundance and nitrogen deficiency conditions. The first group was irrigated every other day with a regular 1/18 strength of Hoagland solution adjusted to pH 7.0; whereas, the second group was irrigated every other day with nitrogen-free 1/18 strength of Hoagland solution adjusted to pH 7.0 for 3 months. Then, 300 mg of young leaves were harvested and placed in liquid nitrogen immediately, and kept in −80°C for RNA extraction. The rest of the green foliage was harvested for mimosine and tannin quantifications.

Mimosine Quantification

Mimosine was isolated as described earlier (Bageel and Borthakur, 2022). Mimosine concentration was determined by using high performance liquid chromatography (Waters 2650) with a C18 column and UV detection at 280 nm, using an isocratic carrier solvent of 0.02 M O-phosphoric acid and a linear flow rate of 1 mL per min for 6 min (Negi et al., 2014).

Tannin Quantification

Total tannin was isolated using the Folin-Ciocalteu's method and quantified as described by Makkar (2003). Briefly, 1.0 g of leaves was placed in the oven for 16 h at 50–52°C to dry; then the sample was grinded to a fine powder. 200 mg of the grinded sample was added to 10 mL of aqueous acetone (70%) in a 50-mL falcon tube. The mixture was subjected to ultrasonic treatment for 5 min at 100 Watt, and then centrifuged for 10 min at 3,000 g at 4°C to remove plant debris.

$$(\text{Total Tannins} = \text{Total Phenolics} - \text{Non-tannin Phenolics})$$

Statistical Analyses

Plants used in this study were grown in pots in a greenhouse and organized in a completely randomized block design with five replications. For gene expression studies, qRT-PCR experiments were conducted with five biological and three technical replicates. For analyses of mimosine and tannin contents in giant leucaena grown with different treatment conditions, Student's *t*-test was used with *p*-value of 0.05. Similarly, statistical significance for differences in gene expression levels between two sets of plants receiving separate treatments was determined using Student's *t*-test with a cutoff *p*-value of 0.05.

RESULTS

Bioinformatic Analyses

The results of quality analyses of the raw data from Illumina sequencing were checked by FASTQC tool, and e-value distribution of the unigenes (**Table 1a** and **Supplementary Figure 2**). The quality of the assembly was also judged by the length of unigenes, GC content, total assembled bases and N50 (**Table 1b**). The number of unigenes identified for giant and common leucaena were 130,825 and 116,332, respectively. On an average, there were around two transcripts for each unigene in the sequences for both giant and common leucaena. The N50 values, indicating the median contig lengths for the unigenes were 895 and 929 for giant and common leucaena, respectively, suggesting that the sequence assemblies were of reliable quality (**Supplementary Figure 3**).

These unigenes were annotated using group blast against sequence databases, including nr, Swissport, GO, and KEGG. For giant leucaena, the annotation with the nr database had the highest aligned unigenes (56,623 unigenes), followed by the annotation with the Swissport database (48,450 unigenes), followed by the annotation with the KEGG database (17,944 unigenes). With the three databases combined, a total of 65,321 unigenes were annotated. For common leucaena, the annotation with the nr database had the highest aligned unigenes (48,983 unigenes), followed by the annotation with the Swissport database (43,895 unigenes), followed by the annotation with the KEGG database (14,883 unigenes). With the three databases combined, a total of 58,848 unigenes were annotated (**Table 1c**).

Among the unigenes annotated in the nr and Swissport databases for giant leucaena, 51.6 and 41.6%, respectively, had an E-value < 1.0E-50. For common leucaena, the unigenes annotated in the nr and Swissport databases, 51.7 and 42%, respectively, had an E-value < 1.0E-50. Transcriptome sequences with very small values are desirable because small E-values indicate high accuracy. On the other hand, high E-values reduce

TABLE 1 | Representing (a) bioinformatics raw data quality, (b) gene assembly, and (c) functional annotation.

a

Subspecies	Raw reads	Raw bases	Valid reads	Valid bases	Valid%	Q20%	Q30%	GC%
Giant	48,680,970	6.86G	47,927,522	6.70G	98.45	98.30	94.38	45.70
Common	43,049,726	6.07G	42,196,800	5.89G	98.02	98.27	94.32	46.95

b

Subspecies	Type	All	GC%	Minimum length	Median length	Maximum length	Total assembled bases	N50
Giant	Transcript	268,663	43.29	186	565	13,399	232,314,065	1325
	Unigene	130,825	42.81	201	360	13,399	81,197,946	895
Common	Transcript	235,658	42.40	184	568	10,258	201,362,860	1,300
	Unigene	116,332	41.62	201	355	10,258	72,153,253	929

c

	Annotated unigenes			
Subspecies	NR	Swissprot	KEGG	Total
Giant	56,623	48,450	17,944	65,321*
Common	48,983	43,895	14,883	58,848*

**Some of the genes were common to all three databases.*

accuracy and increase probability for error. In this study, 3.3% of the transcriptome sequence in the nr database comparison showed an E-value of zero. The highest E-value of 1e-04 to 1e-05 was observed for only 2% of the transcripts in the nr database comparison. The rest of the giant leucaena transcripts had intermediate E-values between these lowest and highest values. The common leucaena transcripts also showed similar E-values (**Figure 2**).

The number of transcripts found for different genes in giant and common leucaena (**Supplementary Table 1**) were different; some of the genes had high numbers of transcripts in giant leucaena and low in common leucaena and vice versa. For example, mannose/glucose-specific lectin-like protein isoform X2 had 11,843 transcripts in giant leucaena compared to only 105 transcripts in common leucaena. On the other hand, abscisic acid receptor PYL6 had only 24 transcripts in giant leucaena compared to 723 transcripts in common leucaena. This suggests differential expression of the genes in the two subspecies. This is further supported by the observation that the transcript per million (TPM) for the same genes also vary greatly between giant and common leucaena (**Supplementary Table 1**). Giant leucaena genes which are up-regulated or down-regulated relative to common leucaena are shown in **Supplementary Table 1**.

Expression of Genes Related to Fodder Quality

From the transcription analysis the nucleotide sequences for two genes for mimosine metabolism, five genes for tannin biosynthesis in giant leucaena were obtained (**Table 2** and **Supplementary Table 2**). It is known that mimosine in the leucaena foliage is degraded under certain environmental conditions. Therefore, the gene sequence for mimosinase was also selected for gene expression studies. Previous studies show that the mimosine contents in the leucaena foliage change under different environmental conditions such as soil salinity, pH and nitrogen content (Honda and Borthakur, 2021; Bageel and Borthakur, 2022). Similarly, the tannin contents of the leucaena foliage are influenced by a number of environmental factors. The effects of some of these environmental factors on expression of genes for mimosine biosynthesis and degradation, and tannin biosynthesis were determined by qRT-PCR analyses.

Expression of Mimosine Synthase and Mimosinase Under Different Salinity Conditions

The highest level of mimosine synthase activity was observed in the absence of salt in the soils (**Figure 3A**). At 50 and 100 mM salt concentrations, the expression of mimosine synthase decreased progressively. Leucaena could not tolerate and grow in 150 mM salt concentration and therefore data could not be recorded. Thus, the presence of salt appears to inhibit mimosine synthase activity. On the other hand, mimosinase expression did not appear to have a direct relationship with salt concentrations; its expression was highest at 50 mM salt concentration and lowest at 100 mM salt concentration (**Figure 3B**). Mimosinase expression was intermediate in the absence of salt in the soil. Thus, it appears that some amount of salt is inducive for mimosinase activity while too much salt is inhibitory for mimosinase activity. Mimosine concentration in the leucaena foliage was negatively correlated (R^2 = 0.88) with salt concentrations in the soil (**Figure 3C**). However, mimosine synthase activities had strong positive correlation with mimosine concentrations in the foliage (R^2 = 0.78). On the other hand, mimosine concentration had no correlation with mimosinase activity in the foliage (R^2 = 0.11).

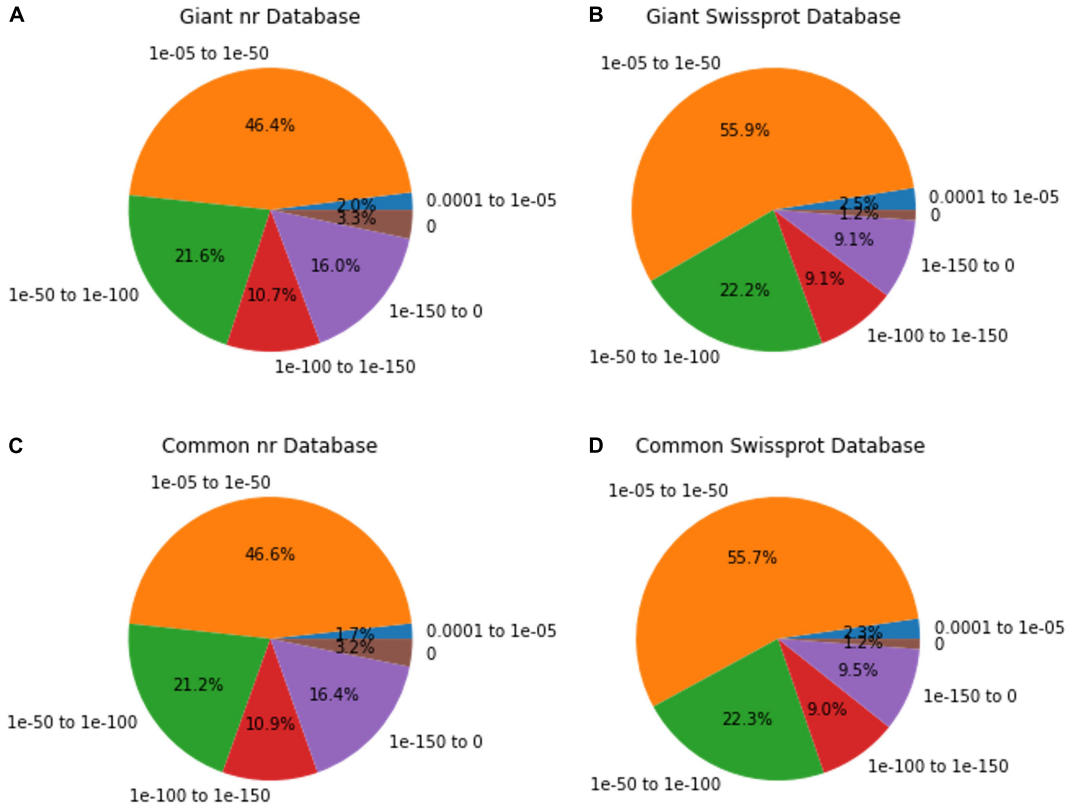


FIGURE 2 | E-value distributions of annotated giant and common leucaena unigenes. The E-values of the highest-scored BLAST hit were identified for each unigene. **(A)** Aligning giant leucaena against the nr protein database, **(B)** aligning giant leucaena against the Swissport protein database **(C)** aligning common leucaena against the nr protein database, and **(D)** aligning common leucaena against the Swissport protein database.

Expression of Mimosine Synthase and Mimosinase Under Different Soil pH Conditions

The expression of mimosine synthase in the leucaena foliage was found to be highest at pH 6.0 and lowest at pH 5.0 (**Figure 4A**). The foliage of plants grown at pHs 6.0–9.0, showed progressively lower levels of mimosine synthase expression activity. Mimosinase expression in the leucaena foliage was relatively low at pHs 5.0, 6.0 and 7.0. and was much higher at pHs 8.0 and 9.0 (**Figure 4B**). Thus, alkaline pH was favorable for mimosinase expression. Mimosine concentrations in the leucaena foliage were highest at pHs 6.0 and 7.0 (**Figure 4C**). Mimosine concentrations in the leucaena foliage decreased with the increase in the soil pH above 7.0. Mimosine concentration in the foliage appears to be directly correlated with mimosine synthase expression ($R^2 = 0.78$).

Expression of Mimosine Synthase and Mimosinase Under N-Abundance and N-Deficiency Conditions

The expression of mimosine synthase was significantly higher in the leucaena foliage under nitrogen abundant condition than in nitrogen deficiency condition (**Figure 5A**). On the other hand, mimosinase expression was significantly higher under nitrogen deficiency condition than in nitrogen abundance condition (**Figure 5B**). Mimosine concentration in the foliage was much higher when the plants were supplied with combined nitrogen compared to nitrogen deficiency condition (**Figure 5C**).

Expression of Five Tannin Biosynthesis Genes Under Different Salinity Conditions

The expression of CHS and F3H were highest when the leucaena plants were grown in the presence of 50 mM salt, while the expressions of DFR, LAR and ANR were highest when the plants were grown in the absence of salt (**Figure 6A**). Tannin concentration in the leucaena foliage was also highest when the plants were grown without adding any salt to the soils (**Figure 6B**). The tannin concentrations progressively decreased with addition of 50 and 100 mM salts to the soils. Thus, tannin concentrations in the leucaena foliage appear to be negatively correlated with the amount of salt present in the soils ($R^2 = 0.79$). CHS, F3H, and DFR had weak correlations with tannin concentrations in the foliage ($R^2 = 0.34$, 0.52, and 0.47), respectively. On the other hand, tannin concentration had strong correlation with LAR and ANR gene expressions in the foliage ($R^2 = 0.74$ and 0.88), respectively.

TABLE 2 | Sequences and primers of the genes of interest.

Gene	Sequence	Length	Forward primer	Reverse primer
Mimosine synthase	GTTTTGAGCTGGACTTCACGGAGCTGCCCTAGATCGCTGTCCATTGT TATCGGCTGCTTCGTGGAGGCAGGGTTAGAGAGTAAGCTATCACCAT GGCAGTGGAACGGAACGGAATTGCCAATGATATTACAGAATTGGTTG GCAAAACCCCATTAGTATATCTTAATAAAGTTGTGGATGGCTGTGTGG GCCGGATTGTTGCCAAACTAGAGTTTATGGAACCCCTCTCTAGTGTC AAAGACAGGATCGGCTATAGTATGATTGCTGATGCAGAGGAAAAAGGT CTTATCACGCCTGGAAAGAGTGTCCTTATTGAACCAACAAGTGGTAATA CTGGCATTGGATTAGCCTTCACGGCAGCAGCAAAAGGTTACAGGCTT ATTATTGTAATGCCTGCTTCAATGACTCTTGAAAGAAGAATGGTTCTGA AAGCTTTAGGGGCTGAGTTGGTGCTGACGGATCCTGAAAAGGGAATA AGAGGTGCATTTCAGAAAGTTGAAGAAATATATGCTAAGACACCTAATT CCTACGTACTTCAACAGTTTGAAAATCCTTCCAATCCCAAGATGCATT ATGACACTACTGGCCCAGAGATTTGGAAAGCCACTGAAGGGAAAATTG ATGTACTTGTAGTAGGGTTAGGTACTGGTGGCACAATAACAGGTTCTGG CAAATTTCTTAAAGAACAAAATCCCAATATTAAGGTTTGTGCTGTTGAAC CTGCTGAAAGTCCCGTGCTCGTTGGTGGACAGCCTGGTTCTCATGG GATTTATGGGATTGGTGCTGGTTTTATACCTAAGGTGGTGGATGTCAGT CTTCTTGACGAAGTTGTTCATGTAACAACCTCTGAAGCCATTGAAACTG CAAAGCTTCTTGCAGTGAAAGAAGGCCTATTTGTGGGAGTTTCATCTGG AGCTGCAGCAGCCGCTGCAATTAAGGTTGCAAAACGACCAGAAATGG CTGGAAAGCTTATTGTAACGATTTTTCCAAGCTGCGGTGAGAGATATCTG TCGTCTGCGTTGTTTGAGTCGATCAGAGAAGAATCTGAGAACATGACTT TTGAGCCATAAAATTTGGATTTCAAGGCTTTCAAGCTACTTTAGTTCTGAT ACTGTTGCAATGCCGGTGGCTTTCGTGGTTGTGATTCTGAAATAAAAATT CGTAGAAAAGTTGTAGCAAAAAACCCCCAAACTTTTAGTGTCATATTTGG TTGTCTTTGACAGTTTGATTCTGTTTGAATAATGTTGGTTATAATCGGCAAA AAAAAAATGTAACAAAAGAAAATAGTGTGGTTCTTGAACATTATTTAATCAA GATGGAACGTTTTTCTGTTTTACTATAATCGGCAGATCGGAAGAGCACAC	1,364 (102)	5′-TCCTTCCAATC CCAAGATGC-3′	5′-CCACCAGTACC TAACCCTACTA-3′
Mimosinase	CTGGCTTTTAATTGTCCAGGTGAAGTTGTGGTTTGCTACACGGCTAATCG AAGAGCATAAGCAAAACCTCTATAGCTCTTTCTTGATCTTCGGAGATGGC TTTGTCATCAACTTTTCTCAACCCTCTCGTTTCTTCGGTCACTGTTAACCC TCACCCTAAGATCACAAATGGAAAGGGGTTCAGAGTAAATTGCTTAATCA GAACACAGCAGACTGTTATCAAAACGGATGCGAAGGAGAATGCTGCGG TGCTTACACCAGGAAAAAAAGTGGAAAAAGAACCAAGTGTGTCTACAGT GTTAGCGAATTACCATGCTGACTGGGATCCTTTCGAGGCAACTTCTACA CCAATTTACCAAAGTGCTACTTTCAGAATGAAAAATGCAACCGAATATAAT GAGTATTACTATTGTAGAGTTGGAAATCCTACCACGAGTACTTTGGAAAA GATATTAGCGGAGATTGAACACGCAGAATATGTCACCTGCTTCACTAGT GGAATGTCTGCTTTGACTGCTGTTTGCGAACTTGTTTCTCCTGGGGATG AAATCCTTACTGTAGAGGATATCTATGGGGGTTCATATGGCTTCATAGAA AATCTAATGGTCAGAAAGTCAGGAGTCACAGTGAAGAGAGTAGATACA AGCAAACTAGAGAACGTGAAAGCTGCTATTACTGATAAGACTAAGTTG GTGTGGCTAGAGTCTCCCAGTAATCCTCAGCTAAAAATCTCTGATATCC GAGAAATAGCCAAGATAGCCCATGCACATGGTGCTATTCTATTTATTGA CAATTGTATAATGTCTCCTCTGTTAAGTCATCCATTGGAATTGGGAGCA GACATCGTTATGCATTCAGCTACAAAATTTATTGCTGGAAACAGCAGT TGCATGGCTGGCAGTCTTGCCACGAACAACAAAGAGTTGGCCGATCT ACTACTCTCTTACCGAAGTGCGACGGGCTGTGGGTTGTCTCCACAGG ACGCCTGGATTTGTTTGGAAGGAATCAAAACATTGCCCCTGCGAGTTG AGGAGAAACAGAAAAATGCACTAACTGTTGCTAATTATCTTCACAACAAT CCTAAAATAACAAAAATAAACTATCCTGGTCTTCCCGATAATCCTGGATA TGAGTTGCATAAATCTCAGTCAAAAGGGCCAGGATCTGTCATGAGCTTT GAAACAGGCTCACTGCCACTCTCAAAACAAATCGTTGAAGATACTAAA TTCTTCAGCAAGATTGTTGGTTTTGGTGGTGTTGGTTCCGCCATATGCT TGCCTTGGTATACATCCCATAAGGCCATTCCGGAGCCAGAGAAAATTAG AATGGGTATAACTAAAGATCTTGTACGTGTCTCTGTGGGAATTGAGGAT GTTCAAGATCTCATTCAAGATCTGGATAATGCAATGTCAACTCCTACGTT CTGAAGTTCGTATGTCAAACGTACTGCTTATTTGCTTCAAATAATGAGATC GATCGAGGTTGCTGATGTTTTGTTGACTCTATTGCCTCTGTTATCTCTTTG CATTGTATTTGGCTGTAACAGTTTGGAACAGTACTAAGCTTATTATTGAAG ATAATCCTATTTTTCATTTGGCACTGGTGTCTTTAAATAGTTTACCTTCCCC TCTCTCTCTCTACCTTCTTGCTTAATTTATTGCATTTTACGTTTCGCGAC GATAATTGTATATAATGATGATG	1,701 (98)	5′-CTGCGAGTTG AGGAGAAACA-3′	5′-ATCGGGAAGA CCAGGATAGT-3′

(Continued)

TABLE 2 | (Continued)

Gene	Sequence	Length	Forward primer	Reverse primer
Chalcone synthase (CHS)	CTCGCAATTTTCTGCAACTTCACATCAAACAAATCCCAATATCTG TAATTAATCTTGTTCATCAAAATGGTGAGCGTTGATGAGATCCGC AATGCCCAGAGAGCAGAAGGGCCGGCCACCGTGATGGCCAT CGGAACGGCCACTCCACCGAACTGTGTTGATCAGAGTACATAC CCTGATTACTACTTCAGAATCACTAAAAGCGAGCACAAGACCG AGCTCAAGGAGAAATTCAAGCGCATGTGCGAAAAGTCAATGA TCAAGAAGAGATACATGCACTTGACGGAGGAGATCTTGAAGG AGAACCCAAACGTGTGTGAGTATATGGCTCCTTCTCTGGATGCA AGGCAAGACATGGTGGTCGTGGAAGTCCCCAAGCTCGGCAAA GAGGCCGCCACCAAAGCCATCAAGGAGTGGGGCCAGCCCAA GTCTAAGATCACTCACCTCATCTTCTGCACCACCAGCGGCGTCG ACATGCCCGGCGCCGACTACCAGCTCACCAAGCTCCTCGGCC TCCGCCCCTCCGTCAAGCGTTACATGATGTACCAGCAGGGCTG CTTCGCCGGCGGCACGGTCCTCCGCCTGGCCAAGGACCTGGC CGAAAACAACAAGGGAGCTCGCGTCCTCGTCGTCTGCTCCGA GATCACCGCCGTCACTTTCCGTGGACCCAGCGACACCCACCTG GACAGCCTCGTGGGCCAAGCCCTCTTCGGAGACGGTGCGGCCG CCGTCATTGTCGGCTCTGACCCCCTCCCTGTGGAGAAGCCTCTG TTTGAGCTTGTGTGGACGGCTCAGACCATTTTACCGGACAGTGAA GGAGCCATTGACGGACACCTCCGCGAGGTTGGCCTCACCTTCC ATCTACTCAAGGACGTTCCCGGGCTGATCTCGAAGAACATAGAG AAGGCATTGGTGGAAGCGTTCCAGCCACTGGGGATATCGGATT ACAACTCCATCTTCTGGATTGCTCACCCGGGAGGGCCTGCCAT TCTTGACCAGGTAGAGGCGAAGCTGAGCTTGAAGCCAGAGAAG ATGAGGGCCACAAGGCACGTGCTGAGTGAGTATGGGAACATGT CGAGTGCTTGCGTGTTGTTCATTTTGGACGAAATGAGGAGGAA GTCTGCAGAAGATGGGCTCAAGACCACTGGTGAAGGACTTG AATGGGGTGTTCTGTTTGGATTCGGGCCTGGGCTCACCGTTGA AACTGTTGTTCTTCACAGTGTGGGTACTAACTAATTGGTCCAA TAGAACGCTGAAAATGAACGGGGAGGCCCAATCGTGAATAAC TTATACTTCTTCTTCCACTTTTTGGGAGACTCTTGGATGTTTTTAT TATTTTTTTAGCATGTACTGTTTATTAAAAGTTAGCCTGTCTTTAG TGTGGTCTGTGTAGAGTGCAAAAGAGTTTGCAATATAAAAATAC GATATTGATATCCAAGCAAGATC	1,446 (95)	5′-GGGCTGATCT CGAAGAACATAG-3′	5′-GGGTGAGCA ATCCAGAAGAT-3′
Flavanone 3ß-hydroxylase (F3H)	CCTCCCCTTCCTTTTCTCCATCTACTCTCACTCCGTCTTCTTCA ATCTTCACTCGTCACCGCACCACTCAGCCCCATGGCGCCCG CCAAGACTCTCACTTCCCTTTCCGACGAGAACACCTTCGACT CTCGCTTCCTCCGCGACGAAGACGAGCGCCCCAAGGTCGCTT ACAACCAATTCAGCAATGATATCCCCATCATCTCTCTCGCCG GCATTGATGACGTCGACGGTCGCCGCGCCCAGATCTGCCA GCAGATTGTTCAGGCTTGCGAGGATTGGGGCATCTTCCAG GTCGTCGATCACGGCGTCGACACTAAACTCGTCTCCAGCAT GACCAGTCTCGCCAAGGAGTTCTTCGCCATGCCCCCCGAA GAGAAGCTCCGCTTCGATATGTCCGGCGGCAAGAAGGGCGGT TTCATCGTCTCCAGCCATCTCCAGGGGGAGATGGTGCAGGATT GGAGAGAGATCGTGACTTACTTCTCGTACCCAATTAGGAACAG GGACTACTCCAGGTGGCCGGACAAGCCGGAATCCTGGAGGA AGGTGACGGAGCGATACAGCGAGGACCTGATGGGGTTGGCAT GTAAGCTGCTGGAGGTTCTGTCGGAGGCGATGGGTCTGGAGA AGGAAGCCCTGACGAAGGCGTGCGTGGACATGGACCAGAAGG TGGTGGTGAATTACTACCCGAAATGCCCGCAACCGGACCTCACC CTCGGACTCAAGCGTCACACTGACCCGGGAACCATCACCCTCC TCCTTCAGGACCAGGTCGGGGGACTCCAGGCCACCAGGGATAA CGGCAAGACGTGGATCACGGTCCAGCCTGTGGAAGGCGCTTTC GTTGTCAATCTCGGGGATCACGGTCATTTCCTGAGCAACGGGAGG TTCAAGAATGCGGATCACCAGGCGGTAGTGAACTCGGAGCACAGC CGTCTGTCGATAGCGACTTTCCAGAACCCGGCGCCAGATGCGATAG TGTACCCGCTGAAGATTAGGGAGGGAGAGAAGTCGGTGATGGAG GAGCCGATAACGTTCGCGGAGATGTACAGGAGGAAGATGAGCAAG GACCTGGAGCTGGCGAGGTTGAAGAAGATGGCTAAGGAGGAGAAG CAATTGCAGGAGATGGAGAAGGCCAAATTGGAAGGGAAGCCCATCGA GCAGATACTTGCCTGATCGACCGAATTTCGATATGTTTCATTTAGTACG TATTAGTTGGGTGGTTATGCTATAGGTGACGTAAACAACAATATGCAGTT			

(Continued)

TABLE 2 | (Continued)

Gene	Sequence	Length	Forward primer	Reverse primer
	GGCGTTGGCAGTGTTTTGCTGGATTTATTTATGGGTCGTGTGTTTTTCC CTGTAAACCACGTTACTTAGTCTTAAACATGAACAAATAAGTTGTGCCC CCCTCATCGTTTTCTTTCCGGTTATCCTTTTGTAACGTATTATATTCCAAA TAGAATTAAGAGAC	1,419 (116)	5′-CCCAATTAGGA ACAGGGACTA-3′	5′-GAACCTCCAGC AGCTTACAT-3′
Leucoantho-cyanidin reductase (LAR)	GAGAGAGAGAGAGAGAGAGAGAGAGAGAGCATACTGAGACCTTTAT CATTTGACTTTTTCGTCTTTTCATCTTCGTCGTCGTTGTCGTTGTCAACAT GACTGTGTTGCCTGCTGTTGTTGGGAAGAAGGACCGGGTGCTCATTGC TGGAGCCACAGGTTTCATCGGTCAGTTCGTGGCGGAGGCGAGCCTGG CAGCGGGTTACGCCACCTATTTGCTCGTACGGCCGGGACCCTTCAGCC AATCCAAGTCTTCCCTTATGAAAGGGCTCCAAGAGAAAGGTGCAGTTGT CATTAATGGGGCGATAAATGAGCAGGAATTAATAGAGAAGATTCTAAGAG ATTATGAGATAGAGATTGTGGTATCAGCTGTGGGAGGCAGAAACTTGCTT GACCAGCTCGTCCTGGTGGAAGCCATGAAATCTGTTAAAACTATCAAGA GGTTCATGCCATCGGAATTCGGGCATGACGTGGATCGGGCTGACCCCG TGGAGCCGGGTTTGGGGATGTACAAACAGAAGCGTCTGGTGAGGCGC GTGGTGGAGAAGTCTGGGATCCCCTACACTTACATCTGCTGCAACTCCA TCGCTTCTTGGCCTTACTACGATAACTGCCACCCTTCCCAGAGTCCTCCG CCCTTGGATCAATTGCACATCTACGGTGATGGCACTGTCAAAGCGTAC TTCGTTGGTGGACCTGACATTGGGAAATTCACAATGAAAGCTGCAAGTGA TATCAGAGCCCTGAACAAGAGCCTCCATTTTCAACCAGCGGCCAACTA CTATAACATGAGCGAGCTCGCTTCCATGTGGGAAACCAAAGTTGGTCG CACCATTCCCAGAGTCATCATCACTGAAGATCACCTCCTTGCTGTTGC GGCTGAGAAAGTTGAGCCGAGGAGCTTTGTGGCATCGTTTACACATG ACATCTTCATCAAAGGGTGCCAGATCTTCAAGCCGGACGGGCCAGG TGATGTGGAAATCGGTTCACTGTATCCTGAAGAACAAATAAGGAGGTT GGAGGATTGCTTTCAGGATTTTGTTCTGATGATACCGCCGGCGGAAAC CAAGAAGACGACAACAACGGAGCTTAACTGTACTGATAACGGCGTTG CGACCAACAACACCGGTGACAGCAACGACACTACTGCCACCGCCAC CGCCCCCACAAACTCATCGGCTGTGGTAAAGCCGGTGCCAATCACGG CTATGTGTTGATGATCGATCAATCATGATGATGATGATGAAATGAGTTCA TATGATATTTAGTAGTATTTGATCTCTCTCTCTCTCTGTTTTACCATTATTA TTGGGAATTTAATTTGTGAGACGGGAATCTTGTTTATTCATATGGTGGTT GGATGTTGCTGGCAGTGTTTGAACAAGAATAAACCCATATGAGTAATAT AATATTTTAAGGCTATTTTGTTCTTGTAGTATCCCACAATTGATATCTCTT CACATCAATTGATGCTGTAATAAATTTTTATTAAAATTCTTATATGACT GTTCCCAAATACGC	1,520 (112)	5′-AAAGGGCTCCA AGAGAAAGG-3′	5′-CCACAGCTGA TACCACAATCT-3′
Dihydroflavonol reductase (DFR)	ATTCAATTGACTTTTCGTCCCCAATAACACCCTCCTCCTCTTTACCCTT TATATTATATTGCTCATTCCGGCCAGTTCCGAGAAAGAGACTGCAGAGT GAGAGATAACCATGGGTTCAGTGTCTGAGACTGTTTGTGTCACTGGC GCTGCCGGCTTTATCGGTTCATGGCTCATCATGAGACTCTTGGAGCG CGGCTACACCGTTCGAGCCACCGTGCGCGACCCTGCTAACATGAAG AAGGTGAAGCATTTGCTACAGCTGCCAAAAGCCGACACACACCTCAC TCTGTGGAAGGCCGACCTTGCCCACGGGGGGAAGTTTCGACGACGCC ATCAATGGCTGCACTGCAGTTTTCCATGTCGCCACCCCCATGGATTTC GAATCCAAGGATCCTGAGAATGAAGTGATACAACCAACGATCAATGGT TTATTGAGCATCATGAAAGCATGCCTCAAGGCCAAAACTAGAAGATTG ATATTCACGTCCTCCGCCGGAACTGTGGATGTCACAGAGCACCAAAA ACCTGTGTTTGACGAGACTTGTTGGAGTGATGTCGAGTTCTGTCGCAG AGTTAAGATGACTGGTTGGATGTATTTCGTTTCAAAAACACTGGCGGAG CAAGAGGCATGGAAATTTGCCAAAGAGAATAACATGGATTTCGTCTCTAT AATTCCTCCTCTCGTTGTTGGCCCCTTTCTCATGGATTCAATGCCTCCC AGTCTCATTACCGGCCTTTCTCTTATCATCGGGAACGAAGCACATTATCC AATCATAAAGCAAGGTCAATTTGTTCACTTGGATGACCTCTGCATGGGT CACATATTCTTGTTAGAGCATCCTGAAGCAAAAGGAAGATACATGTGCT CTTCCCATGATGCCACCATTCATGACGTTGCAAGAATGCTTAGTAAAAA TTACCCTGAATACAATATCCCCACCAAGTTTAAAAACATCCCAGATGAG TTGGAGATCATCCACTTCTCTTCCAAGAAGATTAAAGATTTGGGATTCC AGTTTAAATACAGCTTAGAGGACATGTTCACAGCAGCTGTTGACACCT GCAGGGAAAAAGGGCTTCTTCCAAAGCCTGTTGCAGAAATCCCTGTC AATGGCACCGCAGCTGATCAGAAGTGATTTGGGAAATGTATCCTCGTG TTCGCCATAACCTTTGGCTCTTCTATATATTTAGTTTTTGGTGTTCTAAA TTACGGCTTGGGATATTTATGGGGTCTGGAAATAAGAATGGCTTAGGAC CCGATGGCGGCGGAG	1,270 (204)	5′-CGAGTTCTGTC GCAGAGTTAAG-3′	5′-GCTTCGTTCCC GATGATAAGAG-3′

(Continued)

TABLE 2 | (Continued)

Gene	Sequence	Length	Forward primer	Reverse primer
Anthocyanidin reductase (ANR)	GGCAGTCAGTGGCTACAGAAGAGAGCGCCGGTAGATAGGAGGTGTG ACTGTTTGTGGGATTTGGAATCGTAGATTTGGATTCCCAAGTCTGGGTA GCTGAATCCTAACGAAAAGGGAAAAAATCAAGCCTTCTTCTTGGTGGC GTCGTTGGTCATCACGTGCTCACCTAAGTCCCCTTCCACCTCTGTTG TTCTCACTCAACTCAGCACCCACAAAGCCACCTTATATTAGTTCTGCC GTCTCCCTGCATCAACCAATCGGGAAGAAGAAGAAGAAGATGATGAAG GCGTGCGTAGTGGGAGGAACCGGGTTCGTCGCTTCTGAGCTCGTTA AGCAGTTGCTTCTCAAGGGATATGCTGTTAATACAACTGTTCGAAATCC AGATGATCGCAAAAAAATTTCCCATCTCATAGAACTGCAAAAATTGGG AAAATTGAACATATTTGGAGCAGATCTGACTGAAGAAGACAAATTTGA TGCCCCTATTCTTGACTGTGAACTTGTCTTCCATGTTGCTACTCCTG TGAACTTTGCTTCTGAAGATCCTGAGAATGATATGATCAAACCAGCA ATCCAAGGCGTATTGAATGTCTTGAGAGCATGTGCGCGAGCAAAA GCAGTTAAACGAGTGATCTTGACATCTTCAGCAGCTGCTGTCACA ATAAATGCATTAGAAGGGACTGGTCTGGTTATGGATGAAAACAACT GGACTGACGTTGATTTCTTGAGCACTGCAAAACCACCTACTTGG GGATATCCTGCCTCCAAAGCACTAGCAGAGAAGGCTGCATGGAA ATTTGCCGAAGAAAATCATATAGATCTCATCACTGTGGTACCTACTA TGATGGGTGGTGCTTCTCTCACTCCAGAAATCCCAGGCAGTCT TGGTCTCGCCATGGCCCTAATTACAGGCAATGATTTCCTCATAAA TGCTTTGAAAGGCATGGAGATGTTGTCAGGTTCAATATCCCTTGC TCATGTGAAGGATGTTTGCCGAGCACACATATTTTTGGCTGAG AAAGAATCAGCTTCTGGTCGATATATTTGTTGTGCTCACAACAC AAGTGTTCCTGAACTTGCAAAGTTCCTTAGCAAACGATACCCT CAATACAAAATCCCAACCGAATTTGATGGCTTTCCCTCTAAGGC CAAGTTGGTAATCTCTTCAGAGAAGCTTATCAAAGATGGGTTCAAT TACCAGTTTGGGGGCAGAACAGATCTACGACGAGACTGTGGAGTAC TTCAAGTCCAAGGGAGCCCTTAAAAATTGAATTAGTAGAGGTTCAA AGCTCTAGCAGAAGCTCCACTAGTTATAATTATGTAATGATGATAAAA ATGTGGAGCGGCAGTTCAGTGTGATATATACTGAACTAGTATTTTTAT ATGATGGTATTCTACCCTTCCACCTTGTCCCCTGTAATGCACTTTCT GATATGTTAGTGCACTTTCTGATATGTTAGTGAAAAGCCGAAGAAGA AAAGAAAGAGAAGATTATTGGAAAATAAAAAAATCCAACG	1,516 (481)	5′-CGTCGCTTCT GAGCTCGTTA-3′	5′-CCATGCAGCC TTCTCTGCTA-3′

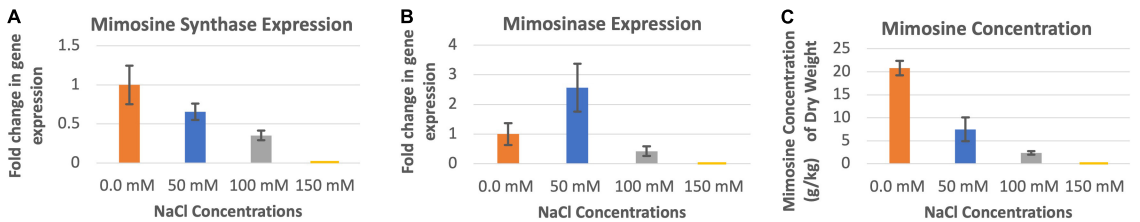


FIGURE 3 | Changes in the gene expression for mimosine synthase **(A)** and mimosinase **(B)** under four different salinity levels. The changes in mimosine production under different salinity conditions **(C)** are also shown. Mimosine production was highest at 0.0 mM salt condition. Therefore, 0.0 mM NaCl was considered as the reference for comparisons; and thus, the expression in panels **(A,B)** values at this salt concentration is one. The error bars in panels **(A–C)** indicate ± SE (n = 15, 5 biological each having three technical replicates).

Effects of Soil pH on Tannin Biosynthesis

Soil pHs between 5.0 and 9.0 did not appear to affect tannin concentration of the leucaena foliage (**Figure 7**).

Expression of Five Tannin Biosynthesis Genes Under N-Abundance and N-Deficiency Conditions

The expression of chalcone synthase and flavanone 3β-hydroxylase were higher when the leucaena plants were grown without providing combined nitrogen, whereas the expressions of dihydroflavonol reductase, leucoanthocyanidin reductase and anthocyanidin reductase were higher when the plants were grown with added combined nitrogen (**Figure 8A**). Overall, the combined expression of all tannin biosynthesis genes resulted in higher tannin concentration under N + condition compared to the N- condition (**Figure 8B**).

DISCUSSION

The concentrations of secondary metabolites such as mimosine and tannin are important aspects of nutritional quality in the leucaena foliage. In the present study, mimosine and tannin

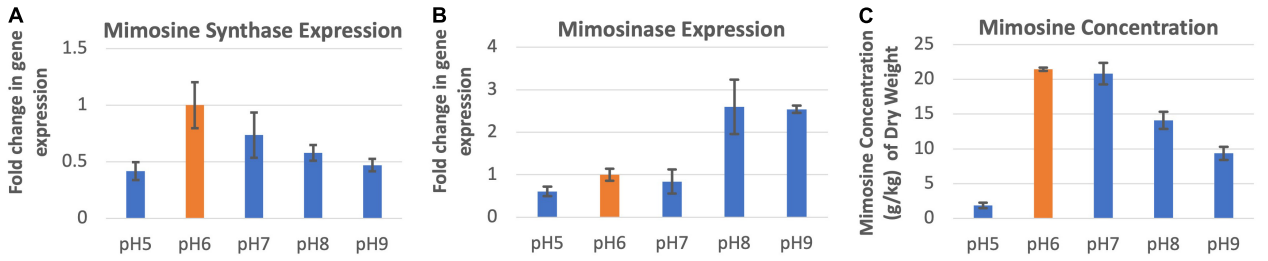


FIGURE 4 | Changes in the gene expression for mimosine synthase **(A)** and mimosinase **(B)** under five different pH levels. The changes in mimosine production under different pH conditions **(C)** are also shown. Mimosine production was highest at pH 6.0. Therefore, pH 6.0 was considered as the reference for comparisons; and thus, the expression in panels **(A,B)** values at this pH is one. The error bars in panels **(A–C)** indicate ± SE (*n* = 15, five biological each having three technical replicates).

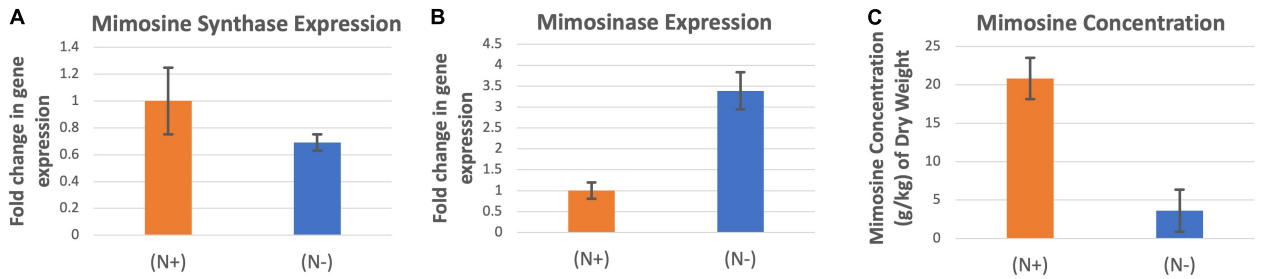


FIGURE 5 | Changes in the gene expression for mimosine synthase **(A)** and mimosinase **(B)** under nitrogen abundant (N +) and nitrogen deficient (N-) conditions. The changes in mimosine production under these two conditions **(C)** are also shown. Mimosine production was higher at (N +) condition. Therefore (N +) was used as the reference for comparisons; and thus, the expression in panels **(A,B)** values under (N +) condition is one. The error bars in panels **(A–C)** indicate ± SE (*n* = 15, five biological each having three technical replicates).

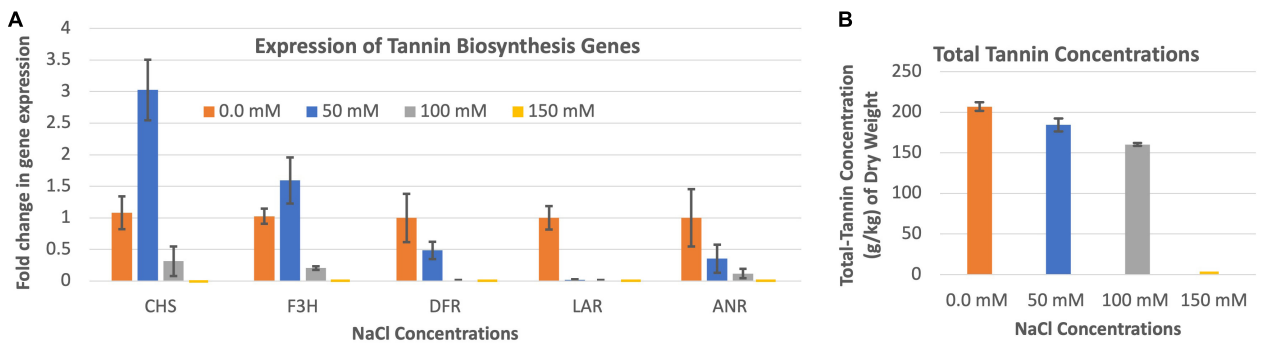


FIGURE 6 | Changes in the gene expression for tannin biosynthesis including chalcone synthase (CHS), flavanone 3ß-hydroxylase (F3H), dihydroflavonol reductase (DFR), leucoanthocyanidin reductase (LAR), and anthocyanidin reductase (ANR) **(A)** under different salinity conditions. The changes in tannins production under different salinity conditions **(B)** are also shown. Total tannins production was highest at 0.0 mM salt condition. Therefore, 0.0 mM NaCl was considered as the reference for comparisons; and thus, the expression in panel **(A)** value at this salt concentration is one. The error bars in panels **(A,B)** indicate ± SE (*n* = 15, five biological each having three technical replicates).

concentrations of leucaena foliage were studied under different salinity, nitrogen availability, and pH conditions with a goal to identify optimum conditions to obtain more nutritious fodder. Other major considerations for forage nutrition are protein content, palatability, digestibility, and total biomass productivity of the foliage. The present study is limited to only mimosine and tannin concentrations in the leucaena foliage because their presence reduces the quality of the fodder.

The expression of two genes for mimosine metabolism, mimosine synthase and mimosinase, were studied. The positions of mimosine synthase and mimosinase in the mimosine metabolism pathways are shown in **Figure 9**. There was a direct correlation between mimosine content and the expression of mimosine synthase in the foliage, whereas there was a negative correlation between mimosine content and mimosinase expression. Thus, mimosine concentration in the foliage at

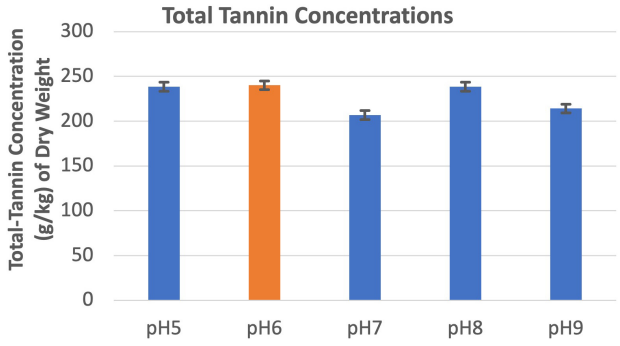


FIGURE 7 | Changes in total tannins concentrations under different pH conditions. The error bars in indicates ± SE (n = 15, five biological each having three technical replicates).

any stage is the result of both synthesis and degradation. The mimosine synthase expression decreases with increase in salt concentration. However, neither low nor very high concentration of salt is favorable for mimosinase expression. Therefore, mimosine concentrations in the foliage is mostly a function of mimosine synthase activity. Recently, mimosine has been described as a "stress response molecule" that is produced in high quantities during favorable environmental conditions and degraded under stress environments such as drought and other nutrient-limited growth conditions (Honda and Borthakur, 2021). Under unfavorable growth conditions, mimosine in the leucaena tissues are converted to produce ammonia and pyruvate that are used in synthesis of other compounds essential for survival.

Among the various pHs tested in this study, the most favorable pH for mimosine biosynthesis was found to be 6.0. Around this favorable soil pH, mimosine is accumulated in the leucaena foliage. The mimosine concentration in the leucaena foliage decreases with increases in pH from 7.0 to 9.0. Under such unfavorable soil pH conditions, mimosine is degraded, and the byproducts are used as sources of nitrogen and carbon for synthesis of other more essential compounds. Similarly, mimosine synthase expression also decreases with increases in pH. On the other hand, mimosinase expression was low under acidic or neutral pH but high under alkaline pH conditions.

Mimosine synthase expression is also positively influenced by high nitrogen availability conditions. On the other hand, mimosinase expression is high under the condition of nitrogen limitation. Honda and Borthakur (2021) observed positive correlation between mimosine contents of foliage with nitrogen availability. The mimosine molecule has two nitrogen atoms, which constitute 14% of its molecular weight. Therefore, it is expected that nitrogen availability conditions are more conducive for mimsoine synthesis. On the other hand, under the condition of nitrogen limitation, leucaena does not have the nitrogen resource to synthesize mimosine. Under such nitrogen limited conditions, the plant will degrade mimosine to obtain nitrogen from it. Therefore, it is not surprising that mimosinase expression is high under nitrogen deficiency conditions. Mimosine has also been described as a storage molecule, which is synthesized under favorable conditions and degraded under stress conditions such as nitrogen deficiency (Honda and Borthakur, 2021).

Tannin biosynthesis pathway is downstream of phenylpropanoid and flavonoid pathways (**Figure 1**). These two upstream pathways contribute to flavonol and tannin biosynthesis pathways. Among the five genes for tannin biosynthesis tested in this study, two of them (CHS and F3H) are required for both tannin and flavonol biosynthesis, whereas the remaining three genes (DFR, LAR, and ANR) are more specific for tannin biosynthesis. Overall, tannin synthesis is reduced under salinity conditions. The expression of chalcone synthase appears to be induced by moderate levels of salt. At 50 mM salt concentration, its expression was three times of that without salt. A similar trend was also observed for F3H, which was induced 1.6 fold at 50 mM salt concentration. On the other hand, the expression of DFR, LAR and ANR is inhibited by salt. In *Arabidopsis*, CHS is regulated by the same transcription factor that regulates flavonol synthase (FLS), thus making a stronger connection between transcription of CHS and FLS (Mehrtens et al., 2005). FLS is a key gene for diverting the flavonoid pathway to flavonol synthesis (**Figure 1**). It has been shown in both *Arabidopsis* and soybean that CHS is involved in conferring resistance to salt stress (Zhu et al., 2021). Therefore, it is not surprising that under salt stress conditions, the leucaena

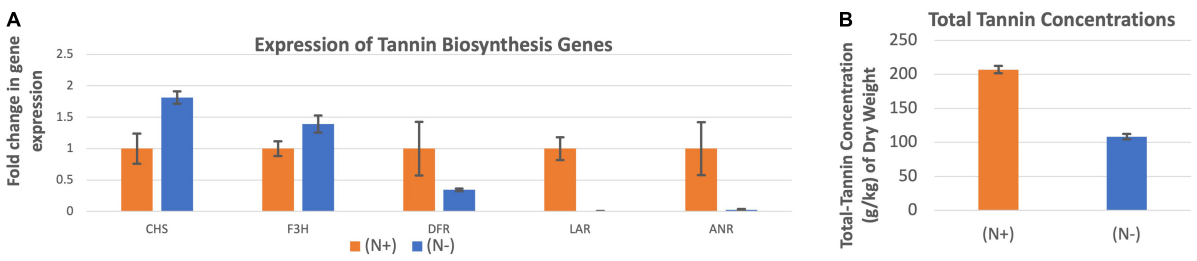


FIGURE 8 | Changes in the gene expression for tannin biosynthesis including chalcone synthase (CHS), flavanone 3ß-hydroxylase (F3H), dihydroflavonol reductase (DFR), leucoanthocyanidin reductase (LAR), and anthocyanidin reductase (ANR) **(A)** under nitrogen abundant (N +) and nitrogen deficient (N-) conditions. The changes in tannins production under these two conditions **(B)** are also shown. Total tannins production was higher at (N +) condition. Therefore (N +) was used as the reference for comparisons; and thus, the expression in panel **(A)** value under (N +) condition is one. The error bars in panels **(A,B)** indicate ± SE (n = 15, five biological each having three technical replicates).

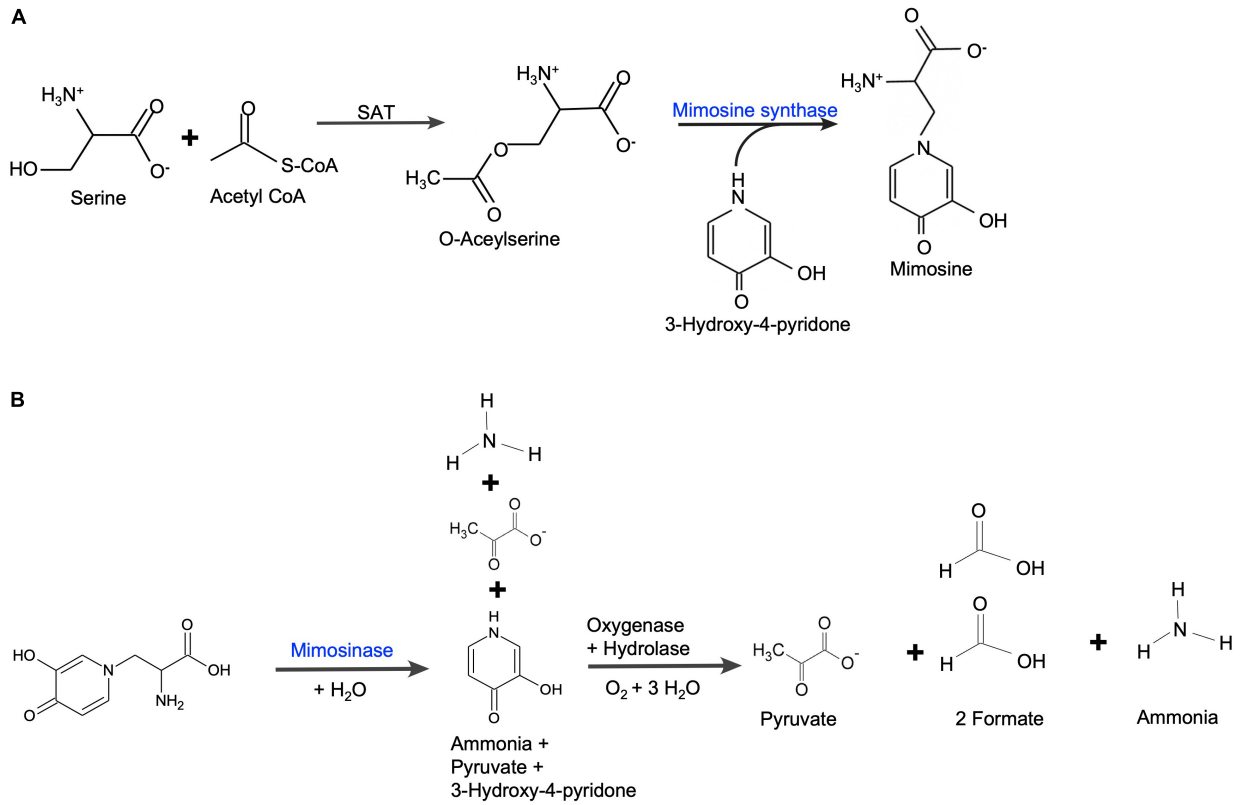


FIGURE 9 | (A) Mimosine biosynthetic pathway in leucaena. Serine acetyltransferase (SAT) catalyzes the formation of *O*-acetylserine from acetyl-CoA and serine. *O*-acetylserine (thiol) lyase (mimosine synthase) catalyzes mimosine biosynthesis from *O*-acetylserine and 3-hydroxy-4-pyridone [based on Yafuso et al. (2014)]. **(B)** Mimosine degradation pathway in leucaena. Mimosine is degraded into pyruvate, formate and ammonia in two steps. First, mimosine is degraded into one molecule each of ammonia, pyruvate and 3-hydroxy-pyridone by mimosinase (Negi et al., 2014). Then, 3-hydroxy-4-pyridone is further degraded by a dioxygenase and a hydrolase into one molecule of ammonia, one molecule of pyruvate and two molecules of formate (Awaya et al., 2005).

CHS gene is induced. Similarly, the overexpression of F3H in transgenic tobacco confers tolerance to salt stress and infection by *Alternaria solani* (Mahajan and Yadav, 2014). On the other hand, the same salt stress conditions that induced transcription of CHS and F3H inhibited transcription of the DFR gene for tannin biosynthesis in *Arabidopsis* (Espley et al., 2009). In the present study also, the salt stress conditions induced the transcription of CHS and F3H but repressed the transcription of the genes that are directly involved in tannin biosynthesis, namely, DFR, LAR, and ANR. Based on the above discussion, it is apparent that flavonols are mostly responsible for salt stress tolerance in leucaena. The amount of tannin is not directly related to salt tolerance.

Tannin molecules do not contain nitrogen; therefore, nitrogen availability is not expected to be a major factor for tannin biosynthesis. However, nitrogen is a general requirement for plant growth and therefore, nitrogen availability may indirectly influence tannin biosynthesis. Under low nitrogen and drought conditions, plants produce condensed tannins with longer polymer structures (Suseela, 2019). In the present study, tannin production increased two-fold under nitrogen availability condition compared to the nitrogen deficiency condition. This increase in tannin production appeared to be affected by increased expression of the three genes (DFR, LAR, and ANR) that are specific for tannin biosynthesis. On the other hand, the expression of other two genes (CHS and F3H) that are common for both flavonol and tannin biosynthesis pathways, did not show the same trend. Their expression levels were induced under nitrogen deficiency conditions. Flavonols are generally involved in biotic and abiotic stress tolerance (Mahajan and Yadav, 2014). The high expression of these two genes might have contributed to increased flavonol synthesis under nitrogen deficiency stress conditions. Knowledge of environmental conditions that promote or inhibit transcription of the genes for mimosine and tannin biosynthesis may be useful in designing environmental conditions that inhibit transcription of these genes, resulting in reduced levels of these compounds in the leucaena foliage.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the NCBI repository, accession numbers SRX14364097 and SRX14364098. The sequence data are released. You may find our entire sequence data for common leucaena at:

https://www.ncbi.nlm.nih.gov/sra/SRX14364097%5Baccn%5D. You may find our entire sequence data for giant leucaena at: https://www.ncbi.nlm.nih.gov/sra/SRX14364098%5Baccn%5D.

AUTHOR CONTRIBUTIONS

AB performed most of the field and laboratory work and wrote the manuscript. AK helped in the bioinformatics analysis of transcriptome data. DB supervised the work and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

AB was supported by a scholarship from the Saudi Cultural Mission, Virginia. This research was supported by a USDA McIntire grant.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.885366/full#supplementary-material

REFERENCES

Awaya, J. D., Fox, P. M., and Borthakur, D. (2005). pyd genes of *Rhizobium* sp. strain TAL1145 are required for degradation of 3-hydroxy-4-pyridone, an aromatic intermediate in mimosine metabolism. *J. Bacteriol.* 187, 4480–4487. doi: 10.1128/JB.187.13.4480-4487.2005

Bageel, A., and Borthakur, D. (2022). The effects of pH, salinity, age of leaves, post-harvest storage duration, and psyllid infestation on nutritional qualities of giant *Leucaena* fodder. *J. Crop Sci. Biotechnol.* doi: 10.1007/s12892-021-00139-9

Bageel, A., Honda, M., Carrilo, J., and Borthakur, D. (2020). Giant *Leucaena* (*Leucaena leucocephala* subsp. glabrata): a versatile tree-legume for sustainable agroforesty. *Agrofor. Syst.* 94, 251–268. doi: 10.1007/s10457-019-00392-6

Espley, R. V., Brendolise, C., Chagné, D., Kutty-Amma, S., Green, S., Volz, R., et al. (2009). Multiple repeats of a promoter segment causes transcription factor autoregulation in red apples. *Plant Cell* 21, 168–183. doi: 10.1105/tpc.108.059329

Harun-Ur-Rashid, M., Iwasaki, H., Parveen, S., Oogai, S., Fukuta, M., Hossain, M. A., et al. (2018). Cytosolic Cysteine Synthase Switch Cysteine and Mimosine Production in *Leucaena leucocephala*. *Appl. Biochem. Biotechnol.* 186, 613–632. doi: 10.1007/s12010-018-2745-z

Honda, M., and Borthakur, D. (2020). Mimosine facilitates metallic cation uptake by plants through formation of mimosine-cation complexes. *Plant Mol. Biol.* 102, 431–445. doi: 10.1007/s11103-019-00956-1

Honda, M., and Borthakur, D. (2021). Mimosine is a stress-response molecule that serves as both an antioxidant and osmolyte in giant *Leucaena* (*Leucaena leucocephala* subsp. glabrata) during environmental stress conditions. *Plant Stress* doi: 10.1016/j.stress.2021.100015

Ishihara, K., Honda, M., Bageel, A., and Borthakur, D. (2018). "*Leucaena leucocephala*: a leguminous tree suitable for eroded habitats of Hawaiian Islands," in *ravine lands: greening for livelihood and environmental security*, eds J. Dagar and A. Singh (Singapore: Springer). 413–431. doi: 10.1007/978-981-10-8043-2

Ishihara, K., Lee, E. K. W., and Borthakur, D. (2016). An improved method for RNA extraction from woody legume species *Acacia koa* and *Leucaena leucocephala*. *Int. J. For. Wood Sci.* 2, 031–037.

Mahajan, M., and Yadav, S. K. (2014). Overexpression of a tea flavanone 3-hydroxylase gene confers tolerance to salt stress and *Alternaria solani* in transgenic tobacco. *Plant Mol. Biol.* 85, 551–573. doi: 10.1007/s11103-014-0203-z

Makkar, H. P. S. (2003). "Measurement of total phenolics and tannins using folin-ciocalteu method," in *Quantification of tannins in tree and shrub foliage*, (Dordrecht: Springer). doi: 10.1007/978-94-017-0273-7_3

McMahon, R., McAllister, A., Berg, P., Majak, W., Acharya, N., Popp, D., et al. (2000). A review of the effects of forage condensed tannins on ruminal fermentation and bloat grazing cattle. *Can. J. Plant Sci.* 80, 469–485.

Mehrtens, F., Kranz, H., Bednarek, P., and Weisshaar, B. (2005). The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. *Plant Physiol.* 138, 1083–1096. doi: 10.1104/pp.104.058032

Min, R., Barry, N., McNabb, C., and Kamp, D. (1998). Effect of condensed tannins on the production of wool and on its processing characteristics in sheep grazing *Lotus corniculatus*. *Aust. J. Agric. Res.* 49, 597–605. doi: 10.1071/a97140

Negi, V. S., Bingham, J. P., Li, Q. X., and Borthakur, D. (2014). A carbon-nitrogen lyase from *Leucaena leucocephala* catalyzes the first step of mimosine degradation. *Plant Physiol.* 164, 922–934. doi: 10.1104/pp.113.230870

Niezen, J.H, Waghorn, T.S, Charleston, W.A.G, and Waghorn, G.C (1995). Growth and gastrointestinal nematode parasitism in lambs grazing either lucerne (*Medicago sativa*) or sulla (*Hedysarum coronarium*) which contains condensed tannins. *J. Agric. Sci.* 125, 281–289. doi: 10.1017/S0021859600084422

Reed, J.D (1995). Nutritional toxicology of tannins and related polyphenols in forage legumes. *J. Anim. Sci.* 73, 1516–1528. doi: 10.2527/1995.7351516x

Suseela, V. (2019). "Potential roles of plant biochemistry in mediating ecosystem responses to warming and drought," in *Ecosystem Consequences of Soil Warming*, (Cambridge, MA: Academic Press). 103–124.

Szyszka, M., and Ter Meulen, U. (1984). The acceptable daily intake for mimosine in productive livestock and its importance for the use of *Leucaena leucocephala* in animal nutrition. *Dtsch. Tieraerztl. Wochenschr.* 91, 260–262.

Xuan, T., Elzaawely, A., and Debah, F. (2006). Mimosine in *Leucaena* as potent bio-herbicide. *Agron. Sustain. Dev.* 26, 89–97.

Yafuso, J. T., Negi, V. S., Bingham, J. P., and Borthakur, D. (2014). An O-Acetylserine (thiol) lyase from *Leucaena leucocephala* is a Cysteine Synthase but not a mimosine synthase. *Appl. Biochem. Biotechnol.* 173, 1157–1168. doi: 10.1007/s12010-014-0917-z

Zhu, J., Zhao, W., Li, R., Guo, D., Li, H., Wang, Y., et al. (2021). Identification and characterization of chalcone isomerase genes involved in flavonoid production in Dracaena cambodiana. *Front. Plant Sci.* 12:616396. doi: 10.3389/fpls.2021.616396

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 21 June 2022
doi: 10.3389/fpls.2022.879874

Comparative Transcriptome and Proteome Analysis Provides New Insights Into the Mechanism of Protein Synthesis in Kenaf (*Hibiscus cannabinus* L.) Leaves

Chao Zhang[1,2†], Yong Deng[1†], Gaoyang Zhang[1,2], Jianjun Li[1], Aiping Xiao[1], Lining Zhao[1], Anguo Chen[1], Huijuan Tang[1], Li Chang[1], Gen Pan[1], Yingbao Wu[2], Jiangjiang Zhang[1], Cuiping Zhang[1], Ziggiju Mesenbet Birhanie[1], Hui Li[1], Juan Wu[1], Dawei Yang[1], Defang Li[1] and Siqi Huang[1]**

[1] Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha, China, [2] School of Life Sciences, Shangrao Normal University, Shangrao, China

Given the rising domestic demand and increasing global prices of corn and soybean, China is looking for alternatives for these imports to produce animal fodder. Kenaf (*Hibiscus cannabinus* L.) has great potential as a new forage source, due to abundant proteins, phenols and flavonoids in its leaves. However, few studies have evaluated the mechanism of protein synthesis in kenaf leaves. In the current work, compared with kenaf material "L332," the percentage of crude protein content in leaves of material "Q303" increased by 6.13%; combined with transcriptome and proteome data, the kenaf samples were systematically studied to obtain mRNA-protein correlation. Then, the genes/proteins related to protein synthesis in the kenaf leaves were obtained. Moreover, this work detected mRNA expression of 20 differentially expressed genes (DEGs). Meanwhile, 20 differentially expressed proteins (DEPs) related to protein synthesis were performed parallel reaction monitoring. Fructose-1,6-bisphosphatase (*FBP*), nitrite reductase (*NirA*), prolyl tRNA synthase (*PARS*) and glycine dehydrogenase (*GLDC*) presented increased mRNA and protein levels within kenaf leaves with high protein content. Based on the obtained findings, *FBP*, *NirA*, *PARS*, and *GLDC* genes may exert a vital function in the protein synthesis of kenaf leaves. The results provide a new idea for further studying the potential genes affecting the quality trait of protein content in kenaf leaves and provide gene resources and a theoretical foundation for further cultivating high protein kenaf varieties.

Keywords: kenaf, transcriptome, gene expression, proteome, protein synthesis

OPEN ACCESS

Edited by:
Chris S. Jones, International Livestock Research Institute, Ethiopia

Reviewed by:
Md Atikur Rahman, Rural Development Administration, South Korea
Parviz Heidari, Shahrood University of Technology, Iran
Muhammad Haneef Kashif, Guangxi University, China

***Correspondence:**
Defang Li
chinakenaf@126.com
Siqi Huang
siqihuang@yahoo.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

Received: *23 March 2022*
Accepted: *02 June 2022*
Published: *21 June 2022*

Citation:
Zhang C, Deng Y, Zhang G, Li J, Xiao A, Zhao L, Chen A, Tang H, Chang L, Pan G, Wu Y, Zhang J, Zhang C, Birhanie ZM, Li H, Wu J, Yang D, Li D and Huang S (2022) Comparative Transcriptome and Proteome Analysis Provides New Insights Into the Mechanism of Protein Synthesis in Kenaf (Hibiscus cannabinus L.) Leaves. Front. Plant Sci. 13:879874. doi: 10.3389/fpls.2022.879874

INTRODUCTION

The impact of global climate change on agricultural production has been a hotspot for academia at home and abroad; the large impact of artificial pollutants on the natural environment plays a leading role in reducing crop productivity and is the real cause of different abiotic stress states (Akcura et al., 2019; Bilen et al., 2019; Tang et al., 2019; Kashif et al., 2020a,b; Li et al., 2021). In China, soybean and corn imports exceeded 110 million tons in 2020, and demand for feed grains is still increasing.

Additionally, the rise in international food prices has diverted the intent of feed producers to look for alternatives to reduce or replace feed grains. Hence, owing to the abundant proteins, phenols, and flavonoids in its leaves, kenaf has great potential as a new forage source. However, few studies have evaluated the molecular mechanisms of protein synthesis in kenaf leaves. Kenaf (*Hibiscus cannabinus* L.), a kind of annual bast fiber crop of Hibiscus in Malvaceae family, represents the heliophilous crop that grows in tropical and temperate zones (Cheng et al., 2004). The kenaf plant can provide 2 fiber types, phloem and core. Bast fiber locates within phloem part in the kenaf stems, which is suitable for textile and industrial purposes, such as in paper, rope, textiles, and carpets (Falasca et al., 2014). The core part is located inside the stem of kenaf, which is rich in cellulose and hemicellulose and can be used as an absorbent material. For example, it can be utilized in indoor acoustic panels, thermal insulation panels, indoor sound insulation boards and heat insulation boards (Okuda and Sato, 2004). Phloem and core can be applied alone or concurrently in producing biological composites or biofuels (Saba et al., 2015). The chemical components of the volatile oil from kenaf leaves were analyzed (Pascoal et al., 2015). A total of 58 components of the volatile oil were identified. Volatile oil has phytotoxicity and antifungal activity. Kenaf have the potential to deal with a wide range of heavy metal pollution with high economic benefits (Deng et al., 2017). Its economic value is reflected not only in fiber utilization but also in important research prospects in the materials, bioenergy, medicine, feed, papermaking and carbon sink trade (Ayadi et al., 2017). In addition, in production practice, kenaf has the characteristics of fast growth and large biomass and can be used as livestock feed. Kenaf has fast growth speed, contains flavonoids and polyphenols, exhibits strong insect resistance, and does not need to apply insecticides, which are natural green protein feeds (Swingle et al., 1978; Birhanie et al., 2021). With the rapid development of animal husbandry and breeding in China, the feed industry has also emerged. High-quality protein feed has become an urgent necessity, especially in southern China. Due to the humid and hot climate and other reasons, there is a lack of high-quality and high-yield plant protein feed. Therefore, carrying out kenaf feed research, breeding kenaf feed varieties and strengthening the research on kenaf feed nutritional value will bring economic and ecological benefits.

Plant protein accumulation is a complex process accompanied by a large amount of protein synthesis and the transcription of many genes (Shen et al., 2019; Xie et al., 2019). The results showed that the crude protein content of different genotypes was significantly different, indicating that genetic factors exerts an essential function in the protein content of soybean (Wang et al., 2021). Studies on genes related to protein synthesis have been carried out in many species, including *Arabidopsis thaliana*, cassava, wheat, radish, phaseolus and soybeans. These studies show that sucrose phosphate synthase (SPS), fructose-1,6-bisphosphatase (FBP), nitrite reductase (NiR), fructose-1,6-bisphosphatase aldolase (ALDO), prolyl tRNA synthase (PARs), and glycine dehydrogenase (GLDC) (Jeannin et al., 1976; Okamura-Ikeda et al., 1993; Sahrawy et al., 2004; Wu et al., 2015; Lv et al., 2017; Huang et al., 2020) were related to protein synthesis. The *A. thaliana NRT* gene regulates NO_3^- absorption and plant dynamic responses to changes in nitrogen content in the environment (Wang et al., 2012). Additionally, Tiwari et al. (2019) found that the post-translational regulation of the *NR* gene highly influences the content of free amino acids and nitrate. Therefore, further genetic analyses on protein synthesis is of necessity for protein accumulation.

Although numerous researches have been reported on the impacts of abiotic stress or cutting methods on the protein content of kenaf leaves, little progress has been made on the molecular mechanism of protein synthesis in kenaf leaves. Most of the 15 reports on kenaf transcriptome sequencing were the study of differentially expressed genes after biological stress (Niu et al., 2015), and there are no reports on kenaf leaf protein synthesis. The understanding of the molecular characteristics of protein synthesis in kenaf leaves is limited, making it difficult to select and breed kenaf varieties with high protein content in leaves. The recent progress in sequencing methods like transcriptome and proteome techniques, brings us a great convenience in the measurement of gene expression and protein abundance. The sequencing methods have turned into a powerful instrument for discovering novel genes and improving the protein content of kenaf leaves (Jérme and Richard, 2008).

Therefore, the objective of this research was to examine the transcriptome and proteome of two kenaf material: "L332" and "Q303," which assistant by high-throughput sequencing technology. We also assessed their difference in the percentage of crude protein content in leaves. Furthermore, we explored the potential genes affecting the quality trait of protein content. The comprehensively analysis strengthen our understanding of kenaf protein synthesis at the molecular level.

MATERIALS AND METHODS

Plant Growth and Sampling

Kenaf seed materials "Q303" and "L332," which are cultivated in southern China, were provided by the Institute of Bast Fiber Crops, Chinese Academy of Agriculture Sciences. In this study, it was observed that the leaves of kenaf material "L332" belong to the lobed-leaf type and the leaves of kenaf material "Q303" belong to the round-like leaf type. Our experimental site was located in the Bairuopu Innovative Experimental Base at the Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences in Changsha. The experimental area had a typical subtropical continental monsoon climate. A field with the same water and fertilizer conditions was selected as the experimental site. Two kenaf materials were sown on May 16, 2019 and harvested on October 20, 2019. The whole growth period lasted for 155 days. At the mature stage (September 17), all leaves within 50 cm below the top of the kenaf plant were collected and mixed for carrying out further analyses. The mixed samples were from the leaves of 20 kenaf plants. This work set 3 biological replicates for every kenaf material. For every kenaf material, this study gathered 30 g leaf samples from an individual plant. We froze sample leaves with liquid nitrogen at once, followed by preservation

under −80°C until the application for RNA isolation, protein separation, and quantitative real-time PCR analyses.

Morphological Character Assays and Extraction of Total Protein of Kenaf Leaves

This study determined kenaf leaf size using the graduated scale, while leaf weight using the electronic balance, and the total protein of kenaf leaves using the Kjeldahl method (Grzeszczuk et al., 2018). The leaves of kenaf were subjected to 0.5-h deactivation under 105°C, followed by drying under 65°C to constant weight. Subsequently, the leaves of kenaf were crushed to determine the protein content. We prepared 3 biological replicates of every kenaf material.

RNA Extraction and cDNA Library Construction

The leaves in the biological replicate for every sample were grinded to powder with liquid nitrogen. According to Chai et al. (2014), extraction of total RNA in two kenaf leaf samples was conducted. Total RNA extracted from kenaf leaves was dissolved in RNase-free water (TIANGEN, China). The A260/A280 ratio was measured using the Nanodrop ND-1000 system (Thermo Scientific, United States) and adopted for checking the RNA concentration in kenaf leaves, and 1.5% agarose gels were adopted for detecting the integrity of the RNA. After the RNA of kenaf samples was qualified, the library was prepared for the qualified samples. cDNA libraries were created with the use of the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, United States) in accordance with specific protocols. Kenaf mRNA purification was conducted using the extracted total RNA (4 μg) with oligo (dT) magnetic beads (Thermo Scientific) following the instructions of the manufacturer. Then, synthesis of first- and second-strand cDNA by reverse transcription was conducted, with mRNA being a template. After purification of double-stranded cDNA and then end repair, polyadenylation, and addition of adaptor sequences, DNA fragment sorting and PCR amplification were performed. Finally, a cDNA library was constructed. After purification of PCR products (AMPure XP system), we assessed cDNA library quality on an Agilent Bioanalyzer 2100 system. Finally, the cDNA library was constructed and sequenced after quality inspection.

RNA Sequencing and Transcriptome Assembly

The cDNA libraries for those 3 biological replicates of kenaf samples were constructed according to Cui et al. (2019). Subsequently, sequencing was carried out by an Illumina NovaSeq 6000 instrument by Shanghai Applied Protein Technology Co., Ltd. Sequencing was completed by an Illumina NovaSeq 6000 instrument from Shanghai Applied Protein Technology Co., Ltd. (Shanghai, China). Besides, through CASAVA base calling analysis, raw imaging data acquired by Illumina high-throughput sequencing were transformed into original sequenced reads. Besides, these findings were kept as the fastq file format. In the fastq file, four lines are regarded as a basic unit and correspond to the sequencing information of a sequence. The clean data (clean reads) were obtained by filter_fq software through filtering out low-quality and adaptor sequences as well as reads that contained the poly-N sequences. High-quality sequence data can be obtained by filtering out sequences that are too short or contain uncertain bases. We calculated the Q20, Q30, GC levels, as well as clean reads sequence duplicate degrees. Using Trinity software package[1] (Grabherr et al., 2011) to assemble and splice the clean reads, data assembly was later conducted. Trinity software workflow was used for *de novo* transcriptome assembly. It contains 3 separate software modules, respectively, Inchworm, Butterfly and Chrysalis. In addition, RNA-seq read data were processed by the three software modules in turn. Transcripts <200 bp were fully removed.

Bioinformatics Analyses

This work annotated *de novo* assembled unigenes with the National Center for Biotechnology Information (NCBI) non-redundant protein sequence database (NR), the database for manual annotation and protein sequence reviewing (Swiss-Prot), a protein family database (Pfam), an integrated database on the basis of Gene Ontology (GO), as well as a genome database on Kyoto Encyclopedia of Genes and Genomes (KEGG), and threshold value was set as an *E*-value $< 10^{-5}$. After gene GO annotation, the annotated genes were categorized in accordance with GO (Biological processes BPs, Cellular Components CCs, Molecular Functions MFs). KEGG database contributes to studying genes and expression information as a whole network. KEGG pathway analysis was accomplished through KOBAS 2.0 test statistical enrichment (Chen et al., 2011).

In this experiment, the Fragments Per Kilobase of exon model per Million mapped fragments (FPKM) value was used for transcription and quantification of gene expression, and DEG-seq was applied to explore the differential expression of samples. We controlled false discovery rate using the *p*-value adjustment methods of Benjamini and Hochberg (1995). The screening criteria for obvious differences were a $p < 0.05$ and $|\log2\ (FC)| \geq 1$. The expression was deemed to be different between the two samples. Cluster Profiler R software was employed to realize GO as well as KEGG analysis on DEGs. Ap, the PlantTFDB[2] was used to analyze transcription factors (TFs) in DEGs.

Protein Separation

Protein was extracted in kenaf leaves of every sample based on previous description, with three biological replicates for each treatment (An et al., 2018). BCA Protein Assay Kit (Bio-Rad, United States) was adopted for determining the protein concentration.

Protein Digestion and TMT Labeling

Protein was digested based on Liu et al.'s method (Liu et al., 2018), with some modifications. In filter-aided sample preparation (FASP Digestion), we poured all samples in the buffer [consisting of 150 mM Tris-HCl (pH 8.0), 4% SDS, 100 mM DTT], followed

[1]https://github.com/trinityrnaseq/trinityrnaseq/releases

[2]http://planttfdb.gao-lab.org/index.php

by 7-min boiling of mixed solution and cooling immediately. The samples were treated following specific protocols of TMT Kit (Thermo Scientific, United States). This work thereafter utilized TMT 6-plex to categorize trypsin (that contained 100 μg protein) for 126-tag (HPC-1), 127-tag (HPC-2), 128-tag (HPC-3), 129-tag (LPC-1) 130-tag (LPC-2) and 131-tag (LPC-3) labeling. The label assimilation was checked, and all the above 6 labeled samples in diverse set were later combined.

HPLC Fractionation and LC–MS/MS Analysis

Thereafter, this work loaded the mixed peptide sample to the reversed-phase trap column (nanoViper C18, 100 μm*2 cm; Thermo Scientific Acclaim PepMap100) equipped with the C18 reversed-phase analytical column (length, 10-cm; 3-μm resin; inner diameter, 75-μm; Thermo Scientific Easy Column) within buffer A consisting of 0.1% formic acid, followed by separation using the gradient buffer B (0.1% formic acid and 84% acetonitrile) at a 300 nl/min flow rate under the control by IntelliFlow Technology. Afterward, peptides were detected using Q-Exactive High-Resolution Mass Spectrometer (Thermo Scientific). This study obtained survey scans at the 70,000 at m/z 200 resolution, whereas HCD spectra were obtained at 17,500 and 35,000 at m/z 200 (TMT 6plex and 10plex, respectively), with isolation width being set at 2 m/z. Meanwhile, this work set the underfill ratio (specifying the minimal target value percentage possibly reached at the maximal fill time) and normalized collision energy as 0.1% and 30 eV separately.

Sequence Database Search and Data Analysis

This work obtained original MS data in RAW files. This work utilized Proteome Discoverer 1.4 software as well as MASCOT engine (Matrix Science, London, United Kingdom; version 2.2) for searching MS/MS spectra against kenaf database with the settings below, enzyme trypsin and two as the maximal missed cleavage allowed; variable oxidation modification (M), fixed carbamidomethyl modification (C), TMT-6plex (K) and TMT-6plex (N-term); and mass tolerance for peptide ions and fragment ions were 20 ppa and 0.1 Da, separately, at both protein and peptide levels of false discovery rate (FDR) < 0.01. Proteins were detected and quantified using specific protein peptides. Differentially abundant proteins (DAPs) were detected by the thresholds of FC ≤ 0.83 or ≥ 1.2 and $p < 0.05$.

The GO annotation based on three categories, including BPs, MFs, and CCs, together with KEGG analysis, was carried out by Fisher's exact test ($p < 0.05$). This work also adopted STRING 9.0 software for constructing the PPI network[3]. Later, Java Treeview software[4] and Cluster 3.0[5] were utilized for hierarchical clustering.

[3]http://string-db.org

[4]http://jtreeview.sourceforge.net

[5]http://bonsai.hgc.jp/~mdehoon/software/cluster/index.html

Quantitative Real-Time PCR

By adopting SYBR Green PCR master mix (Aidlab Biotechnologies, Co., Ltd.), qRT-PCR was conducted to analyze 20 potential DEGs levels according to the manufacturers' instructions on Bio-Rad CFX96 Touch detection system (Bio-Rad, Richmond, CA, United States). The total RNA isolation method and cDNA preparation method were the same as previous description. Genomic DNA was removed using gDNA Eraser (Tsingke), later, the isolated total RNA was prepared into cDNA with Goldenstar^{TM} RT6 cDNA Synthesis Kit Ver. 2 through reverse transcription. Primers were designed with Primer 5.0 software (**Supplementary Table 5**) for qPCR experiment, followed by blasting of these gene sequences to NCBI database. This study utilized around 1 μg total RNA in every reaction. For removing gDNA, we cultured sample for a 2-min period under 42°C and for another 5-min period under 60°C. Reaction buffer (10 μl) in step 1 (20-μl system) was utilized in every reverse transcription. After mixing, the mixture was subject to 30-min incubation under 50°C and 5-min heat shock under 85°C. The PCR conditions were shown below, 30-s under 95°C; 5-s under 95°C and 55-s under 60°C for altogether 40 cycles. Meanwhile, this work utilized gene-specific primers (F/R) of 20 candidate DEGs and the kenaf internal control gene GhACTIN (GenBank accession no. AY305733) (Lu et al., 2018), with action being the control. Further, $2^{-\Delta\Delta Ct}$ approach was utilized for determining 20 gene levels. qRT–PCR procedure was conducted in triplicate.

Parallel Reaction Monitoring Analysis

For verifying the TMT-measured protein levels, a PRM experiment was adopted for further quantifying the expression level of the selected proteins. For every sample, we added the PRTC stable isotope peptide as control. After protein extraction, sample (200 μg) was subject to trypsin hydrolysis. Meanwhile, the C18 cartridge was utilized to desalt peptides; 40 μl of freeze-dried peptide was added to 0.1% formic acid solution, followed by peptide quantification. Peptides of the same mass were mixed and separated by Easy nLC-1200 system (Thermo Fisher Scientific, MA, United States). Meanwhile, this work utilized acetonitrile (ACN) within 40 min with a 1-h liquid chromatographic gradient of 5–30%, and PRM mass spectrometry was carried out using the Q Exactive HF mass spectrometer (Thermo Scientific) in the positive-ion mode, with a full MS1 scanning resolution ratio of 60,000 (200 m/z). This work also set the target values of maximal ion injection time and automatic gain control (AGC) at 200 ms and 3e6, separately; after full MS scanning, 25 PRMs (MS2 scans) were performed, at the resolution and maximal injection time of 120 ms and 30,000 (m/z 200), and agc3e6. In addition, a second window was utilized to isolate target peptides; whereas the normalized collision energy (Grzeszczuk et al., 2018) was adopted for ion activation/dissociation within the high energy dissociation (HCD) collision pool. The 20 target proteins quantified by PRM were imported into Skyline software[6], the peptide settings were selected, the background

[6]https://skyline.ms/

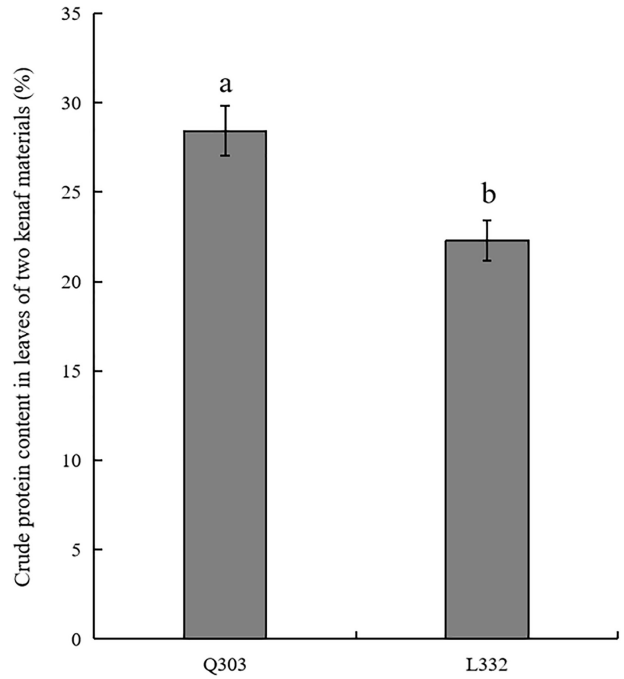


FIGURE 1 | Crude protein content in leaves from two kenaf materials. This assay was carried out in 3 biological replicates, with 3 technical replicates being set each. Student's *t*-test was adopted for data analysis ($p < 0.05$). Error bars = standard error of the mean (SEM). Different letters indicate significant difference at $p < 0.05$.

proteome database file was added, the protein quantitative polypeptide was selected according to the ion signal in the spectral library, the list of related peptides with retention time was derived from Skyline, and the quantitative results of each peptide of target protein were manually checked. Finally, each sample took 1 μg with 20 fmol heavy isotope-labeled peptide section PRTC as the internal label, and the target peptides and proteins were quantitatively analyzed by the same chromatographic conditions and mass spectrometry methods as the previous PRM method. The signal strength of the single peptide sequence of each significantly changed protein was measured for all samples relative to control (Gillet et al., 2012).

RESULTS

Analysis of the Crude Protein Content in Kenaf Feed

Kenaf leaves are applied as livestock feed because of their high protein concentration and large biomass. The crude protein in leaves of two kenaf materials were determined. As a result, mature "Q303" leaves had markedly increased protein levels compared with mature "L332" (**Figure 1**). The average percentage of crude protein content in leaves of kenaf material "Q303" reached 28.43%. Compared with kenaf material "Q303," the percentage of crude protein content in leaves of material "L332" decreased by 6.13%.

Transcriptomic Analysis Overview

For summarizing the transcriptomes of the two kenaf materials in the mature period, this work built six cDNA libraries (i.e., L332 and Q303, three repeats). Altogether 49.33 and 47.74 million raw sequences were generated in cleft Q303 and L332 libraries, separately. Later, adaptor sequences, low-quality reads and those with uncertain base ratios >10% were eliminated, leaving 46.32 and 47.78 million clean reads with Q30 base percentages and GC contents of 94.23-94.38 and 47.07-48.00%, respectively (**Supplementary Table 1**). A total of 222,979 transcripts (range, 201–2,000 bp), with altogether 101,679 unigenes (>200 bp) were discovered (**Table 1**). The size distributions of unigenes and transcripts are presented in **Supplementary Figure 1**.

For determining those identified transcripts' candidate functions, this study utilized Basic Local Alignment Search Tool (BLAST) to annotate unigenes on the basis of 5 databases, like National Center for Biotechnology Information non-redundant protein sequences (NR) (60,171, 59.18% of all the identified unigenes), SwissProt (40,606, 39.94%), Protein families (Pfam) (31,431, 30.91%), Gene Ontology (GO) (34,448, 33.88%), and KEGG (10,456, 10.28%) databases. Based on the above results, NR database had the most functional annotations, which suggested that all the 61141 unigenes corresponded to sequences from one or more public databases. Finally, there were 6879 functionally annotated unigenes from different databases (**Supplementary Table 2**).

Of those, 3,778 DEGs were found by $p < 0.05$ and log2-FC > 1 thresholds in leaves of two kenaf materials. Q303 group showed 1,350 and 2,428 genes with up-regulation and down-regulation, respectively, relative to L332 group. **Supplementary Figure 2A** displays a volcano plot of the above findings. **Table 2** summarized data regarding the identified DEGs. Many of the above genes obtained from annotation had not been characterized before.

TABLE 1 | Summary of the proteomic and transcriptomic data within leaves of two kenaf materials.

RNA-seq data		MS data based on transcriptome	
Total number of transcripts	222,979	Match spectra	66,347
Mean length of transcripts (bp)	1,216	Identified peptides	32,558
Total number of unigenes	101,679	unique peptides	26,972
Mean length of unigenes (bp)	821	Identified proteins	7,196
N50 length of transcripts (bp)	1,919		
N50 length of unigenes (bp)	1,551		

TABLE 2 | Proteins and transcripts measured based on TMT and RNA sequence data.

	Transcriptome	Protein
Unique proteins/genes detected	101,679	7,196
Significantly DEGs/DAPs	3,778	1,320
Up-regulated	1,350	636
Down-regulated	2,428	684
Co-regulated DEGs-DAPs	134	134
Co-regulated DEGs-DAPs with the same trends	122	122
Co-regulated DEGs-DAPs with the opposite trends	12	12

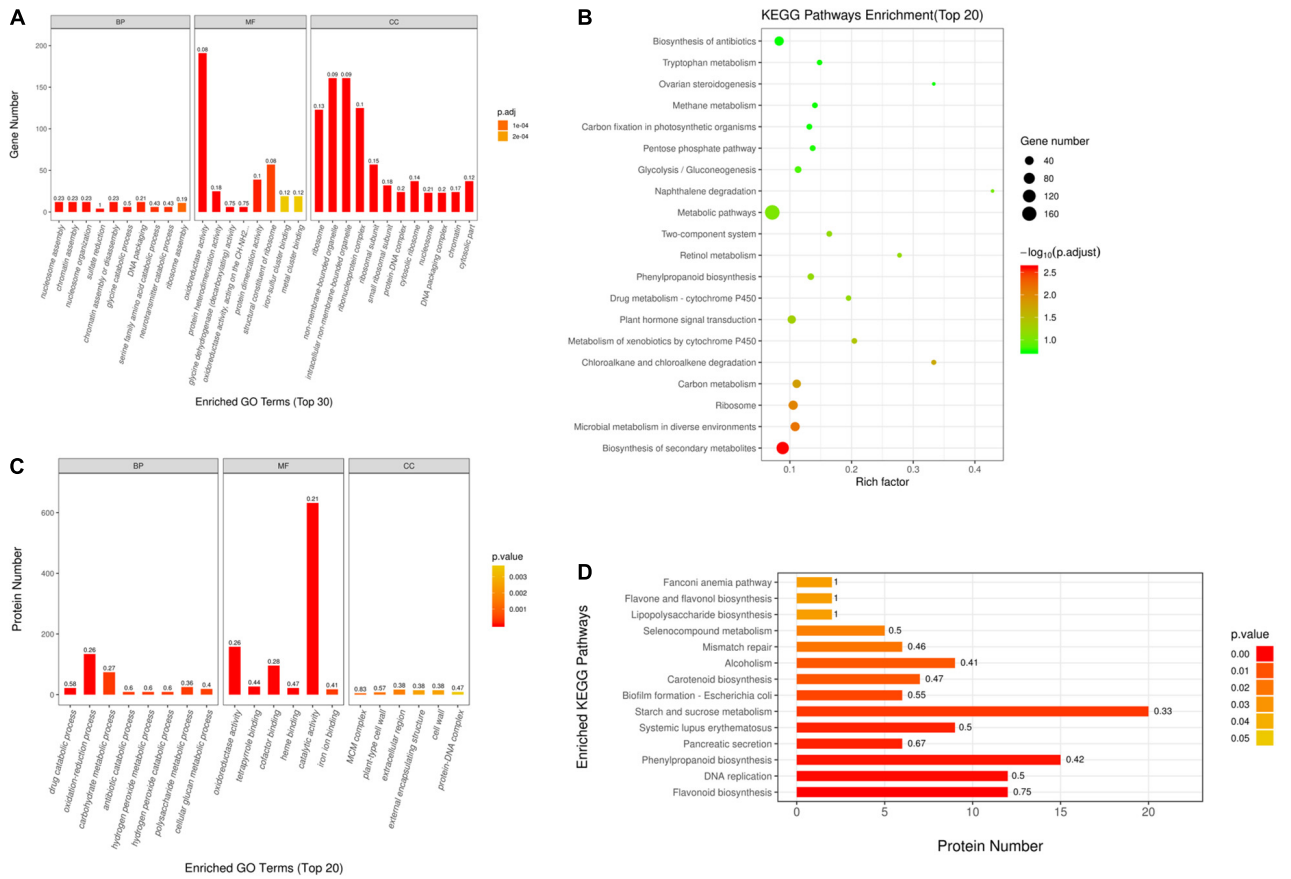


FIGURE 2 | Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses on DAPs and DEGs. **(A)** GO annotations for DEGs within Q303/L332. Those 30 most significant BP, MF, and CC terms are presented in horizontal axis, whereas gene numbers of GO categories are displayed in vertical axis. **(B)** GO annotations for DAPs within Q303/L332. Those 20 most significant BP, MF, and CC terms are presented in horizontal axis, whereas protein numbers of GO categories are displayed in vertical axis. In every histogram, color and height stand for *p*-value and protein/gene count, separately. The histogram label stands for rich factor and it indicates the ratio of DEG/DAP number to overall gene/protein number of GO terms**. (C)** KEGG analysis on DEGs within Q303/L332. **(D)** KEGG pathways enrichment of DAPs in Q303/L332. Enrichment factors of diverse KEGG pathways **(C)** and protein counts in diverse terms **(D)** are displayed in horizontal axis. Those most significant KEGG pathways are presented in vertical axis. The circle color and size stand for *p*-value and gene count of each KEGG term, separately. The histogram color and length represent *p*-value and protein count of each KEGG term. The histogram label stands for rich factor, and it indicates the ratio of DEG/DAP number to overall gene/protein number of KEGG categories.

Functional Classification of the Identified Differentially Expressed Genes

For better understanding DEGs' function, this work carried out bioinformatics analysis through hierarchical clustering and GO analysis. As revealed by GO functional annotation, many DEGs were enriched into serine family amino acid catabolic process and glycine catabolic process with regard to BP terms; oxidoreductase activity and glycine dehydrogenase (decarboxylating) activity with regard to MF terms; and ribosome and intracellular non-membrane-bounded organelle with regard to CC terms (**Figure 2A**).

Additionally, this work conducted KEGG analysis for evaluating DEGs. As a result, most DEGs were associated with secondary metabolite biosynthesis and antibiotic metabolic pathway biosynthesis in the leaves of the two kenaf materials (**Figure 2C**). Besides, this study conducted hierarchical clustering for improving the understanding on alterations of protein synthesis-associated DEG levels (**Figure 3A**). In all, 141 protein synthesis-related pathways, including amino acid biosynthesis and metabolism, energy metabolism, carbohydrate metabolism, genetic information processing, and carbon metabolism, were clustered closely.

Quantitative Proteome Analysis

For exploring the mechanism underlying high protein content in kenaf leaves, this work carried out quantitative proteomics analysis by LC-MS/MS analysis and TMT platform in mature stage, for the sake of complementing transcriptomic analysis. Proteomic analysis identified altogether 66,347 spectra, 26,972 unique peptides and 32,558 identified peptides, among which, we discovered 7,196 proteins (**Table 1** and **Supplementary Table 3**). As for distribution of protein mass, proteins whose molecular weights (MWs) were >9 kDa showed a broad range as well as favorable coverage, and the maximal distribution area reached

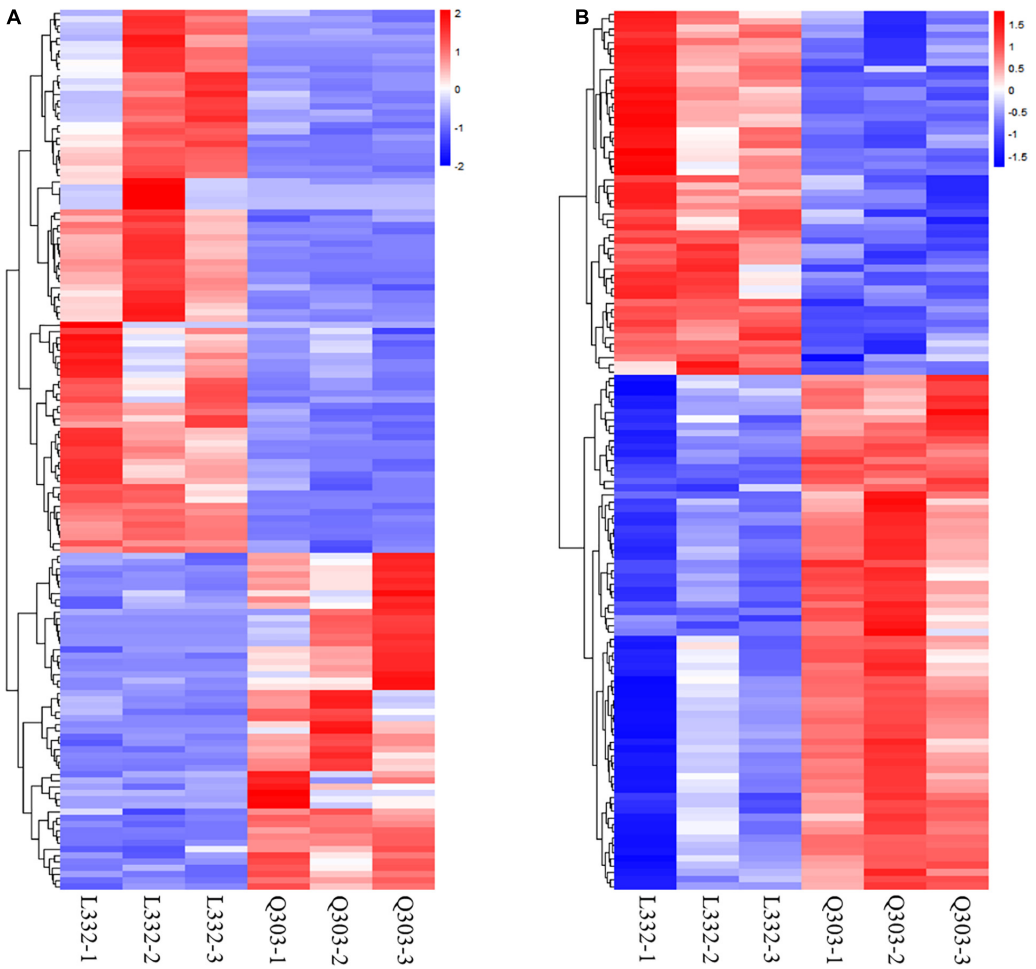


FIGURE 3 | Heatmap showing DAPs and DEGs through proteomic and transcriptomic analyses, and they were related to the protein synthetic metabolic processes. Red, blue, and white stand for markedly up-regulated, down-regulated and non-significant proteins, separately. **(A)** Heatmap of protein synthesis-related gene expression. Amino acid biosynthesis and metabolism, energy metabolism, carbohydrate metabolism, genetic information processing, and carbon metabolism were clustered closely. **(B)** Heatmap of protein synthesis-related protein expression. Including biosynthesis of amino acids, the nitrogen metabolism pathway, starch and sucrose metabolism, carbon metabolism, glycine/serine/threonine metabolism, aminoacyl-tRNA biosynthesis, and the protein synthesis-related proteins were clustered closely.

10–40 kDa (**Supplementary Figure 3A**). According to protein peptide quantification, protein number declined as matching peptide increased (**Supplementary Figure 3B**).

Of the above proteins, we discovered 1,320 differentially abundant protein (DAPs) upon the upregulated and downregulated thresholds being FC $>$ 1.2 as well as $p < 0.05$, separately, including 636 with higher abundance levels, whereas 684 with lower abundance levels between Q303/L332 groups, as displayed by the volcano plot **Supplementary Figure 2B**.

Functional Classification of the Identified Differentially Abundant Proteins

Gene Ontology Analysis

This work also conducted bioinformatics analysis on those identified DAPs according to hierarchical clustering and functional classifications of proteins. According to GO annotation, many DAPs were enriched into oxidation-reduction process and carbohydrate metabolic process with regard to BP terms. As for MF terms, DAPs were mostly enriched into oxidoreductase activity and catalytic activity. In terms of CC category, DAPs were enriched into plant-type cell wall and protein-DNA complex (**Figure 2B**).

Kyoto Encyclopedia of Genes and Genomes Analysis

This work also performed KEGG analysis for evaluating DAPs. Many DAPs were associated with phenylpropanoid biosynthesis, DNA replication, and starch/sucrose metabolism pathways (**Figure 2D**).

Hierarchical Clustering Analysis

This work also conducted hierarchical clustering for better exploring alterations of protein synthesis-associated DAPs

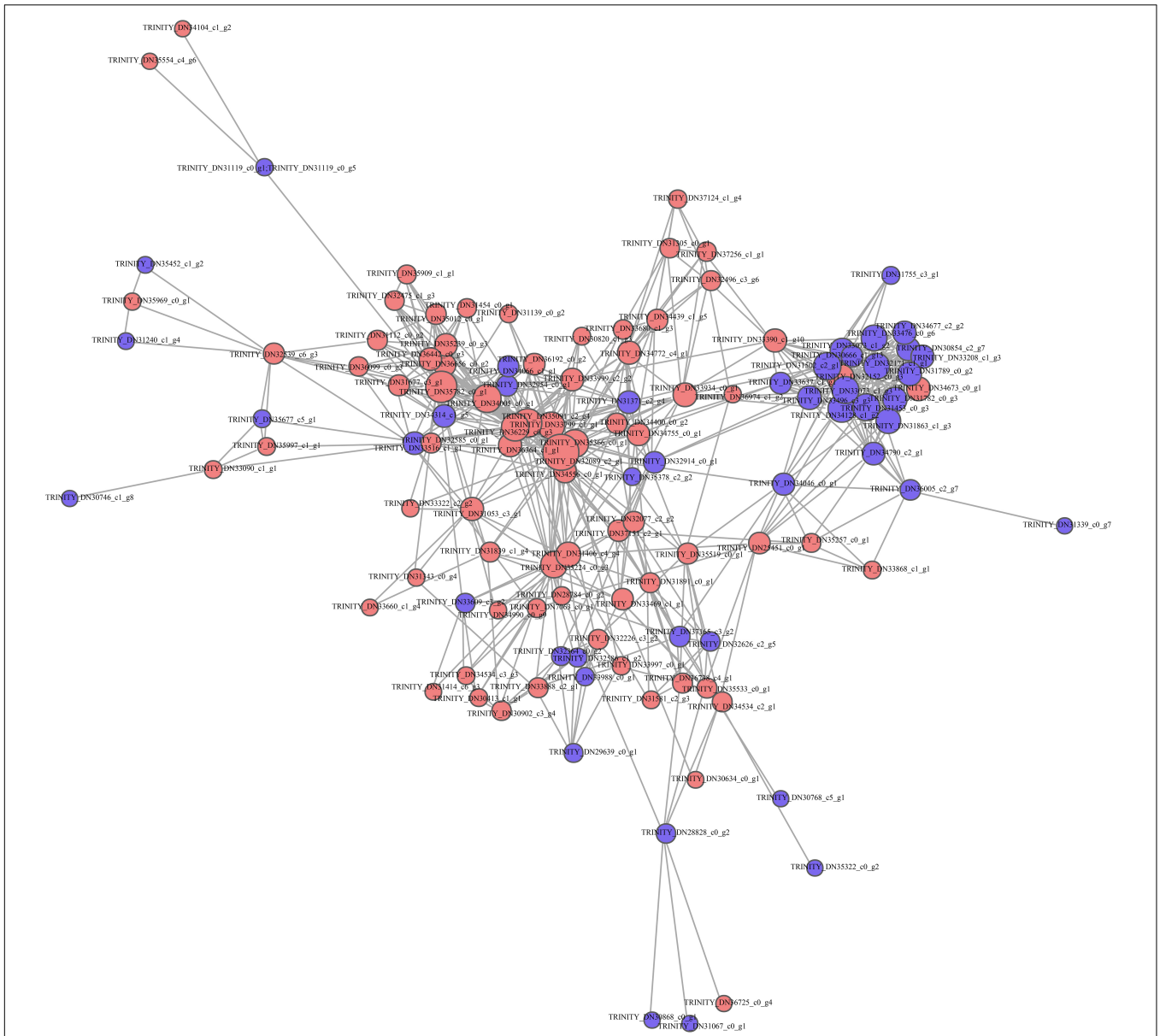


FIGURE 4 | Functional annotation of DAPs based on STRING 9.0. Red and blue stand for markedly up-regulated and down-regulated proteins, separately.

levels. Altogether 128 DAPs related to protein synthesis, which included biosynthesis of amino acids, the nitrogen metabolism pathway, carbon metabolism, starch/sucrose metabolism, glycine/serine/threonine metabolism, protein synthesis-related proteins, and aminoacyl-tRNA biosynthesis, were clustered closely (**Figure 3B**).

PPI Network Analysis

This work established a PPI network based on STRING database to predict the biological function of protein synthesis in kenaf leaves. Most differentially expressed proteins are associated with functional interactions. A total of 128 DAP-related protein syntheses in kenaf leaves, including 75 upregulated and 53 downregulated DAPs, were incorporated in interaction network (**Figure 4**).

Comparison Gene Levels and Protein Levels

A higher number of DEGs (1,350 with up-regulation and 2,428 with down-regulation) was obtained than DAPs (636 with up-regulation and 684 with down-regulation). For evaluating the relation of proteomic and transcriptomic alterations during protein synthesis, correlation analysis was performed using the quantification of DAPs and DEGs. As a result, 134 DAPs and their corresponding DEGs were identified. Of these, 122 DAP abundances increased and 12 DAP abundances decreased

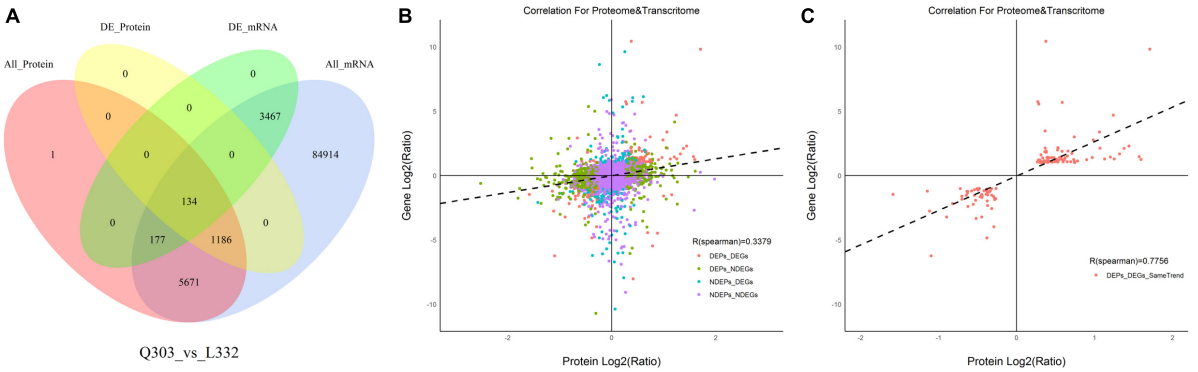


FIGURE 5 | Associations of protein and mRNA levels. **(A)** Venn diagram showing proteome (pink)-, transcriptome (blue)-quantified proteins, DAPs (yellow), and DEGs (green) within Q303/L332. **(B)** Scatterplot of the relationship between genes identified in proteomic and transcriptomic analyses of Q303/L332. **(C)** Scatterplot showing the correlation coefficients of DEGs with DEPs (the same trend) in Q303/L332.

(**Figure 5A**). The results showed that there were more DEGs than DAPs, and there were significant differences in transcription level and protein abundance.

Moreover, Pearson's correlation test suggested that FCs of DAPs showed moderate correlation with specific DEGs ($r = 0.3379$, $p < 0.01$), which indicated that transcript contents were more closely related to protein levels (**Figure 5B**). Meanwhile, FCs in DAPs exhibited positive relation to DEGs ($r = 0.7756$, $p < 0.01$) (**Figure 5C**).

Identification of Differentially Expressed Genes and Differentially Abundant Protein Related to Candidate Pathways

For better understanding those co-expressed DEG-DAP genes for their biological roles, this work performed enrichment analysis on the basis of GO as well as KEGG analysis. Consequently, the markedly enriched BP terms were cellular processes and metabolic processes, whereas cellular parts and binding and catalytic activities were the obviously enriched CC and MF terms, separately (**Supplementary Figure 4A**). According to the association analysis, 134 DAPs were enriched in 160 metabolic pathways, 10 significantly enriched pathways in the transcriptome, and 20 significantly enriched pathways in the proteome. There were fifty-seven common pathways between proteomic and transcriptomic data. The comparison demonstrated 2 common pathways, namely, phenylpropanoid biosynthesis, and starch/sucrose metabolism pathways, in proteomic and transcriptomic data (**Supplementary Figure 4B**). Therefore, these shared metabolic pathways might exert possible functions in the protein synthesis of kenaf leaves.

Validation of Quantitative Real-Time PCR Gene Expression With RNA-Seq Data

For validating whether our TMT and RNA-Seq data were reliable, this work conducted correlation and expression analyses by qPCR, and later acquired fragments per kilobase per million reads mapped (FPKM) values based on proteomic and transcriptomic data. Those 20 most significant DEGs were enriched into starch and sucrose metabolism, carbon fixation in photosynthetic organisms, glutathione metabolism, aminoacyl-tRNA biosynthesis, nitrogen metabolism, glycine/serine/threonine metabolism, and other protein synthesis of kenaf leaves were tested by qPCR experiments, and the results are shown in **Figure 6**. Specifically, most of these DEGs were significantly upregulated, including the starch and sucrose metabolism pathway related genes-sucrose-phosphate synthase (*SPS*), beta-fructofuranosidase (*INV1* and *INV2*), beta-amylase 3 (*BAM3-1*and *BAM3-2*) and sucrose synthase 6 (*SUS6*); glycolysis/gluconeogenesis metabolism related gene-aldehyde dehydrogenase (*ALDH*); carbon fixation in photosynthetic organisms related gene-fructose 1,6-bisphosphate 1-phosphatase activity (*FBP*); nitrogen metabolism related genes-nitrate reductase (*NR*), nitrate transporter (*NRT*), nitrite reductase (*NirA*); serine and threonine metabolism related gene-glycine dehydrogenase (*GLDC*),Glycine dehydrogenase (decarboxylating) A (*GDCSPA*) and aminoacyl-tRNA biosynthesis related gene-proline-tRNA ligase (*PARS*); and cysteine/methionine metabolism associated gene-*S*-adenosylmethionine decarboxylase (*SAMDC*), alanine/aspartate/glutamate metabolism associated gene-glutamate decarboxylase (*GAD*). Translation initiation factor 4E (*EIF4E*), transaldolase-like (*TAL*), glucan endo-1,3-beta-glucosidase (*EGLC*), and L-3-cyanoalanine synthase D-1 (*CYSD1*) were downregulated. Additionally, 14 potential genes, like *FBP* ($r = 0.9349$, $p < 0.01$), *SPS* ($r = 0.8301$, $p < 0.05$), *INV1* ($r = 0.9106$, $p < 0.05$), *INV2* ($r = 0.7154$, $p < 0.05$), *NRT* ($r = 0.6840$, $p < 0.05$), *NR* ($r = 0.6553$, $p < 0.05$), *NirA* ($r = 0.8855$, $p < 0.05$), *PARS* ($r = 0.8863$, $p < 0.01$), *GLDC* ($r = 0.8451$, $p < 0.01$), *GDCSPA* ($r = 0.7583$, $p < 0.01$), *SAMDC* ($r = 0.7693$, $p < 0.05$), *ALDH* ($r = 0.7815$, $p < 0.05$), *EIF4E* ($r = 0.8543$, $p < 0.01$), and *CYSD1* ($r = 0.7442$, $p < 0.05$), were strongly related to RNA-Seq results, whereas 6 genes showed low relation to related protein levels (**Figure 6** and **Table 3**). Generally speaking, qPCR analysis verified gene expression profiles acquired based on proteomic and transcriptomic data, indicating the reliability of our observations.

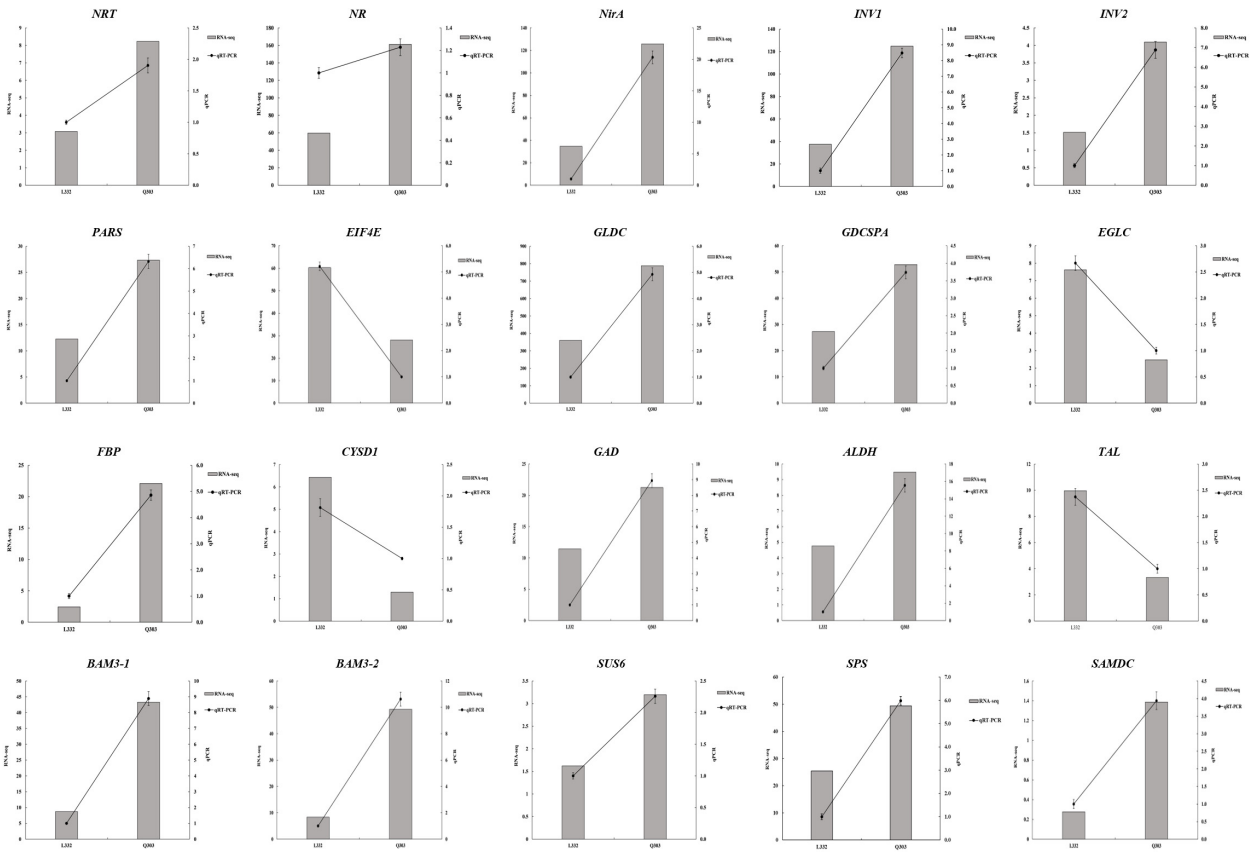


FIGURE 6 | qPCR analysis conducted to validate the chosen gene expression. RNA-seq and qPCR (left and right *y*-axes, separately) were conducted to reveal gene expression. RNA-seq detected gene levels were presented in histograms. qPCR was conducted to validate gene levels, as shown in the line graph.

TABLE 3 | Association of qPCR with FPKM values for those chosen genes.

Accession	Gene description	Gene name	Pearson correlation efficient	*P*-value	Numbers
TRINITY_DN36229_c0_g3	Fructose 1,6-bisphosphate 1-phosphatase activity	*FBP*	0.9349	0.0007	12
TRINITY_DN31677_c3_g1	Sucrose-phosphate synthase	*SPS*	0.8301	0.0392	12
TRINITY_DN32539_c6_g3	Beta-fructofuranosidase, INV1	*INV1*	0.9106	0.0248	12
TRINITY_DN30709_c0_g1	Beta-fructofuranosidase, INV2	*INV2*	0.7154	0.0306	12
TRINITY_DN36622_c1_g1	Nitrate transporter	*NRT*	0.6840	0.0482	12
TRINITY_DN35530_c0_g1	Nitrate reductase	*NR*	0.6553	0.0330	12
TRINITY_DN32077_c2_g2	Nitrite reductase	*NirA*	0.8855	0.0344	12
TRINITY_DN33868_c1_g1	Proline–tRNA ligase	*PARS*	0.8863	0.0018	12
TRINITY_DN37151_c2_g1	Glycine dehydrogenase	*GLDC*	0.8451	0.0002	12
TRINITY_DN34120_c0_g3	Beta-amylase 3-1	*BAM3-1*	0.4584	0.1844	12
TRINITY_DN34120_c0_g1	Beta-amylase 3-2	*BAM3-2*	0.6295	0.1017	12
TRINITY_DN37151_c2_g4	Glycine dehydrogenase (decarboxylating) A	*GDCSPA*	0.7583	0.0004	12
TRINITY_DN33517_c1_g4	*S*-adenosylmethionine decarboxylase	*SAMDC*	0.7693	0.0060	12
TRINITY_DN30966_c3_g1	Aldehyde dehydrogenase	*ALDH*	0.7815	0.0186	12
TRINITY_DN35582_c0_g6	Glutamate decarboxylase	*GAD*	0.5173	0.7802	12
TRINITY_DN32721_c0_g1	Sucrose synthase 6	*SUS6*	0.3489	0.7694	12
TRINITY_DN31962_c2_g2	Translation initiation factor 4E	*EIF4E*	0.8543	0.0003	12
TRINITY_DN32625_c1_g1	L-3-cyanoalanine synthase D-1	*CYSD1*	0.7442	0.0101	12
TRINITY_DN31110_c0_g7	Transaldolase-like	*TAL*	0.3624	0.1873	12
TRINITY_DN31240_c1_g4	Glucan endo-1,3-beta-glucosidase	*EGLC*	0.3975	0.1172	12

Moreover, to further validate the proteomics data, we quantified the protein abundance of 40 randomly selected genes by PRM and analyzed the protein abundance of 20 specific genes associated with protein synthesis, 19 of which were successfully quantified. Fifteen of the 19 (78.9%) proteins exhibited close abundance trends in PRM compared with TMT analysis, including ferredoxin-nitrite reductase (NirA), glutathione *S*-transferase DHAR2 (DHAR2), prolyl-tRNA synthetase (PARS), 3-phosphoshikimate 1-carboxy vinyl transferase (EPSPS), fructose-bisphosphate aldolase (ALD), peroxidase (PRX), ferredoxin-NADP+ reductase (petH), catalase isozyme 2 (CAT2), divinyl chlorophyllide an 8-vinyl-reductase (DVR), 1,4-alpha-glucan-branching enzyme 1 (GBE1), glycine dehydrogenase (GLDC), beta-fructofuranosidase INV1 (INV1), fructose 1,6-bisphosphate 1-phosphatase activity (FBP), cysteine synthase (Csase), and glycine-tRNA ligase (GlyRS). Additionally, four genes [glycerate dehydrogenase (HPR), glyceraldehyde-3-phosphate dehydrogenase GAPA2 (GAPA2), fructokinase-2 (FRK2), glutamine synthetase cytosolic isozyme 2 (GS1-2)] exhibited different levels relative to TMT-measured protein levels (**Table 4**). Thus, the results showed that PRM results were well correlated with proteomics analysis results, besides, some genes related to protein synthesis-associated metabolic pathways were consistently up-regulated and down-regulated within proteome and transcriptome (**Figure 7**).

TABLE 4 | Comparison of PRM and TMT quantification results.

Accession	Protein description	Gene name	Q303/L332 ratio	
			PRM	TMT
TRINITY_DN32077_c2_g2	Ferredoxin-nitrite reductase	NirA	1.253	1.440
TRINITY_DN37151_c2_g1	Glycine dehydrogenase	GLDC	1.4282	1.529
TRINITY_DN32539_c6_g3	Beta-fructofuranosidase INV1	INV1	1.176	1.286
TRINITY_DN36229_c0_g3	Fructose 1,6-bisphosphate 1-phosphatase activity	FBP	1.969	2.714
TRINITY_DN33868_c1_g1	Prolyl-tRNA synthetase	PARS	1.843	1.486
TRINITY_DN33609_c3_g2	Glutathione *S*-transferase DHAR2	DHAR2	0.3936	0.7541
TRINITY_DN33988_c0_g1	3-phosphoshikimate 1-carboxyvinyltransferase	EPSPS	0.9995	0.7456
TRINITY_DN33999_c2_g2	Fructose-bisphosphate aldolase	ALD	1.718	1.676
TRINITY_DN34106_c0_g1	Peroxidase	PRX	2.114	2.380
TRINITY_DN34400_c0_g2	Ferredoxin–NADP+ reductase	petH	1.028	1.387
TRINITY_DN34755_c0_g1	Glycerate dehydrogenase HPR	HPR	0.3105	1.408
TRINITY_DN35224_c0_g3	Catalase isozyme 2	CAT2	4.766	1.442
TRINITY_DN35677_c6_g1	Divinyl chlorophyllide a 8-vinyl-reductase	DVR	1.175	1.392
TRINITY_DN35909_c1_g1	1,4-Alpha-glucan-branching enzyme 1	GBE1	4.591	1.381
TRINITY_DN33469_c1_g1	Glutamine synthetase cytosolic isozyme 2	GS1-2	0.5675	1.219
TRINITY_DN33516_c1_g1	Fructokinase-2	FRK2	1.401	0.6498
TRINITY_DN35366_c0_g1	Glyceraldehyde-3-phosphate dehydrogenase GAPA2	GAPA2	0.7153	1.375
TRINITY_DN31891_c0_g1	Cysteine synthase	CSase	1.413	1.257
TRINITY_DN25451_c0_g1	Glycine-tRNA ligase	GlyRS	1.015	1.283

DISCUSSION

Kenaf is a new plant protein source, but there are few reports on the mechanism of kenaf leaf protein synthesis. In the current work, by determining the crude protein content in the leaves of two kenaf materials, it was determined that the protein content of the kenaf material "Q303" leaves was higher; combined with multiomics data, kenaf samples were systematically studied to obtain mRNA-protein correlations; then, the genes/proteins related to protein synthesis in kenaf leaves were obtained.

General Features of the Transcriptomes and Proteomes of Different Kenaf Leaves

The high protein content in leaves is the key factor affecting kenaf's multi-functional breeding. Illustrating the molecular mechanism of protein synthesis in leaves is helpful for the development of forage products using kenaf leaves. However, the underlying mechanism of kenaf protein synthesis is greatly unknown. According to TMT and RNA-seq data, the current work explored the different proteomic and transcriptomic data of two kenaf materials at the same stage. Since the transcriptome database is applied in identifying proteins, transcriptomic data assembly and sequencing quality is important for the following study. Totally, this work assembled 101,679 unigenes (>200 bp) into kenaf transcriptome, and the total was much higher than that formerly reported for this species as well as other Malvaceae Juss. plants, for example, [55,623 unigenes by Tang et al. (2021); 71,318 unigenes by Li et al. (2016)], and *Hibiscus rosa-sinensis* (30988 unigenes) (Trivellini et al., 2016). The obtained percentages of the Q20 bases (98.08-98.15%), Q30 bases (94.23-94.38%), with close GC concentrations (47.07-48.00%) to other researches on kenaf transcriptome, e.g., (98.19–98.31%), (94.45-94.71%), and (46.15-46.47%) (Tang et al., 2021), respectively. Li et al. (2016) only studied Q20 bases (97.70-97.77%) and GC contents (45.42-46.05%). In total, 60171 (59.18%) and 40606 (39.94%) of all unigenes detected in the present work matched those in SwissProt and NR databases, separately. The findings supported the results obtained by *Hibiscus* L. (20,547 and 13,184; 66.31 and 42.55%) (Trivellini et al., 2016). Therefore, our results provide a wide range of sequence and unigene resources for *Malvaceae* Juss.

In addition, 66,347 match spectra, 26,972 unique peptides, and 7,196 proteins were detected in accordance with the transcriptome data from kenaf leaves. RNA-seq and protein sequencing methods have found and annotated many genes and proteins, providing the foundation for a more correct and thorough description of molecular processes and analysis

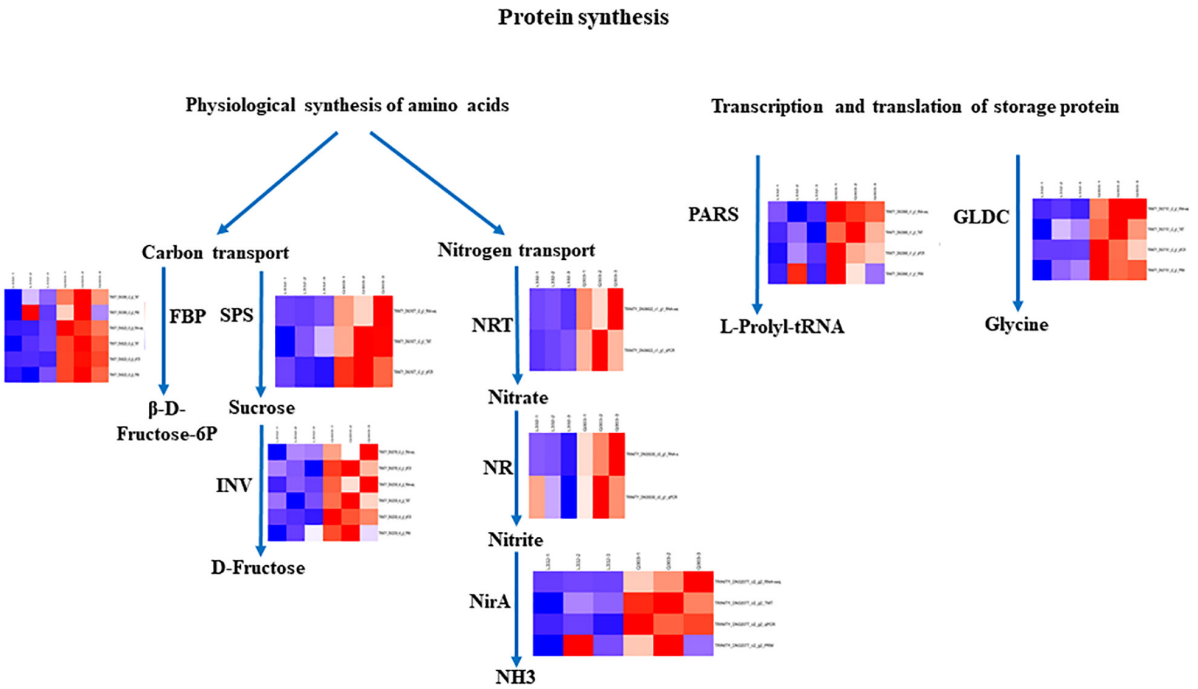


FIGURE 7 | Certain biological pathways related to protein synthesis in kenaf leaves. TRINITY_DN33999_c2_g2 (ALD); TRINITY_DN36622_c1_g1 (NRT), TRINITY_DN35530_c0_g1 (NR); TRINITY_DN30709_c0_g1 (INV2); TRINITY_DN31677_c3_g1 (SPS); TRINITY_DN36229_c0_g3 (FBP); TRINITY_DN32539_c6_g3 (INV1); TRINITY_DN32077_c2_g2 (NirA); TRINITY_DN33868_c1_g1 (PARS); TRINITY_DN37151_c2_g1 (GLDC). Red, blue, and white stand for markedly up-regulated, down-regulated and non-significant proteins, separately.

of genetic regulation mechanisms and complex physiological processes (Vanessa et al., 2012). The data presented here are sufficient and accurate and will provide useful guidance tools for the study of protein synthesis in kenaf leaves and other mallow plants. In addition, there were a lot of "predicted," "putative," or "uncharacterized" transcripts and proteins in the annotation results, suggesting that the nature of the analysis was restricted because of the lack of detailed genomic information (Niu et al., 2020). As a result, the functions of these unknown or unidentified genes and proteins in the protein synthesis of kenaf leaves should be explored in further researches.

Translational and Post-translational Regulation on Protein Accumulation in Kenaf Leaves

The regulation of storage protein gene expression is complex and includes transcription regulation, processing regulation for transcription products, translation regulation, hormone regulation, and so on (Kroj et al., 2003). As a result, the differences in the expression of genes and proteins in a single period alone cannot fully and correctly reflect the cause of protein accumulation in kenaf leaves. In the current work, the comparative analysis of gene expression and protein levels suggested that there existed more DEGs than DAPs. This may be due to protein post-translational modifications that manage the degradation or secretion of the protein, the technical limitations of MS-based proteomics, and the availability of a few samples. These reasons limit the ability to identify proteins (Zhou et al., 2015). Investigating the association between proteins and the network formed by their interaction is significant to show the roles of proteins (Xing et al., 2016). The PPI analysis showed that the interaction between proteins was obvious, and there was a complex network between interacting proteins. The interactions between proteins constitute a major part of the cellular biochemical reaction network, which is of great significance in regulating cells and their signals (Liu et al., 2021). In addition, correlation analysis demonstrated that there existed a weak correlation between the proteome and transcriptome, suggesting that there existed low correlation between transcription level and protein abundance. The findings were similar to those found in former studies, indicating that post-transcriptional and post-translational regulation, reversible phosphorylation, cell splicing events and translation efficiency exert a vital function in the regulation of protein accumulation in leaves (Chen et al., 2019). Therefore, gene translation and post-translational modification could be important methods of protein accumulation in leaves.

Potential Regulators and Metabolic Pathways During the Process of Protein Accumulation in Leaves

The GO and KEGG functional enrichment provides the definition and description of gene and protein functions and integrates genomic, chemical and systemic functional

information. GO term analysis showed that in addition to the co-enrichment of glycine dehydrogenase (decarboxylation) activity and serine family amino acid catabolic processes related to amino acid metabolism, oxidoreductase activity was also co-enriched. Similar to former researches, the target genes can be classified into different categories based on their roles, such as amino acid anabolism and peroxidase activity (Skadsen and Cherry, 1983). Peroxidase participates in the photorespiration of plant cells and redox equilibrium reactions (Jin et al., 2013). Glycine is mainly used as a glycine residue in leaf protein synthesis (Guo et al., 2021). These modifications improved the efficiency of photosynthesis and provided raw materials for protein synthesis. KEGG analysis showed that starch and sucrose metabolism is a shared pathway of DEGs and DAPs and other pathways involved in carbon metabolism, which provide basic substances for protein synthesis. Starch and sucrose metabolism can generate a vital function in the process of protein accumulation in kenaf leaves. In addition to the 21 amino acid synthesis pathways, other metabolic pathways are indirectly involved in amino acid synthesis (Iqbal et al., 2020). These processes include gluconeogenesis, the pentose phosphate pathway, the citrate cycle, carbon fixation in photosynthetic organisms as well as pyruvate metabolism. These metabolic pathways are closely related to the synthesis of amino acids and provide raw materials for protein synthesis. Furthermore, the findings showed that the proteomic data and transcriptome data were complementary and that the proteome could verify the transcriptome data; besides, genes exert the same role at the transcriptome and proteome levels (Muers, 2011). Besides, the functional classification of the transcriptome and proteome is conducive to improving the comprehension of protein synthesis physiology and molecular biology.

Key Candidate Differentially Expressed Genes and Differentially Abundant Proteins Involved in Leaf Protein Synthesis

Protein synthesis is a complex physiological, biochemical, and molecular biological process. It is not directly determined by a single gene but is jointly controlled by a lot of genes with different functions (Skadsen and Cherry, 1983). Fructose-1,6-bisphosphate aldolase (ALD) is a critical enzyme for the conversion from 3C to 6C compounds after CO_2 fixation. The upregulated ALD catalyzes the alcohol aldehyde condensation of dihydroxyacetone phosphate (DHAP) and glyceraldehyde 3-phosphate (GAP) to produce more fructose-1,6-diphosphate (FDP). Fructose-1,6-bisphosphatase (FBP) catalyzes fructose-1,6-diphosphate (FDP) and water with the purpose of generating fructose-6-phosphate and inorganic phosphorus, providing raw material for protein synthesis and has been verified in other crops (Li et al., 2022). Upregulated sucrose phosphate synthase (SPS) catalyzes UDP glucose and fructose 6-phosphate in order to yield more sucrose 6-phosphate (S6P), and sucrose 6-phosphate is hydrolyzed to produce sucrose. The main enzymes of sucrose catabolism are sucrose invertase (INV) and sucrose synthase (SUS), which degrade sucrose to glucose and fructose (Duan et al., 2021). The sucrose synthesized by photosynthesis is converted into reducing sugar by INV to supply the growth of young tissues and provide raw materials for protein synthesis (Deng et al., 2019). Nitrate transporters (NRTs) are a key factor in plant perception, absorption, and transport of nitrate (Zhang et al., 2021). The first step of nitrate degradation occurs in the cytoplasm. Nitrate is lowered to nitrite by nitrate reductase (NR). Nitrite enters chloroplasts or plastids and is degraded to ammonium by nitrite reductase (NirA) to provide nitrogen for protein synthesis (Eady et al., 2016). Aminoacylation of tRNA is the first step in protein synthesis. In this process, aminoacyl tRNA synthetases (aaRSs) connect specific amino acids to homologous tRNA to complete protein synthesis (Smirnova et al., 2012). Some examples include Val tRNA synthetase and threonyl tRNA synthetases (Wang et al., 2016; Zhang et al., 2017). Amino acids play a vital function in plant central metabolism. Amino acids can also play the role of intermediates of final metabolites in some metabolic pathways and be engaged in regulating various metabolic pathways and other physiological and biochemical pathways, thereby influencing protein synthesis. The Aspartic acid (Asp) participates in protein synthesis and provides raw materials for protein synthesis (Yang et al., 2020).

In the current work, 14 DEGs and 15 DAPs which were engaged in protein synthesis presented strong associations with RNA-seq data and protein expression levels, and these correlations are consistent with the results that the metabolic pathway that is directly or indirectly involved in amino acid synthesis may play a key role in protein synthesis (Iqbal et al., 2020). Obviously, the expression levels of four proteins (FBP, NirA, PARS and GLDC) related to carbon fixation in photosynthetic organisms, nitrogen metabolism, aminoacyl-tRNA biosynthesis and glycine, serine and threonine metabolic pathways demonstrated higher protein and gene expression in kenaf leaves with high protein content, indicating FBP, NirA, PARS and GLDC may exert an essential function in protein synthesis of kenaf leaves.

The protein and gene expression levels of ALD, SPS, INV, NR, and NRT in kenaf leaves with a high protein content were higher than those of kenaf leaves with a low protein content. The expression of metabolic proteins related to carbon transport and nitrogen transport increased significantly, indicating that the physiological synthesis of raw amino acids and the transcription and translation of storage protein genes constitute a network in kenaf leaves with high protein content. In future work, transgenic plants are required to overexpress candidate genes to verify the functions of these genes.

CONCLUSION

In the present study, two kenaf cultivars with significant differences in protein content were used as materials. The leaf protein content were determined using the Kjeldahl method during the most important kenaf growth period. More importantly, the potential genes/proteins involved in protein synthesis in kenaf leaves were identified by comparative transcriptomic analysis and proteomic analysis. Protein content is regarded as a typical quality trait regulated by multiple genes,

and the genetic mechanism are complex and related to multiple pathways. The four co-expressed genes (*FBP*, *NirA*, *PARS*, and *GLDC*) identified in this study were suggested involved in carbon transport metabolism, nitrogen transport metabolism and multiple signaling pathways for the transcription and translation of storage protein genes. The research provides new insight into the process of protein synthesis in kenaf leaves.

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

ChZ and YD were in charge of study design and manuscript drafting. GZ, JL, AX, LZ, AC, HT, LC, and GP assisted in data analysis. YW, JZ, CuZ, ZB, HL, JW, and DY contributed to obtaining and interpreting data in this study. SH and DL were responsible for manuscript revision. All authors approved the final manuscript for submission.

FUNDING

This present work was funded by the National Science Foundation of China (32160457), the China Agriculture Technology Research System (CARS-16-E05), the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (ASTIP-IBFC03), the National Science Foundation of Jiangxi (20212BAB205026), and the Science and Technology Project of Education Department of Jiangxi Province (GJJ180890).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.879874/full#supplementary-material

Supplementary Figure 1 | Size distributions of kenaf's unigene and assembled transcript sequences.

Supplementary Figure 2 | Volcano plot showing kenaf's proteomic and transcriptomic data. **(A)** Volcano plot showing transcriptomic data of Q303/L332. **(B)** Volcano plot showing proteomic data of Q303/L332. Absolute log2-FC and log10-FC values serve as *x*- and *y*-axes, separately.

Supplementary Figure 3 | TMT proteomics-identified peptide counts and molecular weights (MWs) based on database. **(A)** Protein distribution across diverse MWs. **(B)** TMT-detected peptide quantity in proteins.

Supplementary Figure 4 | Gene Ontology as well as KEGG analysis of co-expressed proteins and genes within kenaf. **(A)** GO functional annotation of co-expressed proteins and genes within kenaf. **(B)** KEGG analysis on co-expressed proteins and genes within kenaf.

Supplementary Table 1 | Kenaf's sequencing statistics.

Supplementary Table 2 | Unigene annotational statistics.

Supplementary Table 3 | Overall protein and transcript quantities.

Supplementary Table 4 | Differentially abundant proteins (DAPs) and differentially expressed genes (DEGs) discovered.

Supplementary Table 5 | Primer sequences utilized in qPCR assay.

REFERENCES

Akcura, M., Turan, V., Kokten, K., and Kaplan, M. (2019). Fatty acid and some micro element compositions of cluster bean (*Cyamopsis tetragonoloba*) genotype seeds growing under Mediterranean climate. *Ind. Crop Prod*. 128, 140–146. doi: 10.1016/j.indcrop.2018.10.062

An, X., Jin, G. R., Zhang, J. Y., Luo, X. H., Chen, C. L., Li, W. L., et al. (2018). Protein responses in kenaf plants exposed to drought conditions determined using iTRAQ technology. *FEBS Open Bio*. 8, 1572–1583. doi: 10.1002/2211-5463.12507

Ayadi, R., Hanana, M., Mzid, R., Hamrouni, L., Khouja, M. I., and Hanachi, A. S. (2017). *Hibiscus cannabinus* L.-kenaf: a review paper. *J. Nat. Fibers*. 14, 466–484. 2016.1240639 doi: 15440478

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B*. 57, 289–300. 6161.1995.tb02031.x doi: 10.1111/j.2517-

Bilen, S., Bilen, M., and Turan, V. (2019). Relationships between cement dust emissions and soil properties. *Pol. J. Environ. Stud*. 28, 3089–3098. doi: 10.15244/pjoes/92521

Birhanie, Z. M., Xiao, A. P., Yang, D. W., Huang, S. Q., Zhang, C., Zhao, L. N., et al. (2021). Polysaccharides, total phenolic, and flavonoid content from different kenaf (*Hibiscus cannabinus* L.) genotypes and their antioxidants and antibacterial properties. *Plants-Basel* 10:1900. doi: 10.3390/plants10091900

Chai, L. J., Li, Y. M., Chen, S. W., Perl, A., Zhao, F. X., and Ma, H. Q. (2014). RNA sequencing reveals high resolution expression change of major plant hormone pathway genes after young seedless grape berries treated with gibberellin. *Plant Sci*. 229, 215–224. sci. 2014.09. 010 doi: 10.1016/j.plant

Chen, X., Mao, X. Z., Huang, J. J., Yang, D., Wu, J. M., Dong, S., et al. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39, 316–322. doi: 10.1093/nar/gkr483

Chen, X. Q., Tao, Y., Ali, A., Zhuang, Z. H., Guo, D. M., Guo, Q. L., et al. (2019). Transcriptome and proteome profiling of different colored rice reveals physiological dynamics involved in the flavonoid pathway. *Int. J. Mol. Sci*. 20:2463. doi: 10.3390/ijms20102463

Cheng, Z., Lu, B. R., Sameshima, K., Fu, D. X., and Chen, J. K. (2004). Identification and genetic relationships of kenaf (*Hibiscus cannabinus* L.) germplasm revealed by AFLP analysis. *Genet. Resour. Crop Ev.* 51, 393–401. doi: 10.1023/b:gres.0000023454.96401.1c

Cui, Y. Y., Wang, Z. R., Chen, S. W., Vainstein, A., and Ma, H. Q. (2019). Proteome and transcriptome analyses reveal key molecular differences between quality parameters of commercial-ripe and tree-ripe fig (*Ficus carica* L.). *BMC Plant Biol*. 19:146. 12870-019-1742-x doi: 10.1186/s

Deng, S. Y., Mai, Y. T., and Niu, J. (2019). Fruit characteristics, soluble sugar compositions and transcriptome analysis during the development of Citrus maxima "seedless", and identification of *SUS* and INV genes involved in sucrose degradation. *Gene* 689, 131–140. doi: 10.1016/j.gene.2018.12.016

Deng, Y., Li, D. F., Huang, Y. M., and Huang, S. Q. (2017). Physiological response to cadmium stress in kenaf (*Hibiscus cannabinus* L.) seedlings. *Ind. Crop Prod*. 107, 453–457. doi: 10.1016/j.indcrop.2017.06.008

Duan, Y. K., Yang, L., Zhu, H. J., Zhou, J., Sun, H., and Gong, H. J. (2021). Structure and expression analysis of sucrose phosphate synthase, sucrose synthase and invertase gene families in *Solanum lycopersicum*. *Int. J. Mol. Sci.* 22:4698. doi: 10.3390/ijms22094698

Eady, R. R., Antonyuk, S. V., and Hasnain, S. S. (2016). Fresh insight to functioning of selected enzymes of the nitrogen cycle. *Curr. Opin. Chem. Biol.* 31, 103–112. 2016.02.009 doi: 10.1016/j.cbpa

Falasca, S. L., Ulberich, A. C., and Pitta-Alvarez, S. (2014). Possibilities of growing kenaf (*Hibiscus cannabinus* L.) in Argentina as biomass feedstock under dry-subhumidand semi-arid climate conditions. *Biomass. Bioenerg.* 64, 70–80. 03.031 doi: 10.1016/j.biombioe.2014

Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., et al. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell Proteom.* 11:O111.016717. doi: 10.1074/mcp.O111.016717

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Grzeszczuk, M., Stefaniak, A., Meller, E., and Wysocka, G. (2018). Mineral composition of some edible flowers. *J. Elementol.* 23, 151–162. doi: 10.5601/jelem.2017.22.2.1352

Guo, J. H., Gong, B. Q., and Li, J. F. (2021). *Arabidopsis* lysin motif/F−box−containing protein InLYP1 fine−tunes glycine metabolism by degrading glycine decarboxylase GLDP2. *Plant J.* 106, 394–408. doi: 10.1111/TPJ.15171

Huang, T., Luo, X., Wei, M., Shan, Z., Zhu, Y., Yang, Y., et al. (2020). Molecular cloning and expression analysis of sucrose phosphate synthase genes in cassava (*Manihot esculenta Crantz*). *Sci. Rep.* 10:20707. doi: 10.1038/s41598-020-77669-9

Iqbal, A., Dong, Q., Wang, X. R., Gui, H. P., Zhang, H. H., Zhang, X. L., et al. (2020). Transcriptome analysis reveals differences in key genes and pathways regulating carbon and nitrogen metabolism in cotton genotypes under N starvation and resupply. *Int. J. Mol. Sci.* 21:1500. doi: 10.3390/ijms21041500

Jeannin, G., Burkard, G., and Weil, J. H. (1976). Aminoacylation of phaseolus vulgaris cytoplasmic, chloroplastic and mitochondrial trnaspro and trnaslys by homologous and heterologous enzymes. *Biochim. Biophys. Acta* 442, 24–31. 171-174 doi: 10.1016/0005-2787(76)90

Jérme, V., and Richard, D. T. (2008). Transcriptional regulation of storage protein synthesis during dicotyledon seed filling. *Plant Cell Physiol.* 49, 1263–1271. pcn116 doi: 10.1093/pcp/

Jin, X. N., Fu, Z. Y., Ding, D., Li, W. H., Liu, Z. H., and Tang, J. H. (2013). Proteomic identification of genes associated with maize grain-filling rate. *PLoS One* 8:e59353. doi: 10.1371/journal.pone.0059353

Kashif, M. H., Wei, F., Tang, D., Tang, M., Luo, D., Hai, L., et al. (2020b). iTRAQ-based comparative proteomic response analysis reveals regulatory pathways and divergent protein targets associated with salt-stress tolerance in kenaf (*Hibiscus cannabinus* L.). *Ind. Crop Prod.* 153:112566. doi: 10.1016/j.indcrop.2020.112566

Kashif, M. H., Tang, D., Li, Z., Wei, F., Liang, Z., and Chen, P. (2020a). Comparative cytological and gene expression analysis reveals potential metabolic pathways and target genes responsive to salt stress in kenaf (*Hibiscus cannabinus* L.). *J. Plant Growth Regul.* 39, 1245–1260. doi: 10.1007/s00344-019-10062-7

Kroj, T., Savino, G., Valon, C., Giraudat, J., and Parcy, F. (2003). Regulation of storage protein gene expression in *Arabidopsis*. *Development* 130, 6065–6073. 00814 doi: 10.1242/dev

Li, H., Li, D. F., Chen, A. G., Tang, H. J., Li, J. J., and Huang, S. Q. (2016). Characterization of the kenaf (*Hibiscus cannabinus*) global transcriptome using Illumina paired-end sequencing and development of EST-SSR Markers. *PLoS One* 11:e0150548. 0150548 doi: 10.1371/journal.pone

Li, Y. Y., Guo, L. N., Liang, C. Z., Meng, Z. G., Tahira, S., Guo, S. D., et al. (2022). Overexpression of *Brassica napus* cytosolic fructose-1,6-bisphosphatase and sedoheptulose -1,7-bisphosphatase genes significantly enhanced tobacco growth and biomass. *J. Integr. Agr.* 21, 49–59. doi: 10.1016/S2095-3119(20)63438-4

Li, Z., Hu, Y., Chang, M., Kashif, M. H., Tang, M., Luo, D., et al. (2021). 5-azacytidine pre-treatment alters DNA methylation levels and induces genes responsive to salt stress in kenaf (*Hibiscus cannabinus* L.). *Chemosphere* 271:129562. 2021.129562 doi: 10.1016/j.chemosphere

Liu, D. W., Zhan, J., Luo, Z. J., Zeng, N. B., Zhang, W., Zhang, H., et al. (2021). Quantitative proteomics and relative enzymatic activities reveal different mechanisms in two peanut cultivars (*Arachis hypogaea* L.) under waterlogging conditions. *Front. Plant Sci.* 12:716114. doi: 10.3389/fpls.2021.716114

Liu, Y. L., Cao, D., Ma, L. L., Jin, X. F., Yang, P. F., Ye, F., et al. (2018). TMT-based quantitative proteomics analysis reveals the response of tea plant(*Camellia sinensis*) to fluoride. *J. Proteomics.* 176, 71–81. doi: 10.1016/j.jprot.2018.02.001

Lu, P., Magwanga, O. R., Lu, H. J., Kirungu, J. N., Wei, Y., Dong, Q., et al. (2018). A novel G-protein-coupled receptors gene from upland cotton enhances salt stress tolerance in transgenic *Arabidopsis*. *Genes* 9:209. doi: 10.3390/genes9040209

Lv, G. Y., Guo, X. G., Xie, L. P., Xie, C. G., Zhang, X. H., Yang, Y., et al. (2017). Molecular characterization, gene evolution, and expression analysis of the fructose-1, 6-bisphosphate aldolase (FBA) gene family in wheat (*Triticum aestivum* L.). *Front. Plant Sci.* 8:1030. doi: 10.3389/fpls.2017.01030

Muers, M. (2011). Gene expression: transcriptome to proteome and back to genome. *Nat. Rev. Genet.* 12:518. doi: 10.1038/nrg3037

Niu, J., Shi, Y. L., Huang, K. Y., Zhong, Y. C., Chen, J., Sun, Z. M., et al. (2020). Integrative transcriptome and proteome analyses provide new insights into different stages of *Akebia trifoliata* fruit cracking during ripening. *Biotechnol. Biofuels* 13:149. -020-01789-7 doi: 10.1186/s13068

Niu, X. P., Qi, J. M., Chen, M. X., Zhang, G. Y., Tao, A. F., Fang, P. P., et al. (2015). Reference genes selection for transcript normalization in kenaf (*Hibiscus cannabinus* L.) under salinity and drought stress. *Peer J.* 3:e1347. doi: 10.7717/peerj.1347

Okamura-Ikeda, K., Ohmura, Y., Fujiwara, K., and Motokawa, Y. (1993). Cloning and nucleotide sequence of the gcv operon encoding the *Escherichia coli* glycine-cleavage system. *Eur. J. Biochem.* 216, 539–548. doi: 10.1111/j.1432-1033.1993.tb18172.x

Okuda, N., and Sato, M. (2004). Manufacture and mechanical properties of binderless boards from kenaf core. *J. Wood Sci.* 50, 53–61. doi: 10.1007/s10086-003-0528-8

Pascoal, A., Quirantes-Piné, R., Fernando, A. L., Alexopoulou, E., and Segura-Carretero, A. (2015). Phytotoxic and fungitoxic activities of the essential oil of kenaf (*Hibiscus cannabinus* L.) leaves and its composition. *Ind. Crop Prod.* 78, 116–123. 10.028 doi: 10.1016/j.indcrop.2015

Saba, N., Jawaid, M., Hakeem, K. R., Paridah, M. T., Khalina, A., and Alothman, O. Y. (2015). Potential of bioenergy production from industrial kenaf (*Hibiscus cannabinus* L.) based on Malaysian perspective. *Renew. Sust. Energ. Rev.* 42, 446–459. 29 doi: 10.1016/j.rser.2014.10.0

Sahrawy, M., Concepción, Á, Chueca, A., Cánovas, F. M., and López-Gorgé, J. (2004). Increased sucrose level and altered nitrogen metabolism in *Arabidopsis thaliana* transgenic plants expressing antisense chloroplastic fructose-1,6-bisphosphatase. *J. Exp. Bot.* 55, 2495–2503. doi: 10.1093/jxb/erh257

Shen, J. Z., Zhang, D. Y., Zhou, L., Zhang, X. Z., Liao, J. R., Duan, Y., et al. (2019). Transcriptomic and metabolomic profiling of camellia sinensis L. cv. 'Suchazao' exposed to temperature stresses reveals modification in protein synthesis and photosynthetic and anthocyanin biosynthetic pathways. *Tree Physiol.* 39, 1583–1599. 59 doi: 10.1093/treephys/tpz0

Skadsen, R. W., and Cherry, J. H. (1983). Quantitative changes in vitro and in vivo protein synthesis in aging and rejuvenated soybean cotyledons. *Plant Physiol.* 71, 861–868. 104/pp.71.4.861 doi: 10.1

Smirnova, E. V., Lakunina, V. A., Tarassov, I., Krasheninnikov, I. A., and Kamenski, P. A. (2012). Noncanonical functions of aminoacyl-tRNA synthetases. *Biochem. Moscow.* 77, 15–25. doi: 10.1134/S0006297912010026

Swingle, R. S., Urias, A. R., Doyle, J. C., and Voigt, R. L. (1978). Chemical composition of kenaf forage and its digestibility by lambs and in vitro. *J. Anim. Sci.* 46, 1346–1350. 1978.4651346x doi: 10.2527/jas

Tang, D., Wei, F., Qin, S., Khan, A., Kashif, M. H., and Zhou, R. (2019). Polyethylene glycol induced drought stress strongly influences seed germination, root morphology and cytoplasm of different kenaf genotypes. *Ind. Crop Prod.* 137, 180–186. doi: 10.1016/j.indcrop.2019

Tang, M. Q., Li, Z. Q., Luo, D. J., Wei, F., Kashif, M. H., Lu, H., et al. (2021). A comprehensive integrated transcriptome and metabolome analyses to reveal key genes and essential metabolic pathways involved in CMS in kenaf. *Plant Cell Rep.* 40, 223–236. 9-020-02628-7 doi: 10.1007/s0029

Tiwari, J. K., Zinta, R., Kumar, V., Kumar, D., and Chakrabarti, S. K. (2019). Development of molecular marker for nitrate reductase (NR) gene to improve nitrogen use efficiency in potato. *Indian J. Hortic.* 76:80. doi: 10.5958/0974-0112.2019.00012.4

Trivellini, A., Cocetta, G., Hunter, D. A., Vernieri, P., and Ferrante, A. (2016). Satial and temporal transcriptome changes occurring during flower opening

and senescence of the ephemeral hibiscus flower, *Hibiscus rosa-sinensis*. *J. Exp. Bot* 67, 5919–5931. erw295 doi: 10.1093/jxb/

Vanessa, K. M., Liu, P., and Si, Y. Q. (2012). A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. *Am. J. Bot.* 99, 248–256. b.1100340 doi: 10.3732/aj

Wang, J., Mao, L., Zeng, Z., Yu, X., Lian, J., Feng, J., et al. (2021). Genetic mapping high protein content QTL from soybean 'Nanxiadou 25' and candidate gene analysis. *BMC Plant Biol*. 21:388. doi: 10.1186/s12870-021-03176-2

Wang, Y. L., Wang, C. M., Zheng, M., Lyu, J., Xu, Y., Li, X. H., et al. (2016). WHITE PANICLE1, a val-tRNA synthetase regulating chloroplast ribosome biogenesis in rice, is essential for early chloroplast development. *Plant Physiol*. 170, 2110–2123. doi: 10.1104/pp.15.01949

Wang, Y. Y., Hsu, P. K., and Tsay, Y. F. (2012). Uptake, allocation and signaling of nitrate. *Trends Plant Sci*. 17, 458–467. doi: 10.1016/j.tplants.2012.04.006

Wu, Y., Zhang, W., Xu, L., Wang, Y., Zhu, X. W., Li, C., et al. (2015). Isolation and molecular characterization of nitrite reductase (RsNiR) gene under nitrate treatments in radish. *Sci. Hortic. Amsterdam* 193, 276–285. doi: 10.1016/j.scienta.2015.07.01

Xie, J. L., Viviane, D. S. A., Tobias, V. D. H., Keefe, L. O., Lenchine, R. V., Jensen, K. B., et al. (2019). Regulation of the elongation phase of protein synthesis enhances translation accuracy and modulates lifespan. *Curr. Biol.* 29, 737–749. doi: 10.1016/j.cub.2019.01.029

Xing, S. P., Wallmeroth, N., Berendzen, K. W., and Grefen, C. (2016). Techniques for the analysis of protein-protein interactions in vivo. *Plant Physiol.* 171, 727–758. doi: 10.1104/pp.16.00470

Yang, Q. Q., Yu, W. H., Wu, H. Y., Zhang, C. Q., Sun, S. M., and Liu, Q. Q. (2020). Lysine biofortification in rice by modulating feedback inhibition of aspartate kinase and dihydrodipicolinate synthase. *Plant Biotechnol. J.* 19, 490–501. doi: 10.1111/pbi.13478

Zhang, J. Y., Han, Z. J., Lu, Y., Zhao, Y. F., Wang, Y. P., Zhang, J. Y., et al. (2021). Genome-wide identification, structural and gene expression analysis of the nitrate transporters (NRTs) family in potato (*Solanum tuberosum* L.). *PLoS One* 16:e0257383. doi: 10.1371/JOURNALPONE.0257383

Zhang, Y. Y., Hao, Y. Y., Wang, Y. H., Wang, C. M., Wang, Y. L., Long, W. H., et al. (2017). Lethal albinic seedling, encoding a threonyl-tRNA synthetase, is involved in development of plastid protein synthesis system in rice. *Plant Cell Rep.* 36, 1053–1064. doi: 10.1007/s00299-017-2136-x

Zhou, Y., Meng, Z., Edman-Woolcott, M., Hamm-Alvarez, S. F., and Zandi, E. (2015). Multidimensional separation using HILIC and SCX pre-fractionation for RP LC-MS/MS platform with automated exclusion list-based MS data acquisition with increased protein quantification. *J. Proteomics Bioinform*. 8, 260–265. doi: 10.4172/jpb.1000378

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 21 June 2022
doi: 10.3389/fpls.2022.874409

High-Altitude Genetic Selection and Genome-Wide Association Analysis of Yield-Related Traits in *Elymus sibiricus* L. Using SLAF Sequencing

*Zongyu Zhang[1], Yuying Zheng[1], Junchao Zhang[2], Na Wang[1], Yanrong Wang[1], Wenhui Liu[3], Shiqie Bai[4] and Wengang Xie[1]**

[1] The State Key Laboratory of Grassland Agro-Ecosystems, Key Laboratory of Grassland Livestock Industry Innovation, Ministry of Agriculture and Rural Affairs, College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, China, [2] Institute of Qinghai-Tibetan Plateau, Southwest Minzu University, Chengdu, China, [3] Key Laboratory of Superior Forage Germplasm in the Qinghai-Tibetan Plateau, Qinghai Academy of Animal Science and Veterinary Medicine, Xining, China, [4] Sichuan Academy of Grassland Science, Chengdu, China

The genetic adaptations to harsh climatic conditions in high altitudes and genetic basis of important agronomic traits are poorly understood in *Elymus sibiricus* L. In this study, an association population of 210 genotypes was used for population structure, selective sweep analysis, and genome-wide association study (GWAS) based on 88,506 single nucleotide polymorphisms (SNPs). We found 965 alleles under the natural selection of high altitude, which included 7 hub genes involved in the response to UV, and flavonoid and anthocyanin biosynthetic process based on the protein–protein interaction (PPI) analysis. Using a mixed linear model (MLM), the GWAS test identified a total of 1,825 significant loci associated with 12 agronomic traits. Based on the gene expression data of two wheat cultivars and the PPI analysis, we finally identified 12 hub genes. Especially, in plant height traits, the top hub gene (TOPLESS protein) encoding auxins and jasmonic acid signaling pathway, shoot apical meristem specification, and xylem and phloem pattern formation was highly overexpressed. These genes might play essential roles in controlling the growth and development of *E. sibiricus*. Therefore, this study provides fundamental insights relevant to hub genes and will benefit molecular breeding and improvement in *E. sibiricus* and other *Elymus* species.

Keywords: *Elymus sibiricus*, genetic signal selection, GWAS, high altitude, multi-algorithms analysis, yield-related traits, hub genes

OPEN ACCESS

Edited by:
Jorge Fernando Pereira, Embrapa Gado de Leite, Brazil

Reviewed by:
Yaxi Liu, Sichuan Agricultural University, China
Haidong Yan, University of Georgia, United States

****Correspondence:***
Wengang Xie
xiewg@lzu.edu.cn

Specialty section:
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

Received: *12 February 2022*
Accepted: *26 May 2022*
Published: *21 June 2022*

Citation:
Zhang Z, Zheng Y, Zhang J, Wang N, Wang Y, Liu W, Bai S and Xie W (2022) High-Altitude Genetic Selection and Genome-Wide Association Analysis of Yield-Related Traits in Elymus sibiricus L. Using SLAF Sequencing. Front. Plant Sci. 13:874409. doi: 10.3389/fpls.2022.874409

INTRODUCTION

Elymus is the largest and most widely distributed genus of Triticeae (Poaceae) with a valuable gene pool that can be used for the genetic improvement of cereal crops such as wheat (*Triticum aestivum* L.), barley (*Hordeum vulgare* L.), and rye (*Secale cereale* L.) due to its excellent stress tolerance, disease resistance, and good forage quality (Dewey, 1984; Ma et al., 2012a). Siberian wild rye (*Elymus sibiricus* L.), the type species of the genus *Elymus*, is a perennial, sparse-type, self-pollinating, and allotetraploid grass ($2n = 4\times = 28$). *Elymus sibiricus* is commercially useful in stock raising and serves as a major forage grass in natural and artificial grasslands. In addition, it is widely

being used in the restoration of ecosystems, soil stabilization, and erosion control due to its good ecological function (Xie et al., 2015).

Natural selection and artificial domestication in plants have a pivotal role in reshaping the genome, promoting the evolution of populations, high allele frequency in selective sweep regions, local adaptation, and genetically based trait differentiation (Chapman et al., 2013). The geographic distribution of wild *E. sibiricus* accessions mainly extended from Eastern Europe to Siberia, Mongolia, China, Japan, and even to Alaska and Canada, leaving extensive genetic variations and evolutionary signatures in the genome (Ma et al., 2012a). The Qinghai-Tibet Plateau (QTP) is a hot spot area for studying genetic adaptations, as the average elevation of about 4,000 m above sea level caused the extreme ecological conditions, such as high UV-B radiation and low oxygen concentration (∼30% stronger and ∼40% lower than at sea level, respectively; Yang et al., 2017). The genetic basis of harsh climatic adaptations in high-altitude regions has been reported and many genome selective regions and strongly associated environmentally related genes were found for humans, animals, and plants (Li M. et al., 2013; Yang et al., 2017; Wang et al., 2021a,b). *E. sibiricus* grows on wet meadows, shrubs, forests, and subalpine meadows, where altitude gradients range from 0 to 4,000 m above the sea level, yet the selective signatures at molecular level between high-altitude and low-altitude habitats remain unclear.

Recently, increasing attention has been paid to the requirement of high-yield and high-quality forage varieties, such as *E. sibiricus* (Jin et al., 2021). The annual demand for hay and grass seeds in China exceeds 10 million and 150 thousand tons, respectively. However, traditional breeding and phenotypic selection still rely heavily on experience and manual observation, which is time consuming, low efficiency, and environmentally sensitive. The genome-wide association study (GWAS) is a powerful tool for the rapid identification of loci or key genes associated with agronomic traits by analyzing associations between nucleotide polymorphisms and phenotypic diversity. These associations were measured by using a natural or low-structured population, in which a large number of recombination events were found based on linkage disequilibrium (LD; Flint-Garcia et al., 2003). The abundant genetic variations from GWAS facilitate more targeted and accurate genetic selection for marker-assisted selection (MAS) and molecular design breeding when compared with quantitative traits loci (QTL) mapping. Recently, advances in next-generation high-throughput sequencing technologies provide a strategy for high-density genotyping based on large-scale populations at the genome-wide level. An increasing number of GWASs have been reported in crops and forage grasses for important agronomic traits such as heading date, plant height, and panicle length in rice (*Oryza sativa* L.; Yano et al., 2016); fiber quality and yield traits in cotton (*Gossypium hirsutum* L.; Fang et al., 2017a); spike ethylene production and spike dry weight in wheat (Valluru et al., 2017); drought adaptation in maize (*Zea mays* L.; Zhang et al., 2021); salt germination, frost damage, fall dormancy, and unifoliate internode length in alfalfa (*Medicago sativa* L.; Shen et al., 2020); and flowering time in switchgrass (*Panicum virgatum* L.; Grabowski et al., 2017). Limited by high-sequencing costs and difficulties in genome assembly, the genomic information of most species is still unknown, especially for polyploidy plants with large size and complex genomes. Sun et al. (2013) tested an efficient approach (specific-locus amplified fragment sequencing, SLAF-seq) for *de novo* single nucleotide polymorphism (SNP) discovery, namely reduced-representation genome sequencing (RRGS) technology. Several distinguishing characteristics have propelled much of the practicability of SLAF-seq including the generation of millions of high-density SNPs, improvement of efficiency and performance–cost ratio by missing repetitive sequences, capability of detecting novel SNPs compared with SNP arrays, and suitability for species without a reference genome. The technology has successfully applied to many species with haplotype analysis, linkage mapping, and genetic association studies (Li et al., 2017; Zhang et al., 2019a; Zhuo et al., 2021).

Although GWAS test has the potential to detect numerous genetic polymorphisms or QTLs associated with complex agronomic traits, it is difficult to completely avoid possible spurious loci caused by population stratification and unequal relatedness among the studied individuals (Kang et al., 2010; Zhang et al., 2010). The mixed linear model (MLM) has been proven effective as a useful method to control-inflated associations for principal components (or population structure) and kinship modeling as a fixed effect and random effect, respectively (Yu et al., 2006). However, the exact statistics test approach was computationally impractical in dealing with large data sets, for the iterative time of each SNP increased sharply with the number of individuals as a random effect. A P3D (population parameters previously determined) method implemented within the TASSEL (Trait Analysis by aSSociation, Evolution and Linkage) program to reduce the running time from standard MLM proposed an approximate algorithm, which replaced the repeated iteration of population parameters by fixed empirical Bayesian priors (Zhang et al., 2010). The statistical power of P3D has been maintained or improved regardless of the different genetic architectures of the phenotype. Similarly, Efficient Mixed-Model Association eXpedited (EMMAX) simplified the estimation of variance parameters once by using a null hypothesis, based on natural characteristics that the most effective of a single locus is small in complex quantitative traits (Kang et al., 2010). In practice, inaccurate *P* values may result from the approximation approach when dealing with highly structured populations and strongly associated marker effects. Contrarily, an exact association test was provided from GEMMA (Genome-wide Efficient Mixed Model Association) on which the Newton–Raphson optimization method was presented. GEMMA remained the identical results compared with the exact calculations in EMMA, but was roughly *n* times faster, where *n* was the sample size (Zhou and Stephens, 2012). As the GEMMA strictly needs completed or imputed genotype data sets (sample integrity > 0.95), some approximate algorithms remain necessary, especially in the absence of large-scale effective SNPs by using an RRGS. A previous research showed that a conjoint analysis of multiple models favored improving the accuracy and reducing the false positive rate.

Fifty-seven QTLs related to seed size and oil content traits in soybean (*Glycine max* L.) were commonly found using three approaches: mrMLM, GEMMA, and EMMAX (Liu et al., 2020). A similar strategy was reported within maize, barley, flax (*Linum usitatissimum* L.), masson pine (*Pinus massoniana*), spruce (*Picea abies* L. Karst), and pig experiments (Manunza et al., 2014; Gyawali et al., 2017; Xie et al., 2018; Bai et al., 2019; Gao et al., 2019; Chen et al., 2021).

In this study, we performed a genome-wide SNP identification based on SLAF-seq of 210 *E. sibiricus* genotypes. Among these, the genetically selected loci and genes were examined by using two extreme groups from low and high altitudes. Further, the phylogenetic relationship, genetic architecture, and a GWAS of twelve grass yield- and seed-related traits were conducted. Our results provided a new genetic resource for molecular breeding in *E. sibiricus*.

MATERIALS AND METHODS

Plant Materials and Field Trials

The 210 *E. sibiricus* genotypes used in this study were obtained from the Lanzhou University and the U.S. Department of Agriculture Germplasm Resources Information Network (GRIN). The latitude and altitude of the collection site ranged from 32.1°N to 58.8°N and from 0 to 4,300 m, respectively (**Figures 1C,D** and **Supplementary Table 1**). According to diverse geographical origin, these genotypes were divided into seven groups, including the East of Qinghai-Tibet Plateau group (Geo_1, 76 samples), Tien Shan Mountains group (Geo_2, 52 samples), Altay Mountains group (Geo_3, 22 samples), North of Mongolian Plateau group (Geo_4, 30 samples), South of Mongolian Plateau group (Geo_5, 11 samples), Canada and Far East of Russia group (Geo_6, 14 samples), and Unknown group (Geo_7, 5 samples).

The seedlings of all genotypes were planted for phenotypic investigation at Yuzhong experimental field of Lanzhou University (35.95°N, 104.15°E, altitude 1,720 m). Each individual was planted with 50-cm spacing in a randomized design. Twelve agronomic traits, including flag leaf length (FLL), flag leaf width (FLW), leaf length (LL), leaf width (LW), plant height (PH), culm diameter (CD), tiller number (TN), culm node number (CN), spike length (SL), length of seed (LS), width of seed (WS), and 1,000-seed weight (SW1) were assessed from 2017 to 2018. A total of 10 replications for each trait (except PH and TN) in an individual plant were recorded using the method described by the previous study (Zhang et al., 2018). The R package "corrplot" (v0.92) was used for a pairwise correlation analysis. The Shannon's diversity index (H') of each trait was calculated as previously described by Zhang et al. (2018).

Single Nucleotide Polymorphism Genotyping

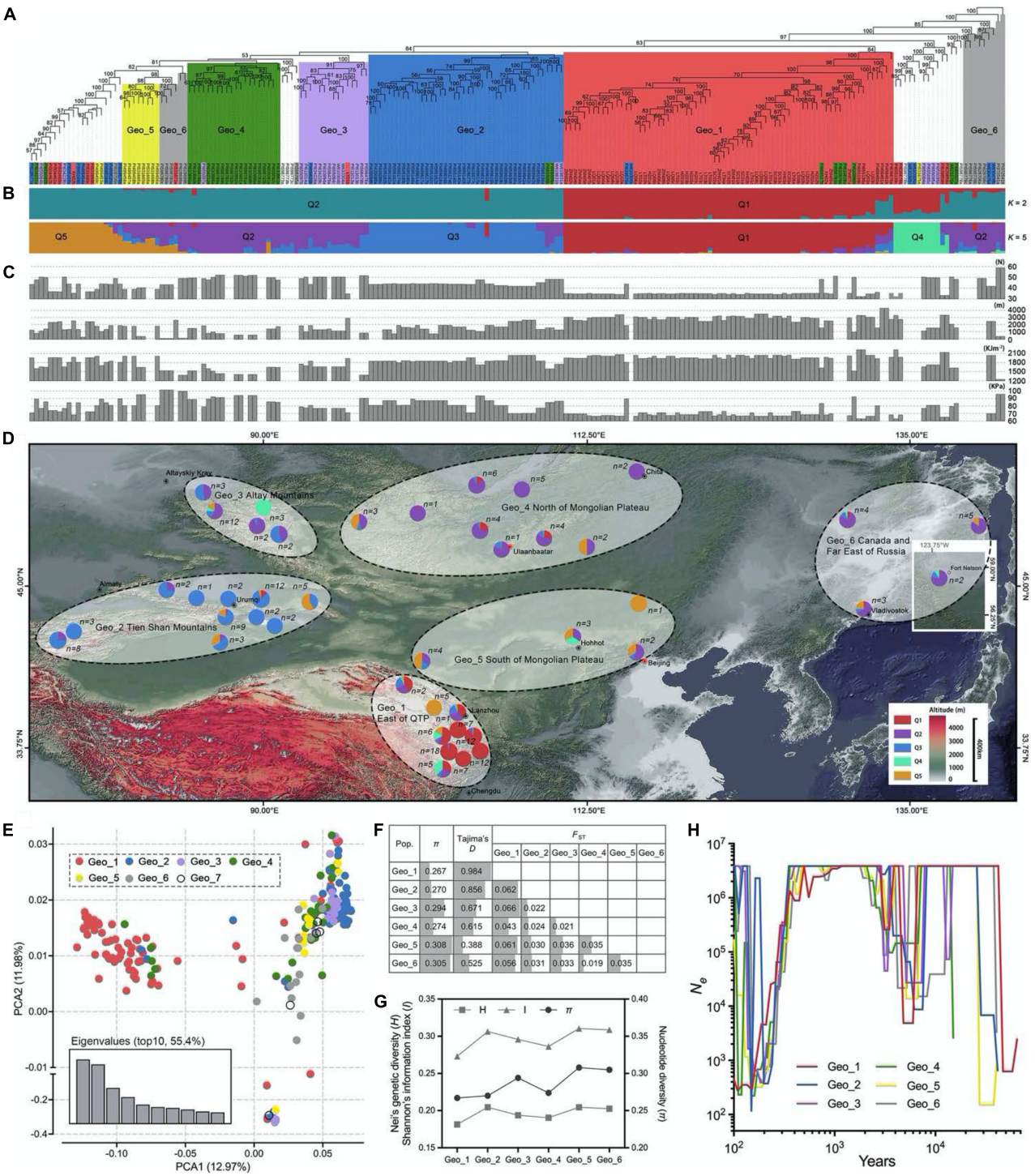
The genomic DNA of each genotype was extracted from young leaf tissues using the Qiagen DNeasy 96-well procedure (QIAGEN, Valencia, Calif). The quantity and quality of samples were measured using a Nano-Drop ND1000 spectrophotometer (NanoDrop, Wilmington, DE, United States). The final concentration of the DNA was adjusted to 50 ng/μl to meet the requirements of SLAF-seq (concentration ≥ 30 ng/μl, DNA quantity ≥ 2 μg).

Elymus sibiricus genome is not available; therefore, the wheat genome was selected as a reference genome[1] (v1.0) for a higher genomic homology (79%) with *E. sibiricus* genetic linkage map (Zhang et al., 2019a). After the digestion prediction analysis of the reference genome, *Hae*III, the appropriate restriction enzyme, was used to digest the genomic DNA and to generate sequencing tags (SLAFs, approximately 464–494 bp in length). This step ensured that the obtained tags were evenly distributed among the chromosomes. The PCR, SLAF library construction, and paired-end sequencing were conducted as previously described (Sun et al., 2013; Zhang et al., 2019a). High-quality raw reads of each sample were mapped onto the reference genome to develop SLAFs by using the Burrows-Wheeler Aligner (BWA) software (v0.7.17; Li and Durbin, 2009). The Genome Analysis Toolkit (GATK, v4.2.0.0) was used for SNP calling (McKenna et al., 2010). The raw DNA sequencing data had been uploaded to NCBI (SRR17714927–SRR17715136). A total of 88,506 highly reliable population SNPs (missing data less than 0.5, minor allele frequency (MAF) greater than 0.05) were selected and used for subsequent analysis. The R package "CMplot" (v4.0.0) was used to show the distribution of SNPs in each chromosome.

Population Structure, Genetic Diversity, and Selective Sweep Analysis

A neighbor-joining phylogenetic tree was constructed using PHYLIP with 500 bootstrap replicates[2] (v3.698) and visualized on the online tool iTOL.[3] The population structure of the tested natural population was assessed by the ADMIXTURE software (v1.3.0; Alexander et al., 2009). The optimum number of subgroups was determined according to cross-validation of the error rate with the presented *K*-value ranging from 1 to 10. A principal component analysis (PCA) was performed using PLINK[4] (v1.90). In addition, the kinship matrixes among all genotypes were calculated by using GEMMA (v0.98.1) and then visualized by the "pheatmap" package (v1.0.12) in R (Zhou and Stephens, 2012). Nucleotide diversity (π), Tajima's *D* value, and genetic differentiation (F_{ST}) for each group were measured with VCFtools (v0.1.16; Danecek et al., 2011). The past evolution of the effective population size (N_e) was estimated by using SMC + + (v1.15.4; Terhorst et al., 2016). A mutation rate of 1.3e-8 per synonymous site per generation was used. Two environmental data sets, surface UV-B radiation and surface pressure, were collected from European Centre for Medium-Range Weather Forecasts (ECMWF[5], ERA5 monthly averaged data on single levels from 1979 to present) from 1979 to 2010 for the genome-wide selective sweep analysis on high-altitude group (greater than 3,000 m) and low-altitude group (less than

[1]https://wheat-urgi.versailles.inra.fr/Seq-Repository/Publications-IWGSC
[2]https://evolution.genetics.washington.edu/phylip.html
[3]https://itol.embl.de/
[4]http://www.cog-genomics.org/plink/1.9/
[5]https://www.ecmwf.int/en/forecasts/datasets/browse-reanalysis-datasets



Pop.	π	Tajima's D	F_{ST}					
			Geo_1	Geo_2	Geo_3	Geo_4	Geo_5	Geo_6
Geo_1	0.267	0.984						
Geo_2	0.270	0.856	0.062					
Geo_3	0.294	0.671	0.066	0.022				
Geo_4	0.274	0.615	0.043	0.024	0.021			
Geo_5	0.308	0.388	0.061	0.030	0.036	0.035		
Geo_6	0.305	0.525	0.056	0.031	0.033	0.019	0.035	

FIGURE 1 | Population structure, geographic distribution, and genetic diversity analysis by using 88,506 high-consistent single nucleotide polymorphisms (SNPs). **(A)** The neighbor-joining (NJ) phylogenetic tree of 210 *Elymus sibiricus* individuals was constructed with 88,506 SNPs. The different colors, respectively, represent the six geographical populations. Only bootstrap values higher than 50% are presented. **(B)** Population genetic structure of *E. sibiricus* germplasms when the *K*-value was two and five. **(C)** Geographic and climatic information for each germplasm, including latitude, altitude, UV-B, and surface pressure. Each bar of **(B,C)** corresponded to every individual in **(A)**. **(D)** Geographical distribution of *E. sibiricus* samples. The pie chart in the map represents the genetic structure proportion of each accession based on the population structure analysis ($K = 5$). **(E)** Scatter plots of the first two principal components. **(F)** Genetic diversity analysis among six populations. **(G)** Genetic diversity relationships of expressed sequence tag-derived simple sequence repeat marker (EST-SSR)-based Nei's genetic diversity *(H)* and Shannon's information index *(I)*, as well as SNP-based nucleotide diversity (π) among six *E. sibiricus* populations. **(H)** Demographic histories of wild *E. sibiricus* groups. Data are shown the estimates of effective population size (N_e) from about 70 Kya for the six groups.

1,000 m). The selected SNPs were annotated to find related genes with response to UV-B and oxidative damage of UV radiation to plants among seven public databases: the NCBI non-redundant protein sequence (Nr, 202009[6]), Gene Ontology (GO, 20200615[7]), Protein family (Pfam, v33.1[8]), Cluster of Orthologous Groups (COG[9]), euKaryotic Orthologous Groups (KOG[10]), Annotated protein sequence database (Swiss-Prot, 202005[11]), and Kyoto Encyclopedia of Genes and Genomes (KEGG, 20191220[12]; **Supplementary Table 4**). The Search Tool for Retrieval of Interacting Genes/Proteins (STRING[13] v11.5) was used to predict and construct the PPI network of the candidate genes based on the minimum required interaction score of medium confidence (Szklarczyk et al., 2019). The intersection genes with a high degree were regarded as candidate hub genes and the PPI network was visualized with the Cytoscape software (v3.8.2).

Genome-Wide Association Study of Agronomic Traits

Three publicly GWAS statistical softwares, Genome-wide Efficient Mixed Model Association (GEMMA, v0.98.1), Efficient Mixed-Model Association eXpedited (EMMAX, beta-07Mar2010 version), and Trait Analysis by aSSociation, Evolution, and Linkage (TASSEL, v5.0), were used together to find out the underlying SNPs or genes associated with 12 grass yield-related traits in *E. sibiricus* (Bradbury et al., 2007; Kang et al., 2010; Zhou and Stephens, 2012). A MLM, with the first three PCA components and the kinship matrix considered a fixed effect and random effect, respectively, was implemented on these programs to carry out GWAS analysis. In addition, the SNP genotyping database was filtered more strictly (marker integrity frequency > 95%, MAF > 0.05) before the GEMMA analysis. Genome-wide significance thresholds were set as $P < 1e\text{-}4$ (EMMAX and TASSEL) and 1e-3 (GEMMA), whereas the Bonferroni-corrected cutoff was 0.1/total applied SNPs (approximately 1.13e-6 for EMMAX and TASSEL and 9.36e-6 for GEMMA; **Supplementary Figures 3–5**). The genes were located within 100-kb flanking regions around the significant trait-associated SNPs based on the public databases. All the genes were selected based on plant development process terms, including root, seed, leaf, flower, cell cycle, cell division, cell proliferation, as well as some related proteins, enzymes, hormones, transcription factors, and so on (**Supplementary Table 6**). The gene expression data from leaf-, culm-, and seed-related traits of two wheat cultivars (Azhurnaya and Chinese Spring) were collected from IWGSC RefSeq Annotations v1.0[14] and used to analyze the gene expression profiles of each identified gene. The candidate genes were identified with an expression value (TPM) of more than 10. In addition, the non-synonymous variations that happened within the gene exon areas were used for the analysis of statistical differences for traits. The PPI network analysis was employed to remove the noise information and focus on the hub genes related to plant growth and development by using the STRING and Cytoscape softwares.

[6]https://ftp.ncbi.nlm.nih.gov/blast/db

[7]http://geneontology.org

[8]http://pfam.xfam.org

[9]https://ftp.ncbi.nih.gov/pub/COG/COG2020/data/

[10]https://ftp.ncbi.nih.gov/pub/COG/KOG/

[11]http://ftp.ebi.ac.uk/pub/databases/swissprot

[12]https://www.kegg.jp/kegg

[13]https://cn.string-db.org/

[14]http://www.wheat-expression.com/

RESULTS

Sequencing and Genome-Wide Polymorphic Single Nucleotide Polymorphism Identification

Using a SLAF high-throughput sequencing approach, a total of 1,176.09 million paired-end reads were generated from the raw sequence data from the panel of 210 *E. sibiricus* samples (**Supplementary Table 2**). The average GC (guanine-cytosine) contents were 45.88% and the Q30 percentage was 93.66%. A bioinformatics analysis aligned to the wheat reference genome revealed that the mapping rate of reads was ranged from 72.83 to 87.92%, with an average of 84.67%. Furthermore, 750,108 SLAF tags were identified with an average mapping depth per sample of 18.01x. A total of 13,643,515 SNP markers were generated. Upon filtering the genotype results with MAFs of 0.05 and sample locus integrity of 0.5, a total of 88,506 highly consistent SNPs were retained. The mean integrity and heterozygosity rate for each sample were 72.39 and 0.2271%, respectively. These obtained SNP markers were not evenly distributed on the genomic regions of 22 chromosomes of wheat (**Table 1** and **Supplementary Figure 1**). The largest number and density of SNPs (6,704.6 and 84.2 kb/SNP) were on chromosome D, followed by chromosome B (3,395.6 and 220.1 kb/SNP) and chromosome A (2,332.0 and 314.2 kb/SNP). The mean number of SNPs was 4,023, ranging from 7,923 (chr7D) to 1,481 (chrUn), and the mean density per marker was 211.5 kb, ranging from 80.4 (chr6D) to 361.1 (chr4A).

Population Structure, Genetic Diversity, and Selective Sweep Analysis

The kinship analysis showed that no direct genetic relationships between each two *E. sibiricus* samples in this study, for the kinship coefficients of more than 97.5% paired individuals, were less than.05 (**Supplementary Figure 2**). Phylogenetic relationships among 210 *E. sibiricus* genotypes were quantified using 88,506 high-quality SNPs (**Figure 1A**). The results showed a monophyletic branch in the neighbor-joining (NJ) tree, indicating a single domestication event happened among all samples. Most of the genotypes of Geo_1, Geo_2, Geo_3, Geo_4, and Geo_5 tended to be grouped according to their geographic origins, and the remains were formed in other admixture clusters. A model-based population structure analysis, reflecting additional inter- and intra-group relationships by the varying *K*-value (the number of ancestry kinship), had assigned the 210 genotypes into two and five subgroups (cross-validation errors of.498 and.521, respectively; **Figure 1B**). Different from others, genotypes that originated from the Qinghai-Tibet Plateau (Geo_1) had a close relationship when *K* was set to two (the

best suitable number from the *K*-value test). When a higher *K*-value was assumed, more subgroups were further detected, including some groups from Tien Shan Mountains (Geo_2), Altay Mountains (Geo_3), and Mongolian Plateau (Geo_4 and Geo_5). Similar to the NJ tree, five structured populations ($K = 5$) were consistent with their geographical distribution characteristics (**Figure 1D**). The PCA mainly split the samples into two distinct groups (QTP and others) at an overall genomic level, in which the first two PCAs had explained 24.95% of the total variation (**Figure 1E**). A continuous distribution showed on the plot indicating no population stratification within the studied genotypes.

The nucleotide diversity (π value) was estimated for six major geographic groups, which was higher for Geo_3, Geo_5, and Geo_6 (average of.302) than Geo_1, Geo_2, and Geo_4 (average of.270; **Figure 1F**). These results were similar to our previous diversity study based on EST-SSR markers (**Figure 1G**; Zhang et al., 2019b). Positive mean Tajima's *D* values existed within all of *E. sibiricus* groups which not only implied a lower probability of population expansion under the balancing selection but also less rare allele frequencies resided in specific habitats, especially in Geo_1 (0.984) and Geo_2 (0.856). The farthest genetic distance (average of.058) was inferred between 76 QTP samples and 129 genotypes from the other five groups using a genetic differentiation analysis (F_{ST}). An evolutionary history analysis among different geo-populations showed similar demographic trajectories: all groups suffered bottleneck events that the effective population size (N_e) decreased first and then increased to a stable size during 2–13 thousand years ago (Kya), and a short common history that the N_e size of 4E6 appeared in.4-2 Kya (**Figure 1H**). Thereafter, in the next bottleneck, effective population size significantly declined and then rebounded except for Geo_1 group.

Genome regions with low nucleotide diversity were considered to be potential selection signatures from human domestication or different natural environment factors. A survey of meteorological information among these studied genotypes within a span of 32 years (from 1979 to 2010) showed significant correlations among UV-B, surface pressure, and altitude (**Figure 2A**). Sixty-six wild genotypes originated from two extreme habitats were used for a selective sweep analysis, of which 38 genotypes originated from high-altitude regions (>3,000 m) and 28 genotypes originated from low-altitude regions (< 1,000 m; **Figure 2B**). Two climate factors were significantly different between the two groups ($p < 0.001$, Welch's test). The mean monthly cumulative UV-B value was 1535.9 and 2036.0 $kJ{\cdot}m^{-2}$ in low and high altitudes, respectively. Correspondingly, the mean surface pressure was 92.9 and 67.0 kPa, respectively. For the selection signals, a total of 965 SNPs were identified from the top 20% θ_π ratio (θ_π, low altitude/θ_π , high altitude) and the top 5% F_{ST} value for high-altitude samples (**Figure 2C** and **Supplementary Table 3**).

TABLE 1 | Distribution and summary of *E. sibiricus* genome-wide variants.

Chr[a]	SLAF number	Polymorphic SLAF	Number of SNPs	Number of SNPs after filtering	Max gap (kb)	Density of SNPs (kb/SNP)
chr1A	23,793	10,235	415,929	1,784	7,945	333.0
chr1B	30,204	13,794	551,062	3,020	12,875	228.4
chr1D	39,202	19,998	745,444	5,815	3,060	85.2
chr2A	31,014	13,938	556,863	3,641	8,880	214.4
chr2B	36,441	16,838	678,080	3,720	8,337	215.4
chr2D	50,737	25,978	968,703	7,480	5,039	87.1
chr3A	28,789	12,631	509,555	2,286	13,917	328.5
chr3B	38,882	17,580	664,931	3,627	11,223	229.1
chr3D	47,808	24,410	913,125	6,907	6,713	89.1
chr4A	28,344	12,405	500,044	2,062	12,589	361.1
chr4B	29,720	14,423	501,873	3,803	8,279	177.1
chr4D	37,143	19,032	689,559	6,020	4,898	84.7
chr5A	27,713	11,993	491,348	2,018	13,412	351.7
chr5B	33,734	15,698	618,957	3,630	6,587	196.5
chr5D	44,405	22,673	858,003	6,887	6,820	82.2
chr6A	23,461	10,219	413,822	1,827	13,586	338.3
chr6B	30,927	14,403	561,932	2,804	11,922	257.1
chr6D	35,973	18,356	685,631	5,891	6,193	80.4
chr7A	30,002	13,253	547,319	2,706	10,635	272.2
chr7B	33,044	15,114	598,225	3,165	8,661	237.2
chr7D	50,656	25,855	973,007	7,932	4,456	80.5
chrUn	18,116	5,127	200,103	1,481	44,055	324.8
Mean	34,096	16,089	620,160	4,023	10,458.3	211.5
Total	750,108	353,953	13,643,515	88,506	–	–

[a]*Chromosome.*

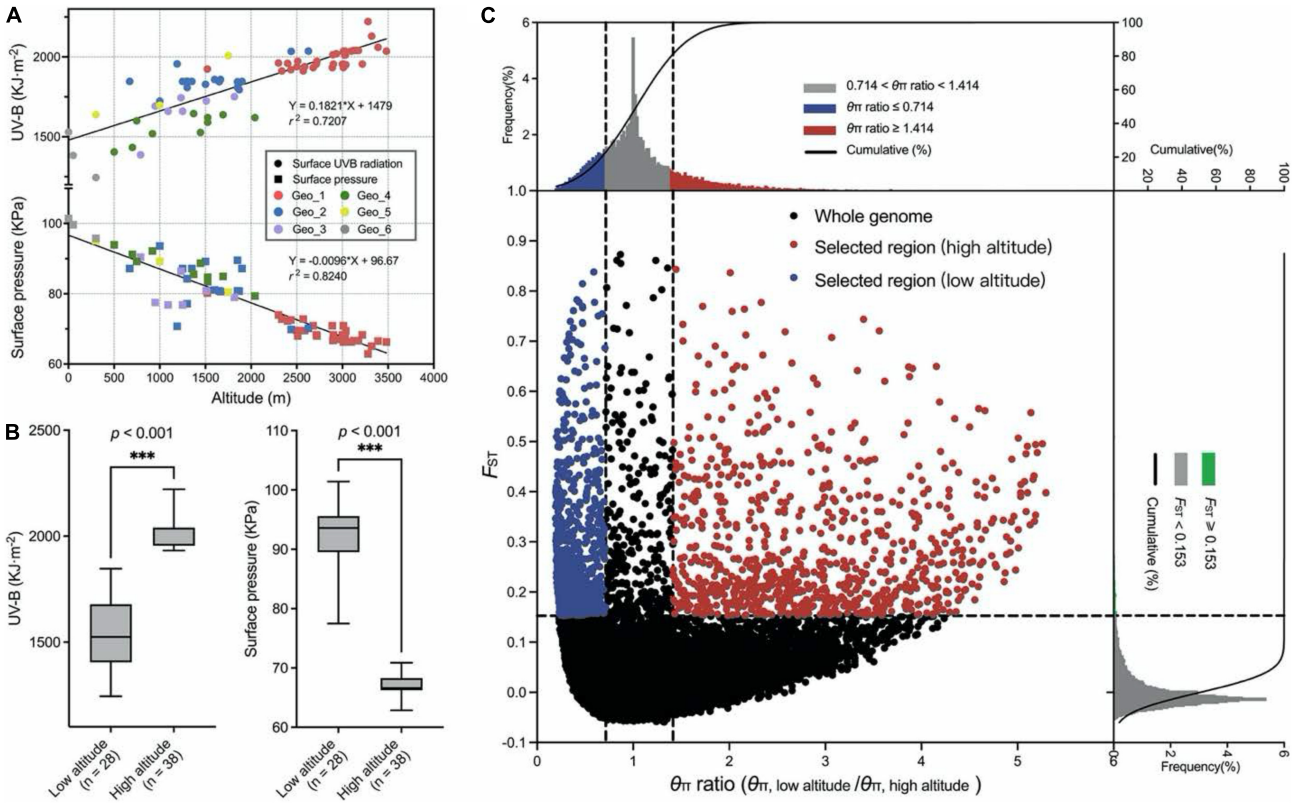


FIGURE 2 | Selective sweep analysis. **(A)** Regression analysis between the UV-B, surface pressure, and altitude for 210 *E. sibiricus* genotypes. **(B)** Two boxplots for UV-B and surface pressure among low-altitude and high-altitude genotypes. Whiskers extend to data ranging from min to max values. Differences between the two pools were statistically analyzed based on the Welch's test. **(C)** Distribution of θ_π ratio (θ_π, low altitude/θ_π, high altitude) and F_{ST} value. The horizontal dashed line represents the top 5% F_{ST} distribution. The left and right dashed lines represent 20% left and right tails of the θ_π ratio distribution. The red points were identified as selected regions for high-altitude genotypes.

Considering the characteristics of SLAF-based molecular markers in this study, the range where the distance between two adjacent markers is less than 100 kb and the distance span is greater than 100 bp was deemed as the selected area. We found 54 selected regions, of which the largest one was on chromosome 2A with eight SNPs and 255.7 kb of physical distance from 459,245,966 to 459,501,653 bp. The mean length was 16.1 kb (**Supplementary Table 3**). Seventy-three genes in response to harsh climate (such as low oxygen and strong ultraviolet radiation) in high-altitude areas were identified, including UV-B receptor UVR8 (3 genes), response to UV-B (5 genes), regulation of photosynthesis (9 genes), photosystem II (PSII) assembly (9 genes), chlorophyll (9 genes), carotenoid (2 genes), response to superoxide (1 gene), reactive oxygen metabolic process (1 gene), hydrogen peroxide (12 genes), peroxidase (15 genes), proline (7 genes), flavonoid biosynthetic process (7 genes), anthocyanin biosynthetic process (9 genes), L-ascorbic acid biosynthetic process (3 genes), and so on (**Supplementary Table 4**). Three haplotype regions with 17 (chr7B, from 267,243,758 to 267,243,793 bp), 8 (chr6A, from 609,091,993 to 609,092,385 bp), and 8 (chr1D, from 81,295,211 to 81,295,288 bp) close SNPs were annotated to gene functions of PSII assembly, anthocyanidin and chlorophyll biosynthetic process, and peroxidase activity, respectively. Five peroxidase activity-related genes were detected within 100-kb flanking regions on a chromosome 2B SNP (793,439,312 bp). Similar genes were found on chr2D (643,719,897 bp).

To find out potential hub candidate genes responding to the harsh climate, a PPI network consisting of 57 nodes and 196 lines was obtained by using the online STRING database (**Figure 3** and **Supplementary Table 4**). The analysis showed a tight PPI network, suggesting that the queried proteins have strong relationships and interactions. There were seven hub genes with the node degree greater than seven, including one with 17 degrees (*TraesCS1A01G442300*), two with 8 degrees (*TraesCS2D01G379400* and *TraesCSU01G146500*), and four with 7 degrees (*TraesCS2D01G462700*, *TraesCS5A01G554500*, *TraesCS5D01G065700*, and *TraesCS7B01G136000*). These genes are mainly involved in the response to UV light, flavonoid biosynthetic process, anthocyanin accumulation, reactive oxygen species (ROS) metabolic process, and cytochrome P450. Furthermore, the potential hub genes were important homologous genes (such as *CYP75B3*, *FabZ*, and *GPA1*) in Poaceae plants, including *Aegilops tauschii*, *Hordeum vulgare*, *Lolium rigidum*, and *Oryza sativa*, and may be important components in the genetic basis of the adaptation to high altitudes for *E. sibiricus*.

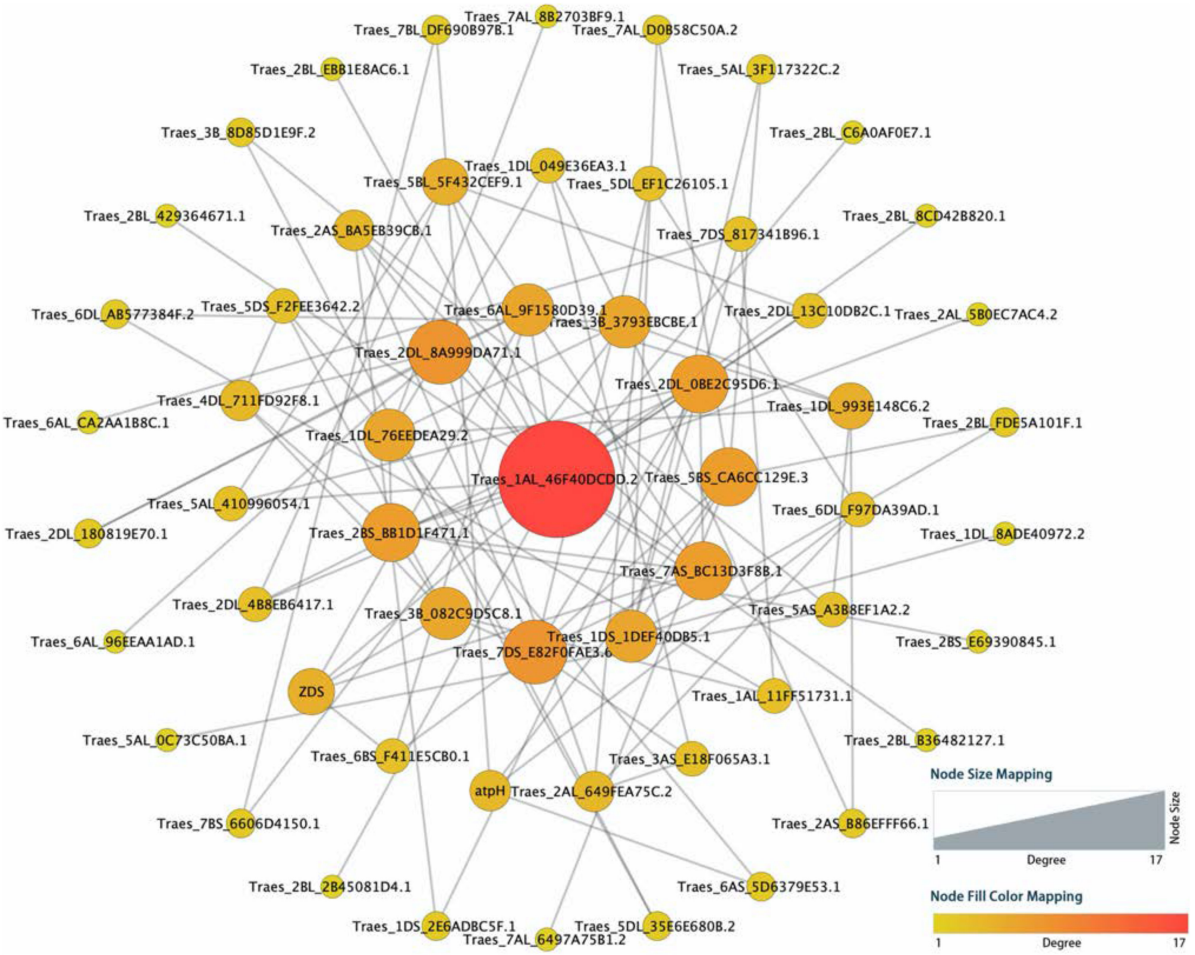


FIGURE 3 | Protein–protein interaction (PPI) regulatory network for the selected candidate genes among high-altitude genotypes. The circles represented queried proteins, and the lines represented protein–protein associations.

Phenotypic Variation and Correlation Analysis

Four leaf-related traits: FLL, FLW, LL, and LW, four culm-related traits: PH, CD, TN, and CN, and four seed-related traits: SL, LS, WS, and SW1 were investigated within 2 years. A wide range of phenotypic variations were found among 210 *E. sibiricus* genotypes. The mean coefficient of variation (CV) between 12 phenotypic traits was 28.2%, varying from 6.8% (LS in 2018) to 84.1% (TN in 2018), and the average Shannon's diversity index (H') was 2.008 varying from 1.798 (TN in 2018) to 2.095 (WS in 2018; **Table 2**).

The phenotypic characters followed a normal or approximate normal distribution without any significant skewness and could be used for the subsequent analysis (**Figures 4A,B**). The Pearson's correlation coefficient analysis showed significantly positive pairwise correlations between four leaf-related traits (FLL, FLW, LL, and LW; 0.60 to.85 in 2017 and.36 to.68 in 2018), as well as three seed-related traits (LS, WS, and SW1; 0.35 to.54 in 2017 and.35 to.39 in 2018; **Figures 4C,D**). These traits may be controlled by similar genetic factors. Two culm-related traits (TN and CD) had a highly significant positive correlation with leaf-related traits, of which the mean phenotypic correlation coefficients of.32,0.60 in 2017 and.44,0.49 in 2018, respectively. Generally, there were weakly positive or negative relations between two seed traits (LS and WS) and two grass yield-related traits (leaf and culm). Eight traits (FLL, FLW, LL, LW, PH, CD, TN, and SW1) had a higher correlation between two consecutive years (**Figure 4E**).

Genome-Wide Association Analysis Reveals Loci Related to Agronomic Traits

Using a MLM with correction of genetic structure (Q) and kinship (K) bias, the multiple GWAS tests (GEMMA, EMMAX, and TASSEL) were performed to confirm significant SNPs associated with twelve grass yield- and seed-related traits in *E. sibiricus* (**Supplementary Figures 3–5**). The quantile-quantile (QQ) plots indicated well-controlled population stratification by using MLM and provided reliability for association analysis. In this study, 1,825 SNP markers were detected above the critical threshold of $-\lg P \geq 4$ (EMMAX and TASSEL) and 3 (GEMMA),

TABLE 2 | Phenotypic variation statistics of genome-wide association study (GWAS) population.

Traits		Year	Min	Max	Mean	SD	CV (%)	H'[a]
Leaf-related	FLL(cm)	2017	5.5	27.3	13.9	4.6	32.9	1.921
		2018	5.0	19.6	10.5	2.6	25.3	1.943
	FLW(mm)	2017	3.05	16.70	6.18	2.0	32.5	1.870
		2018	3.42	11.94	7.25	1.7	23.2	2.055
	LL(cm)	2017	5.5	27.3	15.2	4.2	28.0	2.030
		2018	7.8	23.6	15.3	2.9	18.9	2.059
	LW(mm)	2017	2.77	11.99	6.17	1.8	29.2	1.971
		2018	3.48	9.32	6.07	1.4	22.9	2.073
Culm-related	PH(cm)	2017	12.2	73.2	31.4	12.3	39.1	1.917
		2018	17.3	101.5	55.0	19.5	35.5	2.035
	CD(mm)	2017	0.74	2.58	1.48	0.3	21.5	2.065
		2018	1.10	3.33	2.08	0.4	18.5	2.053
	TN(No.)	2017	3	152	42	26.4	62.2	1.936
		2018	2	310	66	55.6	84.1	1.798
	CN(No.)	2017	1	6	2	0.7	28.7	2.032
		2018	1	6	3	0.9	27.4	2.029
Seed-related	SL(cm)	2017	5.4	23.7	13.4	4.1	31.0	2.026
		2018	6.4	29.1	15.9	3.5	22.2	2.041
	LS(mm)	2017	4.45	11.75	8.87	1.0	11.7	2.006
		2018	7.75	11.02	9.51	0.6	6.8	2.030
	WS(mm)	2017	1.13	1.69	1.42	0.1	8.2	2.074
		2018	1.25	1.81	1.49	0.1	7.3	2.095
	SW1(g)	2017	0.713	5.360	3.208	0.9	28.1	2.047
		2018	0.653	3.986	2.222	0.7	31.9	2.090
Mean							28.2	2.008

[a] *H', Shannon's diversity index.*

respectively (**Table 3**, **Supplementary Table 5**, and **Figure 5A**). More than half of the associated loci (917) were distributed on D subgenomes, while 477, 398, and 33 SNPs were on B, A, and Un, respectively. Three softwares identified a total of 848 SNPs for culm-related traits, followed by leaf-related traits (604 SNPs) and seed-related traits (373 SNPs). Especially, 358, 247, and 235 SNPs were related to PH, FLW, and TN, respectively. The TASSEL test showed that the percentage of phenotypic variance explained (R^2) ranged from 7.62 to 31.85%, with an average of 15.57%.

A gene function annotation analysis was performed to rapidly identify the candidate genes associated with yield-related traits in *E. sibiricus*. There were 1,064 genes identified from 908 significant SNPs within 100-kb flanking regions, of which 344 loci with multigenic effects (**Supplementary Table 6**). In the TN traits, a SNP (6,982,833) on chromosome 1D associated with 8 genes related to the process of recognizing pollen, response to growth hormone, tyrosine kinase, ankyrin repeats, and leucine-rich repeat (LRR). Eight closely linked genes identified from a SNP marker on chr1D (37,459,155), annotated with plasmodesma, cytoskeleton, F-box protein, PPR repeat family, and hydrolases family, may result in multigenic effects for plant height traits. A total of 116 highly consistent SNPs that overlapped from more than two softwares were deemed to be highly reliable SNPs. After a comparison of associated genes with gene expression data of two wheat cultivars (Azhurnaya and Chinese Spring), we finally identified 106 candidate genes (**Figure 5A**). Sixty-two genes were related to PH, CD, TN, and CN traits. Additionally, 20 and 28 genes were related to leaf- and seed-related traits, respectively. Furthermore, functional classifications of these genes were mainly involved in four enriched biological processes: plant development process (70 genes), protein and enzymes (30 genes), hormones (26 genes), and transcription factors (2 genes; **Figure 5B**). The functional annotation showed that the biological process of photosynthesis, root, cytoskeleton, microtubule, photorespiration, flower, pollen, and seed were the most highly represented categories in the plant development process. For the biological regulation of hormone signals, many terms were enriched with auxins, abscisic acid (ABA), jasmonic acid, ethylene, and salicylic acid. Tyrosine kinase, E3 ubiquitin ligase, hydrolase, and LRRs were the major proteins and enzymes. For leaf-related traits, the top four gene functions were photosynthesis, photorespiration, jasmonic acid, and salicylic acid, whereas E3 ubiquitin ligase, meristem, flower, and photosynthesis were enriched for seed-related traits. The candidate genes for culm-related traits covered almost all molecular functional categories, and microtubule, auxins, tyrosine kinase, and E3 ubiquitin ligase were major enrichment terms.

For the PH traits, 40 SNPs with 38 candidate genes were identified (**Figure 5C** and **Supplementary Table 6**). The GEMMA test found a SNP marker (chromosome 1A, 483,033,948) on exon region of *TraesCS1A01G286300*, a plasmodesma-related gene, and the non-synonymous SNP diversity caused amino acid differences between Methionine (Met) and Isoleucine (Ile; **Figure 5D**). The haplotype-based analysis showed that the Hap. B (SNPs of A) with 100 genotypes was significantly positively correlated with plant height (PH; $p < 0.0001$, Welch's test) which increased by 32.7%

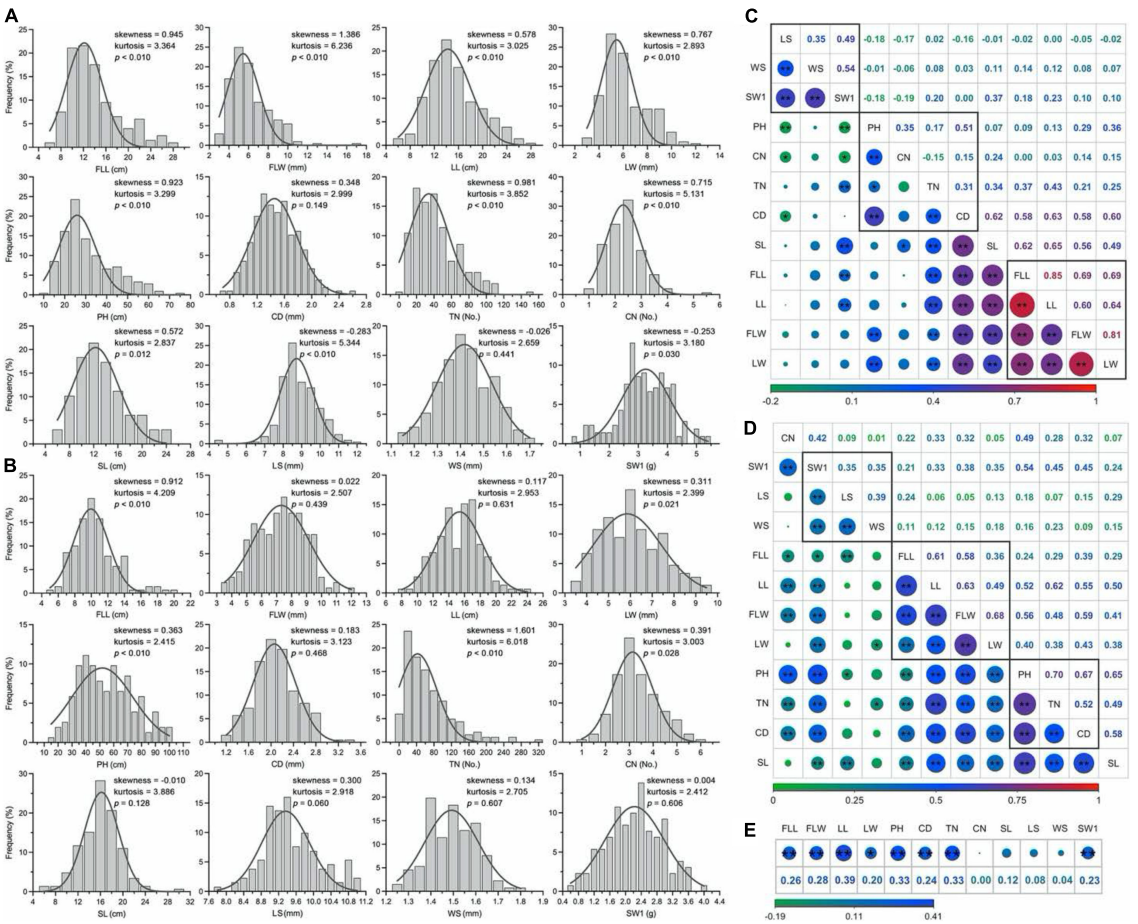


FIGURE 4 | Distribution of phenotypic traits and correlation analysis in *E. sibiricus*. **(A,B)** Frequency distribution of the mean values of twelve phenotypic traits in 2 years. **(C,D)** Pair-wise Person correlations among different traits. The red to green color represents highly positive to highly negative correlations, and the number represents (upper diagonal) the correlation values. The lines in the square indicate the leaf-, culm-, and seed-related traits. Two (**) and single stars (*) indicate significance at 0.01 and 0.05, respectively. **(E)** Correlation analysis of twelve phenotypic traits between 2 years.

when compared with 53 genotypes with Hap. A (SNPs of G; **Figure 5D**). In addition, the gene was highly expressed in the internode, stem axis, and stem of two wheat cultivars, respectively. Furthermore, these genes showed four main functional annotations: (1) ethylene, auxins, and jasmonic acid signaling pathway; (2) tyrosine kinase activity; (3) cytoskeleton organization, microtubule, xylem formation, and plasmodesma; (4) cell cycle and cell division (**Figure 5E**). A total of 12 genes were highly overexpressed across all culm-related traits. Particularly, two SNPs on chromosome 4A (80,134,459) and 6D (378,425,315) were jointly annotated by GEMMA, EMMAX, and TASSEL. A gene *TraesCS4B01G205100* had a higher expression level in internode and stem in two wheat cultivars, which was located within the 100-kb flanking regions of significant loci with the highest –lg(*P*) value (8.929) in the GWAS test.

The PPI network analysis further narrowed down the list of 106 identified yield-related candidate genes in *E. sibiricus*. Based on the degree algorithm, there were three hub genes, *TraesCS4D01G221200* (TPL), *TraesCS4D01G221100* (Traes_4DL_A43826645.1), and *TraesCS7D01G194800* (LIM), with the degree value more than 4 detected within PH traits (**Figure 6A** and **Supplementary Table 7**). The top gene, *TraesCS4D01G221200*, was found to be involved in the biological regulation process of a variety of plant hormones and morphological development, including auxins- and jasmonic acid-mediated signaling pathway; meristem and shoot apical meristem (SAM) specification; xylem and phloem pattern formation. Besides, two (*TraesCS5A01G457500* and *TraesCS1B01G254000*), six (*TraesCS4D01G221200*, *TraesCS7D01G194800*, *TraesCS3D01G190400*, *TraesCS5D01G343000*, *TraesCS1A01G286200*, and *TraesCS5D01G118000*), and four (*TraesCS1A01G402400*, *TraesCS3D01G296900*, *TraesCS6A01G307700*, and *TraesCS7D01G056800*) hub genes were detected in leaf-, culm-, and seed-related traits, respectively (**Figures 6B–D**). These genes mainly played the biological function of photosynthesis (leaf-related traits); auxins, jasmonic acid, and ethylene-mediated signaling pathway, meristem, shoot, and xylem apical meristem specification, cell proliferation, tyrosine kinase and E3 ubiquitin-protein ligase activity (culm-related traits); and tyrosine kinase, flower

TABLE 3 | Genome-wide association study analysis for twelve phenotypic traits using a mixed linear model (MLM) based on three algorithms.

Traits	Year	GEMMA					EMMAX					TASSEL						NJ[g]
		NS[a]	MP[b]	LP[c]	NAS[d]	CG[e]	NS	MP	LP	NAS	CG	NS	MP	LP	R^{2f}	NAS	CG	
FLL(cm)	2017	18	9.35E–06	5.03	6	0	28	8.19E–07	6.09	12	1	38	6.29E-08	7.20	8.21–26.37%	16	1	3
	2018	30	6.05E–05	4.22	12	1	34	3.02E-06	5.52	12	0	13	1.10E-05	4.96	12.87–30.60%	7	1	5
FLW(mm)	2017	17	5.00E-07	6.30	8	0	92	7.05E–08	7.15	50	7	96	9.72E-09	8.01	8.58–30.94%	56	7	22
	2018	16	2.17E–05	4.66	9	0	21	1.51E–06	5.82	13	0	5	1.30E–05	4.89	13.63–26.13%	4	0	6
LL(cm)	2017	20	2.78E–06	5.56	9	1	22	2.79E–06	5.55	7	0	26	1.91E–06	5.72	9.44–30.21%	6	2	1
	2018	8	1.52E–04	3.82	4	0	12	4.95E–07	6.31	6	0	0	–	–	–	0	0	0
LW(mm)	2017	15	5.06E–06	5.30	5	0	25	1.21E–06	5.92	10	1	47	2.45E–07	6.61	9.22–23.85%	25	2	3
	2018	11	9.26E–05	4.03	3	0	10	3.57E–06	5.45	1	0	0	–	–	–	0	0	0
PH(cm)	2017	25	1.61E–05	4.79	12	1	33	1.80E–07	6.74	14	3	246	1.18E–09	8.93	7.62–30.9%	132	26	23
	2018	24	2.48E–06	5.60	13	5	18	3.15E–08	7.50	14	5	12	1.55E–05	4.81	10.74–27.32%	10	2	6
CD(mm)	2017	16	3.08E–05	4.51	5	3	13	1.84E–05	4.73	5	1	11	1.01E–05	5.00	7.82–14.90%	6	4	2
	2018	12	3.49E–07	6.46	2	1	20	4.69E–07	6.33	7	2	10	3.04E–06	5.52	11.13–23.49%	8	2	4
TN(No.)	2017	17	2.17E–06	5.66	3	3	33	8.78E–08	7.06	15	3	6	1.40E–06	5.86	13.33–22.86%	3	1	2
	2018	17	2.24E–06	5.65	7	1	61	3.68E–07	6.43	25	0	101	5.22E–07	6.28	9.33–29.14%	45	5	12
CN(No.)	2017	14	6.36E–06	5.20	10	1	38	1.74E–07	6.76	21	3	13	3.99E–06	5.40	15.30–27.92%	7	0	4
	2018	22	4.92E–07	6.31	11	1	53	1.21E–07	6.92	27	4	33	7.55E–07	6.12	12.02–31.85%	19	0	2
SL(cm)	2017	12	3.99E–05	4.40	7	1	19	8.81E–07	6.05	16	2	0	–	–	–	0	0	3
	2018	23	2.36E–05	4.63	12	0	21	8.72E–07	6.06	11	2	4	2.11E–05	4.68	12.11–27.29%	2	0	1
LS(mm)	2017	23	3.09E–05	4.51	17	1	45	3.00E–06	5.52	27	2	45	8.41E-07	6.08	9.94–24.16%	29	9	12
	2018	50	2.54E–05	4.59	25	5	11	1.80E–05	4.74	9	0	0	–	–	–	0	0	1
WS(mm)	2017	24	3.97E–05	4.40	9	2	8	4.46E–06	5.35	6	1	7	2.88E-05	4.54	10.26–17.30%	4	0	3
	2018	13	3.47E–05	4.46	8	1	5	4.06E–06	5.39	1	0	2	1.28E–05	4.89	22.10–30.58%	1	2	0
SW1(g)	2017	21	5.49E–07	6.26	10	0	24	2.38E–08	7.62	5	1	1	2.18E–05	4.66	12.99–12.99%	1	0	1
	2018	8	8.54E–05	4.07	4	0	7	1.88E–05	4.73	2	3	0	–	–	–	0	0	0
Sum		456			211	28	653			316	41	716				381	64	116

[a] *Number of associated SNPs.*
[b] *The min value of P.*
[c] *The peak value of -lgP.*
[d] *Number of annotated SNPs.*
[e] *Number of candidate genes.*
[f] *Percentage of variance explained.*
[g] *Number of SNPs jointly annotated by more than two software.*

development, photosynthesis, and cytoskeleton organization (seed-related traits).

DISCUSSION

Selective Signature Detection in *Elymus sibiricus* at High Altitudes

In this study, we performed SLAF-seq of 210 *E. sibiricus* genotypes including six geographical populations from the main distribution areas of the species. A total of 88,506 high-consistent SNPs were picked up for genetic research. Phylogenetic relationship and population architecture analysis revealed that the Geo_1 group collected from Qinghai-Tibet Plateau (QTP) regions had different genetic backgrounds from the other groups (**Figures 1A,B,D,E**). A similar result was supported by some genetic diversity studies in *E. sibircus,* including chromosome karyotype analysis, gliadin, and SSR marker study (Ma et al., 2012b; De et al., 2014; Yan et al., 2017). It may indicate that the main climatic factors act as a selective pressure on different geographical populations. The QTP population experienced severe natural pressure to adapt to harsh environmental factors in high-altitude regions, such as low pressure, strong radiation, sharp temperature shifts, and overwintering challenges. To understand the genetic bases of adaptation to high plateaus from a genome perspective, two pools of *E. sibiricus* genotypes collected from high-altitude regions (>3,000 m) and low-altitude regions (<1,000 m) were used to detect the associated genes for altitude adaptation. More than 70 genes, mainly responding to UV-B and superoxide, UV-B receptor UVR8, regulating photosynthesis, PSII, flavonoid biosynthetic process, and peroxidase activity, were identified among 965 strong selected loci. Among them, seven genes were at the hub position in the PPI network analysis, which play pivotal roles in response to UV light, flavonoid biosynthetic process, anthocyanin accumulation, ROS metabolic process, and cytochrome P450.

The selective filter of sunlight in leaf epidermal tissues represented an acclimatizing mechanism by which plants are exposed to strong UV radiation conditions (including UV-B, 280–315 nm) for a long time (Barnes et al., 2016). In response to UV radiations, the common stress-related processes include antioxidant defense and accumulation of UV-absorbing compounds in the vacuoles and cell walls of leaf epidermal cells, such as flavonoids, anthocyanin, and other phenolics, is activated (Siipola et al., 2015; Wu Q. et al., 2016; Berland et al., 2019). Besides, the photomorphogenic responses, gene expression, and accumulation of metabolites were triggered by UV RESISTANCE

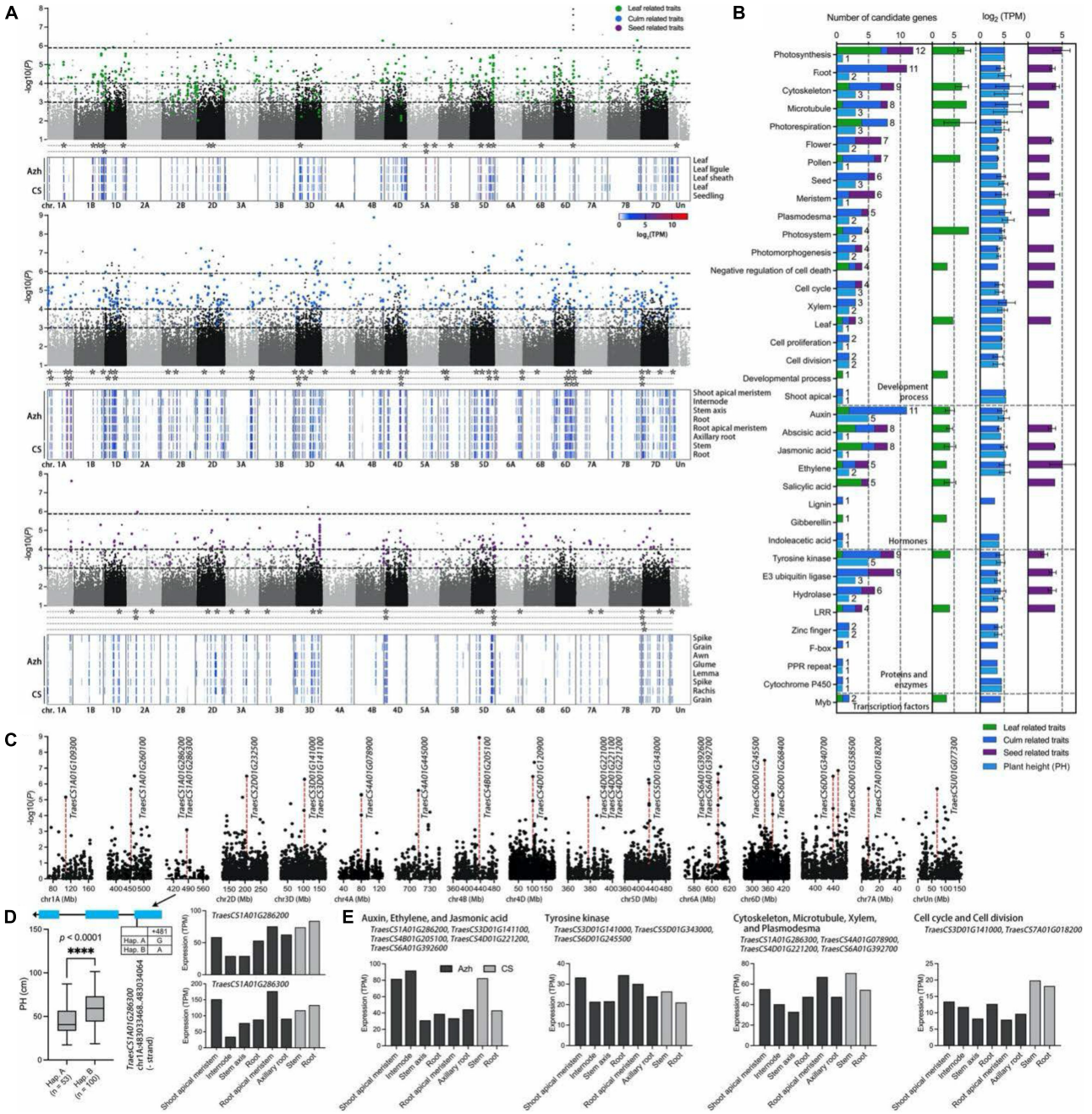


FIGURE 5 | Genome-wide association study for phenotypic traits and identification of the candidate genes for the significant SNPs. **(A)** Manhattan plot for leaf (green points)-, culm (blue points)-, and seed (purple points)-related traits in *E. sibiricus*, respectively. Only SNPs with *P* values smaller than.1 are plotted. Three horizontal dashed lines represent *P* values at of 1.13e-6, 1e-4, and 1e-3, respectively. The gene expression profiles of Azhurnaya and Chinese Spring in leaf-, culm-, and seed-related traits were shown on the heatmap. The stars (*) represent the candidate genes with an expression (TPM) value of more than 10. **(B)** Frequency of category annotations and average TPM value for candidate genes. The green, blue, purple, and wathet blue bars represent leaf, culm, seed, and plant height (PH) traits, respectively. **(C)** Local Manhattan plot for PH. The red dashed lines indicate the significant loci with the highest –lg(P) value. **(D)** Exon–intron structure and DNA polymorphism of *TraesCS1A01G286300*. The difference between the two haplotypes was shown on the boxplots and statistically analyzed based on Welch's test for plant height (PH, 2018). The expression profiles of two candidate genes were shown on the bar plots. **(E)** The expression profiles of twelve candidate genes were identified from PH trait.

LOCUS 8 (UVR8), a specific UV-B photoreceptor in plants (Rizzini et al., 2011). In *Arabidopsis*, transgenic plants with UVR8 showed elevated levels of anthocyanin and enhanced the acclimation-independent UV-B tolerance (Heijde et al., 2013; Podolec et al., 2021). UV-B induced damage to photosynthetic machinery and limit of PSII activity inhibits growth and biomass

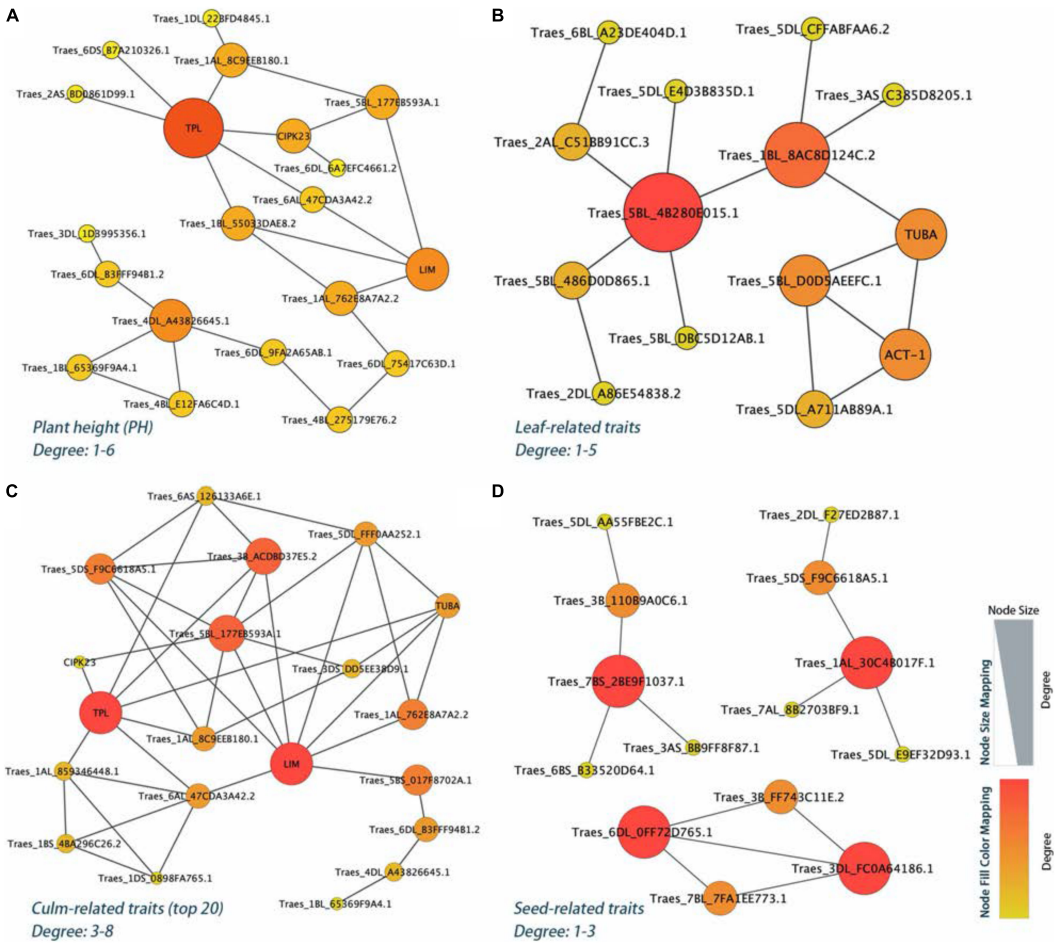


FIGURE 6 | The PPI regulatory network for the yield-related candidate genes in *E. sibiricus*. **(A)** Plant height (PH) traits. **(B–D)** Leaf-, culm-, and seed-related traits, respectively.

accumulation, especially in UV-B sensitive plants. It was reported that ambient UV radiations reduced leaf area by 23% in *Lactuca sativa* L. in mid-latitude locations (Wargent et al., 2011). *E. sibiricus* can thrive in high-alpine areas and shows good tolerance. However, the information related to the effects of ultraviolet radiations on *E. sibiricus* growth among different habitats is limited; thus, future exploration of UV light adaptation mechanisms may provide profound benefits to increase the forage grass production. Furthermore, exposure to UV-B among alpine plants can result in cross-protection against other stress factors at the same time, such as low temperature and hypoxic conditions on Qinghai-Tibet Plateau; thus, a combined stress environment should be considered (Wargent et al., 2011; Schulz et al., 2021). The comprehensive ecological environment on Qinghai-Tibet Plateau drove the adaptation strategies in *Leymus secalinus* and activated antioxidant defense system with multiple protective strategies, including the reduction of chlorophyll contents and increase in carotenoid content to protect plants against photodamage; the changes in antioxidant enzyme activities enhanced the ROS scavenging capacity; the increases in proline and soluble sugar contents adjust osmotic in plants of alpine regions (Cui et al., 2018). In this study, seven hub genes were all related to the antioxidant defense system and may play a key role in the adaptation strategies of *E. sibiricus*. As the significant ambient difference existed between the two groups, the genetic signatures and functional genes we identified may be preferentially amplified to respond to specific environment.

Genome-Wide Association Study in *Elymus sibiricus*

Recently, advances in next-generation sequencing have rapidly increased the density of high-quality molecular markers in many non-model plants (Fang et al., 2017a; Pang et al., 2020; Shen et al., 2020; Wei et al., 2021). As the gigantic genome existed in *E. sibiricus*, with an estimate of 6.86 Gbp, we used a mature reduced whole-genome sequencing technology (SLAF-seq) to investigate the SNP information among 210 genotypes (Sun et al., 2013; Xiong et al., 2021). At present, there is no report on the assembled genome; thus, generated SNPs were mapped on the closest wheat reference genome (v1.0), with 79% of homogeneity

and 84.67% of mapping rate of reads (Zhang et al., 2019a). Our study generated a total of 88,506 highly consistent SNPs, with an average sequencing depth of 18.01-fold and a marker density of 211.5 kb per SNP. Generally, the LD decay in cross-pollinated plants is much faster than in self-pollinated. A precise LD decay distance in *E. sibiricus* was unavailable in the study, but some self-pollinated species provided certain reference information. The physical distance of LD decay in rice, cottons, wheat, and soybean occurred at 445 kb (r^2 = 0.2; Yano et al., 2016), 1,000 kb (r^2 = 0.3; Fang et al., 2017b), 4.4 Mb (r^2 = 0.5; Pang et al., 2020), and 223.2 kb (r^2 = 0.4; Sui et al., 2020), respectively. Similarly, the mean distance between the markers was very less than the LD decay distances in our study, and SNP set provided sufficient resolution for capturing the genetic variation in general GWAS.

The identification of genes associated with important agronomic traits such as biomass and quality in forage grass is indispensable for breeding improvement (Yano et al., 2016). Previously, we identified 29 QTLs of seed-related traits among 14 linkage groups in *E. sibiricus* by using a biparental allelic separation population (F_2; Zhang et al., 2019a). However, the restrictions with limited allelic segregation and recombination in linkage mapping, low genome coverage of molecular markers, low resolution of QTL regions, and small effects of polygenic interaction had hindered the effective identification of critical genes. The GWAS can detect the relationship between natural allelic variations and phenotypic variation and can provide an effective strategy for the rapid identification of key genes responsible for complex quantitative traits. The method overcomes the limitation of biparental populations and offers more variation sites from historical recombination events. The diversity panel of germplasm resources provides abundant variation sites for GWAS. In this study, the PCA and NJ-tree analyses showed a continuous distribution without a strong population structure, and the genotypes we selected were helpful to control false positives in GWAS. The MLM further controlled the population stratification and familial kinship of the accessions and is extensively used for crops and forage grass research as a popular method (Li H. et al., 2013; Valluru et al., 2017; Ramstein et al., 2018; Kang et al., 2019; Chen et al., 2021; Li et al., 2021). In addition, high-throughput sequencing generated more than tens of thousands of markers, and many loci have strongly LD against the estimation of false-positive rate by using independent tests among the actual number of markers in the Bonferroni assumption (Li et al., 2012; Li H. et al., 2013). A secondary identification of GWAS signals using an enlarged candidate region, considering the deviation of observed *P*-values from expected distribution in quantile-quantile (Q-Q) plots, increased some of the false-negative SNPs filtered by too strict and conservative Bonferroni correction (Valluru et al., 2017; Li et al., 2019; Mandozai et al., 2021).

Candidate Genes for Yield-Related Traits

Based on the phenotypic observation and joint analysis with three programs, the GWAS efficiently detected a total of 1,825 significant loci associated with 12 grass yield- and seed-related traits in *E. sibiricus*. To evaluate the effects of identified genes, each of these was used to analyze the gene expression profiles, and we finally generated 106 candidate genes with higher expression levels (TPM values > 10) and indicated different functions (**Figure 5**). For example, 7 photosynthesis genes and 4 photorespiration genes were highly expressed in plant leaves. Four genes involved in plant hormone salicylic acid and 4 genes involved in jasmonic acid play an important role in plant growth and development, as well as plant photosystems in response to light stress (Nitschke et al., 2016; Chen et al., 2020). Furthermore, the PPI network analysis was employed to focus on the hub genes related to yield-related traits in *E. sibiricus* (**Figure 6**).

For PH traits, the top hub gene, *TraesCS4D01G221200*, is involved in a variety of biological regulation processes, such as auxins- and jasmonic acid-mediated signaling pathway, SAM, xylem, and phloem pattern formation. Meanwhile, the corresponding protein of this gene, TOPLESS (TPL), a conserved co-repressor, regulates a wide array of plant physiology process as central to multiple signaling pathways in higher plants, such as development, meristem maintenance, and hormone signaling (auxins, jasmonic acid, and ethylene; Causier et al., 2012; Wang et al., 2013). During post-embryonic development, the shoot and the root apical meristems are responsible for the formation of the main axis of plant growth. In *Arabidopsis* embryogenesis, the *topless-1* (*tpl-1*) mutation failed to form a SAM, for the absence of repression system (TPL), which acts to the expression of root-promoting genes in the top half of the embryo (Long et al., 2006). Cell-to-cell communication is a prerequisite for the process of growth and development for multicellular organisms. Plasmodesmata (PD) connects the adjacent individual cells to the entire organism, allowing transduction and exchange of diverse macromolecules, metabolites, proteins, nutrients, and phytohormones during plant growth and development (Vu et al., 2020). Mutants of multiple genes have been shown to be responsible for PD function. At the stage of embryogenesis in *Arabidopsis*, *ise1* and *ise2* mutants increased the PD-mediated transport and the PD permeability (Kobayashi et al., 2007; Stonebloom et al., 2009; Carlotto et al., 2016). Whereas, *dse1* mutant reduced the transport *via* PD and exhibited developmental defects at all development stages including smaller plants and delayed flower initiation than the wild type (Xu et al., 2012). Induction of a synthetic allele *icals3m* can effectively block the movement of proteins and RNAs *via* PD by increasing callose and decreasing the size of PD aperture. In root meristem of *Arabidopsis*, this interference in cell-to-cell communication significantly caused the loss of symplastic signaling of cells and tissues around and is critical for the regulation of cell divisions and cell expansion (Wu S. et al., 2016). In our study, we associated a non-synonymous SNP within the exon coding region of *TraesCS1A01G286300*, which was functionally related to plasmodesma. Our data showed a phenotype difference of 32.7% (PH) in the two haplotypes. The gene that was highly expressed in the internode, stem axis, and stem of two wheat cultivars may be involved in the development process in *E. sibiricus* and caused the trait differences.

In addition, the candidate genes which encode auxin signaling (9 genes), tyrosine kinase (6 genes), E3 ubiquitin ligase (5 genes), microtubule (6 genes), and cytoskeleton (5 genes) were largely

involved in culm-related traits. Moreover, the genes related to the hormone signaling pathway, tyrosine kinase activity, cytoskeleton, and microtubule structure tend to be in the center position in the PPI analysis. As a multi-functional phytohormone, auxins modulate numerous plant processes including SAM development, vascular elongation, and axillary root formation (De Smet and Jürgens, 2007). The polar auxin transport was impaired in the maize mutant with qph1, a major plant height QTL, and decreased plant height by 20% (Xing et al., 2015). In addition, microtubule growth and cell cytoskeleton rearrangements have been shown to be accompanied by auxin regulation (Winnicki, 2020). In the meristematic cells of the root epidermis of *Arabidopsis thaliana*, the auxin signaling influences actin cytoskeleton, contributing to the vacuolar morphogenesis and leading to the regulation of cell size in plants (Scheuring et al., 2016). Furthermore, the plant receptors play a crucial regulatory role in many cell signaling pathways leading to cell growth, development, and disease resistance. Eukaryotic protein kinases (ePKs) facilitate the signal transduction and activation (or inactivation) of downstream reaction by phosphorylation of substrate protein of serine (Ser)/threonine (Thr)/tyrosine (Tyr) residues (Schwessinger et al., 2011; Jose et al., 2020). In model plants, rice and *Arabidopsis*, the proportions of conserved phosphorylation sites of serine and threonine were about 80–85% and 10–15%, respectively, while tyrosine was less than 5% (Nakagami et al., 2010). Despite lower abundance, Tyr phosphorylation has been demonstrated to be involved in different stages and processes of plant development, including the ABA pathway in seeds (Ghelis et al., 2008), leaf size, flowering time (Oh et al., 2009), plant growth (Jaillais et al., 2011; Nemoto et al., 2017), and root hair growth and development (Yemets et al., 2008).

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

YW, WL, SB, and WX conceived the research. WL and SB provided the germplasm resources and reviewed the manuscript. ZZ, YZ, and NW implemented the experiment. ZZ, YZ, and JZ performed the data curation and analysis. ZZ and WX wrote the manuscript. All authors approved the final manuscript.

FUNDING

This research was funded by the Chinese National Natural Science Foundation (31971751), Gansu Provincial Science and Technology Major Projects (19ZD2NA002), the State's Key Project of Research and Development Plan (2019YFC0507702), the open projects of the Key Laboratory of Superior Forage Germplasm in the Qinghai-Tibetan Plateau (2020-ZJ-Y03), and the Fundamental Research Fund for the Central Universities (lzujbky-2021-ct21).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.874409/full#supplementary-material

Supplementary Figure 1 | Distribution of single nucleotide polymorphisms (SNPs) in a window size of 5 Mb (**A**, with all of 13,643,515 SNPs) and 20 Mb (**B**, with 88,506 filtered SNPs) on 22 chromosomes of wheat. Each color represents different densities.

Supplementary Figure 2 | The heatmap of kinship matrix of the 210 *Elymus sibiricus* genotypes.

Supplementary Figure 3 | Genome-wide association study (GWAS) results of 12 phenotypic traits in *E. sibiricus* by GEMMA test. **(A)** Manhattan plots. The threshold was set as $P < 9.36e\text{-}6$ and 1e-3, respectively. **(B)** The Q-Q plots.

Supplementary Figure 4 | GWAS results of 12 phenotypic traits in *E. sibiricus* by EMMAX test. **(A)** Manhattan plots. The threshold was set as $P < 1.13e\text{-}6$ and 1e-4, respectively. **(B)** The quantile-quantile (Q-Q) plots.

Supplementary Figure 5 | Genome-wide association study results of 12 phenotypic traits in *E. sibiricus* by TASSEL test. **(A)** Manhattan plots. The threshold was set as $P < 1.13e\text{-}6$ and 1e-4, respectively. **(B)** The Q-Q plots.

REFERENCES

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

Bai, Q., Cai, Y., He, B., Liu, W., Pan, Q., and Zhang, Q. (2019). Core set construction and association analysis of *Pinus massoniana* from Guangdong province in southern China using SLAF-seq. *Sci. Rep.* 9:13157. doi: 10.1038/s41598-019-49737-2

Barnes, P. W., Tobler, M. A., Keefover-Ring, K., Flint, S. D., Barkley, A. E., Ryel, R. J., et al. (2016). Rapid modulation of ultraviolet shielding in plants is influenced by solar ultraviolet radiation and linked to alterations in flavonoids. *Plant Cell Environ.* 39, 222–230. doi: 10.1111/pce.12609

Berland, H., Albert, N. W., Stavland, A., Jordheim, M., McGhie, T. K., Zhou, Y., et al. (2019). Auronidins are a previously unreported class of flavonoid pigments that challenges when anthocyanin biosynthesis evolved in plants. *Proc. Natl. Acad. Sci. U.S.A.* 116, 20232–20239. doi: 10.1073/pnas.1912741116

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Carlotto, N., Wirth, S., Furman, N., Ferreyra Solari, N., Ariel, F., Crespi, M., et al. (2016). The chloroplastic DEVH-box RNA helicase INCREASED SIZE EXCLUSION LIMIT 2 involved in plasmodesmata regulation is required for group II intron splicing. *Plant Cell Environ.* 39, 165–173. doi: 10.1111/pce.12603

Causier, B., Ashworth, M., Guo, W., and Davies, B. (2012). The TOPLESS interactome: a framework for gene repression in *Arabidopsis*. *Plant physiol.* 158, 423–438. doi: 10.1104/pp.111.186999

Chapman, M. A., Hiscock, S. J., and Filatov, D. A. (2013). Genomic divergence during speciation driven by adaptation to altitude. *Mol. Biol. Evol.* 30, 2553–2567. doi: 10.1093/molbev/mst168

Chen, Y. E., Mao, H. T., Wu, N., Mohi Ud Din, A., Khan, A., Zhang, H. Y., et al. (2020). Salicylic acid protects photosystem II by alleviating photoinhibition in

Arabidopsis thaliana under high light. *Int. J. Mol. Sci.* 21:1229. doi: 10.3390/ijms21041229

Chen, Z. Q., Zan, Y., Milesi, P., Zhou, L., Chen, J., Li, L., et al. (2021). Leveraging breeding programs and genomic data in Norway spruce (Picea abies L. Karst) for GWAS analysis. *Genome Biol.* 22:179.

Cui, G., Li, B., He, W., Yin, X., Liu, S., Lian, L., et al. (2018). Physiological analysis of the effect of altitudinal gradients on *Leymus secalinus* on the Qinghai-Tibetan Plateau. *PLoS One* 13:e0202881. doi: 10.1371/journal.pone.0202881

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

De Smet, I., and Jürgens, G. (2007). Patterning the axis in plants - auxin in control. *Curr. Opin. Genet. Dev.* 17, 337–343. doi: 10.1016/j.gde.2007.04.012

De, Y., Wang, Q., Mu, H. B., Xie, J. H., and Zhao, L. X. (2014). Study on chromosome karyotype of *Elymus sibiricus*. *Chinese J. Grassl.* 36, 79–83.

Dewey, D. R. (1984). "The genomic system of classification as a guide to intergeneric hybridization with the perennial Triticeae," in *Gene Manipulation in Plant Improvement*, ed. J. P. Gustafson (New York, NY: Plenum Publishers), 209–279.

Fang, L., Wang, Q., Hu, Y., Jia, Y., Chen, J., Liu, B., et al. (2017a). Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat. Genet.* 49, 1089–1098. doi: 10.1038/ng.3887

Fang, L., Gong, H., Hu, Y., Liu, C., Zhou, B., Huang, T., et al. (2017b). Genomic insights into divergence and dual domestication of cultivated allotetraploid cottons. *Genome Biol.* 18:33. doi: 10.1186/s13059-017-1167-5

Flint-Garcia, S. A., Thornsberry, J. M., and Buckler, E. S. (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54, 357–374. doi: 10.1146/annurev.arplant.54.031902.134907

Gao, J., Wang, S., Zhou, Z., Wang, S., Dong, C., Mu, C., et al. (2019). Linkage mapping and genome-wide association reveal candidate genes conferring thermotolerance of seed-set in maize. *J. Exp. Bot.* 70, 4849–4863. doi: 10.1093/jxb/erz171

Ghelis, T., Bolbach, G., Clodic, G., Habricot, Y., Miginiac, E., Sotta, B., et al. (2008). Protein tyrosine kinases and protein tyrosine phosphatases are involved in abscisic acid-dependent processes in *Arabidopsis* seeds and suspension cells. *Plant Physiol.* 148, 1668–1680. doi: 10.1104/pp.108.124594

Grabowski, P. P., Evans, J., Daum, C., Deshpande, S., Barry, K. W., Kennedy, M., et al. (2017). Genome-wide associations with flowering time in switchgrass using exome-capture sequencing data. *New Phytol.* 213, 154–169. doi: 10.1111/nph.14101

Gyawali, S., Otte, M. L., Chao, S., Jilal, A., Jacob, D. L., Amezrou, R., et al. (2017). Genome wide association studies (GWAS) of element contents in grain with a special focus on zinc and iron in a world collection of barley (*Hordeum vulgare* L.). *J. Cereal Sci.* 77, 266–274.

Heijde, M., Binkert, M., Yin, R., Ares-Orpel, F., Rizzini, L., Van De Slijke, E., et al. (2013). Constitutively active UVR8 photoreceptor variant in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20326–20331. doi: 10.1073/pnas.1314336110

Jaillais, Y., Hothorn, M., Belkhadir, Y., Dabi, T., Nimchuk, Z. L., Meyerowitz, E. M., et al. (2011). Tyrosine phosphorylation controls brassinosteroid receptor activation by triggering membrane release of its kinase inhibitor. *Gene. Dev.* 25, 232–237. doi: 10.1101/gad.2001911

Jin, J., Wang, T., Cheng, Y., Wang, L., Zhang, J., Jing, H., et al. (2021). Current situation and prospect of forage breeding in China. *Bull. Chin. Acad. Sci.* 36, 660–665.

Jose, J., Ghantasala, S., and Roy Choudhury, S. (2020). *Arabidopsis* transmembrane receptor-like kinases (RLKs): a bridge between extracellular signal and intracellular regulatory machinery. *Int. J. Mol. Sci.* 21:4000. doi: 10.3390/ijms21114000

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

Kang, Y., Torres-Jerez, I., An, Z., Greve, V., Huhman, D., Krom, N., et al. (2019). Genome-wide association analysis of salinity responsive traits in *Medicago truncatula*. *Plant Cell Environ* 42, 1513–1531. doi: 10.1111/pce.13508

Kobayashi, K., Otegui, M. S., Krishnakumar, S., Mindrinos, M., and Zambryski, P. (2007). Increased Size Exclusion Limit 2 encodes a putative DEVH box RNA helicase involved in plasmodesmata function during *Arabidopsis* embryogenesis. *Plant Cell* 19, 1885–1897. doi: 10.1105/tpc.106.045666

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

Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., et al. (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* 45, 43–50. doi: 10.1038/ng.2484

Li, L., Zhang, C., Huang, J., Liu, Q., Wei, H., Wang, H., et al. (2021). Genomic analyses reveal the genetic basis of early maturity and identification of loci and candidate genes in upland cotton (*Gossypium hirsutum* L.). *Plant Biotechnol. J.* 19, 109–123. doi: 10.1111/pbi.13446

Li, M. X., Yeung, J. M. Y., Cherny, S. S., and Sham, P. C. (2012). Evaluating the effective numbers of independent tests and significant *p*-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum. Genet.* 131, 747–756. doi: 10.1007/s00439-011-1118-2

Li, M., Tian, S., Jin, L., Zhou, G., Li, Y., Zhang, Y., et al. (2013). Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat. Genet.* 45, 1431–1438. doi: 10.1038/ng.2811

Li, T., Ma, X., Li, N., Zhou, L., Liu, Z., Han, H., et al. (2017). Genome-wide association study discovered candidate genes of *Verticillium* wilt resistance in upland cotton (*Gossypium hirsutum* L.). *Plant Biotechnol. J.* 15, 1520–1532. doi: 10.1111/pbi.12734

Li, Y., Cao, K., Zhu, G., Fang, W., Chen, C., Wang, X., et al. (2019). Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. *Genome Biol.* 20:36. doi: 10.1186/s13059-019-1648-9

Liu, J. Y., Zhang, Y. W., Han, X., Zuo, J. F., Zhang, Z., Shang, H., et al. (2020). An evolutionary population structure model reveals pleiotropic effects of *GmPDAT* for traits related to seed size and oil content in soybean. *J. Exp. Bot.* 71, 6988–7002. doi: 10.1093/jxb/eraa426

Long, J. A., Ohno, C., Smith, Z. R., and Meyerowitz, E. M. (2006). TOPLESS regulates apical embryonic fate in *Arabidopsis*. *Science* 312, 1520–1523. doi: 10.1126/science.1123841

Ma, X., Chen, S., Zhang, X., Bai, S., and Zhang, C. (2012a). Assessment of worldwide genetic diversity of Siberian wild rye (*Elymus sibiricus* L.) germplasm based on gliadin analysis. *Molecules* 17, 4424–4434. doi: 10.3390/molecules17044424

Ma, X., Chen, S.-Y., Bai, S.-Q., Zhang, X.-Q., Li, D.-X., Zhang, C.-B., et al. (2012b). RAPD analysis of genetic diversity and population structure of *Elymus sibiricus* (Poaceae) native to the southeastern Qinghai-Tibet Plateau. China. *Genet. Mol. Res.* 11, 2708–2718. doi: 10.4238/2012.June.27.5

Mandozai, A., Moussa, A. A., Zhang, Q., Qu, J., Du, Y., Anwari, G., et al. (2021). Genome-Wide association study of root and shoot related traits in spring soybean (*Glycine max* L.) at seedling stages using SLAF-seq. *Front. Plant Sci.* 12:568995. doi: 10.3389/fpls.2021.568995

Manunza, A., Casellas, J., Quintanilla, R., González-Prendes, R., Pena, R. N., Tibau, J., et al. (2014). A genome-wide association analysis for porcine serum lipid traits reveals the existence of age-specific genetic determinants. *BMC Genomics* 15:758. doi: 10.1186/1471-2164-15-758

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Nakagami, H., Sugiyama, N., Mochida, K., Daudi, A., Yoshida, Y., Toyoda, T., et al. (2010). Large-scale comparative phosphoproteomics identifies conserved phosphorylation sites in plants. *Plant Physiol.* 153, 1161–1174. doi: 10.1104/pp.110.157347

Nemoto, K., Ramadan, A., Arimura, G., Imai, K., Tomii, K., Shinozaki, K., et al. (2017). Tyrosine phosphorylation of the GARU E3 ubiquitin ligase promotes gibberellin signalling by preventing GID1 degradation. *Nat. Commun.* 8:1004. doi: 10.1038/s41467-017-01005-5

Nitschke, S., Cortleven, A., Iven, T., Feussner, I., Havaux, M., Riefler, M., et al. (2016). Circadian stress regimes affect the circadian clock and cause jasmonic acid-dependent cell death in cytokinin-deficient *arabidopsis* plants. *Plant Cell* 28, 1616–1639. doi: 10.1105/tpc.16.00016

Oh, M. H., Wang, X., Kota, U., Goshe, M. B., Clouse, S. D., and Huber, S. C. (2009). Tyrosine phosphorylation of the BRI1 receptor kinase emerges as a component of brassinosteroid signaling in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 658–663. doi: 10.1073/pnas.0810249106

Pang, Y., Liu, C., Wang, D., St Amand, P., Bernardo, A., Li, W., et al. (2020). High-resolution genome-wide association study identifies genomic regions and candidate genes for important agronomic traits in wheat. *Mol Plant* 13, 1311–1327. doi: 10.1016/j.molp.2020.07.008

Podolec, R., Lau, K., Wagnon, T. B., Hothorn, M., and Ulm, R. (2021). A constitutively monomeric UVR8 photoreceptor confers enhanced UV-B photomorphogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2017284118. doi: 10.1073/pnas.2017284118

Ramstein, G. P., Evans, J., Nandety, A., Saha, M. C., Brummer, E. C., Kaeppler, S. M., et al. (2018). Candidate variants for additive and interactive effects on bioenergy traits in switchgrass (*Panicum virgatum* L.) identified by genome-wide association analyses. *Plant Genome* 11, 1–18. doi: 10.3835/plantgenome2018.01.0002

Rizzini, L., Favory, J. J., Cloix, C., Faggionato, D., O'Hara, A., Kaiserli, E., et al. (2011). Perception of UV-B by the *Arabidopsis* UVR8 protein. *Science* 332, 103–106.

Scheuring, D., Löfke, C., Krüger, F., Kittelmann, M., Eisa, A., Hughes, L., et al. (2016). Actin-dependent vacuolar occupancy of the cell determines auxin-induced growth repression. *Proc. Natl. Acad. Sci. U.S.A.* 113, 452–457. doi: 10.1073/pnas.1517445113

Schulz, E., Tohge, T., Winkler, J. B., Albert, A., Schäffner, A. R., Fernie, A. R., et al. (2021). Natural variation among *Arabidopsis* accessions in the regulation of flavonoid metabolism and stress gene expression by combined UV radiation and cold. *Plant Cell Physiol.* 62, 502–514. doi: 10.1093/pcp/pcab013

Schwessinger, B., Roux, M., Kadota, Y., Ntoukakis, V., Sklenar, J., Jones, A., et al. (2011). Phosphorylation-dependent differential regulation of plant growth, cell death, and innate immunity by the regulatory receptor-like kinase BAK1. *PLoS Genet.* 7:e1002046. doi: 10.1371/journal.pgen.1002046

Shen, C., Du, H., Chen, Z., Lu, H., Zhu, F., Chen, H., et al. (2020). The chromosome-level genome sequence of the autotetraploid alfalfa and resequencing of core germplasms provide genomic resources for alfalfa research. *Mol. Plant* 13, 1250–1261. doi: 10.1016/j.molp.2020.07.003

Siipola, S. M., Kotilainen, T., Sipari, N., Morales, L. O., Lindfors, A. V., Robson, T. M., et al. (2015). Epidermal UV-A absorbance and whole-leaf flavonoid composition in pea respond more to solar blue light than to solar UV radiation. *Plant Cell Environ.* 38, 941–952. doi: 10.1111/pce.12403

Stonebloom, S., Burch-Smith, T., Kim, I., Meinke, D., Mindrinos, M., and Zambryski, P. (2009). Loss of the plant DEAD-box protein ISE1 leads to defective mitochondria and increased cell-to-cell transport via plasmodesmata. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17229–17234. doi: 10.1073/pnas.0909229106

Sui, M., Jing, Y., Li, H., Zhan, Y., Luo, J., Teng, W., et al. (2020). Identification of loci and candidate genes analyses for tocopherol concentration of soybean seed. *Front. Plant Sci.* 11:539460. doi: 10.3389/fpls.2020.539460

Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W., et al. (2013). SLAF-seq: an efficient method of large-scale *de novo* SNP discovery and genotyping using high-throughput sequencing. *PLoS One* 8:e58700. doi: 10.1371/journal.pone.0058700

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

Terhorst, J., Kamm, J. A., and Song, Y. S. (2016). Robust and scalable inference of population history from hundreds of unphased whole-genomes. *Nat. Genet.* 49, 303–309. doi: 10.1038/ng.3748

Valluru, R., Reynolds, M. P., Davies, W. J., and Sukumaran, S. (2017). Phenotypic and genome-wide association analysis of spike ethylene in diverse wheat genotypes under heat stress. *New Phytol.* 214, 271–283. doi: 10.1111/nph.14367

Vu, M. H., Iswanto, A. B. B., Lee, J., and Kim, J. Y. (2020). The role of plasmodesmata-associated receptor in plant development and environmental response. *Plants* 9:216. doi: 10.3390/plants9020216

Wang, L., Kim, J., and Somers, D. E. (2013). Transcriptional corepressor TOPLESS complexes with pseudoresponse regulator proteins and histone deacetylases to regulate circadian transcription. *Proc. Natl. Acad. Sci. U.S.A.* 110, 761–766. doi: 10.1073/pnas.1215010110

Wang, X., Gao, Y., Wu, X., Wen, X., Li, D., Zhou, H., et al. (2021a). High-quality evergreen azalea genome reveals tandem duplication-facilitated low-altitude adaptability and floral scent evolution. *Plant Biotechnol. J.* 19, 2544–2560. doi: 10.1111/pbi.13680

Wang, X., Liu, S., Zuo, H., Zheng, W., Zhang, S., Huang, Y., et al. (2021b). Genomic basis of high-altitude adaptation in Tibetan *Prunus* fruit trees. *Curr. Biol.* 31, 3848–3860. doi: 10.1016/j.cub.2021.06.062

Wargent, J. J., Elfadly, E. M., Moore, J. P., and Paul, N. D. (2011). Increased exposure to UV-B radiation during early development leads to enhanced photoprotection and improved long-term performance in *Lactuca sativa*. *Plant Cell Environ.* 34, 1401–1413. doi: 10.1111/j.1365-3040.2011.02342.x

Wei, T., van Treuren, R., Liu, X., Zhang, Z., Chen, J., Liu, Y., et al. (2021). Whole-genome resequencing of 445 *Lactuca* accessions reveals the domestication history of cultivated lettuce. *Nat. Genet.* 53, 752–760. doi: 10.1038/s41588-021-00831-0

Winnicki, K. (2020). The winner takes it all: auxin - the main player during plant embryogenesis. *Cells* 9:606. doi: 10.3390/cells9030606

Wu, Q., Su, N., Zhang, X., Liu, Y., Cui, J., and Liang, Y. (2016). Hydrogen peroxide, nitric oxide and UV RESISTANCE LOCUS8 interact to mediate UV-B-induced anthocyanin biosynthesis in radish sprouts. *Sci. Rep.* 6:29164. doi: 10.1038/srep29164

Wu, S., O'Lexy, R., Xu, M., Sang, Y., Chen, X., Yu, Q., et al. (2016). Symplastic signaling instructs cell division, cell expansion, and cell polarity in the ground tissue of *Arabidopsis thaliana* roots. *Proc. Natl. Acad. Sci. U.S.A.* 11, 11621–11626. doi: 10.1073/pnas.1610358113

Xie, D., Dai, Z., Yang, Z., Tang, Q., Sun, J., Yang, X., et al. (2018). Genomic variations and association study of agronomic traits in flax. *BMC Genomics* 19:512. doi: 10.1186/s12864-018-4899-za

Xie, W., Zhang, J., Zhao, X., Zhang, J., and Wang, Y. (2015). Siberian wild rye (*Elymus sibiricus* L.): genetic diversity of germplasm determined using DNA fingerprinting and SCoT markers. *Biochem. Syst. Ecol.* 60, 186–192.

Xing, A., Gao, Y., Ye, L., Zhang, W., Cai, L., Ching, A., et al. (2015). A rare SNP mutation in *Brachytic2* moderately reduces plant height and increases yield potential in maize. *J. Exp. Bot.* 66, 3791–3802. doi: 10.1093/jxb/erv182

Xiong, Y., Lei, X., Bai, S., Xiong, Y., Liu, W., Wu, W., et al. (2021). Genomic survey sequencing, development and characterization of single- and multi-locus genomic SSR markers of *Elymus sibiricus* L. *BMC Plant Biol.* 21:3. doi: 10.1186/s12870-020-02770-0

Xu, M., Cho, E., Burch-Smith, T. M., and Zambryski, P. C. (2012). Plasmodesmata formation and cell-to-cell transport are reduced in *decreased size exclusion limit 1* during embryogenesis in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 5098–5103. doi: 10.1073/pnas.1202919109

Yan, W. H., Ma, Y. B., Zhang, J. R., Wang, K., and Meng, Q. W. (2017). The analysis of genetic diversity and the construction of core collection for *Elymus sibiricus* L. *J. Grassl. Forage Sci.* 38, 1–12.

Yang, J., Jin, Z.-B., Chen, J., Huang, X.-F., Li, X.-M., Liang, Y.-B., et al. (2017). Genetic signatures of high-altitude adaptation in Tibetans. *Proc. Natl. Acad. Sci. U.S.A.* 114, 4189–4194.

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P., Hu, L., et al. (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* 48, 927–934. doi: 10.1038/ng.3596

Yemets, A., Sheremet, Y., Vissenberg, K., Vanorden, J., Verbelen, J., and Blume, Y. (2008). Effects of tyrosine kinase and phosphatase inhibitors on microtubules in *Arabidopsis* root cells. *Cell Biol. Int.* 32, 630–637.

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702

Zhang, F., Wu, J., Sade, N., Wu, S., Egbaria, A., Fernie, A. R., et al. (2021). Genomic basis underlying the metabolome-mediated drought adaptation of maize. *Genome Biol.* 22:260. doi: 10.1186/s13059-021-02481-1

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42, 355–360. doi: 10.1038/ng.546

Zhang, Z., Xie, W., Zhang, J., Wang, N., Zhao, Y., Wang, Y., et al. (2019a). Construction of the first high-density genetic linkage map and identification of seed yield-related QTLs and candidate genes in *Elymus sibiricus*, an important forage grass in Qinghai-Tibet Plateau. *BMC Genomics* 20:861. doi: 10.1186/s12864-019-6254-4

Zhang, Z., Xie, W., Zhao, Y., Zhang, J., Wang, N., Ntakirutimana, F., et al. (2019b). EST-SSR marker development based on RNA-sequencing of *E. sibiricus* and its application for phylogenetic relationships analysis of seventeen *Elymus* species. *BMC Plant Biol.* 19:235. doi: 10.1186/s12870-019-1825-8

Zhang, Z., Xie, W., Zhang, J., Zhao, X., Zhao, Y., and Wang, Y. (2018). Phenotype- and SSR-based estimates of genetic variation between and within two important *Elymus* species in western and northern China. *Genes* 9:147.

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44, 821–824.

Zhuo, X., Zheng, T., Li, S., Zhang, Z., Zhang, M., Zhang, Y., et al. (2021). Identification of the *PmWEEP* locus controlling weeping traits in *Prunus mume* through an integrated genome-wide association study and quantitative trait locus mapping. *Hortic. Res.* 8:131. doi: 10.1038/s41438-021-00573-4

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 11 July 2022
doi: 10.3389/fpls.2022.913947

A Genome-Wide Association Study Coupled With a Transcriptomic Analysis Reveals the Genetic Loci and Candidate Genes Governing the Flowering Time in Alfalfa (*Medicago sativa* L.)

Fei He[1†], Fan Zhang[1†], Xueqian Jiang[1], Ruicai Long[1], Zhen Wang[1], Yishi Chen[2], Mingna Li[1], Ting Gao[3], Tianhui Yang[3], Chuan Wang[3], Junmei Kang[1], Lin Chen[1] and Qingchuan Yang[1*]*

[1] Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China, [2] Center for Monitoring of Agricultural Ecological Environment and Quality Inspection of Agricultural Products of Tianjin, Tianjin, China, [3] Institute of Animal Science, Ningxia Academy of Agricultural and Forestry Sciences, Yinchuan, China

OPEN ACCESS

Edited by:
Jiyu Zhang,
Lanzhou University, China

Reviewed by:
N. C. Ereful,
National Institute of Agricultural Botany (NIAB), United Kingdom
Ram Kumar Basnet,
Rijk Zwaan Zaadteelt en Zaadhandel B.V., Netherlands

***Correspondence:**
Lin Chen
chenlin@caas.cn
Qingchuan Yang
qchyang66@163.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

Received: 06 April 2022
Accepted: 16 June 2022
Published: 11 July 2022

Citation:
He F, Zhang F, Jiang X, Long R, Wang Z, Chen Y, Li M, Gao T, Yang T, Wang C, Kang J, Chen L and Yang Q (2022) A Genome-Wide Association Study Coupled With a Transcriptomic Analysis Reveals the Genetic Loci and Candidate Genes Governing the Flowering Time in Alfalfa (Medicago sativa L.). Front. Plant Sci. 13:913947. doi: 10.3389/fpls.2022.913947

The transition to flowering at the right time is very important for adapting to local conditions and maximizing alfalfa yield. However, the understanding of the genetic basis of the alfalfa flowering time remains limited. There are few reliable genes or markers for selection, which hinders progress in genetic research and molecular breeding of this trait in alfalfa. We sequenced 220 alfalfa cultivars and conducted a genome-wide association study (GWAS) involving 875,023 single-nucleotide polymorphisms (SNPs). The phenotypic analysis showed that the breeding status and geographical origin strongly influenced the alfalfa flowering time. Our GWAS revealed 63 loci significantly related to the flowering time. Ninety-five candidate genes were detected at these SNP loci within 40 kb (20 kb up- and downstream). Thirty-six percent of the candidate genes are involved in development and pollen tube growth, indicating that these genes are key genetic mechanisms of alfalfa growth and development. The transcriptomic analysis showed that 1,924, 2,405, and 3,779 differentially expressed genes (DEGs) were upregulated across the three growth stages, while 1,651, 2,613, and 4,730 DEGs were downregulated across the stages. Combining the results of our GWAS and transcriptome analysis, in total, 38 candidate genes (7 differentially expressed during the bud stage, 13 differentially expressed during the initial flowering stage, and 18 differentially expressed during the full flowering stage) were identified. Two SNPs located in the upstream region of the *Msa0888690* gene (which is involved in isop renoids) were significantly related to flowering. The two significant SNPs within the upstream region of *Msa0888690* existed as four different haplotypes in this panel. The genes identified in this study represent a series of candidate targets for further research investigating the alfalfa flowering time and could be used for alfalfa molecular breeding.

Keywords: alfalfa, GWAS, flowering time, SNP, haplotypes

INTRODUCTION

Alfalfa (*Medicago sativa* L.) is a global forage legume crop species with a high yield and nutritional value (He et al., 2019). According to the growth and development stages of alfalfa, it is mainly divided into the following stages: greening stage, branching stage, budding stage, flowering stage and maturity stage. The flowering stage period is an important trait that can determine when alfalfa is harvested, and different varieties show extensive variation in life-history characteristics, such as the flowering time. The flowering time is a quantitative trait that is determined by genetic and environmental factors (Koornneef et al., 1998a). In plants, early flowering may be a desirable trait in practical production because it can hasten the growing season, thus avoiding unfavorable climatic conditions (such as drought) (Ridge et al., 2017). The flowering time is also very important for alfalfa because it can determine the harvest time, which affects the forage quality, silage, and yield of alfalfa (Adhikari et al., 2019). Alfalfa differs from other grass species; alfalfa is cut several times a year, and its total digestible nutrients (TDNs) gradually decrease with increasing maturity. Therefore, alfalfa is usually harvested before the seeds are fully mature. Due to these properties, farmers can choose the most suitable grazing periods and harvest dates, thus maximizing economic benefits (Arzani et al., 2004). Although the flowering time is very important, the potential molecular mechanism controlling the flowering time of alfalfa has not been clarified.

The genetic and genomic basis of the flowering time has been investigated extensively in several plant species, such as *Arabidopsis thaliana*, rice (*Oryza sativa* L.), and wheat (*Triticum aestivum* L.) (Grillo et al., 2013; Hori et al., 2016; Ivanièová et al., 2016), whereas such information in alfalfa is scarce. Many flowering time-related genes, such as Flowering Locus T (*FT*), Flowering Locus C (*FLC*), Constans (*CO*), and suppressor of overexpression of constans 1 (*SOC*1), have been cloned. These genes are mainly responsive to pathways involving the photoperiod, temperature and gibberellin (Lyons et al., 2015; Kang et al., 2019). The *FT* gene is generally considered to integrate inputs of several pathways that ultimately result in the floral transition. In *A. thaliana, FT* loss-of-function mutations result in late flowering under long-day conditions (Koornneef et al., 1998b). The genetic basis of the variation in the flowering time of legume species, such as pea (*Pisum sativum*), soybean (*Glycine max* L.) and *Medicago truncatula,* has been studied. In *M. truncatula*, a significant number of genes have been verified to control the flowering time, including *FTa*, *FTb*, and *FTc* (Hecht et al., 2005). Florigen members of the *FT* family (*MtFTa*1, *MtFTb*1, and *MtFTc*) that control the flowering traits of *M. truncatula* were shown to successfully rescue late-flowering mutant plants by inducing early flowering (Laurie et al., 2011). Similarly, the expression of several genes in *M. sativa*, including *MsLFY* (Zhang et al., 2013), *SPL*13 (Gao et al., 2018), and *MsZFN* (Chao et al., 2014), has been measured; these genes have been molecularly cloned, and their involvement in the flowering time of alfalfa was shown via reverse genetic methods.

Knowledge of single-nucleotide polymorphism (SNP) markers of related traits has become the genetic basis of various trait improvements in plant breeding (Battenfield et al., 2018). In recent years, with the introduction of high-density marker arrays, genome-wide association studies (GWASs) have been actively applied to many different crop species and varieties; SNPs related to many agronomic traits, such as the flowering time (Navarro et al., 2017), salt tolerance (Patil et al., 2016), and plant height (Zhang J. P. et al., 2015), have been found, and many candidate genes or genomic regions have been revealed. For example, eight candidate genes related to the flowering time, including *Hd1*, have been identified in 950 different rice varieties (Huang et al., 2012). Ten candidate genes, including *SOC1*, were identified through an association analysis of 309 soybean varieties (Zhang J. P. et al., 2015). Kim et al. conducted a GWAS of 2,662 soybean varieties and found 18 candidate genes involved in 6 main flowering pathways (Kim et al., 2020). Through a GWAS analysis of 137 core germplasms used worldwide, Shen et al. (2020) found genes related to the flowering time in alfalfa; these genes promoted the transition from vegetative growth to flowering (Shen et al., 2020). These findings provide valuable information for various breeding plans focusing on the flowering time. In addition, RNA sequencing (RNA-seq) is a powerful tool used to discover candidate genes and target traits and has been widely applied to many plant species.

Although many GWASs have been performed in alfalfa, most focused only on drought resistance (Zhang T. J. et al., 2015), salt tolerance (Yu et al., 2016), and forage quality (Lin et al., 2020), while few focused on genetic analyses and the molecular regulatory mechanisms underlying flowering time traits in alfalfa. In addition, due to the lack of a reference genome for alfalfa, most previous studies used the genome of *M. truncatula* as a reference (Wang et al., 2020). However, because these two legumes are different species, few functional genes are found in cultivated alfalfa, limiting improvement in cultivars (Chen et al., 2021). With the development of genome assembly technology, the genomes of the Zhongmu No.1, XinJiangDaYe and Zhongmu-4 varieties have been released (Chen H. et al., 2020; Shen et al., 2020; Long et al., 2022). These findings will greatly facilitate the identification of functional genes involved in the alfalfa flowering time and provide important information for the improvement of alfalfa in the future.

To better understand the genetic structure of the alfalfa flowering time, in this study, we first conducted a GWAS of a panel that consisted of 220 alfalfa varieties. Then, RNA-seq analysis was used to identify the genes involved in the flowering time in alfalfa. Finally, we identified candidate genes associated with the flowering time through a combination of GWAS and RNA-seq analysis. Candidate genes or homologous *A. thaliana* genes with known functions were also proposed. This study enriches our knowledge of the genetic basis underlying the flowering time in alfalfa and provides valuable markers for the molecular breeding of alfalfa.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

The plant materials used in this study included 220 accessions collected worldwide. The sources of the materials were reported in a previous study (Chen et al., 2021). In October 2017,

seeds of 220 accessions were planted in a greenhouse of the Chinese Academy of Agricultural Sciences (CAAS) in Langfang, Hebei Province, China (39.59°N, 116.59°E). The greenhouse was maintained such that the photoperiod was 16 h days/8 h nights, the temperature was 22°C, and the relative humidity was 40%. In the experimental field used, the annual average temperature was 11.9°C, the coldest month was January (−4.7°C), and the warmest month was July (26.2°C). The average annual precipitation was 554.9 mm, with large spatial and seasonal variations. During the summer season (July through September), the area received more than 50% of the annual precipitation. The soil was a medium loam that comprised 1.69% organic matter (pH = 7.37). Individual plants were transplanted into the field of the CAAS in April 2018. The field trial employed a randomized complete block design with three replications. Each replication included five cloned plants per individual, and each replicate was spaced 30 cm apart. The spacing was 65 cm between rows and 65 cm between plants. No fertilizer or irrigation was applied, and weeding was performed manually. The remaining 5 cm of mowing was performed in each individual plant before the winter, thus ensuring consistency between individuals.

Phenotyping of the Flowering Time

The flowering time (the date when the first flower appeared) in the first cycle (daily from April to May) in Langfang in 2019, 2020, and 2021 was measured and converted to photothermal units (PTUs) based on the methods described by Grabowski et al. (2017). The average daily temperature (avgT) was calculated as avgT = (minT + maxT)/2. MinT represents the lowest temperature, and maxT represents the highest temperature. In the PTU calculation formula, days were defined as the total days from the date accumulating after 5 consecutive days during which avgT was >10°C to the date the first flower appeared. dayL was defined as the total number of hours between sunrise and sunset in a day. The information of avgT and dayL was obtained from the weather station of Yi Kang Nong (Beijing, China). The random-effects model used was described by Zhang et al. (2020). The frequency distribution, coefficient of variation (CV), broad-sense heritability (H^2), and analysis of variance (ANOVA) of the flowering time traits among the accessions were estimated using SPSS v17.0 according to the methods described by Saxena et al. (2014). The three-year average phenotypic value was recorded as FT, and the data from each individual year were recorded as FT2019, FT2020, and FT2021.

Sequencing and Single-Nucleotide Polymorphisms Calling

Phenotyping was performed in 15 individuals per accession, and one individual exhibiting a typical phenotype was selected for sequencing. Young leaves of plants in the early regrowth stage were collected. A CWBIO Plant Genomic DNA Kit (CoWin Biosciences, Beijing, China) was used to extract DNA from 100 mg of fresh young leaf tissue. Paired-end sequencing (with an insert size of approximately 300 bp) was subsequently performed on an Illumina NovaSeq 6000 instrument by Berry Genomics. For the basic information of the sequencing coverage, read count, mapping read, etc., please refer to Long et al. (2022). The raw sequencing data were first subjected to quality filtering via Trimmomatic software with the default parameters (Bolger et al., 2014). BWA-MEM was used to map paired-end sequencing reads to a haploid alfalfa reference genome with 8 chromosomes (Long et al., 2022). SAMtools software was then used to compare and sort the sequencing data, convert the data to a BAM file, and remove duplicates, and the default parameters were used (Li et al., 2009). For information concerning the SNP detection and filtering and the distribution of the generated SNPs on chromosomes, refer to Chen et al. (2021).

Population Structure, Linkage Disequilibrium and Genome-Wide Association Study

The population structure is an important factor affecting GWAS of complex traits. In this study, ADMIXTURE was used to determine the optimal number of subpopulations (K) for investigating the population genetic structure of the global alfalfa panel, and the LD information was calculated using the software PopLDdecay with the default parameters. The results showed that K = 4 had the lowest value of cross-validation error, suggesting that the 220 alfalfa varieties could be divided into four subgroups (Chen et al., 2021). The GWAS analysis of the flowering time (FT) was performed using BLINK C software using Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK) (Huang et al., 2019). This method can improve the statistical power, reduce the computing time and simultaneously reduce the number of false positives and false negatives. Compared with PLINK or FarmCPU, association analyses of BLINK can reveal more genetic loci and more true positives, including loci previously verified in other studies (Huang et al., 2019). In total, 875,023 high-quality SNPs based on the Zhongmu-4 reference haploid genome were used in this study. The threshold for significantly associated loci was a logarithm of odds (LOD) score ≥5.

RNA-Seq and Transcriptomic Analysis

The total RNA was extracted by using TRIzol reagent (Invitrogen, CA, United States). All RNA extracts were treated with an RNase-Free DNase Set (Qiagen, Valencia, CA, United States) and then cleaned with a RNeasy Mini Kit (Qiagen). The cDNA library construction and sequencing on an Illumina HiSeq 2000 platform were conducted by Novogene (Beijing, China). The flower buds of the early flowering alfalfa variety Cangzhou (the earliest flowering variety among the 220 materials) and the alfalfa representative variety Zhongmu No.1 were sampled at three development stages (the bud stage, initial-flowering stage and full-flowering stage). The flowering time during the three stages occurred at approximately the end of April, the beginning of May and middle and late May. There were two replications per material. The Illumina sequencing of the pooled RNA-seq libraries yielded 24 FASTQ files of sequence data. Clean reads obtained via standard procedures were mapped to the Zhongmu-4 reference haploid genomes by HISAT2 software. Feature counts were used to count the mapped reads before

proceeding to the differential expression analysis. The differential expression analysis was performed using DESeq2 software. The gene expression levels were calculated and normalized by reads per kilobase of exon model per million mapped reads (RPKM) (Mortazavi et al., 2008). RNA-seq data were grouped according to different flowering time, and the two groups were compared. The expression values were normalized by scaling to the default setting of 10 million reads. Moderated t statistics for pairwise contrasts were calculated using the Baggerly's test (Baggerly et al., 2003). The Baggerly's test (beta-binomial) is a weighted t-type test statistic that compares the proportions of counts in a group of samples against those of another group, giving them different weights depending on their sizes (total counts) (Baggerly et al., 2003). Normalized RPKM values were analyzed by Baggerly's test to look for significant differences in expression data of Cangzhou alfalfa and Zhongmu No.1 alfalfa samples. Genes with no counts in all two replicates for at least one of the flowering times were discarded as not detectable above the background. Therefore, baggerly's test is selected for data analysis. The Baggerly's p values were corrected for multiple testing for each contrast separately by means of false discovery rate (FDR) (Benjamini and Hochberg, 1995) for significant genes based on ANOVA. FDR corrected p-value <0.05 and log2 of fold change ≥ 1 was used as a cutoff. Significant differentially expressed genes (DEGs) were identified according to FDR. The functional annotations were performed using the web version of the evolutionary genealogy of genes: Nonsupervised Orthologous Groups (eggNOG) tool.

Candidate Gene Analysis

Reported alfalfa Zhongmu-4 reference haploid genomes were used to identify genes with significantly associated loci. EggNOG[1] databases were used for the gene annotations. All genes within 40 kb (20 kb up- and downstream) of the significant loci were identified according to the linkage disequilibrium (LD) of the association panel. We compared the sequences of the identified genes with the *A. thaliana* genome sequence to identify their orthologues based on sequence similarity and ultimately identified 95 genes related to growth and development. The genes near significantly associated loci received increased attention. We identified a total of 1585 differentially expressed genes from three periods of RNA-seq data. Combining the results of our GWAS and transcriptome analysis, in total, 38 candidate genes (7 differentially expressed during the bud stage, 13 differentially expressed during the initial flowering stage, and 18 differentially expressed during the full flowering stage) were identified. Among the 38 candidate genes, three genes were all differentially expressed at the budding stage, the first flowering stage and the full flowering stage. We used these three genes for subsequent verification.

Pattern Analyses of Candidate Genes by Real-Time Quantitative RT–PCR

The total RNA was extracted from the flowers of normal-growing Zhongmu No.1 and Cangzhou alfalfa with TRIzol reagent according to the manufacturer's instructions. Then, a cDNA library was constructed for the subsequent reactions. qRT–PCR primers were designed via the NCBI website[2]. The alfalfa actin gene was used for data normalization, and for each sample, three technical replicates were included. The data were quantified by the $2^{-(\Delta\Delta CT)}$ method (Livak and Schmittgen, 2001).

RESULTS

Phenotypic Data Analysis

In 2019, 2020, and 2021, the average PTU values were 204.21, 168.29, and 174.98, respectively, while the CVs during the same years were 11.30, 13.70, and 13.10%, respectively (**Table 1**). H^2 was calculated as described in a previous study (Tornqvist et al., 2018) and ranged from 0.51 to 0.59. The P values of the Kolmogorov–Smirnov normality test were lower than the threshold (0.05) within three years, and the results show that the assumption of the non-normal distribution of PTUs in this associated group was correct. The three-year total PTU and single-year PTU are shown in **Supplementary Figures 1**, **2**. In the subsequent GWAS analysis, the average PTU value across the three years was used. The data showed that the characteristics were normally distributed (**Supplementary Figures 1**, **2**). The flower time correlations between accession means are shown in **Supplementary Figure 3**. In 2019, 2020, and 2021, the correlation values ranged between 0.15 and 0.38 (**Supplementary Figure 3**).

From a breeding perspective, our association group could be divided into the following three subgroups: wild, landrace and cultivar subgroups. The statistical analysis of the flowering time showed that there was no significant difference among the three subgroups. Among the plants in the three subgroups, those in the landrace group had the earliest flowering time, followed by those in the cultivar and wild groups (**Figure 1A**). The association panel could also be divided into four subgroups according to geographical origin as follows: America, China, Europe, and Turkey. Regarding the flowering time, the American varieties and Chinese varieties exhibited significant differences ($P < 0.01$), while the other three subgroups did not significantly differ, and the flowering time of the Chinese varieties occurred the earliest, followed by the European and American varieties. The Turkish varieties had the latest flowering time (**Figure 1B**).

Genome-Wide Association Studies and Candidate Genes for Flowering Time

Based on the resequencing results, we obtained 875,023 high-quality SNPs distributed across 8 alfalfa chromosomes. Using three years of phenotypic data to analyze the flowering time, we found a total of 16 significant loci according to the LOD >5.00 threshold (**Figure 2A** and **Supplementary Table 1**). These 16 markers are distributed on all chromosomes, except for chromosome 5. Regarding the flowering time, the proportion of phenotypic variance explained (PVE) by the 16 significant SNPs ranged from 3.52 to 9.58%. To determine the relationship between the average of the three years and a single year, we performed a GWAS of the other three phenotypes (FT2019,

[1]http://eggnog5.embl.de/#/app/home

[2]https://blast.ncbi.nlm.nih.gov/

TABLE 1 | Summary statistics of the PTUs of the flowering time in the GWAS population in 2019, 2020, and 2021.

Year	n	Mean	Range	CV (%)	Kurtosis	Skewness	*P* value	H^2
2019	220	204.21	131.81-271.01	11.30	0.404	0.107	<0.01	0.55
2020	220	168.29	126-254	13.70	0.308	0.662	<0.01	0.59
2021	220	174.98	98.98-291.40	13.10	−0.15	0.597	<0.01	0.51

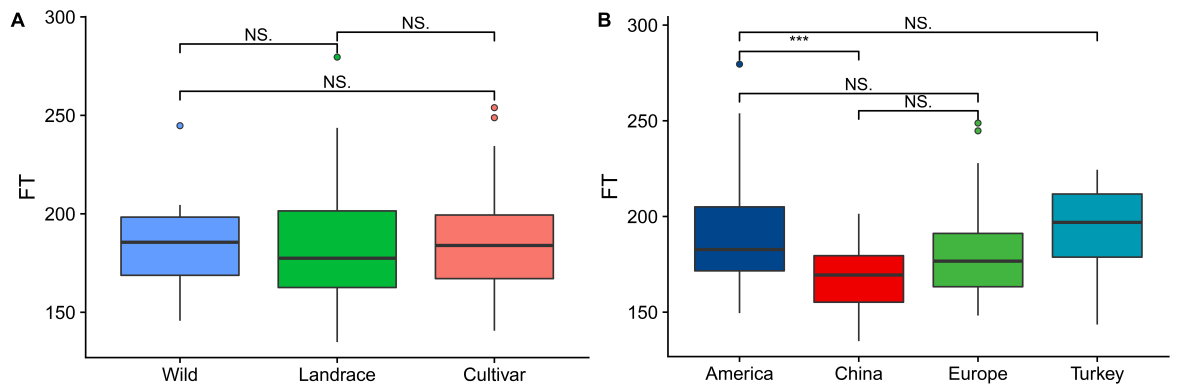


FIGURE 1 | Boxplots showing the variation in the flowering time evaluated in each association panel. **(A)** Boxplot of the flowering time of plants in the subgroups (wild, landrace, and cultivar groups) according to the breeding status. **(B)** Boxplot of the flowering time of plants in the subgroups (America, China, Europe, and Turkey) according to geographical origin. The different asterisks above the boxplots in **(A,B)** indicate significant differences at the P < 0.05 level according to Duncan's multiple comparison test.

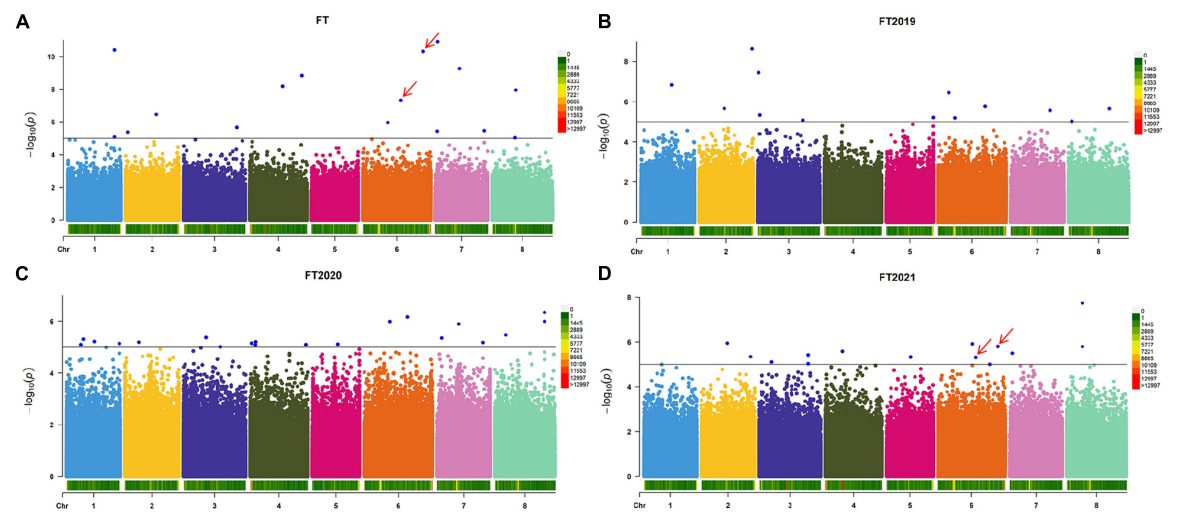


FIGURE 2 | Manhattan plot of the flowering time in different years. **(A)** Manhattan plot of the average flowering time over three years. **(B)** Manhattan plot of the flowering time in 2019. **(C)** Manhattan plot of the flowering time in 2020. **(D)** Manhattan plot of the flowering time in 2021. The GWAS was performed by BLINK C software, and the threshold for significantly associated loci was an LOD score ≥5 (blue line). The red arrow indicates the SNP sites co-located with two years of phenotypic data.

FT2020, and FT2021) to obtain more association data. We detected 13 markers showing a significant association with the FT2019 phenotype, among which chr2__80931183 was the most significant ($-\log p > 8.63$), explaining 7.60% of the variation (**Figure 2B** and **Supplementary Table 1**). We detected 20 markers showing a significant association with the FT2020 phenotype (**Figure 2C** and **Supplementary Table 1**). We detected 14 markers showing a significant association with the FT2021 phenotype, among which chr8__23871223 was the most significant ($-\log p > 7.74$) and explained 15.63% of the variation, and the proportion of PVE ranged from 7.58 to 15.63% (**Figure 2D** and **Supplementary Table 1**). A comparison of the markers detected in the different phenotypes showed that the two markers, namely, chr6__58766748 and chr6__92897252, were associated with a two-year phenotype. These two markers were detected in FT and FT2021 and explained 8% of the phenotypic

variation, indicating that these markers have a major effect on the alfalfa flowering time. The SNP effect size ranged between 19.25 and 31.42.

To observe the effects of different alleles on the flowering time, we selected the eight SNPs with the highest threshold in FT for observations. Raincloud plots of the genotypes of the eight SNPs associated with the flowering time were constructed (**Figure 3**). The differences in the flowering time among the different genotypes can be observed in the eight raincloud plots of the SNPs. There is no homozygous genotype *G/G* for SNPs chr1__71471830 or chr4__50194795, and there is no homozygous genotype *C/C* for SNP chr7__37422136. In addition, there is no homozygous genotype *T/T* for chr8__35962968, chr4__80021920 or chr6__58766748. Among the 8 SNPs, 37.5% (3/8) of the *G/G* genotype was related to early flowering, 50% (4/8) of the genotypes were *T/T*, and flowering occurred relatively late. Therefore, we speculate that these two genotypes may be related to the flowering time of alfalfa. To avoid missing any associations, we selected all 63 significant markers detected in the various phenotypes for further characterization. We used the Zhongmu-4 reference genome and identified all genes within 40 kb (20 kb upstream and downstream) of significant loci according to the LD of the association panel. We compared the sequences of the identified genes with the *A. thaliana* genome sequence to identify their orthologues based on sequence similarity and ultimately identified 95 genes related to growth and development for the subsequent analysis (**Supplementary Table 2**).

The gene annotations showed that the greatest percentage of the candidate genes (40%) had unknown functions, but there were large percentages of genes involved in development (23%) or pollen tube growth (13%) (**Figure 4**). Among the candidate genes, it has been reported that the orthologues of *Msa0041180* and *Msa0041190* regulate plant development (Bekh-Ochir et al., 2013; Ramón et al., 2020). The other genes are involved in pollen tube development and influence flowering. For example, *Msa1016440* (*AT5G41990*) and *Msa1027410* (*AT4G35420*) have been shown to influence flowering during plant development (Tang et al., 2009; Tsuchiya and Eulgem, 2010). In addition, several genes were involved in the auxin pathway, protein modification, and binding proteins, each with a percentage of 2%. The relatively large percentage of genes involved in development and pollen tube growth indicates that our experiments successfully identified genes critical for the alfalfa flowering time.

RNA-Seq With Respect to the Flowering Time

We collected flower buds from two samples at three growth and development stages, constructed a cDNA library and sequenced the transcriptome. The resulting datasets were named CZ_S1 (Cangzhou alfalfa at the bud stage); CZ_S2 (Cangzhou alfalfa at the initial-flowering stage); CZ_S3 (Cangzhou alfalfa at the full-flowering stage); ZM_S1 (Zhongmu No.1 at the bud stage); ZM_S2 (Zhongmu No.1 at the initial-flowering stage); and ZM_S3 (Zhongmu No.1 at the full-flowering stage). In total, 1,924, 2,405, and 3,779 DEGs were identified as upregulated at the bud, initial-flowering, and full-flowering stages, respectively, and 1651, 2613 and 4730 DEGs were downregulated, respectively (**Figure 5A** and **Supplementary Tables 3-8**). Of these DEGs during the three periods, 1585 were common during the flowering period (**Figure 5B**). An overview of the expression of these DEGs is shown in a heatmap, and their expression could be approximately divided into five patterns (**Figure 5C**). The Gene Ontology (GO) analysis showed that these DEGs (**Supplementary Figure 4**) and three

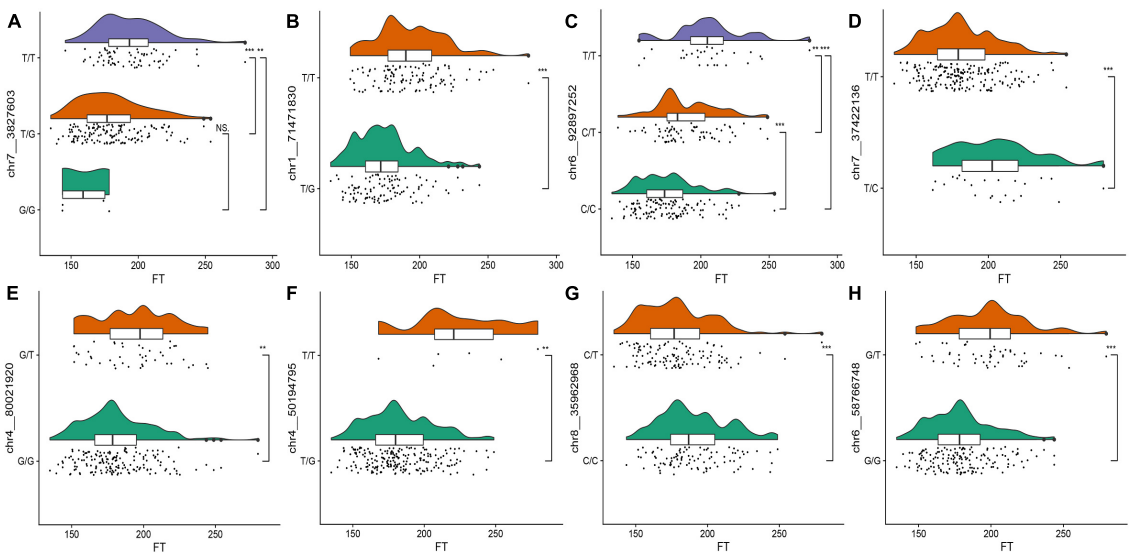


FIGURE 3 | (A–H) Raincloud plots of the highest distribution of the flowering time of the plants with relevant SNP genotypes. The top plots represent the kernel density estimation, the middle plots represent box diagrams, and the bottom plots are dithering scatter diagrams. The different asterisks on the right side of the diagram indicate significant differences at the $P < 0.05$ level according to Duncan's multiple comparison test. The abscissa represents flowering time and the ordinate represents SNP genotype.

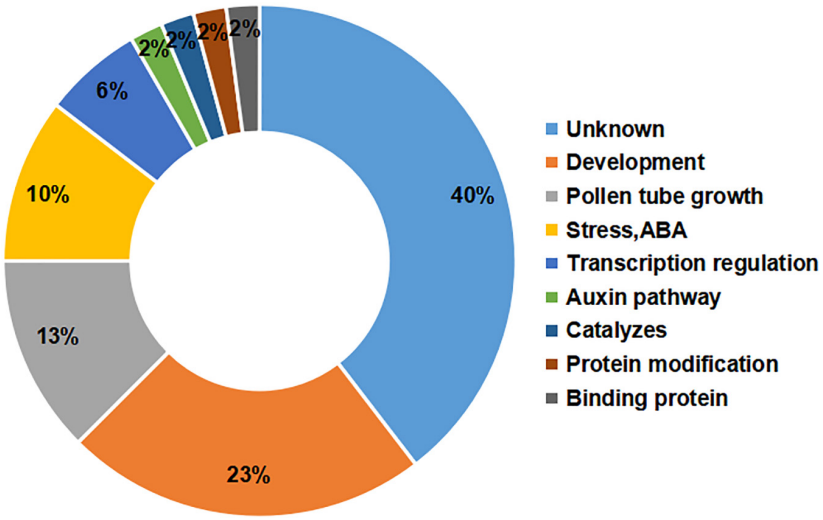


FIGURE 4 | Functional annotations of candidate genes. The number on the left represents the percentage of candidate genes. The text on the right represents the functional classification of the genes.

periods (**Supplementary Figures 5–7**) could be divided into the following three GO categories: biological processes, cellular components and molecular functions. In the biological process category, the two largest subcategories during the bud stage were cellular process and single-organism process, and during the initial flowering and full-flowering stages, the DEGs were mostly associated with cellular processes and metabolic processes. In the cellular components category, the two largest subcategories were cell and cell part during all stages. And the two largest subcategories of the molecular function category were catalytic activity and binding.

Candidate Gene Analysis of the Flowering Time in Alfalfa

We combined the results of the Genome-Wide Association Studies and Differentially Expressed Genes analysis to further analyze the genes associated with the flowering time in alfalfa. According to the LD, in total, 95 candidate genes were found in 39 significant SNPs (17 genes were found simultaneously near different SNP sites) (**Supplementary Table 9**). Among the 95 genes associated with the flowering time, 7 genes were differentially expressed during the bud stage, 13 genes were differentially expressed during the initial flowering stage, and 18 genes were differentially expressed during the full flowering stage. Most significant SNPs were located in intergenic spacer regions. To further identify the candidate genes for flowering time in alfalfa, the DEGs closest to the significantly associated SNPs were analyzed by candidate gene association mapping. According to this criterion, 3 DEGs were included in the subsequent analysis. To verify the accuracy of these results, primers were designed for these three genes for qRT–PCR verification (**Supplementary Table 10**). The results show that the qRT–PCR results are consistent with the RNA-seq results. The expression levels of these three genes gradually increased in the early flowering cultivars and were relatively decreased in the late-flowering cultivars, indicating that these three genes may be closely related to the growth and development of plants (**Figure 6**).

Among the three candidate genes, *Msa0888690*, which encodes a protein involving isoprene, was selected for analysis because there were enough SNP sites around this gene. This gene is highly homologous (59.2% identical) to the *A. thaliana AT2G37540* gene, which is involved in the regulation of the development of plant pollen tubes. In addition, we found that two SNPs (chr6__53758710 and chr6__53758715) located approximately 30 kb upstream of the *Msa0888690* gene were significantly related to the flowering time (**Figure 7A**). The two significant SNPs in the upstream region of *Msa0888690* existed as four different haplotypes in the association panel (Hap1-Hap4) (**Figure 7B**). Among the four haplotypes, Hap2 (*T/C*, *T/T*) had the lowest frequency; it was the first to bloom, followed by Hap4 (*T/C*, *T/C*) and Hap1 (*T/T*, *T/C*), and the latest was Hap3 (*T/T*, *T/T*) (**Figure 7C**). Regarding the different geographical origins, the frequency of Hap3 was increased among the American, European and Chinese varieties. In addition, Hap1 was not found among the Turkish varieties, and its frequency gradually increased among the other varieties (**Figure 7D**). This finding suggests that Hap3 and Hap1 may be under selection in alfalfa breeding.

DISCUSSION

In plants, the flowering time is an important characteristic that affects yield and quality. The identification of genetic loci for flowering time traits is helpful for clarifying the genetic basis of the growth period, which is of great value for cultivars suitable for different geographical regions. Currently, association analyses and linkage analyses are the two most commonly used methods for locating genes underlying complex traits, and these methods have been widely used in plant breeding. Linkage analyses can

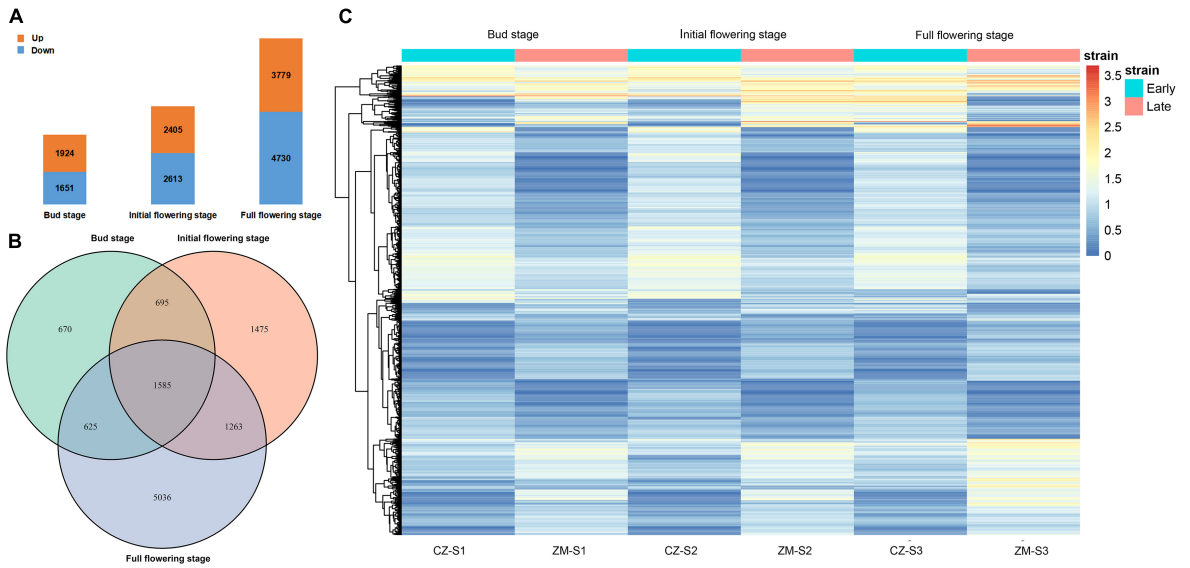


FIGURE 5 | Transcriptomic analysis. **(A)** Number of upregulated genes (orange) and downregulated genes (blue) during alfalfa flowering at three different periods. **(B)** Upregulated and downregulated DEGs in the three periods. **(C)** Heatmap clustering of all DEGs according to their expression level. The different colors indicate different levels of expression. The blue box above represents the early flowering varieties, and red represents the late-flowering varieties.

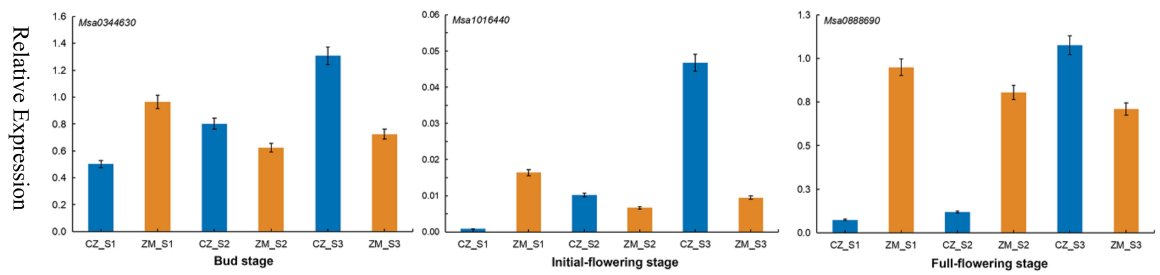


FIGURE 6 | qRT–PCR analysis of three candidate genes in three stages associated with the flowering time in alfalfa.

be used to eliminate false positives of related loci. Indeed, several studies have investigated the genetic basis of alfalfa flowering (Adhikari et al., 2019; Zhang et al., 2020). However, the intervals between positions are usually large in alfalfa, and association analyses can greatly reduce these intervals. Therefore, accuracy and efficiency can be improved by the combination of these two methods. These studies are important but did not provide enough information to determine the mechanisms underlying the alfalfa flowering time. In this study, we used 875,023 high-quality, high-density SNP markers to analyze the genetic basis of the flowering time via a GWAS of 220 alfalfa materials. We detected numerous significantly associated loci and genes. In addition, through the combination of a transcriptomic analysis and GWAS, several candidate genes related to the flowering time were analyzed. Our results revealed many promising genes for further studies investigating plant growth and development and for breeding new alfalfa varieties suitable for harvesting.

Mastering the proper flowering time can maximize economic benefits (Koornneef et al., 1998a). However, we still do not understand the mechanism underlying the flowering time, and research investigating the influence of the flowering period on alfalfa varieties in different geographical regions and improvement in breeding conditions is limited. The results of the present study showed that there were no significant differences in the flowering time of alfalfa varieties in different geographical regions. However, compared with the varieties from other geographical areas, there are significant differences between Chinese varieties and American varieties. The flowering of Chinese varieties occurs obviously earlier than that of American varieties. The reason for this difference is not only due to climatic conditions but also closely related to the germplasm history. Chinese alfalfa varieties have been cultivated for more than 2000 years and have continuously experienced domestication selection, resulting in several local varieties (Annicchiarico et al., 2015) that are suitable for different ecological environments. In contrast, the genetic diversity of American germplasm materials is low, which may be due to the short introduction time of alfalfa in the United States, which was introduced approximately 200∼300 years ago (Prosperi et al., 2014). The short cultivation history and low diversity of germplasm materials at the time of

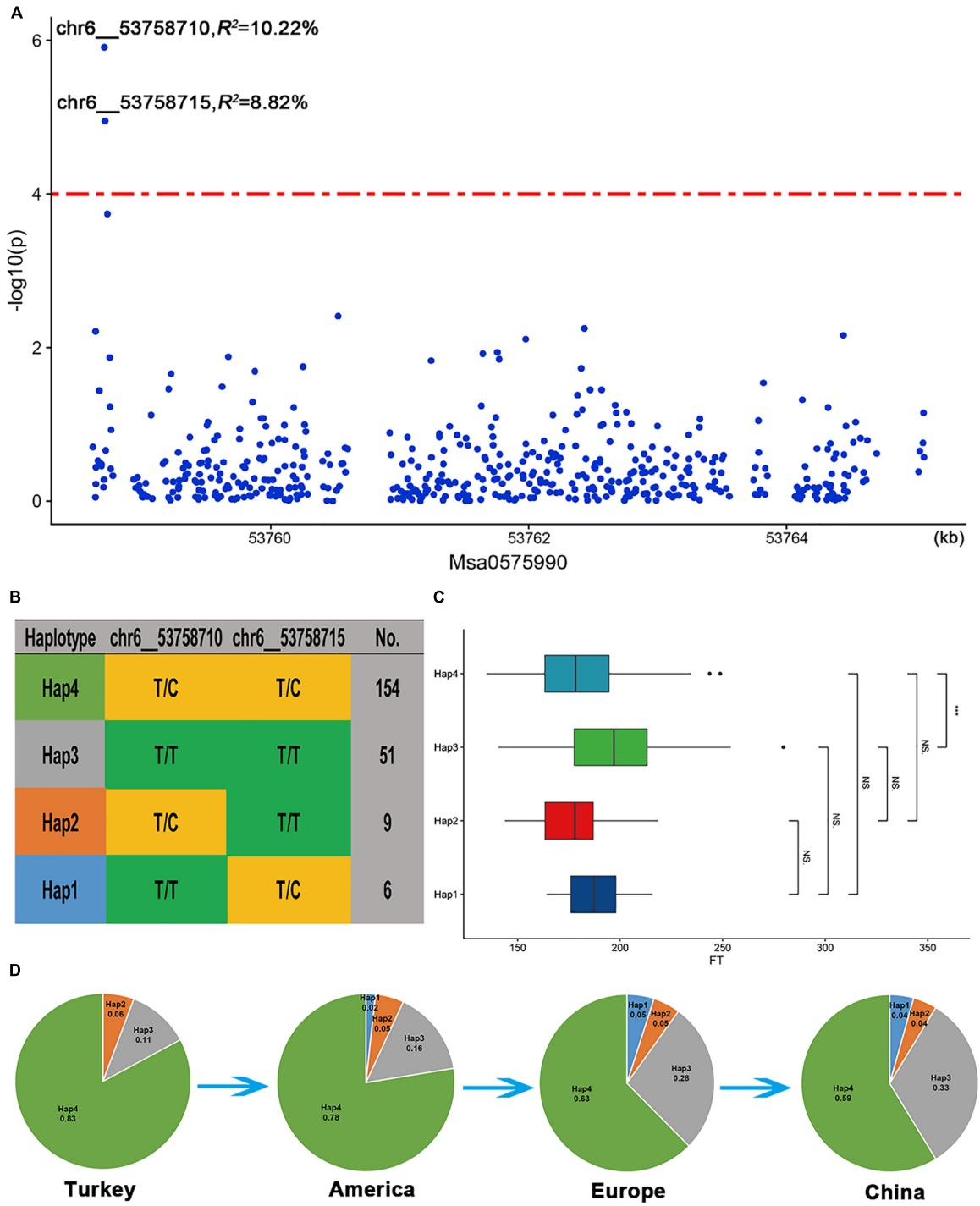


FIGURE 7 | *Msa0888690* involvement in the flowering time in alfalfa. **(A)** Association analysis of SNPs located within the *Msa0888690* gene region. **(B)** Haplotype analysis of significant SNPs associated with the flowering time. **(C)** Comparison of the flowering time among the four haplotypes. A boxplot of the four haplotypes is shown on the left, and the symbols represent the correlations. **(D)** Allele frequency of the four haplotypes in the Turkish, American, European and Chinese subgroups.

introduction may lead to great differences in the flowering time between American varieties and Chinese varieties. Our previous research results show that the decay of LD with distance varied as a function of geographical origin: The LD decay distance of American alfalfa varieties is 40.1 kb, and that of Chinese varieties is 17.7 kb (Chen et al., 2021), which further indicates that the breeding improvement status and geographical origin may affect the allelic diversity and Germplasm history of alfalfa.

However, research concerning the flowering time is very scarce, and there is no reference genome for alfalfa; thus, the important markers are aligned only with the reference genome (*M. truncatula*). Therefore, we are concerned that markers associated with important traits may be lost. In the past two years, several alfalfa reference genomes have been released, which has made it possible to mine candidate genes related to flowering traits based on GWAS results (Chen H. et al., 2020; Shen et al., 2020; Long et al., 2022). In this study, we used the average flowering time as an indicator to investigate the flowering period in alfalfa association populations and used three-year data. In total, 67 significant SNP markers were identified in this study. Among these significant SNP makers, we found some SNPs have the higher LOD value but with lower explained variance for example the SNP (chr6__72110460). In general, if a QTL or significant SNP (GWAS) which have a higher LOD value may also have a higher explained variance. The reason for this results in our study maybe that this SNP maybe is false positive and need to be confirmed in the future study. In addition, two SNPs were colocalized in the two environments. Our results revealed a small number of identical SNPs across different phenotypes, which may be due to the large variation in temperature and different day lengths (average day length over three years differed by nearly an hour, approximately 40 min) influencing the alfalfa flowering time during the three years or the absence of major flowering-related alleles in alfalfa. The number of important SNP markers detected by our GWAS indicates that this phenotype can be effectively used to study the flowering period of alfalfa. Moreover, 95 candidate genes were located in 67 markers. Therefore, most markers we found were detected by our GWAS at a single-gene resolution. The annotation analyses of these candidates identified potential mechanisms for the genetic regulation of the alfalfa flowering time. Of these candidate genes, 23% play a role in plant growth and development, and 13% are involved in pollen tube growth (**Figure 4**). Therefore, these genes may be the main genes that cause flowering time changes in related populations. According to a previous report, plant hormones, such as auxin, cytokinin (CK), abscisic acid (ABA), gibberellic acid (GA), salicylic acid (SA), and jasmonic acid (JA), are involved in the floral transition process and plant flowering (Martínez et al., 2004; Davis, 2009; Bartrina et al., 2011). In our study, 10% of the candidate genes were involved in the response to stress and ABA, suggesting that these processes are also associated with important mechanisms regulating the growth and development of the plants composing this population.

Transcriptomic analyses have been shown to be a powerful tool for identifying genes related to plant growth and development. In recent years, RNA-seq technology has also been used to study alfalfa, and many genes related to important traits have been identified, such as traits associated with defoliation (Cheng et al., 2018), drought stress (Arshad et al., 2018), and salt stress (Postnikova et al., 2013). In the present study, 1,924, 2,405, and 3,779 DEGs were upregulated at different flowering stages, and 1,651, 2,613, and 4,730 DEGs were downregulated. These genes can help us understand the molecular mechanism of alfalfa flowering. In addition, we identified three DEGs in three different developmental stages, all of which play a role in flower growth and development. For example, *Msa1016440* (*AT5G41990*), encoding ATWNK8, regulates the flowering time by regulating the photoperiodic pathway in *A. thaliana* (Wang et al., 2008). ABA is a sesquiterpenoid phytohormone that plays crucial roles in plant development, growth, and responses to stresses (Chen K. et al., 2020; Nonogaki, 2020). Phaseic acid (PA), the main catabolite of ABA, is structurally related to ABA and exhibits ABA-like hormone activity (Nambara and Marion-Poll, 2005). In *A. thaliana* plants overexpressing Cinnamoyl-coA: NADP oxidoreductase-like 1 (CRL1), which encodes a PA reductase that catalyses the formation of dihydrophaseic acid from PA, both the seed germination rate and flowering time were increased (Yin et al., 2022). *Msa0344630* also encodes cinnamoyl-coA, and its expression was significantly increased according to the RNA-seq data. Therefore, we speculate that this gene can also promote the early flowering of plants. This finding shows that these DEGs play an important role in alfalfa flowering.

Isoprenoids constitute the largest class of natural plant products and perform diverse biological functions, including functions in plant growth and development (Lange et al., 2020). Through the overexpression of isoprenoid homologue 3-hydroxy-3-methylglutaryl CoA reductase (HMGR) in *Solanum tuberosum*, it was found that the flowering time was advanced, and biomass was increased. Here, we also found that *Msa0888690*, which encodes a protein involving isoprene, was significantly associated with the flowering time of alfalfa. The expression of *Msa0888690* in alfalfa flowers was significantly upregulated at different flowering stages. Interestingly, we found four haplotypes at this location, with Hap3 flowering the latest and increasing in frequency in different geographical origins, and among the varieties in China, the proportion of Hap3 was the greatest. Hap1 germplasm was not present among the Turkish cultivars but gradually increased in frequency in several other cultivars with different geographical origins. Therefore, we hypothesize that Hap3 and Hap1 may be among the targets selected during the domestication of alfalfa. The generation of transgenic alfalfa with *Msa0888690* knocked out via miR156 or *Msa0888690* overexpressed could facilitate the elucidation of its functions in growth and development. The genetic materials, candidate genes, and haplotypes identified in our study may potentially be used for cultivating alfalfa varieties suitable for local needs.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly avlailable. RNA-seq data of the flowers at three development periods of two species of alfalfa grown in the field have been submitted to the NCBI Sequence Read Archive (BioProject: PRJNA817856).

AUTHOR CONTRIBUTIONS

JK, RL, and QY planned and designed the research. FH, FZ, and LC analyzed the data and wrote the manuscript. ML, XJ, YC, TG, TY, and CW performed the field work and collected the

phenotypic data. ZW, RL, and LC supervised the research. FH and FZ contributed equally. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (31971758), the National Natural Science Foundation of China (32071868), the Agricultural Science and Technology Innovation Program (ASTIP-IAS14), the Central Public-interest Scientific Institution Basal Research Fund (No. 2022-YWF-ZYSQ-04), and the Agricultural Science and Technology Innovation Program (ASTIP No. CAAS-ZDRW202201).

ACKNOWLEDGMENTS

We thank all colleagues in our laboratory for providing useful discussions and technical assistance. We are very grateful to the editor and reviewers for critically evaluating the manuscript and providing constructive comments for its improvement.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.913947/full#supplementary-material

Supplementary Figure 1 | Normal distribution of the average values of the flowering time over three years among the association panel.

Supplementary Figure 2 | Normal distribution of the average values of the flowering time in a single year among the association panel.

Supplementary Figure 3 | Correlations between different years. The correlation coefficient (R) and P value are presented at the top of the figure. **(A)** Correlation information in 2019 and 2020. **(B)** Correlation information in 2019 and 2021. **(C)** Correlation information in 2020 and 2021. The gray part represents the 95% confidence interval.

Supplementary Figure 4 | GO analysis of all DEGs.

Supplementary Figure 5 | GO analysis of the bud stage.

Supplementary Figure 6 | GO analysis of the initial flowering stage.

Supplementary Figure 7 | GO analysis of the full-flowering stage.

Supplementary Table 1 | Significant SNPs identified via BLINK.

Supplementary Table 2 | Genes associated with growth and development identified by sequence similarity.

Supplementary Table 3 | Upregulated DEGs in the bud stage under normal growth conditions.

Supplementary Table 4 | Downregulated DEGs in the bud stage under normal growth conditions.

Supplementary Table 5 | Upregulated DEGs in the initial flowering stage under normal growth conditions.

Supplementary Table 6 | Downregulated DEGs in the initial flowering stage under normal growth conditions.

Supplementary Table 7 | Upregulated DEGs in the full-flowering stage under normal growth conditions.

Supplementary Table 8 | Downregulated DEGs in the full-flowering stage under normal growth conditions.

Supplementary Table 9 | Candidate genes identified via GWAS and a transcriptomic analysis.

Supplementary Table 10 | Primers used in this study.

REFERENCES

Adhikari, L., Makaju, S. O., and Missaoui, A. M. (2019). QTL mapping of flowering time and biomass yield in tetraploid alfalfa (*Medicago sativa* L.). *BMC Plant Biol.* 19:359. doi: 10.1186/s12870-019-1946-0

Annicchiarico, P., Barrett, B., Brummer, E. C., Julier, B., and Marshall, A. H. (2015). Achievements and challenges in improving temperate perennial forage legumes. *Crit. Rev. Plant Sci.* 34, 327–380.

Arshad, M., Gruber, M. Y., and Hannoufa, A. (2018). Transcriptome analysis of microRNA156 overexpression alfalfa roots under drought stress. *Sci. Rep.* 8, 1–13. doi: 10.1038/s41598-018-27088-8

Arzani, H., Zohdi, M., Fish, E., Amiri, G. H. Z., Nikkhah, A., and Wester, D. (2004). Phenological effects on forage quality of five grass species. *J. Range Manage.* 57, 624–629.

Baggerly, K. A., Deng, L., Morris, J. S., and Aldaz, C. M. (2003). Differential expression in SAGE: accounting for normal between-library variation. *Bioinformatics* 19, 1477–1483. doi: 10.1093/bioinformatics/btg173

Bartrina, I., Otto, E., Strnad, M., Werner, T., and Schmülling, T. (2011). Cytokinin regulates the activity of reproductive meristems, flower organ size, ovule formation, and thus seed yield in *Arabidopsis thaliana*. *Plant Cell* 23, 69–80. doi: 10.1105/tpc.110.079079

Battenfield, S. D., Sheridan, J. L., Silva, L. D. C. E., Miclaus, K. J., Dreisigacker, S., Wolfinger, R. D., et al. (2018). Breeding-assisted genomics: applying meta-GWAS for milling and baking quality in CIMMYT wheat breeding program. *PLoS One* 13:e0204757. doi: 10.1371/journal.pone.0204757

Bekh-Ochir, D., Shimada, S., Yamagami, A., Kanda, S., Ogawa, K., Nakazawa, M., et al. (2013). A novel mitochondrial DnaJ/Hsp40 family protein BIL2 promotes plant growth and resistance against environmental stress in brassinosteroid signalling. *Planta* 237, 1509–1525. doi: 10.1007/s00425-013-1859-3

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 57, 289–300.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.

Chao, Y. H., Zhang, T. J., Yang, Q. C., Kang, J. M., Sun, Y., Gruber, M. Y., et al. (2014). Expression of the alfalfa CCCH-type zinc finger protein gene MsZFN delays flowering time in transgenic *Arabidopsis thaliana*. *Plant Sci.* 215, 92–99. doi: 10.1016/j.plantsci.2013.10.012

Chen, H., Zeng, Y., Yang, Y., Huang, L., and Qiu, Q. (2020). Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. *Nat. Commun.* 11:2494. doi: 10.1038/s41467-020-16338-x

Chen, K., Li, G. J., Bressan, R. A., Song, C. P., Zhu, J. K., and Zhao, Y. (2020). Abscisic acid dynamics, signalling, and functions in plants. *J. Integr. Plant Biol.* 62, 25–54.

Chen, L., He, F., Long, R. C., Zhang, F., Li, M., Wang, Z., et al. (2021). A global alfalfa diversity panel reveals genomic selection signatures in Chinese varieties and genomic associations with root development. *J. Integr. Plant Biol.* 63, 1937–1951. doi: 10.1111/jipb.13172

Cheng, Q. M., Bai, S. Q., Ge, G. T., Li, P., Liu, L. Y., Zhang, C. D., et al. (2018). Study on differentially expressed genes related to defoliation traits in two alfalfa varieties based on RNA-Seq. *BMC Genomics* 19:807. doi: 10.1186/s12864-018-5180-1

Davis, S. J. (2009). Integrating hormones into the floral-transition pathway of *Arabidopsis thaliana*. *Plant Cell Environ.* 32, 1201–1210. doi: 10.1111/j.1365-3040.2009.01968.x

Gao, R., Gruber, M. Y., and Amyot, L. A. (2018). Hannoufa, SPL13 regulates shoot branching and flowering time in *Medicago sativa*. *Plant Mol. Biol.* 96, 119–133. doi: 10.1007/s11103-017-0683-8

Grabowski, P. P., Evans, J., Daum, C., Deshpande, S., Barry, K. W., Kennedy, M., et al. (2017). Genome-wide associations with flowering time in switchgrass using exome-capture sequencing data. *New Phytol.* 213, 154–169. doi: 10.1111/nph.14101

Grillo, M. A., Li, C., Hammond, M., Wang, L. J., and Schemske, D. W. (2013). Genetic architecture of flowering time differentiation between locally adapted populations of *Arabidopsis thaliana*. *New Phytol.* 197, 1321–1331. doi: 10.1111/nph.12109

He, F., Kang, J. M., Zhang, F., Long, R. C., Yu, L. X., Wang, Z., et al. (2019). Genetic mapping of leaf-related traits in autotetraploid alfalfa (*Medicago sativa* L.). *Mol. Breed.* 39, 1–12.

Hecht, V., Foucher, F., Ferrándiz, C., Macknight, R., Navarro, C., Morin, J., et al. (2005). Conservation of Arabidopsis flowering genes in model legumes. *Plant Physiol.* 137, 1420–1434.

Hori, K., Matsubara, K., and Yano, M. (2016). Genetic control of flowering time in rice: integration of *Mendelian* genetics and genomics. *Theor. Appl. Genet.* 129, 2241–2252. doi: 10.1007/s00122-016-2773-4

Huang, M., Liu, X., Zhou, Y., Summers, R. M., and Zhang, Z. (2019). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* 8:giy154. doi: 10.1093/gigascience/giy154

Huang, X. H., Zhao, Y., Wei, X. H., Li, C. Y., Wang, A., Zhao, Q., et al. (2012). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat. Genet.* 44, 32–39. doi: 10.1038/ng.1018

Ivanièová, Z., Jakobson, I., Reis, D., Šafáø, J., and Milec, Z. (2016). Characterization of new allele influencing flowering time in bread wheat introgressed from *Triticum militinae*. *New Biotech* 33, 718–727. doi: 10.1016/j.nbt.2016.01.008

Kang, J. M., Zhang, T. J., Guo, T., Ding, W., Long, R. C., Yang, Q. C., et al. (2019). Isolation and functional characterization of msfta, a flowering locus t homologue from alfalfa (*Medicago sativa* L.). *Int. J. Mol. Sci.* 20:1968. doi: 10.3390/ijms20081968

Kim, K. H., Kim, J. Y., Lim, W. J., Jeong, S., Lee, H. Y., Cho, Y., et al. (2020). Genome-wide association and epistatic interactions of flowering time in soybean cultivar. *PLoS One* 15:e0228114. doi: 10.1371/journal.pone.0228114

Koornneef, M., Alonso-Blanco, C., Blankestijn-de, V. H., and Hanhart, C. J. (1998a). Genetic interactions among late-flowering mutants of *Arabidopsis*. *Genetics* 148, 885–892. doi: 10.1093/genetics/148.2.885

Koornneef, M., Alonso-Blanco, C., Peeters, A. J., and Soppe, W. (1998b). Genetic control of flowering time in *Arabidopsis*. *Annu. Rev. Plant Biol.* 49, 345–370.

Lange, I., Lange, B. M., and Navarre, D. A. (2020). Altering potato isoprenoid metabolism increases biomass and induces early flowering. *J. Exp. Bot.* 71, 4109–4124. doi: 10.1093/jxb/eraa185

Laurie, R. E., Diwadkar, P., Jaudal, M., Zhang, L. L., Hecht, V., Wen, J. Q., et al. (2011). The *Medicago* flowering locus t homologue, mtfta1, is a key regulator of flowering time. *Plant Physiol.* 156, 2207–2224.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079.

Lin, S., Medina, C. A., Boge, B., Hu, J., Fransen, S., Norberg, S., et al. (2020). Identification of genetic loci associated with forage quality in response to water deficit in autotetraploid alfalfa (*Medicago sativa* L.). *BMC Plant Biol.* 20:303. doi: 10.1186/s12870-020-02520-2

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

Long, R., Zhang, F., Zhang, Z., Li, M., Chen, L., Wang, X., et al. (2022). Assembly of chromosome-scale and allele-aware autotetraploid genome of the Chinese alfalfa cultivar Zhongmu-4 and identification of SNP loci associated with 27 agronomic traits. *Genomics Proteomics Bioinformatics* doi: 10.1101/2021.02.21.428692

Lyons, R., Rusu, A., Stiller, J., Powell, J., Manners, J. M., and Kazan, K. (2015). Investigating the association between flowering time and defence in the *Arabidopsis thaliana-Fusarium oxysporum* interaction. *PLoS One* 10:e0127699. doi: 10.1371/journal.pone.0127699

Martínez, C., Pons, E., Prats, G., and León, J. (2004). Salicylic acid regulates flowering time and links defence responses and reproductive development. *Plant J.* 37, 209–217. doi: 10.1046/j.1365-313x.2003.01954.x

Mortazavi, A., Williams, B. A., Mccue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628.

Nambara, E., and Marion-Poll, A. (2005). Abscisic acid biosynthesis and catabolism. *Annu. Rev. Plant Biol.* 56, 165–185.

Navarro, J. A. R., Willcox, M., Burgueño, J., Romay, C., Swarts, K., Trachsel, S., et al. (2017). A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat. Genet.* 49, 476–480.

Nonogaki, H. (2020). A repressor complex silencing ABA signalling in seeds? *J. Exp. Bot.* 71, 2847–2853.

Patil, G., Do, T., Vuong, T. D., Valliyodan, B., Lee, J. D., Chaudhary, J., et al. (2016). Genomic-assisted haplotype analysis and the development of high-throughput SNP markers for salinity tolerance in soybean. *Sci. Rep.* 6, 1–13. doi: 10.1038/srep19199

Postnikova, O. A., Shao, J., and Nemchinov, L. G. (2013). Analysis of the alfalfa root transcriptome in response to salinity stress. *Plant Cell Physiol.* 54, 1041–1055.

Prosperi, J. M., Jenczewski, E., Muller, M. H., Fourtier, S., Sampoux, J. P., and Ronfort, J. (2014). Alfalfa domestication history, genetic diversity and genetic resources. *Legume Per.* 4, 13–14. doi: 10.1111/j.1365-294X.2006.02851.x

Ramón, P. F., Edith, M. P., Salvador, B. O., Randy, O. C., Jorge, S. M., María, A. O. A., et al. (2020). The cysteine-rich receptor-like protein kinase CRK28 modulates Arabidopsis growth and development and influences abscisic acid responses. *Planta* 251, 1–12. doi: 10.1007/s00425-019-03296-y

Ridge, S., Deokar, A., Lee, R., Daba, K., Macknight, R. C., Weller, J. L., et al. (2017). The chickpea early flowering 1 (Efl1) locus is an Orthologue of *Arabidopsis* ELF3. *Plant Physiol.* 175, 802–815. doi: 10.1104/pp.17.00082

Saxena, M. S., Bajaj, D., Das, S., Kujur, A., Kumar, V., Singh, M., et al. (2014). An integrated genomic approach for rapid delineation of candidate genes regulating agro-morphological traits in chickpea. *DNA Res.* 21, 695–710. doi: 10.1093/dnares/dsu031

Shen, C., Du, H., Chen, Z., Lu, H., Zhu, F., Chen, H., et al. (2020). The chromosome-level genome sequence of the autotetraploid alfalfa and resequencing of core germplasms provide genomic resources for alfalfa research. *Mol. Plant.* 13, 1250–1261. doi: 10.1016/j.molp.2020.07.003

Tang, L. K., Chu, H., Yip, W. K., Yeung, E. C., and Lo, C. (2009). An anther-specific dihydroflavonol 4-reductase-like gene (DRL1) is essential for male fertility in *Arabidopsis*. *New Phytol.* 181, 576–587. doi: 10.1111/j.1469-8137.2008.02692.x

Tornqvist, C. E., Taylor, M., Jiang, Y., Evans, J., Buell, C. R., Kaeppler, S. M., et al. (2018). Quantitative trait locus mapping for flowering time in a lowland × upland switchgrass pseudo-F2 population. *Plant Genome* 11:169. doi: 10.3835/plantgenome2017.10.0093

Tsuchiya, T., and Eulgem, T. (2010). The Arabidopsis defence component EDM2 affects the floral transition in an FLC-dependent manner. *Plant J.* 62, 518–528. doi: 10.1111/j.1365-313X.2010.04169.x

Wang, Y., Liu, K., Liao, H., Zhuang, C., Ma, H., and Yan, X. (2008). The plant WNK gene family and regulation of flowering time in *Arabidopsis*. *Plant Biol.* 10, 548–562. doi: 10.1111/j.1438-8677.2008.00072.x

Wang, Z., Wang, X., Zhang, H., Ma, L., Zhao, H., Jones, C. S., et al. (2020). A genome-wide association study approach to the identification of candidate

genes underlying agronomic traits in alfalfa (*Medicago sativa* L.). *Plant Biotechnol. J.* 18, 611. doi: 10.1111/pbi.13251

Yin, X. M., Bai, Y. L., Ye, T. T., Yu, M., Wu, Y., and Feng, Y. Q. (2022). Cinnamoyl coA: NADP oxidoreductase-like 1 regulates abscisic acid response by modulating phaseic acid homeostasis in *Arabidopsis thaliana*. *J. Exp. Bot.* 73, 860–872. doi: 10.1093/jxb/erab474

Yu, L. X., Liu, X., Boge, W., and Liu, X. P. (2016). Genome-wide association study identifies loci for salt tolerance during germination in autotetraploid alfalfa (*Medicago sativa* L.) using genotyping-by-sequencing. *Front. Plant Sci.* 7:956. doi: 10.3389/fpls.2016.00956

Zhang, T. J., Chao, Y. H., Kang, J. M., Ding, W., and Yang, Q. C. (2013). Molecular cloning and characterization of a gene regulating flowering time from alfalfa (*Medicago sativa* L.). *Mol. Biol. Rep.* 40, 4597–4603. doi: 10.1007/s11033-013-2552-0

Zhang, J. P., Song, Q. J., Cregan, P. B., Nelson, R. L., Wang, X. Z., Wu, J. X., et al. (2015). Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics* 16:217. doi: 10.1186/s12864-015-1441-4

Zhang, T. J., Yu, L. X., Zheng, P., Li, Y. J., Rivera, M., Main, D., et al. (2015). Identification of loci associated with drought resistance traits in heterozygous autotetraploid alfalfa (*Medicago sativa* L.) using genome-wide association studies with genotyping by sequencing. *PLoS One* 10:e0138931. doi: 10.1371/journal.pone.0138931

Zhang, F., Kang, J. M., Long, R. C., Yu, L. X., Sun, Y., Wang, Z., et al. (2020). Construction of high-density genetic linkage map and mapping quantitative trait loci (QTL) for flowering time in autotetraploid alfalfa (*Medicago sativa* L.) using genotyping by sequencing. *Plant Genome* 13, e20045. doi: 10.1002/tpg2.20045

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

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

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 05 August 2022
DOI 10.3389/fpls.2022.975456

OPEN ACCESS

EDITED BY
Jiyu Zhang,
Lanzhou University,
China

REVIEWED BY
Yushan Jia,
Inner Mongolia Agricultural University,
China
Qinghua Qiu,
Jiangxi Agricultural University,
China

*CORRESPONDENCE
Min Wang
mwang@isa.ac.cn

SPECIALTY SECTION
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

RECEIVED 22 June 2022
ACCEPTED 19 July 2022
PUBLISHED 05 August 2022

CITATION
Yi S, Zhang X, Zhang J, Ma Z, Wang R, Wu D, Wei Z, Tan Z, Zhang B and Wang M (2022) *Brittle Culm 15* mutation alters carbohydrate composition, degradation and methanogenesis of rice straw during *in vitro* ruminal fermentation. *Front. Plant Sci.* 13:975456. doi: 10.3389/fpls.2022.975456

Brittle Culm 15 mutation alters carbohydrate composition, degradation and methanogenesis of rice straw during *in vitro* ruminal fermentation

Siyu Yi[1,2], Xiumin Zhang[1], Jianjun Zhang[1], Zhiyuan Ma[1], Rong Wang[1], Duanqin Wu[3], Zhongshan Wei[4], Zhiliang Tan[1], Baocai Zhang[5] and Min Wang[1,2]*

[1]CAS Key Laboratory of Agro-Ecological Processes in Subtropical Region, National Engineering Laboratory for Pollution Control and Waste Utilization in Livestock and Poultry Production, Institute of Subtropical Agriculture, The Chinese Academy of Sciences, Changsha, Hunan, China, [2]College of Animal Science and Technology, Guangxi University, Nanning, Guangxi, China, [3]Institute of Bast Fiber Crops and Center of Southern Economic Crops, Chinese Academy of Agricultural Sciences, Changsha, China, [4]Institute of Hunan Animal and Veterinary Science, Changsha, Hunan, China, [5]State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, The Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing, China

Brittle Culm 15 (*BC15*) gene encodes a membrane-associated chitinase-like protein that participates in cellulose synthesis, and *BC15* gene mutation affects cell wall composition in plant, such as cellulose or hemicellulose. The present study was designed to investigate the changes of carbohydrates composition in *bc15* mutant straw, and the resulting consequence on rumen fermentation, methanogenesis, and microbial populations (qPCR) during *in vitro* ruminal fermentation process. Two substrates, *bc15* mutant and wild-type (WT) rice straws, were selected for *in vitro* rumen batch culture. The first experiment was designed to investigate the kinetics of total gas and CH_4 production through 48-h *in vitro* ruminal fermentation, while the second experiment selected incubation time of 12 and 48h to represent the early and late stage of *in vitro* ruminal incubation, respectively, and then investigated changes in biodegradation, fermentation end products, and selected representative microbial populations. The *bc15* mutant straw had lower contents of cellulose, neutral detergent fiber (NDF) and acid detergent fiber (ADF), and higher contents of water-soluble carbohydrates, neutral detergent solubles (NDS) and monosaccharides. The *bc15* mutant straw exhibited a distinct kinetics of 48-h total gas and CH_4 production with faster increases in early incubation when compared with WT straw. The *bc15* mutant straw had higher DM degradation, NDF degradation and total volatile fatty acid concentration at 12h of incubation, and lower NDF degradation and CH_4 production at 48h of incubation, together with lower acetate to propionate ratio and ADF degradation and higher butyrate molar percentage and NDS degradation at both incubation times. Furthermore, the *bc15* mutant straw resulted in greater 16S gene copies of *F. succinogenes*, with lower 18S gene copies of fungi at

both incubation times. These results indicated that the *BC15* gene mutation decreased fibrosis of cell wall of rice straw, enhanced degradation at the early stage of rumen fermentation, and shifts fermentation pattern from acetate to propionate and butyrate production, leading to the decreased volume and fractional rate of CH_4 production. However, *BC15* gene mutation may enhance hardenability of cell wall structure of rice straw, which is more resistant for microbial colonization with decreased fiber degradation. Thus, this study modified rice straw by manipulating a cell wall biosynthesis gene and provides a potential strategy to alter degradation and CH_4 production during *in vitro* ruminal fermentation process.

KEYWORDS

rice straw, *Brittle Culm*, methane, hydrogen, rumen fermentation

Introduction

Rice straw is one of the major cropping by-products and can be used as a roughage source for ruminant animals (Zhang et al., 2020). However, rice straw cannot satisfy the maintenance needs of ruminants because of its high fiber content, low protein, and low energy level (Xu et al., 2018; Alemu et al., 2020). Thus, the resource of rice straw has not been effectively utilized although its yield is most abundant, and the import of alfalfa hay and other high-quality forages are increasing annually in China (Xu et al., 2018). Therefore, it has been a hot issue in the animal husbandry to find ways to reduce the fiber content of rice straw and increase its crude protein (CP) and non-fiber carbohydrate levels. In recent years, changing the nutritional components of rice straw has been widely investigated through physical, chemical, and microbial treatments (Zhang et al., 2018, 2020; Oskoueian et al., 2021). However, these technologies are usually limited in terms of cost or environment risk.

Brittle Culm (*BC*) genes are involved in cell wall formation of rice plants, and mutation of *BC* genes can modulate integrity of the cell wall (Li et al., 2019), which is accompanied by the changes of cell wall composition, such as cellulose or hemicellulose. The *Brittle Culm 15* (*BC15*) gene encodes a membrane-associated chitinase-like protein that participates in the synthesis of cellulose (Wu et al., 2012). The *BC15* gene mutation causes a reduction in cellulose content and mechanical strength and increases hemicellulose content without changed in plant growth (Wu et al., 2012). It seems that mutation of *BC15* gene can remodel cell wall through disturbing cellulose synthesis in rice straw, leading to change in cell wall structure and reduction in cellulose level. Such changes in cell wall composition of *bc15* mutant may alter nutritional composition of rice straw in ruminants, which has not been investigated before.

Carbohydrates are degraded to produce volatile fatty acids (VFA) in rumen microbial ecosystem for the host animals, accompanied with hydrogen (H_2) production (Wang et al., 2016a). Ruminal H_2 is mainly used by methanogens to produce methane (CH_4; Janssen, 2010), which represents a loss of dietary energy and contributes to global anthropogenic greenhouse gas emissions (Wang et al., 2021a, 2022). Compared with forage fiber, non-fiber carbohydrates generally exhibit greater rate of fermentation and cause a shift in the VFA fermentation pattern from acetate to propionate or butyrate production, leading to a decrease in efficiency of H_2 and CH_4 production (Ma et al., 2015; Zhang et al., 2020). It seems that mutation of *BC15* gene can alter carbohydrate compositions of rice straw, which may lead to varied rumen fermentation and methanogenesis.

We hypothesized that *BC15* gene mutation possesses increased content of the soluble carbohydrates compared with the wild-type (WT) straw, which would alter rumen fermentation, methanogenesis, and microbial populations during ruminal fermentation process. To test this hypothesis, the *bc15* mutant and WT rice straws were employed for *in vitro* ruminal batch culture. We then firstly measured the kinetics of total gas and CH_4 production through 48-h *in vitro* ruminal fermentation. Incubation time of 12 and 48 h were then selected to represent the early and late stage of *in vitro* ruminal incubation, respectively, and the second experiment was designed to investigate changes in biodegradation, fermentation end products, and selected representative microbial populations.

Materials and methods

The experiment was approved (No. ISA W202101) by the Animal Care Committee, Institute of Subtropical Agriculture, the Chinese Academy of Sciences, Changsha, China.

Substrates of *bc15* mutant and wild-type rice straw

The *bc15* mutant rice was isolated from tissue-cultured plants generated from the callus of the rice japonica cultivar Zhonghua

8 as described by Wu et al. (2012). These two genotypes of rice used in this study were grown in the experimental fields in Lingshui (Hainan Province, China) during the natural growing seasons. The wild-type (Zhonghua 8) and *bc15* mutant rice straw were harvested from mature plants after removing the grains under the same conditions and growing period.

The rice straws (not crushed) were dried in the sun under natural conditions. Then, approximately 100-g dried rice straws were crushed with a multi-functional pulverizer and screened through a 1-mm stainless steel flour sieve to ensure that the samples are homogeneous and have similar particle size, then the samples packed in hermetical plastic bags for chemical composition analysis, cell wall composition analysis, and *in vitro* ruminal fermentation. Beyond that, another 100-g dried rice straws from similar part of the stem in each rice straw were selected to a length of 2 cm, and then employed to investigate the changes in fiber structure after ruminal microbial degradation by using scanning electron microscopy.

In vitro ruminal batch incubation and sampling

This study contained two *in vitro* experiments. The first experiment was designed to investigate the kinetics of total gas and methane (CH_4) production after 48-h *in vitro* ruminal batch incubation. The second experiment was to compare the degradation and fermentation profiles of two straws at the early (12 h) and late (48 h) stages of *in vitro* ruminal incubation. Each experiment was conducted by a completely randomized block design, which included 3 runs with each treatment containing two fermentation bottles (replicates), and each run was incubated with mixed rumen fluid from 2 of 3 donor goats on different days. For the second experiment, each treatment had six fermentation bottles with two bottles for volatile fatty acids (VFA) and microbial samples, two bottles for scanning electron microscopy samples, and the other two bottles for substrates degradation measurement.

The *in vitro* ruminal batch culture was performed according to Wang et al. (2016b). Approximately 1.20 g of substrates were weighed into each of 135-ml serum bottles for the *in vitro* incubation. Rumen fluid was collected through permanent rumen cannula before morning feeding, which were fed by TMR diet containing the CP content of 127 g/kg of DM and the NDF content of 380 g/kg of DM. The rumen fluid was filtered through four layers of cheesecloth into a pre-warmed insulated bottle and taken to the laboratory. Then buffered rumen fluid was prepared by mixing 12-ml rumen fluid with 48-ml McDougall's buffer (Cone and Becker, 2012), and then added into bottle under a stream of CO_2 at 39.5°C. Bottles were immediately placed into an automatic incubation system (39.5°C, 55 r/min). Each bottle was connected to a pressure sensor, from which a signal operated a computer-controlled three-way solenoid valve. Venting pressure was set at 10 kPa, and vented gas was transferred to gas chromatographer (GC, Agilent 7890A; Agilent Inc., Palo Alto, CA, United States) to measure the CH_4 concentration. Total gas production (GP) and CH_4 production were calculated using the equations described by Wang et al. (2013).

Samples were collected at 12 and 48 h to represent the early and late stage of *in vitro* ruminal incubation, respectively. The 2 ml of liquid without visible particles were collected from two bottles and centrifuged at 12,000*g* for 10 min at 4°C. The supernatant (1.5 ml) was acidified using 0.15 ml of 25% (w/v) metaphosphoric acid, and stored at-20°C for analysis of VFA. The 1.5 ml of microbial samples were collected after intense shaking of the same two bottles to ensure that representative portions of liquid and particle fractions were included, and were immediately frozen in liquid N_2 and stored at −80°C until DNA extraction. The pH was measured immediately with a portable pH meter (Starter 300; Ohaus Instruments Co. Ltd., Shanghai, China). Then, the 2-cm length of rice straw samples were taken from the other two fermentation bottles and dried in 65°C, then packed in hermetical plastic bags for scanning electron microscopy. Solid residues were filtered into pre-weighed Gooch filter crucibles from the other two bottles, dried at 105°C to determine degradation of incubated substrates.

Sample analysis

The cell wall composition was determined according to the methods described by Huang et al. (2015). In brief, the feed samples were treated with 70% ethanol and a mixture of chloroform and methanol (1:1 v/v) twice to prepare alcohol insoluble residues (AIRs). After acid hydrolysis in 2 mol/l trifluoroacetic acid and derived into alcohol acetates, samples were analyzed by an Agilent 7,890 series GC equipped with a 5,975\u00B0C MS detector (Agilent, Palo Alto, California, United States). The cellulose content was quantified by anthrone assay using the remains after TFA treatment (Updegraff, 1969).

Contents of DM (method 945.15), OM (method 942.05) and CP (method 945.01, total $N \times 6.25$) were analyzed according to methods of AOAC (2005). Neutral detergent fiber and ADF contents were determined with inclusion of a heat stable α-amylase and sodium sulfite, and were expressed as residual ash (Van Soest et al., 1991). Gross energy was determined by an isothermal automatic calorimeter (5E-AC8018; Changsha Kaiyuan Instruments Co, Changsha, China). The water-soluble carbohydrate (WSC) fraction was measured using the anthrone method (Yemm and Willis, 1954).

Alterations of fiber structure were obtained by field emission scanning electron microscopy (FESEM; model SU8010, Hitachi, Japan). Briefly, the similar part of the stem in each rice straw was selected and coated with gold before scanning (Zhang et al., 2018). The images representative of the average characteristics of each treatment group were screened with magnification of 1,000 or 5,000 times.

The VFA samples were recentrifuged at 15,000 g, and supernatants were collected to measuring the molar concentration of individual VFA by a GC (Agilent 7,890 A, Agilent Inc., Palo Alto, California, United States). Details of GC configurations were set according to the Wang et al. (2014). Molar percentage of individual VFA was then calculated based on their molar concentrations to better represent the fermentation pattern.

DNA extraction and quantification of microbial groups

The microbial DNA was extracted by using a modified RBB + C methodology (Yu and Morrison, 2004) with sand beating according to Ma et al. (2020). The quality of the DNA extracts was assessed using agarose gel (0.8%) electrophoresis. Total DNA extracted was quantified with Nano Drop ND1000 (NanoDrop Technologies, Wilmington, DE, United States), and then stored at −80°C until further analyses.

Total DNA was then diluted to 10 ng/μl to quantify selected groups of microorganisms, including total bacteria, protozoa, fungi and methanogens, *Fibrobacter succinogenes, Ruminococcus albus, Ruminococcus flavefaciens, Selenomonas ruminantium, Prevotella ruminicola, Ruminobacter amylophilus, Butyrivibrio fibrisolvens, Methanobacteriales,* and *Methanobrevibacter* (Table 1). Quantitative PCR was performed according the procedures described by Jiao et al. (2014). Final absolute amounts of target groups or species were estimated by relating the C_T value to the standard curves and expressed as $\log_{10}$ copies/mL rumen contents.

Data analysis

The logistic-exponential model (Wang et al., 2011) was performed to analyze the kinetics of total gas production by using the Nonlinear Regression Analysis Program (NLREG, version 5.4; Sherrod, 1995), and expressed as follows:

$$GPt = VF\frac{1-\exp(-kt)}{1+\exp(b-kt)},$$

where GP_t is the accumulated gas or CH_4 production at time t (ml/g); *VF* is the final asymptotic gas or CH_4 production (ml/g);

TABLE 1 List of primers used for qPCR assay.

Microbial species	Primer sets (5′-3′)	Product size, bp	References
Protozoa	F: GCTTTCGWTGGTAGTGTATT;	223	Sylvester et al., 2004
	R: CTTGCCCTCYAATCGTWCT		
Fungi	F: GAGGAAGTAAAAGTCGTAACAAGGTTTC;	121	Denman and McSweeney, 2006
	R: CAAATTCACAAAGGGTAGGATGATT		
Bacteria	F: CGGCAACGAGCGCAACCC;	146	Denman and McSweeney, 2006
	R: CCATTGTAGCACGTGTGTAGCC		
Methanogens	F: GGATTAGATACCCSGGTAGT;	192	Hook et al., 2010
	R: GTTGARTCCAATTAAACCGCA		
Selected groups of bacteria			
Fibrobacter succinogenes	F: GTTCGGAATTACTGGGCGTAAA;	121	Denman and McSweeney, 2006
	R: CGCCTGCCCCTGAACTATC		
Ruminococcus albus	F: CCCTAAAAGCAGTCTTAGTTCG;	176	Koike and Kobayashiy, 2001
	R: CCTCCTTGCGGTTAGAACA		
Ruminococcus flavefaciens	F:GAACGGAGATAATTTGAGTTTACTTAGG;	132	Denman and McSweeney, 2006
	R:CGGTCTCTGTATGTTATGAGGTATTACC		
Selenomonas ruminantium	F: CAATAAGCATTCCGCCTGGG	138	Stevenson and Weimer, 2007
	R: TTCACTCAATGTCAAGCCCTGG		
Prevotella ruminicola	F:GAAAGTCGGATTAATGCTCTATGTTG	74	Stevenson and Weimer, 2007
	R: CATCCTATAGCGGTAAACCTTTGG		
Ruminobacter amylophilus	F: CTGGGGAGCTGCCTGAATG	102	Stevenson and Weimer, 2007
	R: GCATCTGAATGCGACTGGTTG		
Selected groups of methanogens			
Methanobacteriales	F:CGWAGGGAAGCTGTTAAGT;	343	Yu et al., 2005
	R:TACCGTCGTCCACTCCTT		
Methanobrevibacter	F: CCTCCGCAATGTGAGAAATCGC;	230	Huang et al., 2016
	R: TCWCCAGCAATTCCCACAGTT		

k is the fractional rate of gas or CH_4 production (/h); b is the shape parameter. Then, initial fractional rate of degradation at 0 h (FRD_0, /h) was calculated using the following equation:

$$FRD_0 = k / [1 + \exp(b)].$$

The net hydrogen production relative to the amount of total VFA produced (R_{NH2}, mol/100 mol of VFA) was calculated using stoichiometric equations (Wang et al., 2014), and was expressed as follows:

$$R_{NH2} = [2(Ace + But + Isobut) - (Pro + Val + Isoval)] / \text{total VFA}.$$

where Ace, But, Pro, Val, Isobut, and Isoval indicate the concentrations (mmol/L) of acetate, propionate, valerate, isobutyrate, and isovalerate, respectively.

The neutral detergent solubles (NDS) content, hemicellulose content, and *in vitro* degradation was calculated according to the following equations (Zhang et al., 2020), and was expressed as follows:

$$\text{NDS (g/kg of DM)} = 1{,}000 - \text{NDF (g/kg of DM)}.$$
$$\text{hemicellulose (g/kg of DM)} = \text{NDF} - \text{ADF (g/kg of DM)}.$$
$$\text{degradation (g/kg of DM)} = [1 - (W_2 \times C_2) / (W_1 \times C_1)] \times 1{,}000.$$

where C_1 is NDS, NDF, or ADF content in the substrate before incubation; C_2 is NDS, NDF, or ADF content in the residue after 12 or 48 h of incubation; W_1 is DM weight of substrate before incubation; W_2 is DM weight of residue after 12 or 48 h of incubation.

The values from replicated bottles for each run were averaged for statistical analysis. The data were then analyzed using a linear mixed model of SPSS 26.0 software (Chicago, IL, United States). The first analytic model included mutant ($n = 2$) as a fixed effect, and run ($n = 3$) as random effect. When sampling time was included, the second analytic model included mutant ($n = 2$) and interaction of mutant and sampling time as the fixed effect, sampling time ($n = 2$) as a repeated measurement, and run ($n = 3$) as the random effect. When significant interaction of mutant and sampling time was observed, statistical analysis was re-performed for each sampling time point by using the first model. Statistical significance value and a trend toward difference were set at levels of $p \leq 0.05$ and $0.05 < p \leq 0.10$, respectively.

Results

The *bc15* mutant straw had lower NDF (−7.33%) and ADF (−35.3%) contents, and higher CP (+97.4%) hemicellulose (+38.1%), WSC (+277%), and NDS (+20.5%) contents (Table 2). The *bc15* mutant straw had higher rhamnose (+17.2%, $p < 0.001$), fucose (+8.87%, $p = 0.001$), arabinose (+43.4%, $p < 0.001$), xylose (+24.8%, $p < 0.001$), mannose (+6.25%, $p = 0.07$) and galactose (+52.8%, $p < 0.001$) contents, and lower cellulose content (−37.3%, $p < 0.001$), although no difference was observed for glucose content between WT and *bc15* mutant straw (Table 3).

TABLE 2 Chemical composition of wild-type (WT) and *brittle culm 15* (*bc15*) mutant rice straws (expressed in g/kg of dry matter; $N = 3$).

Items	WT	*bc15*
OM	846	862
CP	23.0	45.4
NDF	737	683
ADF	456	295
Hemicellulose	281	388
WSC	12.2	46.0
NDS	263	317
Gross energy, MJ/kg DM	15.2	16.2

OM, Organic matter; CP, Crude protein; NDF, Neutral detergent fiber; ADF, Acid detergent fiber; WSC, Water-soluble carbohydrates; NDS, Neutral detergent solubles; GE, Gross energy.

TABLE 3 Cell wall composition of wild-type (WT) and *brittle culm 15* (*bc15*) mutant rice straws (expressed in g/kg of DM; $N = 3$).

Items	WT	*bc15*	SEM	Value of *p*
Rhamnose	2.15	2.52	0.033	<0.001
Fucose	1.24	1.35	0.014	0.001
Arabinose	24.9	35.7	0.59	<0.001
Xylose	145	181	2.1	<0.001
Mannose	2.40	2.55	0.049	0.07
Galactose	10.8	16.5	0.39	<0.001
Glucose	41.3	42.7	1.08	0.40
Cellulose	418	262	9.0	<0.001

SEM, Standard error of mean.

The *bc15* mutant straw displayed an altered kinetics of *in vitro* gas and CH_4 production, with a faster increase in gas and CH_4 production at first 24 or 30 h of incubation, and a reduction in gas and CH_4 production after 36 h of incubation (Figures 1A, 2A). Further analysis of kinetics indicated that *bc15* mutant straw had greater initial fractional rate of degradation ($p < 0.001$), but lower fractional rate of CH_4 production ($p = 0.048$, Figure 2B), together with a tendency of lower fractional rate of gas production ($p = 0.09$, Figure 1B). Significant interactions ($p < 0.001$) between mutant and time were observed for GP and CH_4 production. The *bc15* mutant straw resulted in greater total gas ($p < 0.001$) and CH_4 ($p = 0.01$) productions at 12 h of incubation, and lower total gas ($p = 0.06$) and CH_4 ($p = 0.01$) productions at 48 h of incubation (Table 4).

The *bc15* mutant straw had greater NDS degradation ($p < 0.001$) and lower ADF degradation ($p = 0.001$) at both 12 and 48 h of incubation (Table 4). Interactions between mutant and time were observed for DM ($p = 0.03$) and NDF degradation ($p < 0.001$). The *bc15* mutant straw had higher ($p < 0.05$) DM and NDF degradation at 12 h of incubation, and lower NDF degradation ($p = 0.01$) at 48 h of incubation. The scanning electron microscopy analyses showed that a greater number of attached rumen microorganism were observed on *bc15* mutant straw at 12 h of incubation but with less small holes destroyed by rumen microorganism at 48 h of incubation when compared with WT straw (Figure 3).

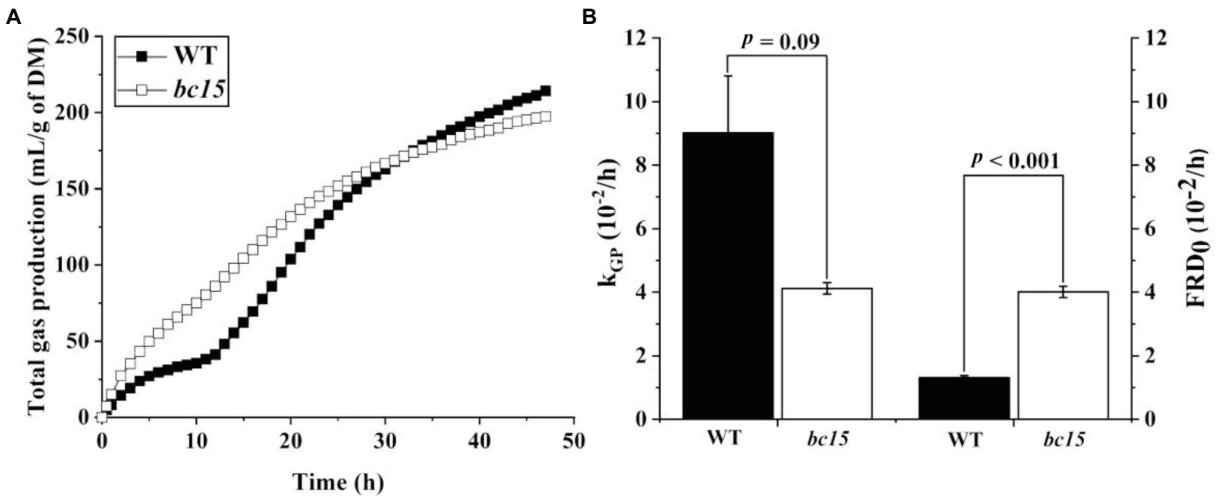


FIGURE 1
The kinetics of ruminal gas production **(A)** and its fractional rate of gas production (k_{GP}, 10^{-2}/h) and initial fractional rate of degradation at 0h (**B**, FRD_0, 10^{-2}/h) of wild-type (WT) and *brittle culm 15* (*bc15*) mutant rice straws after 48-h *in vitro* ruminal incubation (n =3). Data with error bars are expressed as mean±standard error.

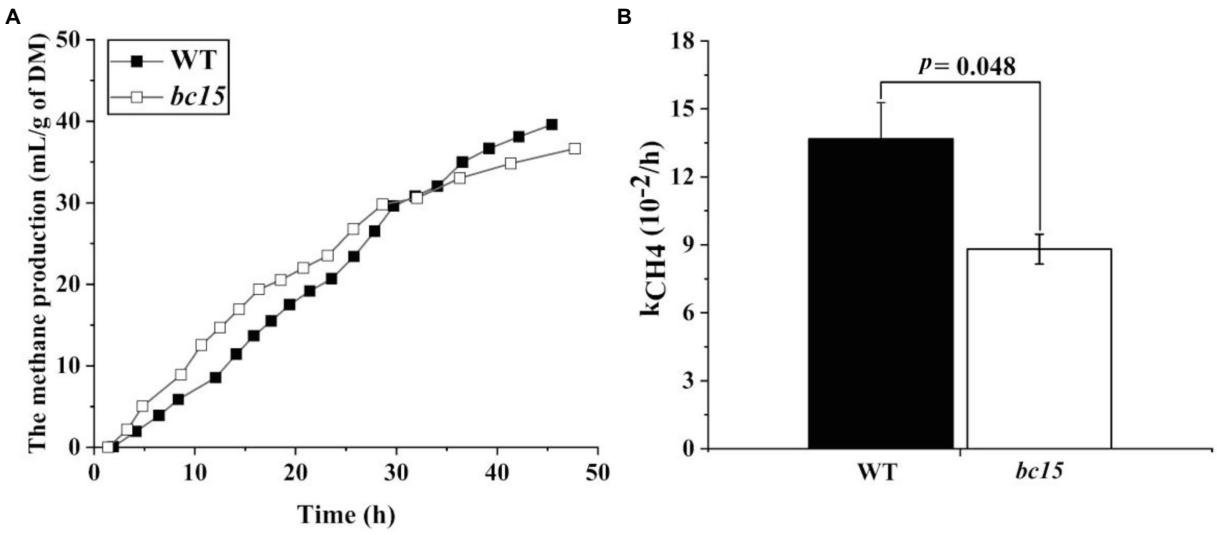


FIGURE 2
The kinetics of ruminal methane (CH_4) production **(A)** and its fractional rate of CH_4 production (k_{CH4}, 10^{-2}/h; **B**) of wild-type (WT) and *brittle culm 15* (*bc15*) mutant rice straws after 48-h *in vitro* ruminal incubation (n =3). Data with error bars are expressed as mean±standard error.

Interactions ($p < 0.05$) between mutant and time were also observed for the changes of pH and VFA profile. The *bc15* mutant straw had lower pH ($p < 0.001$) and higher total VFA concentration ($p = 0.03$) at 12 h of incubation, although such differences were not observed ($p > 0.1$) at 48 h of incubation. The *bc15* mutant straw had lower ($p = 0.002$) acetate to propionate ratio at both 12 and 48 h of incubations, and higher ($p = 0.008$) propionate molar percentage at 12 h of incubation. The *bc15* mutant straw had lower ($p = 0.004$) acetate molar percentage and R_{NH2} ($p = 0.03$) and higher ($p < 0.001$) molar percentages of butyrate and valerate at both 12 and 48 h of incubations (Table 5).

The further qPCR results showed that *bc15* mutant straw had greater ($p = 0.002$) 16S gene copy number of *F. succinogenes*, and lower 18S gene copy number of protozoa ($p = 0.09$) and fungi ($p = 0.05$) at both 12 h and 48 h of incubation. However, there was no difference ($p > 0.1$) in 16S gene copy number of bacteria, methanogens, *R. albus*, *R. flavefaciens*, *S. ruminantium*, *P. ruminicola*, *R. amylophilus*, *Methanobacteriales*, and *Methanobrevibacter* (Table 6).

TABLE 4 Gases production and degradation of wild-type (WT) and *brittle culm 15* (*bc15*) mutant rice straws after 12- and 48-h *in vitro* ruminal incubation (n =3).

Items	12 h		48 h		SEM	Value of p[1]		
	WT	*bc15*	WT	*bc15*		Mutant	Time	Mutant × Time
Gases production, ml/g of DM								
GP	48.4	88.3*	218	204	2.69	0.001	<0.001	<0.001
CH_4	4.10	11.9*	40.8	34.7*	1.80	0.24	<0.001	<0.001
Degradation, g/kg of DM								
DM	145	201*	476	472	12.9	0.04	<0.001	0.03
NDS	299	392	554	636	15.3	<0.001	<0.001	0.74
NDF	62.2	85.5*	431	380*	5.13	0.03	<0.001	<0.001
ADF	107	75.7*	451	344*	12.6	0.001	<0.001	0.02

GP, Total gas production; DM, Dry matter; NDS, Neutral detergent solubles; NDF, Neutral detergent fiber; ADF, Acid detergent fiber; SEM, Standard error of mean. [1]Mutant, Effects of *bc15* mutant and WT rice straws; Time, Effects of different incubation time; Mutant × Time, Interaction between mutant and incubation time.
*Indicates significant difference ($p < 0.05$) between bc15 mutant and WT rice straws at 12 or 48 h of incubation, when there was significant interaction between mutant and time ($p < 0.05$).

Discussion

Mutation of *BC15* gene has been reported to down-regulate cellulose biosynthesis in rice (Wu et al., 2012). The *bc15* mutant employed in our study exhibited higher CP, WSC, hemicellulose, and NDS contents, and lower NDF and ADF contents in rice straw. Furthermore, *bc15* mutant straw had lower cellulose content, and higher rhamnose, arabinose, xylose, mannose and galactose contents, in comparison with WT straw. Such changes in monosaccharides composition of cell wall are consistent with the study of Wu et al. (2012), which report that cellulose content is decreased through remodeling plant cell wall in *bc15* mutant rice straw. Furthermore, *bc15* mutant straw also exhibits higher CP content (45.4 vs. 23.0 g/kg of DM), which might be due to the activation of expression of specific genes in cellulose-deficient mutants (Raggi et al., 2015; Li et al., 2019), and such related mechanism needs further investigation.

The kinetics of 48-h total gas and CH_4 production indicated a distinct ruminal fermentation process between *bc15* mutant and WT straws. The *bc15* mutant straw exhibited a higher gas and CH_4 production than WT straw at early stage of incubation (<30 h). Such enhancement of fermentation at early stage of incubation was further supported by the greater FRD_0, and was consistent with higher CP, WSC, NDS and monosaccharides content in *bc15* mutant straw. However, this was not the case for the late stage of incubation, as *bc15* mutant straw exhibited lower total gas and CH_4 production than WT straw, which was consistent with lower NDF and ADF content in *bc15* mutant straw.

The *bc15* mutant straw consisted of changed carbohydrate compositions, and could alter substrates hydrolysis and degradation during *in vitro* ruminal fermentation, resulting in the significant interactions between mutant and time for substrates degradation. The *bc15* mutant straw had greater DM degradation at 12 h of incubation, but there was no difference in DM degradation at 48 h of incubation compared with WT straw. These results indicate that *BC15* gene mutation alters straw structure and/or composition, which enhances substrates degradation with increased initial fractional rate of degradation at early stage of incubation. Neutral detergent solubles, includes WSC, protein and ether extract, can be rapidly degraded by rumen microorganisms (Leiva et al., 2000). Previous study indicates elevated NDS content can promote degradation and gas production *in vitro* (Zhang et al., 2018, 2020). Forages with greater WSC content also have greater DM degradation rate and gas production *in vitro* (Amer et al., 2012). In our study, the elevated NDS and WSC content in *bc15* mutant straw contributed to the greater rate of degradation and gas production of rice straw at early stage of *in vitro* incubation, and enhanced NDS degradation at both incubation times.

Plant cell walls are mainly composed of cellulose, hemicellulose, lignin and pectin (Wang et al., 2016c), and its structure and composition may affect ruminal fiber digestibility (Zhang et al., 2020; Wang et al., 2021b). We observed significant interactions between mutant and time for NDF degradation with higher 12-h NDF degradation but lower 48-h NDF degradation and lower ADF degradation at both incubation times of the *bc15* mutant straw during *in vitro* incubation. Such enhancement of NDF degradation at early stage of incubation can be due to the greater monosaccharides (e.g., rhamnose, fucose, xylose, mannose, and galactose) content in *bc15* mutant straw, as monosaccharides in hemicellulose are usually more rapidly degraded in rumen than cellulose (Sun et al., 2007; Janssen, 2010). However, fermentation of *bc15* mutant straw exhibited the reduction in fiber degradation, which was supported by less small holes destroyed by rumen microorganism at 48 h of incubation. In cellulose-deficient mutants with defects in cell wall composition, plant cells often activate the expression of specific genes, such as peroxidase as a response to cell wall damage, which leads to the strengthening and hardening of cell walls and limits cell expansion (Raggi et al., 2015; Li et al., 2019). Such enhanced

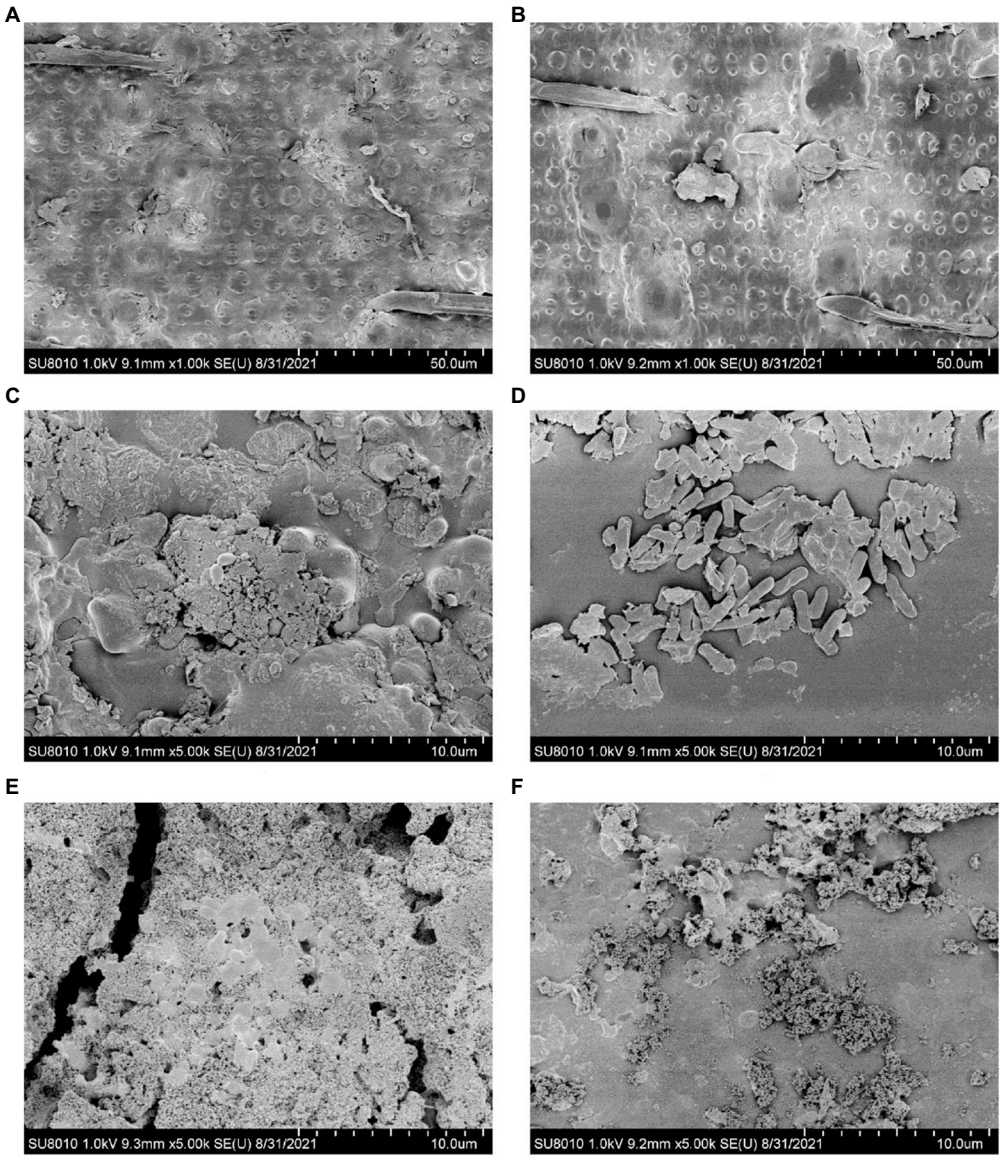


FIGURE 3
Scanning electron microscopy images of wild-type **(A,C,E)** and *brittle culm 15* **(B,D,F)** mutant rice straws before and after 12- and 48-h *in vitro* ruminal incubation.

hardenability of cell wall structure in the *bc15* mutant straw likely causes greater difficulty to be degraded by rumen microorganisms compared with WT straw, leading to the reduction in fiber degradation.

Volatile fatty acid is mainly produced during ruminal carbohydrate fermentation, thus the VFA fermentation profile is greatly associated with carbohydrate composition. Cellulose is fermented to lesser propionate than acetate, whereas readily fermentable carbohydrates is fermented to less acetate but more butyrate and propionate (Janssen, 2010; Oba, 2011). The fermentation of *bc15* mutant straw resulted in greater butyrate molar percentage, and lower acetate molar percentage and acetate to propionate ratio at both 12 and 48 h of incubation times. The results are consistent with previous studies, which indicate that elevated WSC content in forages can increase butyrate or propionate molar percentage and decrease acetate to propionate ratio *in vitro* (Purcell et al., 2014; Rivero et al., 2020). It seems that increased content of soluble carbohydrates in the *bc15* mutant alters fermentation pathway of rice straw, leading to a less favorability of acetate production.

Hydrogen is produced during carbohydrate fermentation and is mainly utilized by methanogens to produce CH_4 (Wang et al., 2016d; Martinez-Fernandez et al., 2017). The formation of acetate is associated with relative more H_2 production, while the formation of propionate and butyrate are associated with net H_2 consumption and relative less H_2 production, respectively

TABLE 5 Fermentation characteristics of wild-type (WT) and *brittle culm 15* (*bc15*) mutant rice straws after 12- and 48-h *in vitro* ruminal incubation (n =3).

Items	12 h		48 h		SEM	Value of p[1]		
	WT	*bc15*	WT	*bc15*		Mutant	Time	Mutant × Time
pH	6.76	6.62*	6.44	6.41	0.019	0.003	<0.001	0.02
Total VFA, mM	32.3	47.1*	82.4	79.1	4.51	0.15	<0.001	0.04
Molar percentage of individual VFA, mol/100 mol								
Acetate	67.6	64.8	67.7	64.3	0.78	0.004	0.78	0.63
Propionate	17.6	19.7*	23.8	23.6	0.26	0.007	<0.001	0.004
Butyrate	8.86	10.2	5.61	8.16	0.267	<0.001	<0.001	0.06
Isobutyrate	1.39	1.16*	0.80	1.05*	0.060	0.82	<0.001	0.004
Valerate	1.35	1.53	0.89	1.16	0.083	0.03	0.002	0.61
Isovalerate	3.17	2.54*	1.18	1.74*	0.249	0.88	0.001	0.03
Acetate to propionate ratio	3.84	3.29*	2.85	2.72	0.077	0.002	<0.001	0.02
R_{NH2} (mol/mol)	1.34	1.29	1.23	1.20	0.015	0.03	<0.001	0.29

R_{NH2}, Estimated net hydrogen production relative to the amount of total VFA produced; SEM, Standard error of mean. [1]Mutant, Effects of *bc15* mutant and WT rice straws; Time, Effects of different incubation time; Mutant × Time, Interaction between mutant and incubation time.
*Indicates significant difference ($p < 0.05$) between bc15 mutant and WT rice straws at 12 or 48 h of incubation, when there was significant interaction between mutant and time ($p < 0.05$).

TABLE 6 Select microbial groups (determined by RT-PCR) of wild-type (WT) and *brittle culm 15* (*bc15*) mutant rice straws after 12- and 48-h *in vitro* ruminal incubation, log_{10} (copies/ml; n =3).

Items	12 h		48 h		SEM	Value of p[1]		
	WT	*bc15*	WT	*bc15*		Mutant	Time	Mutant × Time
Protozoa	10.1	9.90	8.37	8.13	0.118	0.09	<0.001	0.88
Fungi	7.86	7.68	8.62	8.11	0.150	0.05	0.004	0.30
Bacteria	11.7	11.7	11.6	11.6	0.07	0.74	0.21	0.76
Methanogens	10.1	10.1	9.97	9.91	0.163	0.85	0.32	0.87
Selected bacteria								
Fibrobacter succinogenes	9.01	9.63	9.06	9.27	0.092	0.002	0.14	0.06
Ruminococcus albus	8.52	8.60	7.85	7.69	0.197	0.73	<0.001	0.28
Ruminococcus flavefaciens	8.24	7.85	7.73	7.01	0.569	0.30	0.22	0.76
Selenomonas ruminantium	9.98	10.0	9.72	9.63	0.073	0.56	0.001	0.36
Prevotella ruminicola	10.0	10.1	10.1	10.0	0.15	0.99	0.92	0.69
Ruminobacter amylophilus	6.83	6.68	8.58	8.56	0.145	0.52	<0.001	0.67
Selected methanogens								
Methanobacteriales	9.34	9.26	9.16	9.12	0.195	0.68	0.30	0.89
Methanobrevibacter	9.22	9.34	9.20	9.08	0.145	0.99	0.34	0.43

SEM, Standard error of mean. [1]Mutant, Effects of *bc15* mutant and WT rice straws; Time, Effects of different incubation time; Mutant × Time, Interaction between mutant and incubation time.

(Ungerfeld, 2020). Shifting fermentation pathway from acetate to propionate or butyrate production can be associated with a reduction in CH_4 production (Wang et al., 2016a; Zhang et al., 2020). Significant interactions between mutant and time for CH_4 production was observed. The *bc15* mutant straw had greater CH_4 production at 12 h of incubation, which could be caused by greater extent of initial degradation. However, *bc15* mutant straw had lower CH_4 production without any difference in DM degradation compared to WT straw at 48 h of incubation, indicating that the *BC15* gene mutation altered carbohydrate composition with less favorable methanogenesis, which was also reflected by the reduced R_{NH2} of *bc15* mutant straw. This result is consistent with previous studies, which found that fiber fermentation produces more CH_4 than soluble carbohydrates (Pirondini et al., 2015; Bougouin et al., 2018; Zhang et al., 2020). Further studies are still needed to confirm the effect of *bc15*

mutant straw on enteric CH_4 emissions through feeding ruminant animals.

Efficient degradation of fibrous substrate can be achieved by a consortium of major cellulolytic groups. Fungi, protozoa, methanogens, and fibrolytic bacteria, such as *F. succinogenes*, *R. albus*, and *R. flavefaciens,* are the major cellulolytic microorganisms (Wang and McAllister, 2002; Koike and Kobayashi, 2009; Fuma et al., 2012). Fungi have been regarded as the primary colonizers and have the ability to destroy plant cell walls for subsequent attack by fibrolytic bacteria (Wang and McAllister, 2002; Zhang et al., 2007; Terry et al., 2019). In our study, the *bc15* mutant straw displayed lower 18S gene copy number of fungi, which is consistent to decreased ADF degradation and less fiber destruction observed during *in vitro* ruminal fermentation. Previous studies also indicate that increased fiber degradation can be closely associated with the increases in fungi population (Zhang et al., 2020; Wang et al., 2021b). However, the *bc15* mutant straw displayed greater 16S gene copy number of *F. succinogenes*. Although *F. succinogenes* is widely attributed to fiber degraders, it can also utilize different energy sources, such as structural and nonstructural carbohydrates (Ghali et al., 2017; Neumann et al., 2018). We propose that the fiber of *bc15* mutant straw would be more difficult for degradation with a reduction in population of major cellulolytic microorganisms.

Conclusion

The *BC15* gene mutation altered carbohydrate compositions of rice straw with reduction in cellulose content and increase in hemicellulose, WSC, and NDS content. Such changes in carbohydrate composition altered the process of ruminal biodegradation, such as promoted degradation of rice straw at early stage (12 h) of *in vitro* rumen fermentation. However, cell wall of *bc15* mutant straw seems to be more resistant for microbial degradation, as demonstrated by lower fiber degradation and cellulolytic fungi population at 48 h of incubation. Furthermore, the *BC15* gene mutation also altered rumen fermentation of rice straw, which displayed shifted fermentation pattern from acetate production to propionate and butyrate production, leading to a reduction in efficiency of H_2 production and 48-h CH_4 production. Thus, the *BC15* gene mutation provides an excellent case for alter the nutritional components, degradation, and CH_4 production of rice straw during *in vitro* ruminal fermentation process. Further researches are still needed to investigate the effects of mutant straws with more gene types on animal feeding.

Data availability statement

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

Ethics statement

The animal study was reviewed and approved by CAS Key Laboratory of Agro-Ecological Processes in Subtropical Region, National Engineering Laboratory for Pollution control and Waste Utilization in Livestock and Poultry Production, Institute of Subtropical Agriculture, The Chinese Academy of Sciences, Changsha, Hunan, 410125, China.

Author contributions

SY conducted the experiments, analyzed the data, and wrote the initial manuscript. MW designed the research. MW and XZ performed funding acquisition. MW, XZ, and BZ reviewed the manuscript. BZ provided the experimental materials. JZ, ZM, and RW participated in experiment and sampling. DW and ZW performed formal analysis and investigation. ZT performed supervision and management. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Strategic Priority Research Program of the Chinese Academy of Sciences (grant no. XDA26040203), National Natural Science Foundation of China (grant no. 31922080 and 32002204), Hunan Province Science and Technology Plan (2020NK2066 and 2022NK2021), the Science and Technology Innovation Program of Hunan Province (2021RC2102), China Agriculture Research System of MOF and MARA, Youth Innovation Promotion Association CAS (grant no. Y202078), and Open Fund of Key Laboratory of Agro-ecological Processes in Subtropical Region Chinese Academy of Sciences (grant no. ISA2021203).

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

Alemu, D., Tegegne, F., and Mekuriaw, Y. (2020). Comparative evaluation of effective microbe-and urea molasses-treated finger millet (*Eleusine coracana*) straw on nutritive values and growth performance of Washera sheep in northwestern Ethiopia. *Tropl. Anim. Health Prod.* 52, 123–129. doi: 10.1007/s11250-019-01986-z

Amer, S., Hassanat, F., Berthiaume, R., Seguin, P., and Mustafa, A. F. (2012). Effects of water soluble carbohydrate content on ensiling characteristics, chemical composition and *in vitro* gas production of forage millet and forage sorghum silages. *Anim. Feed Sci. Technol.* 177, 23–29. doi: 10.1016/j.anifeedsci.2012.07.024

AOAC (2005). *International. Official Methods of Analysis. 18th Edn.* AOAC International, Gaithersburg, MD.

Bougouin, A., Ferlay, A., Doreau, M., and Martin, C. (2018). Effects of carbohydrate type or bicarbonate addition to grass silage-based diets on enteric methane emissions and milk fatty acid composition in dairy cows. *J. Dairy Sci.* 101, 6085–6097. doi: 10.3168/jds.2017-14041

Cone, J. W., and Becker, P. M. (2012). Fermentation kinetics and production of volatile fatty acids and microbial protein by starchy feedstuffs. *Anim. Feed Sci. Technol.* 172, 34–41. doi: 10.1016/j.anifeedsci.2011.12.006

Denman, S. E., and McSweeney, C. S. (2006). Development of a real-time PCR assay for monitoring anaerobic fungal and cellulolytic bacterial populations within the rumen. *FEMS Microbiol. Ecol.* 58, 572–582. doi: 10.1111/j.1574-6941.2006.00190.x

Fuma, R., Oyaizu, S., Nukui, Y., Ngwe, T., Shinkai, T., Koike, S., et al. (2012). Use of bean husk as an easily digestible fiber source for activating the fibrolytic rumen bacterium Fibrobacter succinogenes and rice straw digestion. *Anim. Sci. J.* 83, 696–703. doi: 10.1111/j.1740-0929.2012.01017.x

Ghali, I., Sofyan, A., Ohmori, H., Shinkai, T., and Mitsumori, M. (2017). Diauxic growth of *Fibrobacter succinogenes* S85 on cellobiose and lactose. *FEMS Microbiol. Lett.* 364, 1–9.

Hook, S. E., Wright, A. D., and McBride, B. W. (2010). Methanogens: methane producers of the rumen and mitigation strategies. *Archaea* 2010:945785. doi: 10.1155/2010/945785

Huang, X. D., Martinez-Fernandez, G., Padmanabha, J., Long, R., Denman, S. E., and McSweeney, C. S. (2016). Methanogen diversity in indigenous and introduced ruminant species on the Tibetan plateau. *Archaea* 2016:5916067, –5916010. doi: 10.1155/2016/5916067

Huang, D., Wang, S., Zhang, B., Shang-Guan, K., Shi, Y., Zhang, D., et al. (2015). A gibberellin-mediated DELLA-NAC signaling Cascade regulates cellulose synthesis in Rice. *Plant Cell* 27, 1681–1696. doi: 10.1105/tpc.15.00015

Janssen, P. H. (2010). Influence of hydrogen on rumen methane formation and fermentation balances through microbial growth kinetics and fermentation thermodynamics. *Anim. Feed Sci. Technol.* 160, 1–22. doi: 10.1016/j.anifeedsci.2010.07.002

Jiao, J., Wang, P., He, Z., Tang, S., Zhou, C., Han, X., et al. (2014). *In vitro* evaluation on neutral detergent fiber and cellulose digestion by post-ruminal microorganisms in goats. *J. Sci. Food Agric.* 94, 1745–1752. doi: 10.1002/jsfa.6485

Koike, S., and Kobayashi, Y. (2009). Fibrolytic rumen Bacteria: their ecology and functions. *Asian Aust. J. Anim. Sci.* 22, 131–138. doi: 10.5713/ajas.2009.r.01

Koike, S., and Kobayashiy, Y. (2001). Development and use of competitive PCR assays for the rumen cellulolytic bacteria-*Fibrobacter succinogenes*, Ruminococcus albus and *Ruminococcus flavefaciens*. *FEMS Microbiol. Lett.* 204, 361–366. doi: 10.1016/s0378-1097(01)00428-1

Leiva, E., Hall, M. B., and Van Horn, H. H. (2000). Performance of dairy cattle fed Citrus pulp or corn products as sources of neutral detergent-soluble carbohydrates. *J. Dairy Sci.* 83, 2866–2875. doi: 10.3168/jds.S0022-0302(00)75187-3

Li, P., Liu, Y., Tan, W., Chen, J., Zhu, M., Lv, Y., et al. (2019). Brittle culm 1 encodes a COBRA-like protein involved in secondary Cell Wall cellulose biosynthesis in Sorghum. *Plant Cell Physiol.* 60, 788–801. doi: 10.1093/pcp/pcy246

Ma, T., Tu, Y., Zhang, N. F., Deng, K. D., and Diao, Q. Y. (2015). Effect of the ratio of non-fibrous carbohydrates to neutral detergent Fiber and protein structure on intake, digestibility, rumen fermentation, and nitrogen metabolism in lambs. *Asian Aust. J. Anim. Sci.* 28, 1419–1426. doi: 10.5713/ajas.15.0025

Ma, Z. Y., Zhang, X. M., Wang, R., Wang, M., Liu, T., and Tan, Z. L. (2020). Effects of chemical and mechanical Lysis on microbial DNA yield, integrity, and downstream amplicon sequencing of rumen Bacteria and Protozoa. *Front. Microbiol.* 11:581227. doi: 10.3389/fmicb.2020.581227

Martinez-Fernandez, G., Denman, S. E., Cheung, J., and McSweeney, C. S. (2017). Phloroglucinol degradation in the rumen promotes the capture of excess hydrogen generated from Methanogenesis inhibition. *Front. Microbiol.* 8:1871. doi: 10.3389/fmicb.2017.01871

Neumann, A. P., Weimer, P. J., and Suen, G. (2018). A global analysis of gene expression in *Fibrobacter succinogenes* S85 grown on cellulose and soluble sugars at different growth rates. *Biotechnol. Biofuels* 11:295. doi: 10.1186/s13068-018-1290-x

Oba, M. (2011). Review: effects of feeding sugars on productivity of lactating dairy cows. *Can. J. Anim. Sci.* 91, 37–46. doi: 10.4141/CJAS10069

Oskoueian, E., Jahromi, M. F., Jafari, S., Shakeri, M., Le, H. H., and Ebrahimi, M. (2021). Manipulation of Rice straw silage fermentation with different types of lactic acid Bacteria inoculant affects rumen microbial fermentation characteristics and methane production. *Vet. Sci.* 8:100. doi: 10.3390/vetsci8060100

Pirondini, M., Colombini, S., Mele, M., Malagutti, L., Rapetti, L., Galassi, G., et al. (2015). Effect of dietary starch concentration and fish oil supplementation on milk yield and composition, diet digestibility, and methane emissions in lactating dairy cows. *J. Dairy Sci.* 98, 357–372. doi: 10.3168/jds.2014-8092

Purcell, P. J., Boland, T. M., and Kiely, P. (2014). The effect of water-soluble carbohydrate concentration and type on *in vitro* rumen methane output of perennial ryegrass determined using a 24-hour batch-culture gas production technique. *Irish J. Agr. Food Res.* 53, 21–36. doi: 10.2307/24369733

Raggi, S., Ferrarini, A., Delledonne, M., Dunand, C., Ranocha, P., De Lorenzo, G., et al. (2015). The Arabidopsis class III peroxidase AtPRX71 negatively regulates growth under physiological conditions and in response to Cell Wall damage. *Plant Physiol.* 169, 2513–2525. doi: 10.1104/pp.15.01464

Rivero, M. J., Keim, J. P., Balocchi, O. A., and Lee, M. R. F. (2020). *In vitro* fermentation patterns and methane output of perennial ryegrass differing in water-soluble carbohydrate and nitrogen concentrations. *Animals* 10:1076. doi: 10.3390/ani10061076

Sherrod, P. H. (1995). *Nonlinear Regression Analysis Program, NLREG Version 5.0.* Phillip H. Sherrod, Nashville, TN.

Stevenson, D. M., and Weimer, P. J. (2007). Dominance of Prevotella and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. *Appl. Microbiol. Biotechnol.* 75, 165–174. doi: 10.1007/s00253-006-0802-y

Sun, X. Z., Hoskin, S. O., Joblin, K. N., Andrew, I. G., and Harris, P. J. (2007). Degradation of pectic polysaccharides in forage chicory (*Cichorium intybus L.*) by rumen bacteria. *J. Anim. Feed Sci.* 16, 101–106. doi: 10.22358/jafs/74463/2007

Sylvester, J. T., Karnati, S. K., Yu, Z., Morrison, M., and Firkins, J. L. (2004). Development of an assay to quantify rumen ciliate protozoal biomass in cows using real-time pcr. *J. Nutr.* 134, 3378–3384. doi: 10.1089/jmf.2004.7.498

Terry, S. A., Badhan, A., Wang, Y. X., Chaves, A. V., and McAllister, T. A. (2019). Fibre digestion by rumen microbiota - a review of recent metagenomic and metatranscriptomic studies. *Can. J. Anim. Sci.* 99, 678–692. doi: 10.1139/cjas-2019-0024

Ungerfeld, E. M. (2020). Metabolic hydrogen flows in rumen fermentation: principles and possibilities of interventions. *Front. Microbiol.* 11:589. doi: 10.3389/fmicb.2020.00589

Updegraff, D. M. (1969). Semimicro determination of cellulose in biological materials. *Anal. Biochem.* 32, 420–424. doi: 10.1016/S0003-2697(69)80009-6

Van Soest, P. J., Robertson, J. B., and Lewis, B. A. (1991). Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. *J. Dairy Sci.* 74, 3583–3597. doi: 10.3168/jds.S0022-0302(91)78551-2

Wang, R., Bai, Z., Chang, J., Li, Q., Hristov, A. N., Smith, P., et al. (2022). China's low-emission pathways toward climate-neutral livestock production for animal-derived foods. *Innovation* 3:100220. doi: 10.1016/j.xinn.2022.100220

Wang, X., Cheng, Z., Zhao, Z., Gan, L., Qin, R., Zhou, K., et al. (2016c). BRITTLE SHEATH1 encoding OsCYP96B4 is involved in secondary cell wall formation in rice. *Plant Cell Rep.* 35, 745–755. doi: 10.1007/s00299-015-1916-4

Wang, F., Harindintwali, J. D., Yuan, Z., Wang, M., Wang, F., Li, S., et al. (2021a). Technologies and perspectives for achieving carbon neutrality. *Innovation* 2:100180. doi: 10.1016/j.xinn.2021.100180

Wang, M., Janssen, P. H., Sun, X. Z., Muetzel, S., Tavendale, M., Tan, Z. L., et al. (2013). A mathematical model to describe *in vitro* kinetics of H_2 gas accumulation. *Anim. Feed Sci. Technol.* 184, 1–16. doi: 10.1016/j.anifeedsci.2013.05.002

Wang, Y., and McAllister, T. A. (2002). Rumen microbes, enzymes and feed digestion—A review. *Asian-Aust. J. Anim. Sci.* 15, 1659–1676. doi: 10.5713/ajas.2002.1659

Wang, M., Sun, X. Z., Janssen, P. H., Tang, S. X., and Tan, Z. L. (2014). Responses of methane production and fermentation pathways to the increased dissolved hydrogen concentration generated by eight substrates in *in vitro* ruminal cultures. *Anim. Feed Sci. Technol.* 194, 1–11. doi: 10.1016/j.anifeedsci.2014.04.012

Wang, M., Tang, S. X., and Tan, Z. L. (2011). Modeling *in vitro* gas production kinetics: derivation of logistic–exponential (LE) equations and comparison of models. *Anim. Feed Sci. Technol.* 165, 137–150. doi: 10.1016/j.anifeedsci.2010.09.016

Wang, R., Wang, M., Lin, B., Ma, Z. Y., Ungerfeld, E. M., Wu, T. T., et al. (2021b). Association of fibre degradation with ruminal dissolved hydrogen in growing beef bulls fed with two types of forages. *Br. J. Nutr.* 125, 601–610. doi: 10.1017/S0007114520002962

Wang, M., Wang, R., Tang, S. X., Tan, Z. L., Zhou, C. S., Han, X. F., et al. (2016b). Comparisons of manual and automated incubation systems: effects of venting procedures on *in vitro* ruminal fermentation. *Livest. Sci.* 184, 41–45. doi: 10.1016/j.livsci.2015.12.002

Wang, M., Wang, R., Xie, T. Y., Janssen, P. H., Sun, X. Z., Beauchemin, K. A., et al. (2016a). Shifts in rumen fermentation and microbiota are associated with dissolved Ruminal hydrogen concentrations in lactating dairy cows fed different types of carbohydrates. *J. Nutr.* 146, 1714–1721. doi: 10.3945/jn.116.232462

Wang, M., Wang, R., Yang, S., Deng, J. P., Tang, S. X., and Tan, Z. L. (2016d). Effects of three methane mitigation agents on parameters of kinetics of total and hydrogen gas production, ruminal fermentation and hydrogen balance using *in vitro* technique. *Anim. Sci. J.* 87, 224–232. doi: 10.1111/asj.12423

Wu, B., Zhang, B., Dai, Y., Zhang, L., Shang-Guan, K., Peng, Y., et al. (2012). Brittle culm 15 encodes a membrane-associated chitinase-like protein required for cellulose biosynthesis in rice. *Plant Physiol.* 159, 1440–1452. doi: 10.1016/j.biombioe.2016.08.005

Xu, N. N., Wang, D. M., Wang, B., Wang, J. K., and Liu, J. X. (2018). Different endosperm structures in wheat and corn affected *in vitro* rumen fermentation and nitrogen utilization of rice straw-based diet. *Animal* 13, 1607–1613. doi: 10.1017/S1751731118003257

Yemm, E. W., and Willis, A. J. (1954). The estimation of carbohydrates in plant extracts by anthrone. *Biochem. J.* 57, 508–514. doi: 10.1042/bj0570508

Yu, Y., Lee, C., Kim, J., and Hwang, S. (2005). Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnol. Bioeng.* 89, 670–679. doi: 10.1002/bit.20347

Yu, Z., and Morrison, M. (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* 36, 808–812. doi: 10.2144/04365ST04

Zhang, Y., Gao, W., and Meng, Q. (2007). Fermentation of plant cell walls by ruminal bacteria, protozoa and fungi and their interaction with fibre particle size. *Arch. Anim. Nutr.* 61, 114–125. doi: 10.1080/17450390701204020

Zhang, X. M., Wang, M., Wang, R., Ma, Z. Y., Long, D., Mao, H., et al. (2018). Urea plus nitrate pretreatment of rice and wheat straws enhances degradation and reduces methane production *in vitro* ruminal culture. *J. Sci. Food Agr.* 98, 5205–5211. doi: 10.1002/jsfa.9056

Zhang, X. M., Wang, M., Yu, Q., Ma, Z. Y., Beauchemin, K. A., Wang, R., et al. (2020). Liquid hot water treatment of rice straw enhances anaerobic degradation and inhibits methane production during *in vitro* ruminal fermentation. *J. Dairy Sci.* 103, 4252–4261. doi: 10.3168/jds.2019-16904

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 August 2022
DOI 10.3389/fpls.2022.970957

OPEN ACCESS

EDITED BY
Jiyu Zhang,
Lanzhou University, China

REVIEWED BY
Kai Luo,
Hainan University, China
Xiao Ma,
Sichuan Agricultural University, China

*CORRESPONDENCE
Jing Zhou
zhoujing_lz@hotmail.com
Fulin Yang
fulin.yang@aliyun.com

SPECIALTY SECTION
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

RECEIVED 16 June 2022
ACCEPTED 25 July 2022
PUBLISHED 18 August 2022

CITATION
Zhou J, Li Y, Wang X, Liu Y, David-Schwartz R, Weissberg M, Qiu S, Guo Z and Yang F (2022) Analysis of *Elymus nutans* seed coat development elucidates the genetic basis of metabolome and transcriptome underlying seed coat permeability characteristics. *Front. Plant Sci.* 13:970957. doi: 10.3389/fpls.2022.970957

Analysis of *Elymus nutans* seed coat development elucidates the genetic basis of metabolome and transcriptome underlying seed coat permeability characteristics

Jing Zhou[1]*, Yan Li[2], Xun Wang[3], Yijia Liu[2], Rakefet David-Schwartz[4], Mira Weissberg[4], Shuiling Qiu[2], Zhenfei Guo[5] and Fulin Yang[2]*

[1]National Engineering Research Center of Juncao Technology, Fujian Agriculture and Forestry University, Fuzhou, China, [2]College of Animal Sciences (College of Bee Science), Fujian Agriculture and Forestry University, Fuzhou, China, [3]Qinghai University, Academy of Animal Science and Veterinary Medicine, Xining, China, [4]Volcani Center, Agriculture Research Organization, Institute of Plant Sciences, Beit Dagan, Israel, [5]College of Agro-Grassland Science, Nanjing Agricultural University, Nanjing, China

The seed coat takes an important function in the life cycle of plants, especially seed growth and development. It promotes the accumulation of nutrients inside the seed and protects the seed embryo from mechanical damage. Seed coat permeability is an important characteristic of seeds, which not only affects seed germination, but also hinders the detection of seed vigor by electrical conductivity (EC) method. This research aimed to elucidate the mechanism of seed coat permeability formation through metabolome and transcriptome analysis of *Elymus nutans*. We collected the samples at 8, 18, and 28 days post-anthesis (dpa), and conducted a seed inclusion exosmosis experiment and observed the seed coat permeability. Moreover, we analyzed the changes in the metabolome and transcriptome during different development stages. Here, taking 8 dpa as control, 252 upregulated and 157 downregulated differentially expressed metabolites (DEMs) were observed and 886 upregulated unigenes and 1170 downregulated unigenes were identified at 18 dpa, while 4907 upregulated unigenes and 8561 downregulated unigenes were identified at 28 dpa. Meanwhile, we observed the components of ABC transporters, the biosynthesis of unsaturated fatty acids, and phenylalanine metabolism pathways. The key metabolites and genes affecting seed coat permeability were thiamine and salicylic acid. Furthermore, there were 13 and 14 genes with correlation coefficients greater than 0.8 with two key metabolites, respectively, and the $|\log_2$Fold Change$|$ of these genes were greater than 1 at different development stages. Meanwhile, pathogenesis-related protein 1 and phenylalanine ammonia-lyase play an important role in regulating the formation of compounds. Our results

outline a framework for understanding the development changes during seed growth of *E. nutans* and provide insights into the traits of seed coat permeability and supply a great significance value to seed production and quality evaluation.

KEYWORDS

seed coat, permeability, development, metabolome, transcriptome, *Elymus nutans*

Introduction

As part of the seeds, the seed coat or testa, takes an important function in the growth and development of plants (Ben-Tov et al., 2015). It can protect the embryo from mechanical damage and pathogen infection in mature seeds and store nutrients to prevent loss in developing seeds (Mendu et al., 2011; Ezquer et al., 2016). The capacity of the testa to defend embryo tissue away from mechanical and oxidative stress is important for seed viability (Chen et al., 2014b; Loubéry et al., 2018). The seed coat may be used in an aqueous solution as a immerse or during priming to inject active substance into the embryo (Salanenka and Taylor, 2011). In seed coats, permeability is a common phenomenon. The permeability of the seed coat is generally considered to control the exchange of water, gas, and nutrients between the seed embryo and the outside world and plays a prominent part in seed germination, dormancy, storage, and species reproduction (Dongen et al., 2001; Briggs and Morris, 2008; Li et al., 2010). Different plant seed coats have different osmotic characteristics. According to the osmotic capacity of different substances, seed coat permeability is generally divided into three types: non-permeable, semi-permeable, and permeable. Non-permeable seed coats are one of the main factors causing seed dormancy (physical dormancy), which is mainly because the tight seed coat structure makes it difficult for the seed to absorb water and air (Jayasuriya et al., 2007; Orozco-Segovia et al., 2007; Brancalion et al., 2010). Semi-permeable seed coats ordinarily have selective permeability to the exchange of materials inside and outside the seed coat. They can absorb water and swap gas, but they can limit the permeation of macromolecules (Beresniewicz et al., 1995; Amritphale et al., 2010; Zhou et al., 2013b). Permeable seed coats allow not only the exchange of water and gas, but also the infiltration of some large numbers of macromolecular substances (Barua et al., 2012).

Seed coat permeability is affected by the environment of seed development, such as maternal temperature signaling, which regulates seed coat permeability properties. In soybeans, high temperatures during grain filling raise the occurrence of impermeable seed coat (hard seed), which is related to disunion and deferred emergence and germination (Kebede et al., 2014). On the contrary, in *Arabidopsis thaliana*, lower maturation temperatures augments testa permeability to tetrazolium, and mutants with superior seed coat permeability cannot easily enter strongly dormant states (MacGregor et al., 2015). Meanwhile, low-temperature plasma treatment modifies *A. thaliana* testa permeability due to the change of lipid compounds and structure (Bafoil et al., 2019). In addition to temperature, water content is another main factor influence on permeability. The testa permeability of *Astragalus adsurgens* was induced by reduced moisture content, which caused seed dormancy (Jaganathan et al., 2019). The seed coat permeability is also determined by testa thickness and composition. The cell wall thickness in the outer integuments varies according to seed development. Along with a build-up of polyphenolic compounds, callose, and lipids cause changes in seed coat permeability (Janská et al., 2019).

Seed coat permeability was generally thought to contribute toward seed treatment, especially in seed detection. Under eco-friendly pest management, the penetration ability of tomato seed coat is the key factor for exogenous jasmonic acid to successfully mediate the pest defense system (Mouden et al., 2020). Different permeability characteristics make it difficult to measure seed vigor using the electrical conductivity (EC) and tetrazolium chloride (TTC) staining methods. In the detection of orchid seed vigor, the TTC method cannot accurately identify whether the seed is actually alive; thus, a rapid method to evaluate the permeability level of seed coat needs to be developed (Magrini et al., 2019). In previous studies, the TTC and EC methods were used to determine the seed vigor of rice seeds with whole testa and poaceae seeds with punctured seed coat. The results demonstrated that the seed coat permeability was the primary factor limiting the application of seed vigor detection method (Sun et al., 2015, 2018).

Many studies on seed coat permeability have been conducted. However, most of them focus on understanding its characteristics, and few have investigated the regulation mechanism of seed coat permeability. In *Arabidopsis*, the content of condensed-tannin decreased seed permeability to tetrazolium dye, while *rgl2-1* mutants showed increased seed coat permeability; suberin acted as a hydrophobic obstacle to control the movement of gases, water, and solutes and as an antimicrobial obstruction (Vishwanath et al., 2013; Chen et al., 2014a). Transparent testa (*tt*) mutations of *Arabidopsis* have

anomalously high seed coat permeability because the cuticles of developing *tt* mutant integument have meaningful structural faultiness, which are related to increasing cuticle permeability (Loubéry et al., 2018). Phenylpropanoid gene expression in seeds was correlated with high concentrations of seed coat procyanidins, which leads to decreased seed coat permeability (MacGregor et al., 2015). Molecular mapping and inheritance studies identified a single dominant gene *isc* (impermeable seed coat) that may be important in developing cultivars of soybean with permeable seed coat (Kebede et al., 2014). Seed coat impermeability is controlled by the product of genes (*GmHs1-1*, *qHS1*) that encode calcineurin-like transmembrane proteins, endo-1,4-β-glucanase and β-1,4-glucans, which reinforce testa impermeability (Jang et al., 2015; Sun et al., 2015; Janská et al., 2019).

Elymus nutans is a kind of perennial gramineous forage, which is widely distributed in the alpine region at an altitude of 3,000–5,000 m (Fu et al., 2014; Liu et al., 2019). It has strong cold resistance and can safely overwinter at low temperature (Ding et al., 2020). *E. nutans* has high yield, long growing season, and drought and trample resistance, which makes it an important grass species for grassland establishment and ecological environment construction in the alpine region (Fu et al., 2014). Previous studies have found that *E. nutans* seed coat is semi-permeable, making it difficult for EC and TTC methods to detect seed vigor. The semi-permeable property of the seed coat is owing to the existence of a special structure called the semi-permeable layer. This layer can restrict or impede the exchange of solutes while permitting the penetration of external and internal water and gas, which provides valuable protection to support plant health and protect plant growth, development, and germination.

However, in *E. nutans*, the semi-permeable layer is situated in the outermost layer of the seed coat and originates from the outer layer of the inner integument at 10 days post-anthesis (dpa) (Zhou et al., 2013b). The purpose of this study is to reveal the main substances and key genes associated with *E. nutans* seed coat permeability variation using metabolome and transcriptome methods and to provide a theoretical basis for the formation mechanism of this characteristic. The results will supply a great significance value to seed production and quality evaluation.

Materials and methods

Experimental materials

Seed collection

The seeds of wild *E. nutans* were taken from a field place: Zeku Country, Qinghai Province, China (100°58′E, 35°23′N), from July to August 2020. The region belongs to a typical semi-temperate climate and the general characteristics are low temperature, insufficient heat, short frostless period and strong solar radiation. Totally 1,000 spikelets with flowers were tagged at anthesis, and samples were collected at 8, 18, and 28 dpa. The seeds at each development stage were divided into three parts. One part was dried naturally to test its physiological indicators, including lanthanum trace analysis. In the second part, the lemmas were removed from fresh seeds, and the seeds were fixed in 4% glutaraldehyde for observation of seed coat structure. In the third part, the seed coat was removed from the naked seed with a sharp scalpel and immediately fixed in liquid nitrogen for metabolome and transcriptome analysis.

Seed quality index

Thousand-seed weight was measured from the average weight of eight repetitions of 100 seeds. The petri dish paper germination method was used to test seed germination percentage on four repetitions of 50 seeds in each development stage incubated in a light incubator (GXZ-280B, Ningbo, China) at 25°C according to GB/T 2930.4-2017. The number of germinating seeds was counted every day during the experiment, and the seedlings were kept in the petri dish. The seed germination rate was calculated at the end of 12 days. On the basis of the above germination test, the germination energy, vigor index, and seedling and root length of seedlings grown in the first 5 days were calculated. The germination energy is the cumulative number of germinations in the first 5 days divided by the sum of tested seeds multiplied by 100%. The vigor index is the sum of the average seedling length and average root length multiplied by the germination percentage. The seed moisture content was determined using the high-constant-temperature oven method described by the International Seed Testing Association.

Seed coat permeability characteristic

Electrical conductivity and imbibition rate testing

The EC and imbibition rate were determined using four repetitions of 50 seeds from three development stages. According to the Hampton and TeKrony method, each replicate was weighed, shifted to 100 mL distilled water, and stored at 20°C for 24 h. A conductivity meter (DDSJ-319L, Shanghai, China) was used to test conductivity. Then, the seeds were surface desiccated and weighed. The imbibition rate was indicated to the wet weight minus the dry weight divided by the dry weight.

Light microscopy and lanthanum tracer energy dispersive X-ray analyses

The fixed caryopsis coat including endosperm tissue was embedded in Technovit7100 and cut into 1 μm semi-thin slices according to the methods described by Zhou et al. (2013a). The

slices were stained with 0.05% aniline blue in 0.1 M phosphate buffer (pH 8.2) for 20 min. The structure of the caryopsis coat was examined, and images were taken by means of a compound microscope (Eclipse E 100, Nikon, Tokyo, Japan).

In order to detect the penetration characteristics of the seed coat, lanthanum nitrate tracer method combined with energy dispersive X-ray (EDX) technology was used to observe the permeability of lanthanum in seeds. The dried seeds without lemmas were soaked in distilled water for 20 h at 20°C, and the seeds with intact seed coat were selected using a stereomicroscope (Zoom-600, Shanghai, China). The intact seeds were incubated in 1% (W/V) lanthanum nitrate solution at 20°C. After 24 h, the seeds were taken out and dried naturally. The dried seeds were cut into halves with a razor blade and fixed to an aluminum shell. The structure of the seeds was observed using the JEM 5600 LV scanning electron microscope (SEM) with EDX spectroscopy (KEVE2, United States). Then, the seeds were sputter coated and viewed under the backscattered electron imaging mode at 20 kV. The snowflake crystals located in seed samples were analyzed for 20 s. Lanthanum nitrate amounts were based on the area of L-series.

Fluorescence staining indicated seed coat permeability

To further examine the seed coat permeability, the dry seeds without lemmas were immersed in 20°C distilled water for 24 h. Then, the seeds of three development stages with intact seed coat were placed in 1% Rhodamine B prepared with absolute alcohol for 24 h. Thereafter, the penetration of the fluorescent agent into the seeds was examined using a fluorescence microscope (Leica MDG41, Singapore).

Non-targeted metabolite analysis of seed coat permeability

Metabolite extraction

About 50 mg of seed coat of *E. nutans* was placed in a 1.5 mL eppendorf tube for non-targeted metabolite profiling with four replicates per development stage. All samples were added with 800 μL of methanol and ground for 90 s by using a high-throughput tissue grinding mill (SCIENTZ-48, Zhejiang, China). The mixture was blended by ultrasound for 30 min and centrifuged at 4°C for 15 min at 12,000 rpm. Then, 5 μL DL-o-Chlorophenylalanine was added to the supernatant and transferred into the sample vials.

Non-targeted metabolite analysis by LC-MS/MS

Metabolites were separated using an ultra-high-performance liquid phase chromatograph (UHPLC, Waters ACQUITY Ultra Performance LC, United States) equipped with the ACQUITY UPLC HSS T3 Column (100 mm× 2.1 mm × 1.8 μm). The mobile phase A was 0.05% formic acid, and the mobile phase B was acetonitrile. The mobile phase was used to separate the metabolites while the flow rate was maintained at 300 μL/min. The parameters of gradient elution were set as follows: 0 min, 95% A, 5% B; 1 min, 95% A, 5% B; 12 min, 5% A, 95% B; 13.50 min, 5% A, 95% B; 13.60 min, 95% A, 5% B; 16 min, 95% A, 5% B. The injection volume was 4 μL and the automatic temperature was 4°C.

Then, the analytes were detected by using a Q Exactive™ Mass Spectrometer (MS/MS, Thermo Fisher Scientific, United States) with electron spray ionization including positive and negative ion mode to obtain mass spectral data, which were collected by full scan mode with m/z 70–1050 and dd-MS2 mode with TopN = 10, respectively. The following settings were used: sheath gas flow rate, 45 Arb; auxiliary gas flow rate, 15 Arb; capillary temperature, 350°C; full ms resolution, 70,000; MS/MS resolution, 17,500; collision energy, 15/30/45 eV; spray voltage, 3.0 kV (positive) or −3.2 kV (negative). In order to enhance the accuracy of identified substance, the following criteria were used: accurate mass with variation less than 10 ppm and MS/MS spectra with high forward and reverse scores based on comparisons of the ions present in the experimental and library spectrum entries (Shen et al., 2020). The metabolites were identified with mzCloud, FiehnLib and Chinese Natural Product databases.

Data preprocessing and annotation

Preprocessing and quality assessment of raw data were performed. Single peaks and total ion current of each sample were filtered and normalized with R package metaX. The missing value in the original data was simulated by half of the minimum value. Then, the low mass ions were removed, and ions with relative standard deviation >30% were filtered out. After obtaining the matrix data, the original data were annotated into the KEGG,[1] HMDB,[2] METLIN,[3] ChEBI,[4] and PubChem[5] databases.

The resulting three-dimensional data involving the sample name, peak number, and normalized peak area were supplied to R package metaX for principal component analysis (PCA) and orthogonal partial least square-discriminant analysis (OPLS-DA). PCA exhibited the distribution of original data. To acquire a higher level of group separation and a better understanding of the variables responsible for classification, supervised OPLS-DA was applied. The first principal component of variable importance in projection (VIP) was obtained.

1 http://www.kegg.jp
2 http://www.hmdb.ca/
3 http://metlin.scripps.edu/index.php
4 http://www.ontobee.org/ontology/CHEBI
5 https://pubchem.ncbi.nlm.nih.gov/

Identification of differential metabolites and enrichment analysis

We used a combination of the p-values of Student's t-tests with the VIP value to filter DEMs, with extra screening criteria ($p < 0.05$, VIP > 1). In addition, commercial databases including KEGG were used to search for the pathways of DEMs, which were identified by the second mass spectrometry.

Characteristics of transcriptome in seed coat permeability

Total ribonucleic acid extraction and sequencing library preparation for transcriptome

High-quality total ribonucleic acid (RNA) was extracted from the *E. nutans* seed coat using the TRIzol reagent (TransGen, Beijing, China) following the manufacturer's instructions. The purity and quality of RNA were gauged and checked. The RIN value of every sample was greater than 8.4. Every development stage had four replicates. About 1.5 μg of RNA was used to build the RNA-seq library with Illumina's NEBNext® Ultra™ RNA Library Prep Kit (Illumina Inc., San Diego, CA, United States). The purified mRNA was carried out by divalent cations. First and second-strands cDNAs were synthesized using random hexamer primer and DNA Polymerase I and RNase H (Invitrogen, Carlsbad, CA, United States), respectively. The remaining molecules were transformed into blunt ends via exonuclease/polymerase activities. The library preparations were sequenced on an Illumina HiSeq 4000 platform and generated paired-end reads of 150 bp.

Sequence read mapping and *de novo* assembly

Raw data in FASTQ format were first processed to filter out rRNA by using bowtie2 (Langmead and Salzberg, 2013). Clean data were acquired by removing reads containing poly-N, adapter, and low-quality reads from raw data through in-house Perl scripts. At the same time, GC-content, Q20, Q30, and sequence duplication level of clean data were calculated. According to the Trinity method including the Inchworm, Chrysalis and Butterfly program, the spliced transcript was used as the reference sequence. Clean reads were separately mapped to them using SALMON software (version 0.14.1)[6] and *de novo* assembled with parameters of Kmer = 25 (Grabherr et al., 2011). Then, CD-HIT,[7] Corset and BUSCO software were used to classify transcripts, remove redundancy, and evaluate the quality of splicing, respectively.

Gene functional annotation, differential expression analysis, and transcription factor identify

All the assembled transcripts were analyzed by BLAST with seven public databases, including non-redundant protein sequence (Nr), nucleotide database (Nt), Pfam protein families database (Pfam), manually annotated and reviewed protein sequence database (Swiss-Prot), eukaryotic orthologous groups database (KOG), Gene Ontology database (GO), and Kyoto Encyclopedia of Genes and Genomes pathway database (KEGG). The expression levels of unigenes were standardized and calculated as the values of fragments per kilobase of transcripts per million mapped fragments (FPKM) during the assembly and clustering process (Trapnell et al., 2010). After normalizing the read count data with DESeq (Anders and Huber, 2010), the p-values of DESeq2 package analyses were adjusted via the Benjamini–Hochberg method to determine the false discovery rate (FDR) and identify differentially expressed genes (DEGs). Then, the p-value obtained from the test was corrected to obtain the q-value. The standard of differential gene expression screening is q-value < 0.05 (Robinson et al., 2010). To study the seed development significance of DEGs, the GO database was used for GO enrichment analysis of DEGs during the different processing of *E. nutans* using the GOseq (v1.22) software (Young et al., 2010). By estimating the bias of gene length, the probability of GO term enriched by different genes can be calculated more accurately. The assorted metabolic pathways of DEGs were analyzed by using the KEGG database (Mao et al., 2005). The statistical enrichment of DEGs was tested using KOBAS 2.0 web server, and a corrected p-value < 0.05 was considered to be significantly enriched in KEGG. All identified DEGs were blasted with PlantTFDB 4.0.[8] We chose prediction interface to input unigenes sequence in fasta format. Both the hmmscan and blast e-value were set as $1 \times e^{-5}$.

Time series analysis in seed coat development

To explore the new important unigenes that can affect seed coat penetration, time series analysis was conducted to excavate the transcription data. We removed genes whose FPKM values were less than 1 in all samples. Then, we used the log normalize data strategy to transform the data and selected STEM (v1.3.13) clustering for time series analysis. Subsequently, we drew the time sequence diagram. The maximum number of modules is 20 (the default is 50), the maximum change multiple is 2, and other parameters are default.

Quantitative RT-PCR validation

Quantitative RT-PCR analysis was processed on a CFX Connect qPCR detection system to verify the accuracy of the RNA-seq results. Six genes were randomly detected, and the

6 https://github.com/COMBINE-lab/salmon

7 http://www.bioinformatics.org/cd-hit/

8 http://planttfdb.gao-lab.org/

EnACT gene was used as a house-keeping gene (the primers are shown in **Supplementary Table 1**; Niu et al., 2017). The qPCR program run consisted of an initial denaturation step at 95°C for 15 min followed by amplification and quantification cycles repeated 40 times at 95°C for 10 s, 58°C for 20 s and 72°C for 30 s. We used the $2^{-\Delta\ \Delta\ CT}$ method to calculate the relative expression levels of genes.

Co-expression analysis of metabolome and transcriptome

The obtained DEMs and DEGs data were applied in the calculation of correlation coefficients (Spearman rank correlation test) using R through inhouse Perl script. Positive and negative correlations >0.8 and <−0.8 were considered for the construction of a dynamic network. The results were visualized using Cytoscape (version 3.7.1) and Origin 2018.

Statistical analysis

Statistical analyses of physiological data were performed using SPSS 20.0 (SPSS Inc., Chicago, IL, United States). The significance of differences in each processing was tested by one-way ANOVA and Duncan's multiple comparative analysis ($p < 0.05$).

Results

Physiological indexes and permeability characteristic during seed development of *Elymus nutans*

The seed viability and 1000-seed weight showed an upward trend with seed development, while moisture content, imbibition rate, and EC showed a downward trend (**Figure 1**). From 8 to 28 dpa, the seed viability increased from 27.50 to 95.50%, while the seed weight increased from 1.91 to 4.14 g. The seed moisture content decreased from 65.19% at 8 dpa to 25.33% at 28 dpa (**Figure 1A**), and the imbibition rate and EC decreased from 103.15 to 54.42% and 210.90–76.85 $\mu s \cdot cm^{-1} \cdot g^{-1}$, respectively (**Figure 1B**). A pericarp, two layers of inner integument, primary aleurone cell layer and free endosperm cells were observed in the seed coat at 8 dpa (**Supplementary Figure 1A**). Several parenchymal cells were disappeared, and the inner integument was developed to form the compact seed coat located outside of the aleurone layer at 18 dpa (**Supplementary Figure 1B**). Upon reaching maturation, the seeds contained a pericarp, seed coat, aleurone layer, and endosperm. The pericarp became a membranous tissue without cellular structure, and the seed coat still adhered tightly to the aleurone cell layer at 28 dpa (**Supplementary Figure 1C**).

The structure of the seed coat at 18 dpa was observed using SEM (**Supplementary Figure 2A**). Parts a, b, and c are the outside of the seed coat, and parts d and e are the inside contents of the seed. Combined with EDX, the lanthanum peaks detected on the L-series in parts a, b, and c showed that the lanthanum was deposited on the outside of the seed (**Supplementary Figures 2B–D**). On the contrary, the lanthanum peak was not detected in the seed interior (**Supplementary Figures 2E,F**). Seed coat permeability was detected based on penetration of Rhodamine B into seeds during seed development. Fluorescence in the seeds could be observed at 8 dpa, indicating that the dye was penetrated into the inner seed (**Figure 2A**), but the fluorescent dye was blocked on the outside of the seed which formed a circle at 18 and 28 dpa (**Figures 2B,C**), indicating that the seed coat permeability decreased.

Metabolomic changes and key metabolites associated with seed coat permeability during seed development

To characterize the metabolites associated with responsible to seed coat development in *E. nutans*, the varieties of metabolites were detected by UHPLC-MS/MS. MS signals from positive and negative ionization modes were analyzed for each sample to exclude commonly known adduct and fragment ions and to reduce signal redundancy. A total of 1887 metabolites were identified in the positive ion mode (POS) and negative ion mode (NEG) (**Supplementary File 1**). All the identified metabolites were classified into four groups by the hierarchical clustering heatmap (**Figure 3A**). There were 470, 442, and 443 upregulated and 359, 387, and 386 downregulated metabolites observed under the POS testing at 8, 18, and 28 dpa, respectively (**Figure 3B**). Taking 8 dpa as control, 252, 277 upregulated and 157, 193 downregulated DEMs were observed under the POS testing, respectively (**Figure 3C**).

KEGG analysis was performed on the DEMs identified by secondary mass spectrometry (**Figure 3D**). Under the POS, only 101 and 90 upregulated DEMs were annotated into 45 and 41 upregulated KEGG pathways (**Figure 3D**). However, under the NEG, 75 and 167 downregulated DEMs were annotated into 34 and 54 downregulated KEGG pathways (**Figure 3D**).

PCA was performed on the basis of peak areas. The first principal component (PC1), which accounts for 86.1 and 53.6% of the variance separates the different development stages, while PC2 (5.5 and 31.2%) effectively separates the differential samples under the POS and NEG (**Figures 4A,B**). For OPLS-DA, the predicted principal component scores of the first principal component were 48.8 and 41.4%, while the orthogonal principal component scores were 10.1 and 9.4% under the POS and NEG testing, respectively (**Figures 4C,D**).

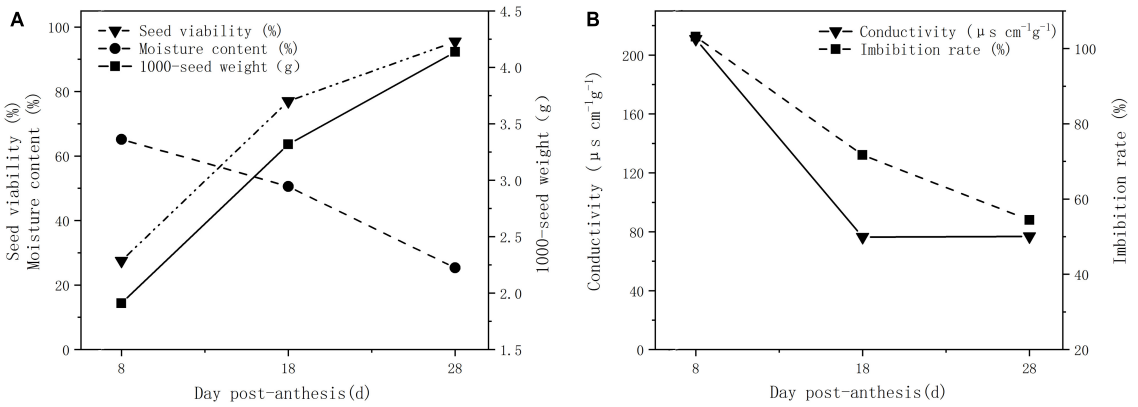


FIGURE 1
Physiological indexes of plant seeds at different developmental stages. **(A)** Seed viability, moisture content, and 1,000-seed weight. **(B)** EC and imbibition rate.

To accurately identify the metabolites affecting seed coat permeability, we further analyzed the substances identified by secondary mass spectrometry. According to the p-value < 0.05 of the KEGG annotation, the ABC transporters; biosynthesis of unsaturated fatty acids (UFA); and sulfur relay system pathways were significantly enriched under the POS analysis (Table 1). Meanwhile, under the NEG analysis, citrate cycle (TCA cycle); alanine, aspartate and glutamate metabolism; C5-branched dibasic acid metabolism; lysine degradation; and glyoxylate and dicarboxylate metabolism were the significantly enriched pathways at 8 and 18 dpa (Table 2).

To determine the key metabolites that affected seed coat permeability, we focused on the upregulated metabolites that were significantly enriched in KEGG, and 11 metabolites exhibited different development stages under the POS and NEG (Supplementary Table 2). We performed heat-map analysis on the differential metabolites in each treatment combination and screened out compounds with more significant expression at 18 and 28 dpa than 8 dpa. Taking 8 dpa as control, 4-coumaric acid (241), L-glutamic acid (659), DL-alpha-tocopherol (2395), thiamine (1753), alpha-lactose (2308), and salicylic acid (SA) (123) were visualized in a plot, and their contents showed an upward trend at 18 dpa (Figure 5A). With seed development, thiamine (1753) and SA (123) continued to be deposited in the seed coat, which contributed to the enhancement the tightness and permeability of the seed coat at 8 dpa to physiological maturity (Figure 5B).

Transcriptional analysis revealed the key genes affecting seed coat permeability

To reveal the mechanism underlying the development stage of seed coat for *E. nutans*, 12 RNA-seq libraries from the early stage of seed coat formation (8 dpa), seed coat maturity (18 dpa), and seed physiological maturity (28 dpa) were constructed. Each development stage had four biological replicates. An overview of the sequencing is presented in Supplementary Table 3. More than 96 and 90% of bases had a q-value more than 20 and 30 (an error probability of 0.03%), respectively. The GC-content ranged between 56.35 and 60.53% (Supplementary Table 3). All original FASTQ data files were submitted to the NCBI Sequence Read Archive (SRA) under accession number PRJNA773127. After removing low-quality reads, a total of 570 million clean reads were produced from three samples (Supplementary Table 3). Trinity was used to generate 240,172 transcripts with N50 of 1,418 bp and N90 of 466 bp (Supplementary Table 4). Among them, 128,900 were unigenes, where less than 300 bp has 14,633 unigenes, 300–500 bp has 35,944 unigenes, 500–1,000 bp

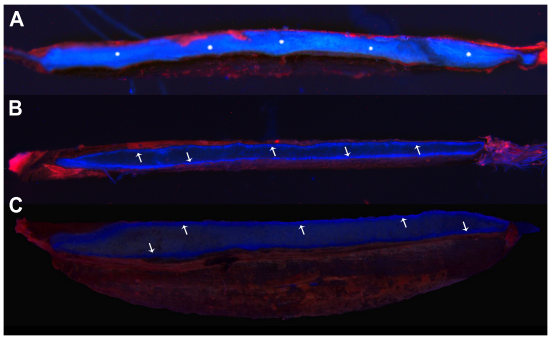


FIGURE 2
Fluorescence staining indicates the osmotic properties of the seed coat. **(A–C)** Samples at 8, 18, and 28 dpa, respectively. The snowflakes showed that the fluorescent dye infiltrated into the inside seed, and the arrows indicated that the dye was blocked outside the seed, forming a fluorescent line.

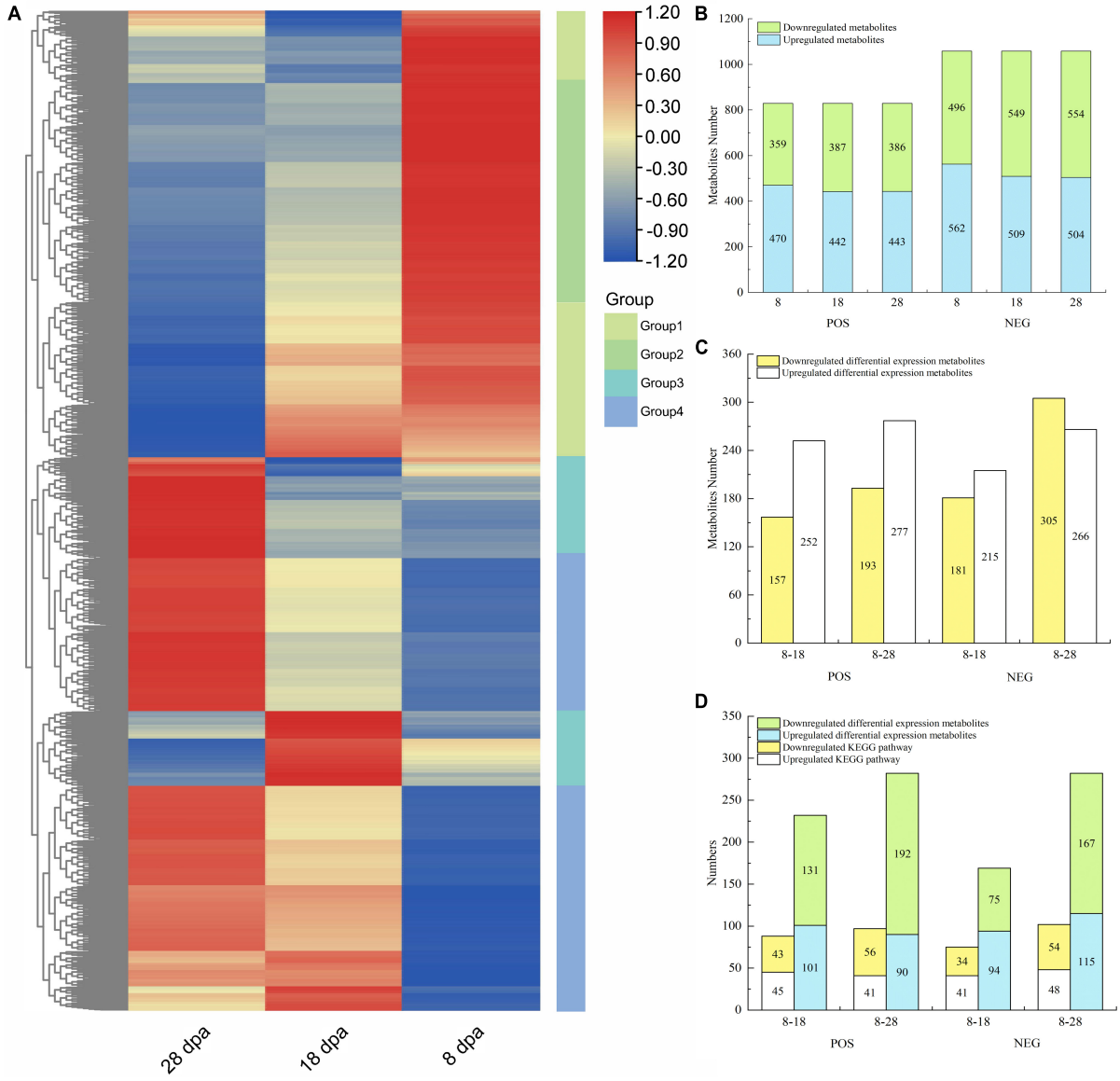


FIGURE 3
Metabolite abundance differences between three different development stages of *E. nutans* studied on the basis of LC-MS analysis. **(A)** Hierarchical clustering heatmap of all metabolites under the POS and NEG. Each column corresponded to 8, 18, and 28 dpa, and four groups were separated; **(B)** up- and downregulated metabolites under the POS and NEG; **(C)** up- and downregulated DEMs under the POS and NEG; **(D)** all DEMs enriched in KEGG and the KEGG numbers.

has 41,569 unigenes, 1,000–2,000 bp has 24,818 unigenes, and >2,000 bp has 11,936 unigenes. A total of 64,988, 83,297, 8,609, 36,154, 48,273, 43,644 and 28,740 unigenes were annotated into the above seven databases, which occupied 50.42, 64.62, 6.68, 28.05, 37.45, 33.86, and 22.3%, respectively (Supplementary Table 5).

After obtaining the transcriptome data, PCA was performed on 12 samples. PC1 accounted for 66.2%, while PC2 (11.5%) effectively separated the differential samples under the three development stages (Figure 6A). DEGs between the three developmental stages were calculated and normalized using the DESeq method. Adjusted p-value < 0.05 was set as the threshold for significant differential expression. We found 2,056 and 13,468 unigenes, which showed differential expression in development samples. Compared with 8 dpa, 886 upregulated unigenes and 1,170 downregulated unigenes were identified at 18 dpa, while 4,907 upregulated unigenes and 8,561 downregulated unigenes were identified at 8 and 28 dpa (Figure 6B). To identify DEGs in different development stages, the overlaps in each comparison were exhibited in a Venn diagram (Figures 6C,D). The overlap of up- and down-regulated genes was between 8 and 18 dpa, which showed

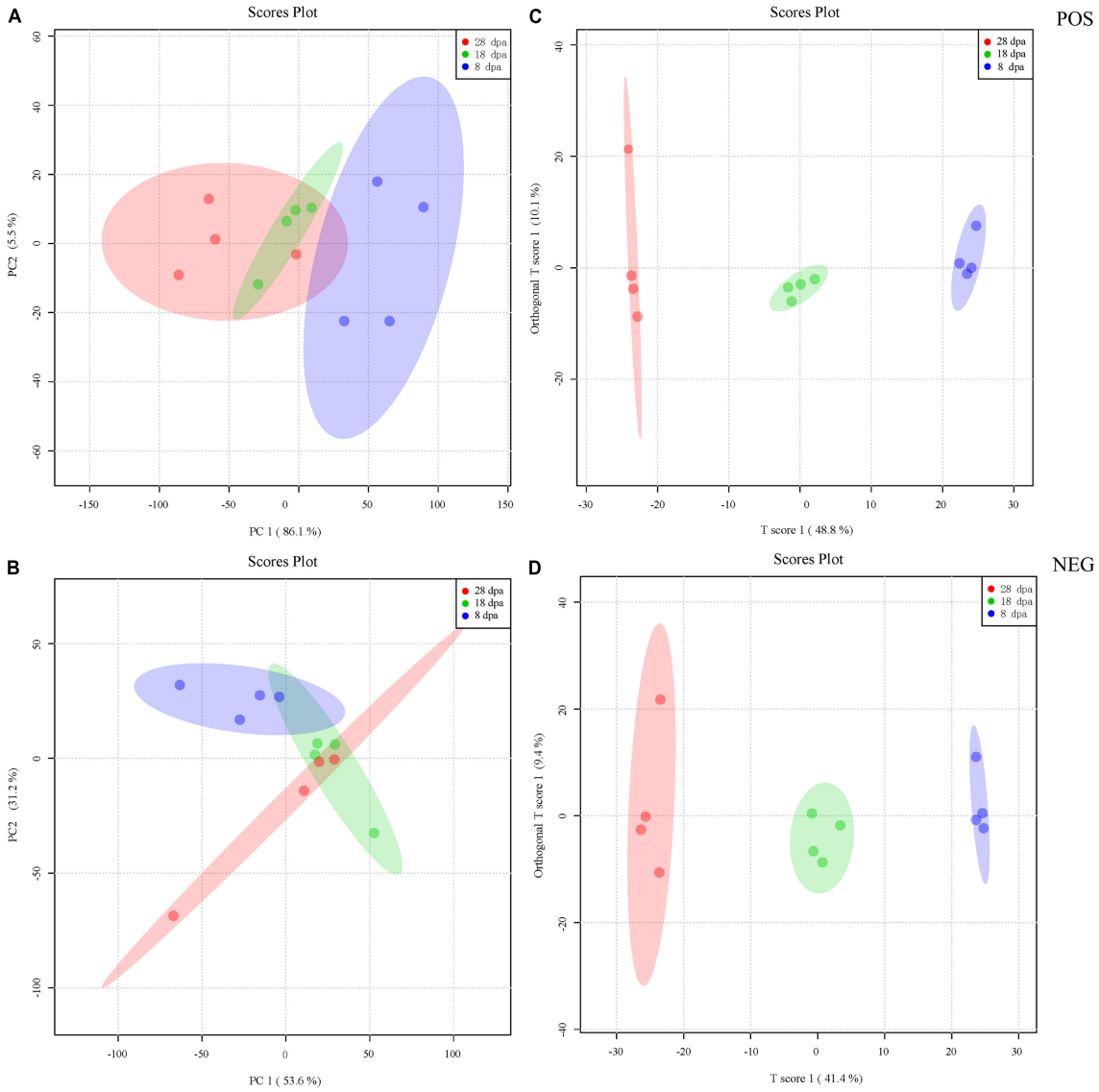


FIGURE 4
PCA and OPLS-DA score plot derived from the LC-MS spectra of extracts obtained from seed development stages at 8 (blue), 18 (green), and 28 dpa (red). **(A,B)** PCA was first used to obtain an overview of metabolic profiles of the 8, 18, and 28 dpa conditions under the POS and NEG. The different plots represent combinations of the first and second principal components (PC1 and PC2). The proportions of explained variance are shown in brackets. **(C,D)** The OPLS-DA score plot was then employed to maximize the separation according to the classification of samples at different stages of seed development acquired under the POS and NEG. The proportions of explained variance are shown in brackets. Each development stage contains four replicates.

111 and 19 DEGs, suggesting that seed coat development has greatly changed (Figures 6C,D). The overlap of the up- and downregulated unigenes between 8 and 28 dpa were 4,132 and 7,410 DEGs (Figures 6C,D). The results showed that these DEGs were the main factors causing the changes in seed coat permeability.

To identify the primary biological processes that were expressed at different development stages, we showed each GO enrichment analysis of up- and down-regulated DEGs at different physiological stages depending on FDR < 0.05 (Supplementary File 2). On the basis of GO enrichment, we chose the top five most significant accessions to interpret the seed coat permeability changes. At 8 and 18 dpa, the upregulated genes were largely located in the pathway associated with chitin binding and metabolic process while the downregulated genes were mainly located in the pathway associated with cell wall organization (Supplementary Table 6). With seed development, cell wall organization or biogenesis was the significantly enriched pathway that affected testa development. At the same time, xanthine dehydrogenase activity, xanthine

10.3389/fpls.2022.970957

TABLE 1 Positive analysis significant enrichment of KEGG pathway.

	POS up				POS down			
	Pathway	Hits	Raw_p	FDR	Pathway	Hits	Raw_p	FDR
8–18	ABC transporters	5	0.0006851	0.01781	Aminoacyl-tRNA biosynthesis	4	0.000117	0.00351
					Tropane, piperidine and pyridine alkaloid biosynthesis	4	0.0003347	0.003565
					Biosynthesis of amino acids	5	0.0003565	0.003565
					Lysine degradation	3	0.002006	0.01504
					2-Oxocarboxylic acid metabolism	4	0.004287	0.02322
					ABC transporters	4	0.004644	0.02322
					Valine, leucine and isoleucine biosynthesis	2	0.006163	0.02641
					Lysine biosynthesis	2	0.01397	0.05241
					Valine, leucine and isoleucine degradation	2	0.0198	0.06599
					Cyanoamino acid metabolism	2	0.02256	0.06767
					Nicotinate and nicotinamide metabolism	2	0.03282	0.08952
					Flavonoid biosynthesis	2	0.03156	0.1326
8–28	Biosynthesis of unsaturated fatty acids	2	0.04329	0.2215	Aminoacyl-tRNA biosynthesis	4	0.0002518	0.007554
	Sulfur relay system	1	0.04923	0.2215	Tropane, piperidine and pyridine alkaloid biosynthesis	4	0.0007107	0.009118
					Biosynthesis of amino acids	5	0.0009118	0.009118
					ABC transporters	5	0.001243	0.009321
					Lysine degradation	3	0.003465	0.02079
					2-Oxocarboxylic acid metabolism	4	0.008608	0.03793
					Valine, leucine and isoleucine biosynthesis	2	0.008849	0.03793
					Pantothenate and CoA biosynthesis	2	0.01482	0.05558
					beta-Alanine metabolism	2	0.01677	0.05592
					Lysine biosynthesis	2	0.0199	0.0597
					Valine, leucine and isoleucine degradation	2	0.02805	0.07651
					Cyanoamino acid metabolism	2	0.0319	0.07974
					Glycine, serine and threonine metabolism	2	0.03874	0.0894
					Nicotinate and nicotinamide metabolism	2	0.0461	0.09878

TABLE 2 Negative analysis significant enrichment of KEGG pathway.

	NEG up				NEG down			
	Pathway	Hits	Raw_p	FDR	Pathway	Hits	Raw_p	FDR
8–18	Phenylalanine metabolism	2	0.01433	0.09876	Citrate cycle (TCA cycle)	2	0.004671	0.09251
	Ubiquinone and other terpenoid-quinone biosynthesis	2	0.01975	0.09876	Alanine, aspartate and glutamate metabolism	2	0.009069	0.09251
	Plant hormone signal transduction	1	0.03739	0.1246	C5-Branched dibasic acid metabolism	2	0.01322	0.09251
					Lysine degradation	2	0.02749	0.1443
					Glyoxylate and dicarboxylate metabolism	2	0.04092	0.1719
8–28	Phenylalanine metabolism	2	0.02522	0.2726	Citrate cycle (TCA cycle)	2	0.005316	0.09995
	Plant hormone signal transduction	1	0.04957	0.2726	Alanine, aspartate and glutamate metabolism	2	0.0103	0.09995
					C5-Branched dibasic acid metabolism	2	0.01499	0.09995
					Lysine degradation	2	0.03107	0.1553
					Glyoxylate and dicarboxylate metabolism	2	0.04612	0.1845

catabolic process, and xanthine metabolic process were the three most significantly upregulated pathways during seed coat development (Supplementary Table 6).

KEGG pathway enrichment analysis showed that 396 and 1,160 upregulated DEGs were enriched in 66 and 101 different pathways while comparing 8 with 18 dpa and 8 with 28 dpa (Supplementary File 3). However, 503 and 3,035 downregulated DEGs were involved in 83 and 120 pathways in *E. nutans*, respectively (Supplementary File 3). According to the KEGG pathway analyses, a heatmap reflecting the main enrichment process of seed coat development was constructed (Figure 7). In these pathways, the 8 dpa samples were used as control. Amino sugar and nucleotide sugar metabolism and MAPK signaling pathway—plant pathways had the most upregulated DEGs at 18 dpa, suggesting improved pathway progress in response to seed coat permeability. Meanwhile, the starch and sucrose metabolism, phagosome, and fatty acid elongation pathways had enriched DEGs at 28 days (Figure 7).

Transcription factor (TF) analysis was performed on all identified DEGs, and the genes were randomly selected for

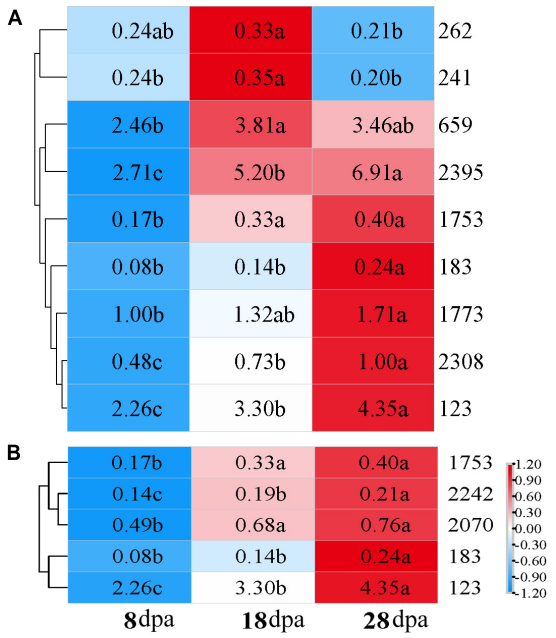


FIGURE 5
Expression of key metabolites in different samples. Distribution of key identified metabolites across different samples compared with the different development stages. Values represented in the scale correspond to the relative quantitative value of metabolites. Letter indicate the significant difference of relative quantitative values of three stages. **(A)** 8 dpa with 18 dpa; **(B)** 8 dpa compared with 28 dpa. 262: choline, 241: 4-coumaric acid, 659: L-glutamic acid, 2,395: DL-alpha-tocopherol, 1,753: thiamine, 183: hydrocinnamic acid, 1,773: adenosine, 2,308: alpha-lactose, 123: salicylic acid (SA), 2,242: docosapentaenoic acid, 2,070: eicosapentaenoic acid.

quantitative RT-PCR verification (**Supplementary Figure 3**). The largest group of TFs was the basic helix–loop–helix family, followed by B3, whereas other TFs belonged to the FAR1, NAC, and WRKY families, which including 404, 325, 300, 291, and 284 unigenes (**Supplementary Figure 3** and **Supplementary File 4**). To validate the accuracy of TF sequencing, the relative expression level of six randomly selected unigenes was examined using qRT-PCR analysis (**Supplementary Figure 4**). These results demonstrated alike expression patterns to the results of RNA-seq.

To identify the key gene that contributed to seed permeability, we used time series analysis to explore the RNA-seq results of three developmental stages of *E. nutans*. The analysis produced 16 profiles including 8 profiles that were significantly different ($p < 0.001$). The 8 profiles were separated into 3 groups, and the same color showed the same group. The black lines showed the changing trend of genes with seed developments (**Figure 8A**). At 8, 18, and 28 dpa, we focused on the gene showing an increasing trend. The results showed that profile 8, 12, and 13 had the same trends, and all the clustered genes were analyzed by KEGG (**Figures 8B–D**). Tryptophan metabolism was the most significantly enriched pathway, which is similar to that of heatmap analysis.

Correlation between differentially expressed metabolites and differentially expressed genes and their insights into seed coat development in *Elymus nutans*

To explore the regulatory genes affecting metabolites, we conducted the analysis into two parts. One is to use the significant pathway enriched by metabolites to find the same pathway with RNA-seq and then identify the DEGs in this pathway. The other is to analyze the correlation between all DEMs identified by MS2 and DEGs obtained by RNA-seq analysis. The relative quantitation and FPKM values of the two omics were used to find the genes.

For the first analysis, 27 genes were annotated into ABC transporters, phenylalanine metabolism, ubiquinone and other terpenoid-quinone biosynthesis, and plant hormone signal transduction at 8 with 18 dpa (**Figure 9A**). When we compared the immature seeds with mature seeds, 110 DEGs were annotated in the biosynthesis of UFA, sulfur relay system, phenylalanine metabolism, and plant hormone signal transduction (**Figure 9B**). To explore the regulatory mechanism of the formation of key compounds affecting seed coat permeability, pearson correlation coefficient was used to analyze the correlation between metabolites and genes among samples. We built a metabolite-gene correlation network for two comparisons using 76 and 78 DEMs under the POS and 67 and 80 DEMs under the NEG and 2,056 and 13,468 DEGs in each comparison (8 and 18 dpa and 8 and 28 dpa), respectively. The DEMs in key pathways and the corresponding DEGs were analyzed and visualized by Cytoscape. The results show that depending on different combination, 659 (L-glutamic acid), 1753 (thiamine), 2308 (alpha-lactose), 2395 (DL-alpha-tocopherol) and 123 (SA) had the relationship with those genes which included in 27 DEGs (**Figure 9C**). However, 2070 (eicosapentaenoic acid), 2242 (docosapentaenoic acid), and 123 (SA) had the relationship with 110 DEGs (**Figure 9D**).

Response of key unigenes to thiamine and salicylic acid formation

On the basis of the correlation coefficient between metabolites and genes, we chose *p*-value less than 0.05 and correlation coefficient $|r| > 0.8$. Thiamine and SA had the corresponding genes that met the screening criteria. In the comparison combination of 8 and 18 dpa, 14 and 4 genes had a correlation coefficient $|r| > 0.8$, which were related

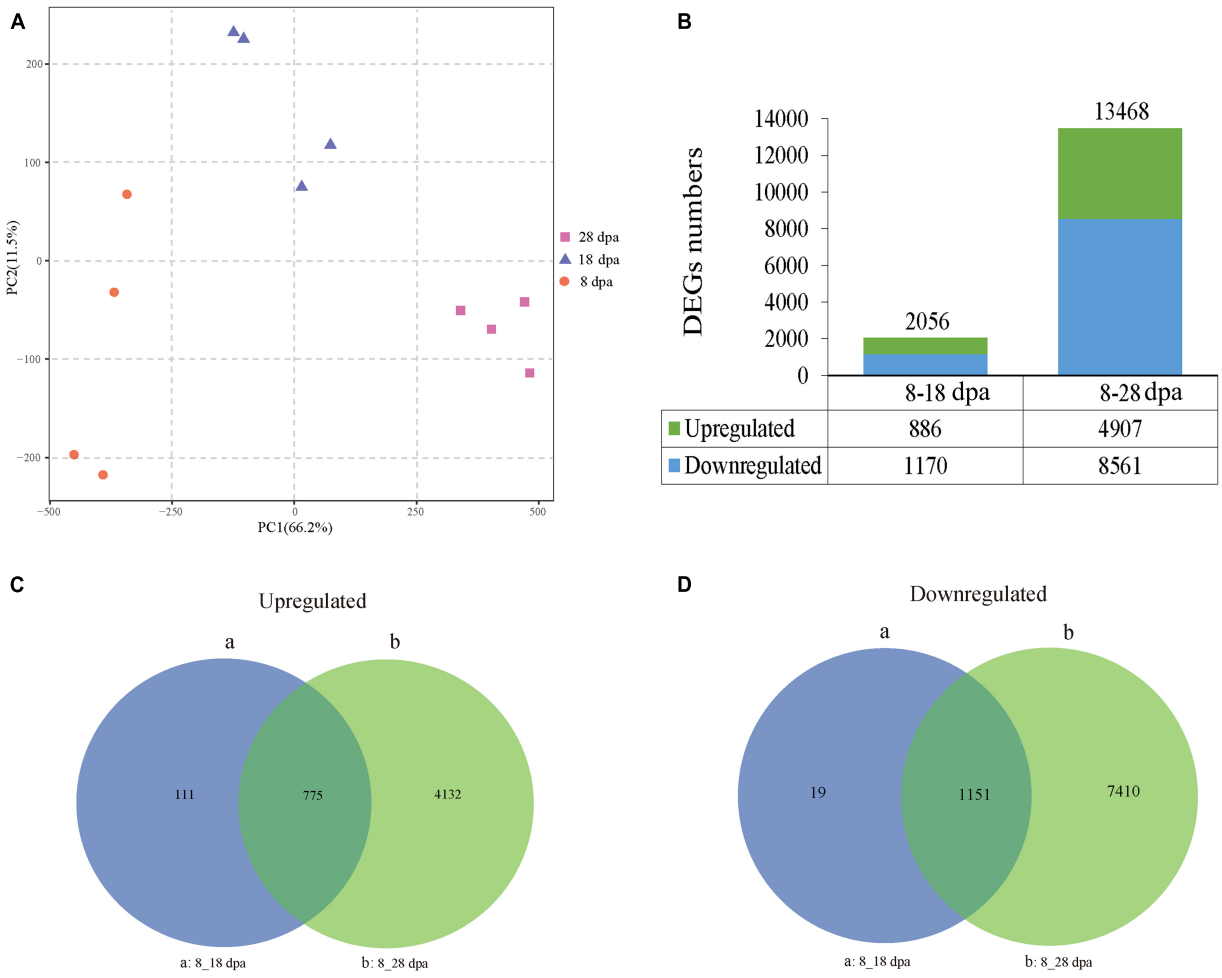


FIGURE 6
PCA and DEGs during seed development. **(A)** PCA was first used to obtain an overview of RNA-seq of the 8, 18, and 28 dpa conditions. Different plots represent combinations of the first and second principal components (PC1 and PC2). The proportions of explained variance are shown in brackets. **(B)** Numbers of DEGs at 8–28 dpa. The upper and lower parts of each bar represent upregulated and downregulated genes. The total numbers of DEGs are located at the top of the bar in each group. The numbers of up- and downregulated expressed genes are listed at the bottom of the figure. **(C,D)** Venn diagrams show the number of DEGs across three comparisons. (a,b) Show the comparisons of different unigenes identified between 8 dpa vs. 18 dpa and 8 dpa vs. 28 dpa, respectively.

to thiamine and SA, respectively (Supplementary File 5). For the immature and mature seed coats (8 and 28 dpa), 14 genes contributed to seed coat permeability development when SA were significantly deposited in the seed coat (Supplementary File 5).

After deduplication of all DEGs obtained from the analysis, the fold change of each gene in different comparisons was analyzed, and selected genes with |Log_2 Fold Change| greater than 1 were mapped (Figure 10). Taking 8 dpa as control, 14 and 15 genes regulated thiamine and SA, respectively. Meanwhile, pathogenesis-related protein 1 (*PR1*) which regulated thiamine formation had Log_2 Fold Change of −3.07 (Figure 10A). In the meantime, phenylalanine ammonia-lyase (*PAL*) which regulated SA formation had the |Log_2 Fold Change| greater than 3 especially when compared 8 with 28 dpa (Figure 10B).

Discussion

Changes in seed coat permeability during seed development

With seed development, the seed viability and 1,000-seed weight significantly increased, while the moisture content, conductivity, and imbibition rate showed the opposite results. In particular, EC did not change at a certain stage of seed development (Figure 1). This result is similar to that of some gramineous forages such as *E. sibiricus*, *Roegneria nutans*, *Festuca sinensis*, *Achnatherum inebrians*, and *Hordeum vulgare* var. *nudum* (Zhou et al., 2013a). The seed coat is believed to contain

10.3389/fpls.2022.970957

KEGG-Pathway	Upregulated		Downregulated	
	8-18 dpa	8-28 dpa	8-18 dpa	8-28 dpa
Fatty acid elongation	0	0	20	31
Biosynthesis of unsaturated fatty acids	0	0	8	0
Biosynthesis of secondary metabolites	0	0	79	0
Fatty acid metabolism	0	0	10	0
Ribosome	0	0	0	0
Starch and sucrose metabolism	0	0	0	72
Phagosome	0	0	0	46
Amino sugar and nucleotide sugar metabolism	18	0	0	0
MAPK signaling pathway - plant	16	34	0	0
Glyoxylate and dicarboxylate metabolism	9	0	0	0
Tryptophan metabolism	8	0	0	0
Linoleic acid metabolism	4	0	0	0

FIGURE 7
Heatmap diagram reflecting the dynamics of enriched biological process in KEGG analysis during seed development. Enriched processes with FDR < 0.05 were included in the heatmap. Red and green colors represent upregulated and downregulated DEGs, respectively. The intensity of the color reflects the number of DEGs as indicated at each development stage.

a semi-permeable layer, which leads to decreased seed permeability. According to anatomical microscopic observation, this structure is formed by the thickening of the peripheral cell wall of lateral cells of the inner integument (Zhou et al., 2013b). In our study, this layer prevents the infiltration of large molecules, including lanthanum nitrate and fluorescent dyes (**Figure 2** and **Supplementary Figure 2**).

The semi-permeable layer is a common structure in grass seeds, which affects seed vigor detection using the EC method (Lv et al., 2017). This structure is located in the seed coat. Considering that the semi-permeable layer is a very thin layer and is difficult to separate from the testa, it can't be extracted for analysis. Therefore, we regarded the barrier function of the semi-permeable layer to be a function of the seed coat and considered that the formation of this layer is the main reason for the rapid decline of seed coat permeability. However, previous studies have demonstrated that the semi-permeable layer is a lifeless tissue, and its function is due to the deposition of chemical substances on the structure (Zhou et al., 2013b). Histochemical staining showed that the substances causing the change in seed coat permeability included lipid, pectin, and cellulose (Zhou et al., 2013a). However, detailed reports on the chemical composition of the seed coat are lacking. Therefore, this study focused on the specific components that caused changes in seed coat permeability through high-throughput metabolomic analysis. The results showed that the main substances causing the decrease of seed coat permeability are thiamine and SA (**Figure 5**).

Differential metabolites causing changes in seed coat permeability

Through metabolomics analysis, a total of 1,887 metabolites were identified, including 4-coumaric acid, L-glutamic acid, DL-alpha-tocopherol, thiamine, alpha-lactose, and SA which were the main differentially abundant compounds with seed development. In *Styrax tonkinensis* seeds developing, a total of 187 and 1,556 metabolites were obtained and carboxylic acids and derivatives, flavonoids, fatty acid were the main components (Wu et al., 2020). However, we found that the significant increase of thiamine and SA during early seed development was the main reason for the decrease of seed coat permeability (**Figure 5A**). As one of the B-complex vitamins, thiamine is indispensable to provide energy for organs with active cell division and plays predominantly antioxidant roles in cellular physiological metabolic processes under environmental stresses (Asensi-Fabado and Munné-Bosch, 2010; Rosado-Souza et al., 2020). Thiamine biosynthesis is a highly regulated process involving light, stress, circadian rhythms, and the cofactor pyrophosphate (Khozaei et al., 2015). It is known to enhance plant defense by priming, but its effect on seed development is rarely reported (Nagae et al., 2016; Wang et al., 2016). In *Lathyrus sativus*, *Lotus japonicus*, and *Phyllanthus amarus*, the thiamine content will increase with seed development, which is similar to the results in this experiment (Lisiewska et al., 2003; Okiki et al., 2015; Nagae et al., 2016). In the whole process of seed growth and development, a good defense system

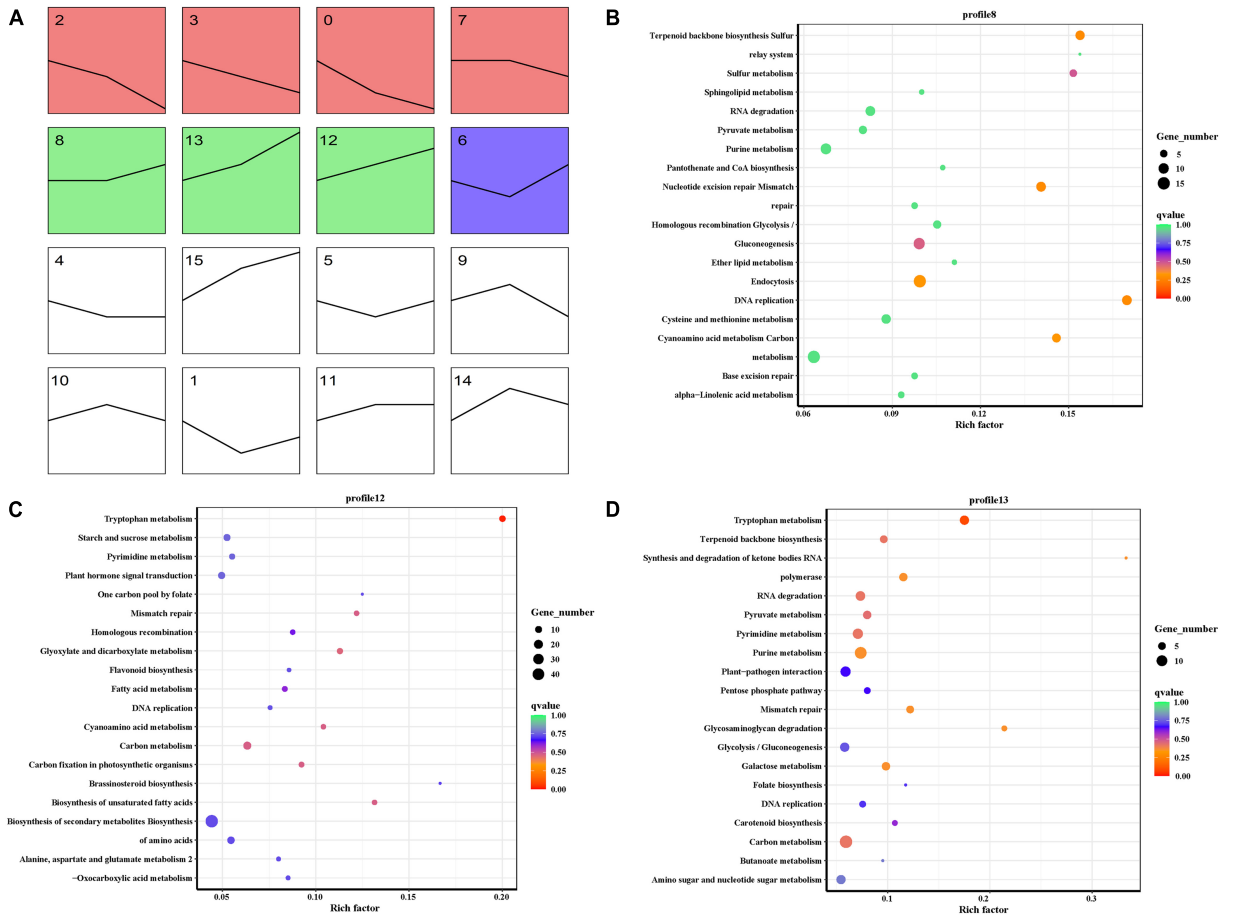


FIGURE 8
Time series analysis of RNA-seq. **(A)** Overall rendering of gene clustering. **(B)** Profile 8 KEGG pathway enrichment scatterplot. **(C)** Profile 12 KEGG pathway enrichment scatterplot. **(D)** Profile 13 KEGG pathway enrichment scatterplot.

is the key to ensure the healthy growth of seeds. As a physical barrier, the seed coat plays a pioneering role in protection. The presence of thiamine indicates that the seed coat is capable of resisting the external environment during seed growth.

In addition to thiamine, SA is also the main chemicals deposited in the seed coat during the early stages of seed development. SA is a kind of small molecular phenols widely distributed in higher plants and is also an important plant hormone, which can participate in the response to stress through complex signal transduction network (Ma et al., 2017). SA is known to increase in many patho systems upon infection with viruses, fungi, insects, and bacteria (Lefevere et al., 2020). In our research, the relative quantitative value of SA showed a significant upward trend among different comparisons (**Figure 5**). This result reflects that the seed coat accumulates endogenous hormones during the development process to enhance its role in protecting the seed embryo from external environmental interference, so as to ensure the healthy development of the seed.

Key genes regulating seed coat permeability

According to the transcriptome analysis, 2056 and 13,468 unigenes were observed, which showed differential expression in development samples (**Figure 6B**). Through joint analysis of key genes identified by transcription and key substances identified by metabolism, we found the key genes that regulated the changes in seed coat permeability, including *PR1* and *PAL* (**Figure 10** and **Supplementary File 5**). PRs are a great protein superfamily in plants and play important roles in plant response to various biotic and abiotic stresses (Tang et al., 2019; Ghorbel et al., 2021). In tobacco, thiamine homeostasis in plants was regulated by the transketolase protein which provided the

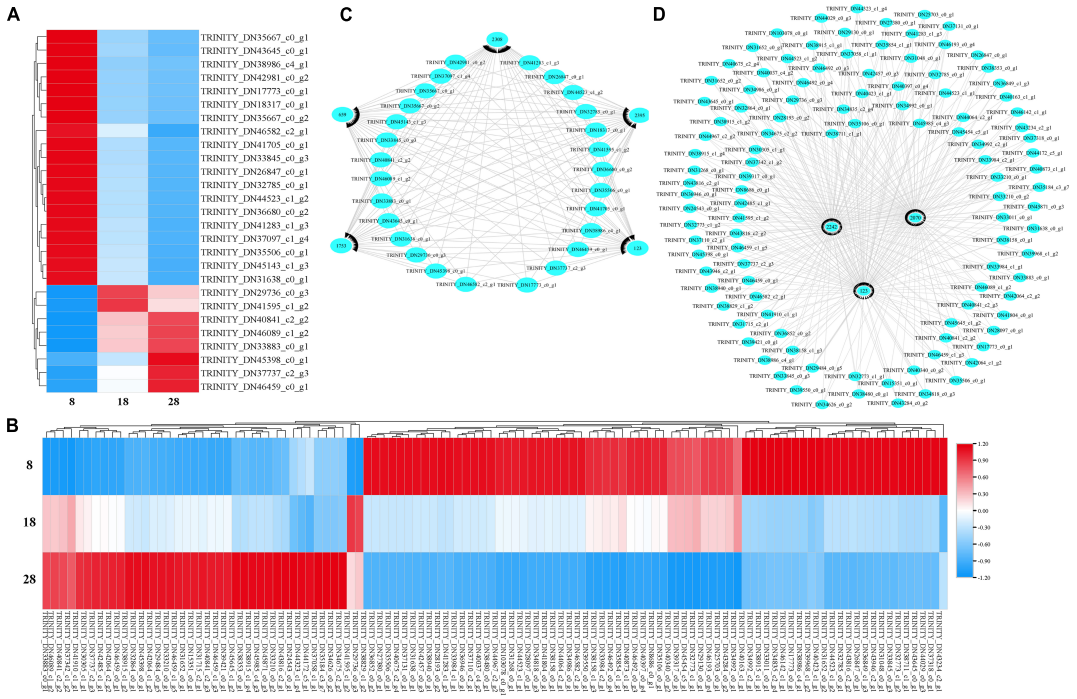


FIGURE 9
Heatmap and cytoscape shows the DEGs in KEGG pathways and the interaction between the key metabolites and their associated DEGs that were significantly enriched under metabolic analysis. **(A,B)** All DEGs identified at 8 vs. 18 dpa and 8 vs. 28 dpa, respectively. **(C,D)** Key genes between 8 and 18 dpa, respectively. 1,753: thiamine, 2,308: alpha-lactose, 123: salicylic acid (SA), 2,242: docosapentaenoic acid, 2,070: eicosapentaenoic acid, 659: L-glutamic acid, 2,395: DL-alpha-tocopherol.

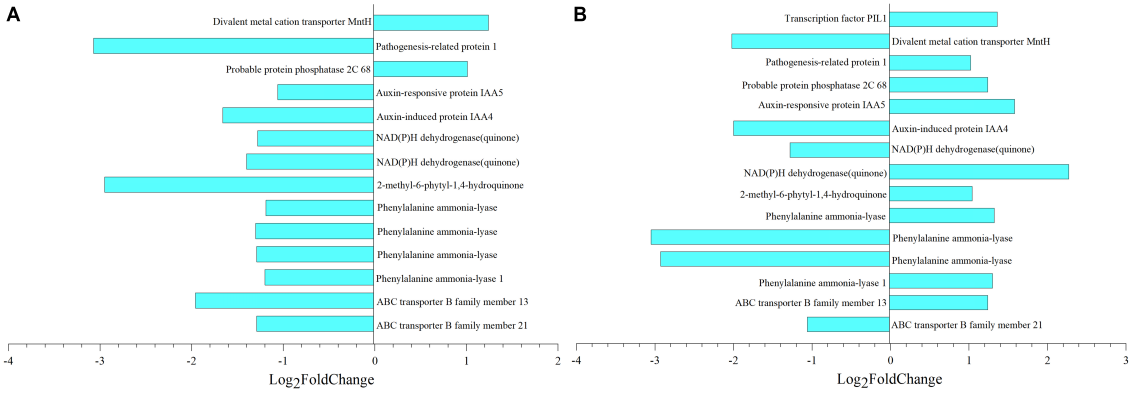


FIGURE 10
Metabolic and gene expression changes associated with seed coat development of *E. nutans*. Column diagram representing the accumulation of key metabolites that were significantly associated with DEGs. **(A,B)** Differential multiples of key genes between 8 vs. 18 dpa and 8 vs. 28 dpa, respectively. 1,753, thiamine; 123, salicylic acid (SA).

precursor for synthesis of intermediates and to enable plants to produce thiamine and thiamine pyrophosphate for growth and development (Khozaei et al., 2015). In our research, *PR1* has the highest |log$_2$Fold Change| multiple of 3.07, and its correlation coefficient with thiamine was −0.88 (**Figure 10A** and **Supplementary File 5**). The results showed that the gene was closely related to the substance that caused the change of seed coat permeability, so it might be a key gene in regulating the change of seed coat permeability.

At the same time, *PAL* is one of the genes with the greatest difference in seed coat development. It occurs mainly in mature seed coat and regulates the formation of SA (**Figure 10B** and **Supplementary File 5**). It is widely accepted that plants possess

PAL pathway to synthesize SA because PAL is an upstream enzyme that leads to many other possibly defense-related compounds (Lefevere et al., 2020). *PAL* genes are differentially expressed between plant tissues and induction upon infection with a distinct pathogen (Tonnessen et al., 2015). Seed coat development plays an important role in the whole plant growth process, which makes seeds have strong adaptability to adverse natural environment. The reason may be the existence of a large number of defense proteins regulating the course of disease, protecting the development of seeds, and this result is similar to the previous studies.

In this study, the formation mechanism of seed coat permeability was elucidated using metabolome and transcriptome data in *E. nutans*. Two key metabolites including thiamine and SA were found to affect seed coat permeability. Meanwhile, *PR1* and *PAL* play an important role in regulating the formation of compounds. This study provides not only insights into the changes in seed coat permeability caused by compounds deposited on the seed coat, but also candidate gene resources for seed coat development. The results will supply a great significance value to seed production and quality evaluation.

Data availability statement

The RNA-Seq data and the datasets presented in this study can be found at the NCBI repository, accession number PRJNA773127, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA773127.

Author contributions

JZ conceived, designed and performed the experiments, and wrote the manuscript. YL, XW, and YJL took the samples and performed parts of the experiments. RD-S revised and edited the manuscript. MW checked the experiments data. SQ made substantial contributions to the data analysis. ZG provided some suggestions. FY conceived the experiments and modified the manuscript. All authors have read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (32101418).

Acknowledgments

We thank Hua Shi and Libo Han, Fujian Agriculture and Forestry University, for their professional advice and use of electron microscope facilities.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.970957/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Anatomical microsturcture of seed coating at different developmental stages of E. nutans. **(A)** The section of the ovary at 8 dpa with nucellus cell (nu), two layers of the inner integument (ii), and the pericarp (pe). **(B)** The seed coat (sc) had formed at 18 dpa. The nucellus cell differentiated into aleurone (al) and endosperm (en) cells. **(C)** In addition to the embryo, mature seeds include pericarp, seed coat, aleurone layer, and endosperm cells at 28 dpa.

SUPPLEMENTARY FIGURE 2
SEM image of seeds and EDX detection of different parts of seeds. **(A)** Scanning electron microscopy image of seed structure. **(B–F)** Distribution of lanthanum deposition in different positions. Oval boxes represent the presence of lanthanum.

SUPPLEMENTARY FIGURE 3
TFs that were differentially expressed under different development stages in Elymus nutans.

SUPPLEMENTARY FIGURE 4
qRT-PCR analysis of the relative expression level of six unigenes.

References

Amritphale, D., Ramakrishna, P., Singh, B., and Sharma, S. K. (2010). Solute permeation across the apoplastic barrier in the perisperm-endosperm envelope in cucumber seeds. *Planta* 231, 1483–1494. doi: 10.1007/s00425-010-1146-5

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11:106. doi: 10.1186/gb-2010-11-10-r106

Asensi-Fabado, M. A., and Munné-Bosch, S. (2010). Vitamins in plants: Occurrence, biosynthesis and antioxidant function. *Trends Plant Sci.* 15, 582–592. doi: 10.1016/j.tplants.2010.07.003

Bafoil, M., Ru, A. L., Merbahi, N., Eichwald, O., Dunand, C., and Yousfi, M. (2019). New insights of low-temperature plasma effects on germination of three genotypes of *Arabidopsis* thaliana seeds under osmotic and saline stresses. *Sci. Rep.* 9:8649. doi: 10.1038/s41598-019-44927-4

Barua, D., Butler, C., Tisdale, T. E., and Donohue, K. (2012). Natural variation in germination responses of *Arabidopsis* to seasonal cues and their associated physiological mechanisms. *Ann. Bot.* 109, 209–226. doi: 10.1093/aob/mcr264

Ben-Tov, D., Abraham, Y., Stav, S., Thompson, K., Loraine, A., Elbaum, R., et al. (2015). COBRA-LIKE2, a member of the glycosylphosphatidylinositol-anchored COBRA-LIKE family, plays a role in cellulose deposition in Arabidopsis seed coat mucilage secretory cells. *Plant Physiol.* 167, 711–724. doi: 10.1104/pp.114.240671

Beresniewicz, M. M., Taylor, A. G., Goffinet, M. C., and Terhune, B. T. (1995). Characterization and location of a semipermeable layer in seed coats of leek and onion (*liliaceae*), tomato and pepper (solanaceae). *Seed Sci. Technol.* 23, 123–134.

Brancalion, P. H. S., Novembre, A. D. L. C., Rodrigues, R. R., and Filho, J. M. (2010). Dormancy as exaptation to protect mimetic seeds against deterioration before dispersal. *Ann. Bot.* 105, 991–998. doi: 10.1093/aob/mcq068

Briggs, C. L., and Morris, E. C. (2008). Seed-coat dormancy in Grevillea linearifolia: little change in permeability to an apoplastic tracer after treatment with smoke and heat. *Ann. Bot.* 101, 623–632. doi: 10.1093/aob/mcn006

Chen, M. X., Xuan, L. J., Wang, Z., Zhou, L. H., Li, Z. L., Du, X., et al. (2014a). Transparent testa 8 inhibits seed fatty acid accumulation by targeting several seed development regulators in *Arabidopsis*. *Plant Physiol.* 165, 905–916. doi: 10.1104/pp.114.235507

Chen, M., MacGregor, D. R., Dave, A., Florance, H., Moore, K., Paszkiewicz, K., et al. (2014b). Maternal temperature history activates flowering locus t in fruits to control progeny dormancy according to time of year. *Proc. Natl. Acad. Sci. U.S.A.* 111, 18787–18792. doi: 10.1073/pnas.1412274111

Ding, Z. T., Bai, J., Xu, D. M., Li, F. H., Zhang, Y. X., and Guo, X. S. (2020). Microbial community dynamics and natural fermentation profiles of ensiled alpine grass *Elymus nutans* prepared from different regions of the qinghai-tibetan plateau. *Front. Microbiol.* 11:855. doi: 10.3389/fmicb.2020.00855

Dongen, J.T.v, Laan, R. G. W., Wouterlood, M., and Borstlap, A. C. (2001). Electrodiffusional uptake of organic cations by pea seed coats. further evidence for poorly selective pores in the plasma membrane of seed coat parenchyma cells. *Plant Physiol.* 126, 1688–1697. doi: 10.1104/pp.126.4.1688

Ezquer, I., Mizzotti, C., Nguema-ona, E., Gotté, M., Beauzamy, L., Viana, V. E., et al. (2016). The developmental regulator seedstick controls structural and mechanical properties of the *Arabidopsis* seed coat. *Plant Cell* 28, 2478–2492. doi: 10.1105/tpc.16.00454

Fu, J. J., Sun, Y. F., Chu, X. T., Xu, Y. F., and Hu, T. M. (2014). Exogenous 5-aminolevulenic acid promotes seed germination in Elymus nutans against oxidative damage induced by cold stress. *PLoS One* 9:e107152. doi: 10.1371/journal.pone.0107152

Ghorbel, M., Zribi, I., Missaoui, K., Drira-Fakhfekh, M., Azzouzi, B., and Brini, F. (2021). Diferential regulation of the durum wheat pathogenesis-related protein (pr1) by calmodulin TdCaM1.3 protein. *Mol. Biol. Rep.* 48, 347–362. doi: 10.1007/s11033-020-06053-7

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Jaganathan, G. K., Li, J. J., Biddick, M., Han, K., Song, D. P., Yang, Y. S., et al. (2019). Mechanisms underpinning the onset of seed coat impermeability and dormancy-break in *Astragalus adsurgens*. *Sci. Rep.* 9, 129–138. doi: 10.1038/s41598-019-46158-z

Jang, S. J., Sato, M., Sato, K., Jitsuyama, Y., Fujino, K., Mori, H., et al. (2015). A single-nucleotide polymorphism in an endo-1,4-β-glucanase gene controls seed coat permeability in soybean. *PLoS One* 10:e0128527. doi: 10.1371/journal.pone.0128527

Janská, A., Pecková, E., Sczepaniak, B., Smýkal, P., and Soukup, A. (2019). The role of the testa during the establishment of physical dormancy in the pea seed. *Ann. Bot.* 123, 815–829.

Jayasuriya, K. M. G. G., Baskin, J. M., Geneve, R. L., and Baskin, C. C. (2007). Seed development in *Ipomoea lacunosa* (*convolvulaceae*), with particular reference to anatomy of the water gap. *Ann. Bot.* 100, 459–470. doi: 10.1093/aob/mcm137

Kebede, H., Smith, J. R., and Ray, J. D. (2014). Identification of a single gene for seed coat impermeability in soybean PI 594619. *Theor. Appl. Genet.* 127, 1991–2003. doi: 10.1007/s00122-014-2355-2

Khozaei, M., Fisk, S., Lawson, T., Gibon, Y., Sulpice, R., Stitt, M., et al. (2015). Overexpression of plastid transketolase in tobacco results in a thiamine auxotrophic phenotype. *Plant Cell* 27, 432–447. doi: 10.1105/tpc.114.131011

Langmead, B., and Salzberg, S. L. (2013). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lefevere, H., Bauters, L., and Gheysen, G. (2020). Salicylic acid biosynthesis in plants. *Front. Plant. Sci.* 11:338. doi: 10.3389/fpls.2020.00338

Li, M., Han, J., Wang, Y., Sun, J., and Haferkamp, M. (2010). Different seed dormancy levels imposed by tissues covering the caryopsis in zoysiagrass (*Zoysia japonica steud.*). *Seed Sci. Technol.* 38, 320–331. doi: 10.15258/sst.2010.38.2.05

Lisiewska, Z., Korus, A., and Kmiecik, W. (2003). Changes in chemical composition during development of grass pea (*Lathyrus sativus L.*) seeds. *Mol. Nutrl. Food Res.* 47, 391–396. doi: 10.1002/food.200390088

Liu, J. J., Wang, X. P., Gao, Y. F., Rong, Z. L., Zhang, G. D., Wang, W. B., et al. (2019). Leaf gas exchange and photosynthesis curves of *Elymus nutans* and *Potentilla anserina* under fencing and grazing conditions in the qilian mountains, northwest china. *J. Arid. Land* 11, 431–445. doi: 10.1007/s40333-019-0009-6

Loubéry, S., Giorgi, J. D., Utz-Pugin, A., Demonsais, L., and Lopez-Molina, L. (2018). A maternally deposited endosperm cuticle contributes to the physiological defects of transparent testa seeds. *Plant Physiol.* 177, 1218–1233. doi: 10.1104/pp.18.00416

Lv, Y. Y., He, X. Q., Hu, X. W., and Wang, Y. R. (2017). The seed semipermeable layer and its relation to seed quality assessment in four grass species. *Front. Plant Sci.* 8:1175. doi: 10.3389/fpls.2017.01175

Ma, X. H., Zheng, J., Zhang, X. L., Hu, Q. D., and Qian, R. J. (2017). SA alleviates the adverse effects of salt stress on *Dianthus superbus* (Caryophyllaceae) by activating photosynthesis, protecting morphological structure, and enhancing the antioxidant system. *Front. Plant Sci.* 8:600. doi: 10.3389/fpls.2017.00600

MacGregor, D. R., Kendall, S. L., Florance, H., Fedi, F., Moore, K., Paszkiewicz, K., et al. (2015). Seed production temperature regulation of primary dormancy occurs through control of seed coat phenylpropanoid metabolism. *New Phytol.* 205, 642–652. doi: 10.1111/nph.13090

Magrini, S., Barreca, D., and Zucconi, L. (2019). A rapid double-staining technique to improve seed viability testing in terrestrial orchids. *Plant Biosyst.* 153, 877–882.

Mao, X. Z., Cai, T., Olyarchuk, J. G., and Wei, L. P. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21, 3787–3793. doi: 10.1093/bioinformatics/bti430

Mendu, V., Griffiths, J. S., Persson, S., Stork, J., Downie, A. B., Voiniciuc, C., et al. (2011). Subfunctionalization of cellulose synthases in seed coat epidermal cells mediates secondary radial wall synthesis and mucilage attachment. *Plant Physiol.* 157, 441–453. doi: 10.1104/pp.111.179069

Mouden, S., Kappers, I. F., Klinkhamer, P. G. L., and Leiss, K. A. (2020). Cultivar variation in tomato seed coat permeability is an important determinant of jasmonic acid elicited defenses against western flower thrips. *Front. Plant Sci.* 11:576505. doi: 10.3389/fpls.2020.576505

Nagae, M., Parniske, M., Kawaguchi, M., and Takeda, N. (2016). The thiamine biosynthesis gene THI1 promotes nodule growth and seed maturation. *Plant Physiol.* 172, 2033–2043. doi: 10.1104/pp.16.01254

Niu, K., Shi, Y., and Ma, H. (2017). Selection of candidate reference genes for gene expression analysis in kentucky bluegrass (*Poa pratensis* L.) under abiotic stress. *Front. Plant Sci.* 8:193. doi: 10.3389/fpls.2017.00193

Okiki, P. A., Olatunji, B. P., Egbebi, A. A. S., and Ojo, C. (2015). A comparative study of nutritional and phytochemical composition of *Phyllanthus amarus* leaf and seed. *Am. Eurasian J. Toxicol. Sci.* 7, 321–327. doi: 10.5829/idosi.aejts.2015.7.4.9699

Orozco-Segovia, A., Márquez-Guzmán, J., Sánchez-Coronado, M. E., Gamboa de Buen, A., Baskin, J. M., and Baskin, C. C. (2007). Seed anatomy and water uptake in relation to seed dormancy in *Opuntia tomentosa* (Cactaceae, Opuntioideae). *Ann. Bot.* 99, 581–592. doi: 10.1093/aob/mcm001

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Rosado-Souza, L., Fernie, A. R., and Aarabi, F. (2020). Ascorbate and thiamin: Metabolic modulators in plant acclimation responses. *Plants (Basel)* 9:101. doi: 10.3390/plants9010101

Salanenka, Y. A., and Taylor, A. G. (2011). Seed coat permeability: Uptake and post-germination transport of applied model tracer compounds. *HortScience* 46, 622–626.

Shen, B., Yi, X., Sun, Y. T., Bi, X. J., Du, J. P., Zhang, C., et al. (2020). Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell* 182, 59–72. doi: 10.1016/j.cell.2020.05.032

Sun, L. J., Miao, Z. Y., Cai, C. M., Zhang, D. J., Zhao, M. X., Wu, Y. Y., et al. (2015). GmHs1-1, encoding a calcineurin-like protein, controls hard-seededness in soybean. *Nat. Genet.* 47:939. doi: 10.1038/ng.3339

Sun, Q. J., Lv, Y. Y., and Wang, Y. R. (2018). Study on the semipermeable characteristics of seven poaceae seeds. *Seed Sci. Technol.* 46, 327–340. doi: 10.15258/sst.2018.46.2.13

Tang, B. F., Li, X., Pu, L. M., Zhao, Q., Cui, X. M., Ge, F., et al. (2019). A pathogenesis-related protein 10 gene PnPR10–3 was involved in molecular interaction between *Panax notoginseng* and *Fusarium solani*. *Australas. Plant Pathol.* 48, 447–456. doi: 10.1007/s13313-019-00644-0

Tonnessen, B. W., Manosalva, P., Lang, J. M., Baraoidan, M., Bordeos, A., Mauleon, R., et al. (2015). Rice phenylalanine ammonia-lyase gene OsPAL4 is associated with broad spectrum disease resistance. *Plant Mol. Biol.* 87, 273–286. doi: 10.1007/s11103-014-0275-9

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., and van Baren, M. J. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–518. doi: 10.1038/nbt.1621

Vishwanath, S. J., Kosma, D. K., Pulsifer, I. P., Scandola, S., Pascal, S., Joubes, J., et al. (2013). Suberin-associated fatty alcohols in *Arabidopsis*: distributions in roots and contributions to seed coat barrier properties. *Plant Physiol.* 163, 1118–1132. doi: 10.1104/pp.113.224410

Wang, L. C., Ye, X. F., Liu, H. C., Liu, X. J., Wei, C. C., Huang, Y. Q., et al. (2016). Both overexpression and suppression of an *Oryza sativa* NB-LRR-like gene OsLSR result in autoactivation of immune response and thiamine accumulation. *Sci. Rep.* 6:24079. doi: 10.1038/srep24079

Wu, Q., Zhao, X., Chen, C., Zhang, Z., and Yu, F. (2020). Metabolite profiling and classification of developing *Styrax tonkinensis* kernels. *Metabolites* 10:21. doi: 10.3390/metabo10010021

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biol.* 11:R14. doi: 10.1186/gb-2010-11-2-r14

Zhou, J., Wang, Y. R., and Jahufer, Z. (2013b). Location and chemical composition of semi-permeable layer of forage seeds. *Bangladesh J. Bot.* 42, 23–29. doi: 10.3329/bjbv42i1.15802

Zhou, J., Wang, Y. R., and Trethewey, J. (2013a). Semi-permeable layer formation during seed development in *Elymus nutans* and *Elymus sibiricus*. *Acta. Soc. Bot. Pol.* 82, 165–173. doi: 10.5586/asbp.2013.012

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 September 2022
DOI 10.3389/fpls.2022.1005895

OPEN ACCESS

EDITED BY
Jiyu Zhang,
Lanzhou University,
China

REVIEWED BY
Hao Lin,
Biotechnology Research Institute (CAAS),
China
Chunxiang Fu,
Qingdao Institute of Bioenergy and
Bioprocess Technology (CAS), China

*CORRESPONDENCE
Qingchuan Yang
qchyang66@163.com

SPECIALTY SECTION
This article was submitted to
Crop and Product Physiology,
a section of the journal
Frontiers in Plant Science

RECEIVED 28 July 2022
ACCEPTED 19 August 2022
PUBLISHED 06 September 2022

CITATION
Wang X, Wei C, He F and Yang Q (2022)
MtPT5 phosphate transporter is involved in
leaf growth and phosphate accumulation
of *Medicago truncatula*.
Front. Plant Sci. 13:1005895.
doi: 10.3389/fpls.2022.1005895

MtPT5 phosphate transporter is involved in leaf growth and phosphate accumulation of *Medicago truncatula*

Xue Wang, Chunxue Wei, Fei He and Qingchuan Yang*

Institute of Animal Science, The Chinese Academy of Agricultural Sciences, Beijing, China

Phosphorus (P) is an indispensable mineral nutrient for plant growth and agricultural production. Plants acquire and redistribute inorganic phosphate (Pi) via Pi transporters (PHT1s/PTs). However, apart from MtPT4, functions of the *M. truncatula* (*Medicago truncatula*) PHT1s remain unclear. In this study, we evaluated the function of the PHT1 family transporter MtPT5 in *M. truncatula*. MtPT5 was closely related to AtPHT1; 1 in *Arabidopsis* (*Arabidopsis thaliana*) and GmPT7 in soybean (*Glycine max*). *MtPT5* was highly expressed in leaves in addition to roots and nodules. Ectopic expression of *MtPT5* complemented the Pi-uptake deficiency of *Arabidopsis pht1;1Δ4Δ* double mutant, demonstrating the Pi-transport activity of MtPT5 in plants. When overexpressing *MtPT5* in *M. truncatula*, the transgenic plants showed larger leaves, accompanying with higher biomass and Pi enrichment compared with wild type. All these data demonstrate that MtPT5 is important for leaf growth and Pi accumulation of *M. truncatula* and provides a target for molecular breeding to improve forage productivity.

KEYWORDS

phosphate, PHT1 transporter, MtPT5, *Medicago truncatula*, leaf growth, Pi accumulation

Introduction

Phosphorus (P) is an essential mineral nutrient for plant growth. It plays various biological functions and is a major determinant of crop production (Raghothama, 1999). Inorganic phosphate (Pi) is the main form of P that can be absorbed by plant roots (Chiou and Lin, 2011; Vincent et al., 2012; López-Arredondo et al., 2014). The total P level in soil is high, but the soluble Pi is always limited due to its low mobility, as well as precipitation and fixation (Marschner and Rimmington, 1988; Yan et al., 2021). It has been reported that about 70% of cultivated land in the world is deficient in plant-available Pi. P is one of the limiting factors for cultivated plants (Smith and Schindler, 2009; Péret et al., 2011; López-Arredondo et al., 2013). To maintain crop yield, the usage of P fertilizer is increased annually (Dobre et al., 2014; Heuer et al., 2017). However, excessive fertilizer is not only a

waste, but also leads to environmental issues (Zhang et al., 2013; Zak et al., 2018; Che et al., 2020).

Plants absorb and translocate Pi via Pi transporters (PHT1s/PTs; Versaw and Garcia, 2017; Dai et al., 2022). Hence, *PHT1* genes are potential targets for improving plants Pi efficiency and benefiting yields (Veneklaas et al., 2012). Most of the *PHT1* genes are root-specific, while some are highly expressed in the aerial part or nodules and involved in Pi redistribution (Chen et al., 2019; Wang et al., 2020). The firstly identified *PHT1* gene was *Pho84* cloned from *Saccharomyces cerevisiae* (Bun-ya et al., 1991). Then, numerous PHT1s have been identified in plants including *Arabidopsis* (*Arabidopsis thaliana*; Muchhal et al., 1996; Shin et al., 2004), *M. truncatula* (*Medicago truncatula*; Liu et al., 2008), rice (*Oryza sativa* L.; Liu et al., 2011), maize (*Zea mays* L.; Wang et al., 2020) and soybean (*Glycine max*; Chen et al., 2019). Nine PHT1 members were identified in *Arabidopsis* (Mudge et al., 2002). Among them, AtPHT1;1 and AtPHT1;4 play predominant roles in Pi uptake (Shin et al., 2004). GmPT7 was reported to be responsible for Pi uptake from soil into nodules and distribution to the fixation zones. Overexpression of *GmPT7* promotes plant growth and soybean yield (Chen et al., 2019). When overexpressing *OsPT1* in rice, transgenic plants accumulated more Pi in shoots and displayed increased tiller numbers compared with wild-type plants (Seo et al., 2008). Thus, investigation of the functions of PHT1s provides an efficient route for improving plants nutrient efficiency.

Currently, 11 PHT1s were identified in *M. truncatula* (Liu et al., 2008). Yeast kinetics assays showed that MtPT1, MtPT2, MtPT3, and MtPT4 are low-affinity Pi transporters. MtPT1, MtPT2, MtPT3, and MtPT5 share 84% sequence identity, but only MtPT5 displayed high affinity for Pi (Liu et al., 2008). MtPT4 is highly expressed in mycorrhizal roots, responsible for Pi acquisition from arbuscules (Harrison et al., 2002). It is also expressed in plant root tip in the absence of the arbuscular mycorrhizal (AM) fungus and modulates root branching, whereas it does not significantly affect Pi accumulation in plants without AM symbiosis (Cao et al., 2020). Recently, MtPT6 was reported to be involved in Pi uptake by heterologous expression of *MtPT6* in *Arabidopsis pht1;1* or *pht1;4* mutant. However, the role of MtPT6 in *M. truncatula* is unknown (Volpe et al., 2016). Information on the functions of PHT1s in *Medicago* is still limited.

In this study, we identified the role of MtPT5 in leaf growth and Pi accumulation of *M. truncatula*. *MtPT5* is highly expressed in roots, leaves, and nodules and is low-Pi inducible. MtPT5 can rescue the Pi-uptake deficiency of *Arabidopsis pht1;1Δ4Δ* double mutant, indicating the Pi transport activity of MtPT5 in plants. When overexpressing *MtPT5* in *M. truncatula*, the transgenic plants displayed larger leaf size and higher Pi content. These data demonstrate that MtPT5 plays important roles in *M. truncatula* vegetable growth and Pi nutrition.

Materials and methods

Plant materials and growth conditions

Medicago truncatula ecotype R108, *Arabidopsis thaliana* ecotype Wassilewskija ecotype (Ws) and *pht1;1Δ4Δ* mutant were used in this study. For germination of *M. truncatula,* seeds were placed on wet filter paper at 4°C for 2 days. Then, the imbibed seeds were transferred to chamber with illumination of 120 μmol m^{-2} s^{-1}, temperature 24°C, and 16 h light/8 h dark photoperiod for 4 days. The seedlings were grown in 1/2 Hoagland or in soil for experiments. For nodulation, four-day-old seedlings were incubated with Sm1021 resuspended in 1/2 Hoagland, then transferred to soil and injected with Sm1021 every 2 days for 1 month. Different organs were harvested separately for RNA extraction. For *Arabidopsis* germination, the seeds were kept at 4°C for 2 days for imbibition, then transferred to medium with 200 μM arsenate or 1/2 MS under normal conditions (120 μmol m^{-2} s^{-1}, 22°C, 16 h light/8 h dark).

Measurement of Pi content

For *Arabidopsis*, the 20-day-old seedings grown on 1/2 MS medium were harvested for Pi content measurement. For *M. truncatula*, the top leaflets of 3-month-old plants grown in soil and 20-day-old seedlings grown in 1/2 Hoagland were collected. The measurement was assayed as described in the previous report (Ames, 1996). Briefly, different samples were collected and frozen in liquid nitrogen immediately. Pi was extracted in the buffer containing acetic acid at 42°C for 30 min. Pi concentration was measured at 820 nm wavelength using universal microplate spectrophotometer (BioTek Power Wave XS2). The Pi content was calculated based on the concentration and fresh weight of different samples.

Plasmid construction and plant transformation

The full-length coding sequence (CDS) of *MtPT5* was cloned into *pTOPO-TA* Simple vector (Science Tool) for sequencing. Then the sequence-verified *MtPT5* CDS was constructed into *BamH* I-linearized *pCAMBIA1302* vector to generate *35S:MtPT5* plasmid via homologous recombination. The recombinant vector was used for plant transformation. For *Arabidopsis* (*pht1;1Δ4Δ* mutant), floral dip method was used as described (Clough and Bent, 1998) using *Agrobacterium tumefaciens* strain GV3101. The transformants were obtained on MS medium containing 50 mg/l hygromycin. For *M. truncatula*, the construct was introduced into R108 leaves via *Agrobacterium* EHA105-mediated transformation as described previously (Cosson et al., 2006).

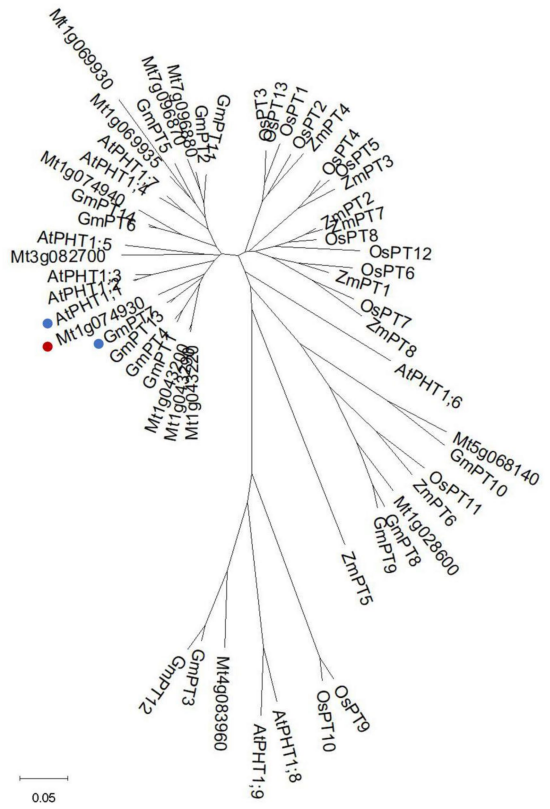


FIGURE 1
Phylogenetic analysis of PHT1s from different species. Phylogenetic tree of PHT1s from *Medicago truncatula*, *Arabidopsis*, soybean, maize, and rice. The tree was generated as described in materials and methods. Mt1g074930 (MtPT5) was labeled with a red spot. AtPHT1;1 and GmPT7 were labeled with blue spots. The bar shows 0.05 amino acid substitutions per site.

The transgenic *M. truncatula* plants were identified by PCR using vector-specific primers. T_2 and T_3 transgenic lines were used for *Arabidopsis* and *M. truncatula* separately in this study.

qRT-PCR and RT-PCR analysis

For quantification of gene expression, total RNA was isolated using Eastep Super Total RNA Extraction KIT (Promega) and quantified by nanodrop. 1 μg RNA was used for reverse-transcription using the PrimeScript II 1st Strand cDNA Synthesis Kit (Takara). qRT-PCR was performed using 2 × EasyTaq® PCR SuperMix (TransGen Biotech) on CFX96 system (Bio-Rad). *MtActin11* was used to calculate the relative quantitative results for *M. truncatula*. The transcripts of *MtPT5* in R108, *pht1;1Δ4Δ* mutant and *pht1;1Δ4Δ/MtPT5* were tested by RT-PCR using cDNAs as templates. *EF1a* was amplified as a quantitative control.

Sequence alignment and construction of phylogenetic tree

PHT1 amino acid sequences were obtained from NCBI[1] and EnsemblPlants.[2] Amino acid sequences were firstly aligned using CluxtalX. The neighbor-joining tree was conducted in MEGA5 using bootstrap method (900 replicates) on poisson model.

Statistical analysis

Significant differences were determined by One-way ANOVA with Tukey test or Student's *t*-test using SigmaPlot 12.5 software.

Results

Phylogenetic analysis of PHT1s from different species

It has been reported that there are 11 PHT1 transporters in *M. truncatula* (Cao et al., 2020). We identified another two members (Mt4g083960 and Mt5g068140) by searching Ensemblplants.[3] All members shared the common secondary structures with 11 predicted transmembrane domains (TM) separated by a large hydrophilic loop between TM6 and TM7 (Supplementary Figure S1). The signature GGDYPLSATIxSE (Karandashov and Bucher, 2005; Loth-Pereda et al., 2011) was identified and conserved among all MtPHT1s, except two of them. The signature of Mt1g069930 was modified with a Thr (T) replaced by a Val (V), and Mt1g074940 was modified with an Ala (A) replaced by a Ser (S; Supplementary Figure S1). The amino acid sequences of PHT1 proteins from *M. truncatula*, *Arabidopsis*, soybean maize and rice were used for constructing the neighbor-joining tree (Figure 1). The analysis showed that Mt1g074930 (MtPT5) was clustered phylogenetically to AtPHT1;1 and GmPT7, showing 80% and 86% amino acid sequence identities, respectively.

Expression pattern of *MtPT5* in *Medicago truncatula*

MtPT1, MtPT2, and MtPT3 are paralogues of MtPT5 in *M. truncatula* (Liu et al., 2008). The coding sequences of *MtPT1*, *MtPT2*, and *MtPT3* share 97% identity. A single pair of primers were used to test the expression of these three genes. Quantitative RT-PCR (qRT-PCR) analysis showed that *MtPT1/2/3* was

1 https://www.ncbi.nlm.nih.gov/
2 http://plants.ensembl.org/Medicago_truncatula/Info/Index
3 http://plants.ensembl.org/Medicago_truncatula/Info/Index

predominantly expressed in roots and nodules and nearly undetectable in shoots (Figure 2A). The transcription abundance of *MtPT5* was around four-to six-fold higher than that of *MtPT1/2/3* in the underground tissues. In addition, *MtPT5* was also highly expressed in shoots (Figure 2A). The expression pattern of *MtPT5* in the aerial part was further tested. qRT-PCR results showed that *MtPT5* was mainly expressed in leaves (Supplementary Figure S2). Pi starvation analysis showed that *MtPT5* was induced under Pi-deficient condition (Figure 2B), in accordance with the previous report (Liu et al., 2008). The expression profile suggests that MtPT5 probably have multiple functions in plants and crucial roles in leaves. Leaf is important for plant yield. Hence, MtPT5 was further analyzed in Pi nutrition and development of *M. truncatula*.

MtPT5 rescued the Pi-uptake deficiency of *Arabidopsis pht1;1Δ4Δ* double mutant

AtPHT1;1 and AtPHT1;4 are two major Pi transporters in *Arabidopsis*. The double mutant *pht1;1Δ4Δ* displayed dramatic reduction in Pi uptake capacity and Pi content compared with wild type (Shin et al., 2004). To examine the Pi uptake activity of MtPT5 in plants, the coding sequence of *MtPT5* driven by 35S promoter (*35S:MtPT5*) was introduced into *pht1;1Δ4Δ*. Two independent transgenic lines, *35S:MtPT5/pht1;1Δ4Δ-1* and *35S:MtPT5/pht1;1Δ4Δ-2*, were used in this study. RT-PCR analysis showed that the *MtPT5* transcripts were present in the two transgenic lines and not detectable in wild type (Ws) and *pht1;1Δ4Δ* mutant (Figure 3A). The fresh weight (FW) measurement showed that loss of *PHT1;1* and *PHT1;4* led to about 27% reduction in *pht1;1Δ4Δ* mutant biomass compared with wild type, similar to the previous report (Shin et al., 2004). Meanwhile, the biomasses of *35S:MtPT5/pht1;1Δ4Δ* transgenic lines could be rescued to the level of wild type (Supplementary Figure S3). This indicates that MtPT5 can rescue the morphological defects of *pht1;1Δ4Δ* mutant. Next, we tested the Pi contents in different genotypic *Arabidopsis* seedlings grown under Pi-sufficient condition (1/2 MS). Pi content in *pht1;1Δ4Δ* mutant was significantly reduced compared with wild type, while the two overexpression lines exhibited similar Pi contents with wild type (Figure 3B). These data suggest that MtPT5 can complement the Pi-uptake deficiency of *pht1;1Δ4Δ* mutant.

Arsenate is a toxic metalloid structurally analogous of Pi and is transported into root cells mainly via PHT1 transporters (Catarecha et al., 2007; Castrillo et al., 2013; Wang et al., 2014). Phenotypes of wild type, *pht1;1Δ4Δ* mutant and *35S:MtPT5/pht1;1Δ4Δ* transgenic plants were compared on the medium with or without arsenate. When grown on the medium with 200 μM arsenate, the *pht1;1Δ4Δ* mutant showed an arsenate-tolerant phenotype as previously reported (Shin et al., 2004), while the wild type and *35S:MtPT5/pht1;1Δ4Δ* seedlings were hypersensitive to arsenate with dramatically shorter roots and smaller shoots (Figure 3C). Taken together, these data indicate that MtPT5 has Pi transport capacity and positively modulates Pi uptake in plants.

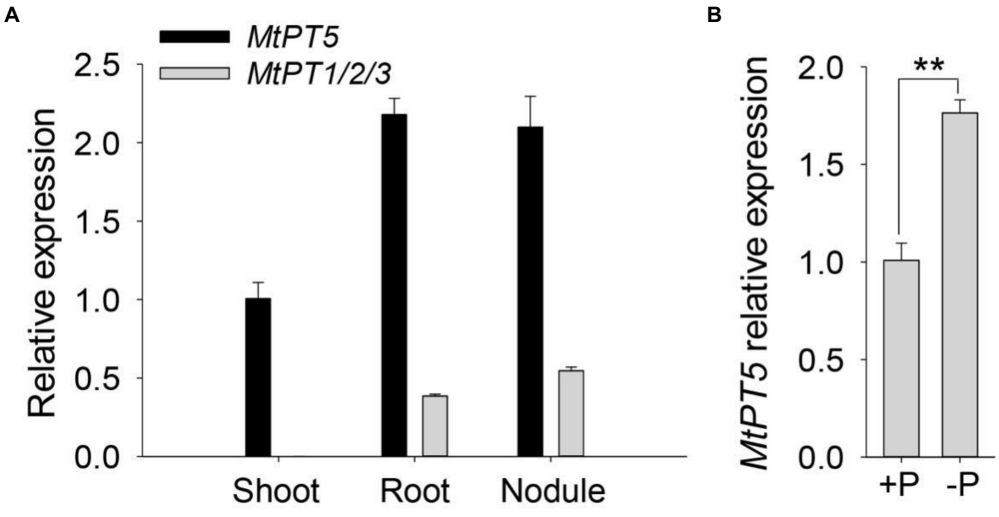


FIGURE 2
Expression profiles of *MtPTs* in *M. truncatula*. **(A)** qRT-PCR analysis of *MtPT1*/2/3 and *MtPT5* in different tissues of *M. truncatula*. Four-day-old wild-type seedlings (R108) were incubated with Sm1021 resuspended in 1/2 Hoagland and then transferred to soil and injected with Sm1021 every 2 days for 1 month. Shoots, roots, and nodules were harvested, respectively, for RNA extraction. Data represent mean±SE (n=3). **(B)** qRT-PCR analysis of *MtPT5* in wild-type seedlings (R108) during phosphate starvation. Four-day-old *M. truncatula* seedlings were transferred to hydroponic solution with Pi (+P) or solution without Pi (−P) for 5 days. The whole seedlings were used for RNA extraction. Data represent mean±SE (n=3). ** indicates significant difference at p<0.01 (Student's t-test).

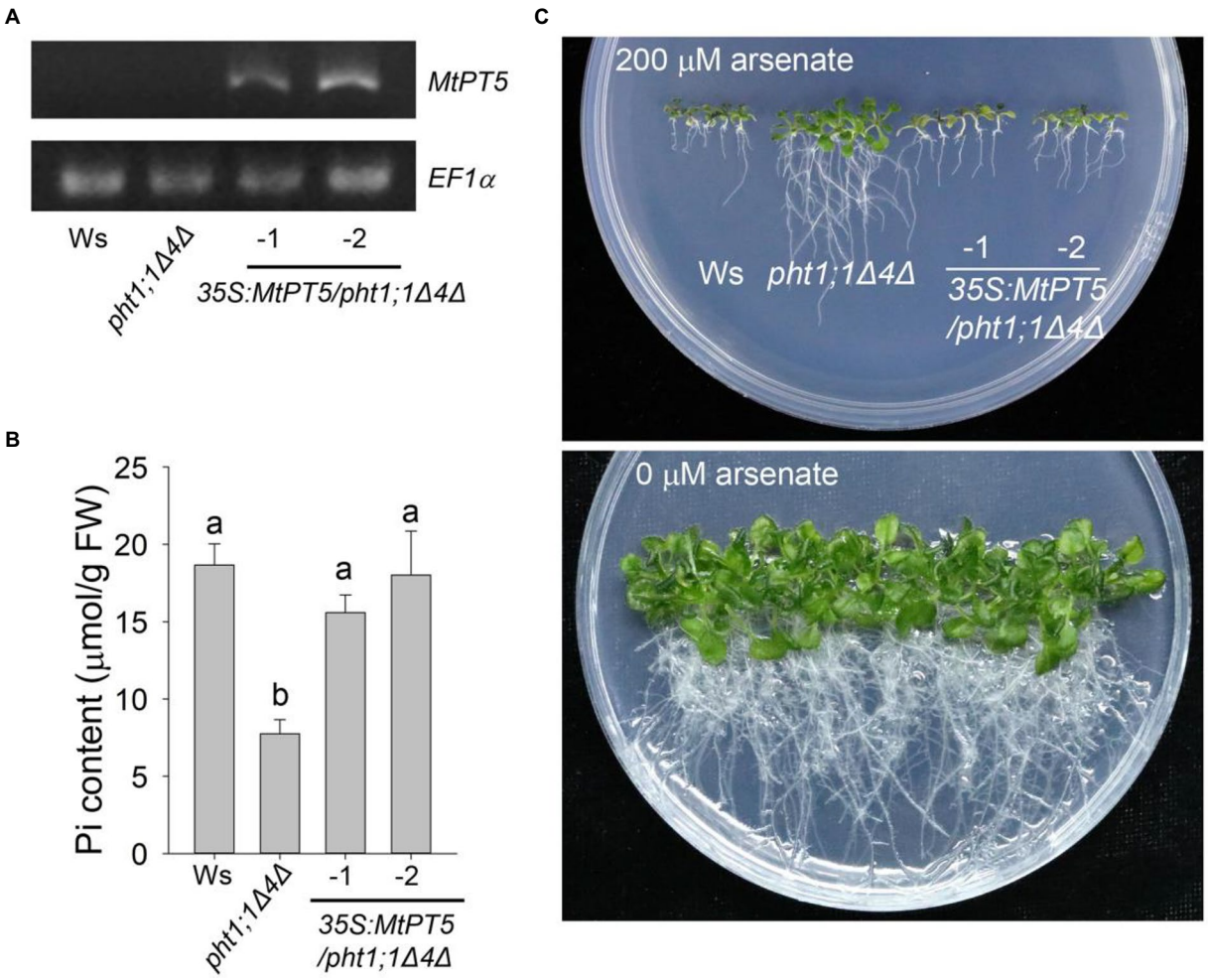


FIGURE 3
Ectopic expression of *MtPT5* enhanced Pi acquisition in *Arabidopsis*. **(A)** RT-PCR analysis of *MtPT5* in wild-type *Arabidopsis* (Ws), *pht1;1Δ4Δ* double mutant and *35S:MtPT5/pht1;1Δ4Δ* transgenic plants using *MtPT5*-specific primers. The housekeeping gene *EF1α* was used as an internal control. **(B)** Pi content measurement of wild-type *Arabidopsis* (Ws), *pht1;1Δ4Δ* double mutant and *35S:MtPT5/pht1;1Δ4Δ* transgenic plants. The whole seedlings grown on 1/2 MS for 20 days were collected for Pi extraction. Data represent mean±SE (n=3). Different letters indicate significant difference at p<0.05 (One-way ANOVA, Tukey test). **(C)** Image of wild-type *Arabidopsis* (Ws), *pht1;1Δ4Δ* double mutant and *35S:MtPT5/pht1;1Δ4Δ* transgenic plants grown on 1/2 MS containing 200μM arsenate or without arsenate for 20 days.

MtPT5 promotes leaves growth of *Medicago truncatula*

Given that *MtPT5* was induced by low-Pi stress, two independent *MtPT5*-overexpressing lines, *35S:MtPT5-1* and *35S:MtPT5-2*, were generated to examine the physiological role of *MtPT5* in *M. truncatula*. qRT-PCR analysis showed that both *MtPT5*-overexpressing lines had significantly increased *MtPT5* transcripts compared with wild-type *M. truncatula* (Figure 4A). We performed phenotypic tests on wild type and *MtPT5*-overexpressing plants. In both hydroponic culture and soil pots, the *MtPT5*-overexpressing lines displayed larger leaves compared with wild type (Figures 4B–D). Quantifications of leaf area confirmed this phenotype (Figure 4E). Meanwhile, leaf biomasses of the *MtPT5*-overexpressing plants were significantly higher than that of wild type (Figure 4F). These morphological traits indicate that overexpression of *MtPT5* promotes leaves growth in *M. truncatula*.

Overexpression of *MtPT5* enhances Pi accumulation of *Medicago truncatula*

To explore the function of MtPT5 in *M. truncatula* Pi nutrition, we measured the Pi content in leaves of wild type and *MtPT5*-overexpressing lines. The top leaflets of plants grown in soil for 3 months were collected for Pi extraction. The measurement showed that relative to wild type, the Pi content in *MtPT5*-overexpressing plants increased dramatically, especially in *35S:MtPT5-2* line (Figure 5A). The Pi contents of

10.3389/fpls.2022.1005895

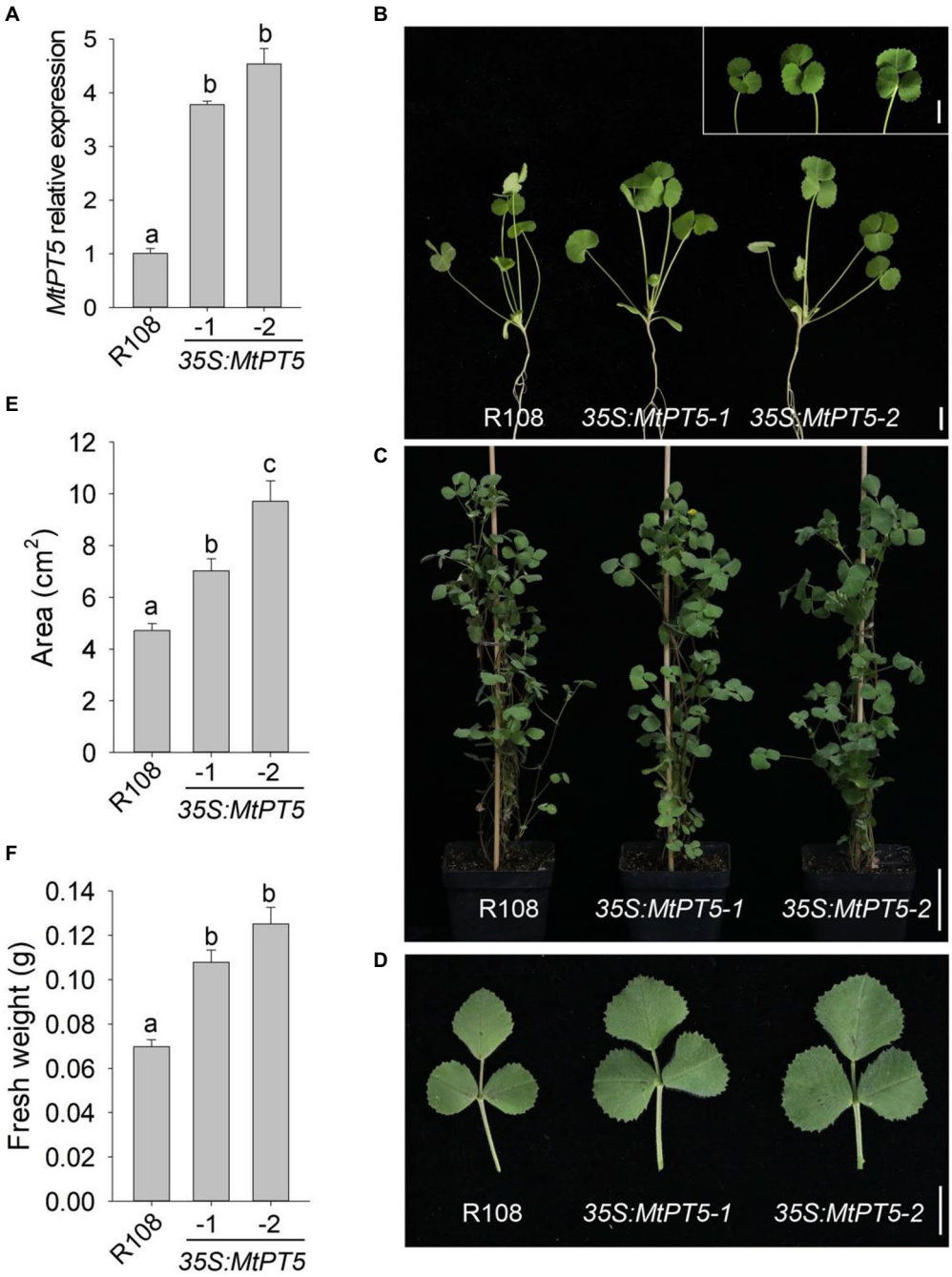


FIGURE 4
Overexpression of *MtPT5* promotes leaves growth of *M. truncatula*. **(A)** qRT-PCR analysis of *MtPT5* in wild-type *M. truncatula* (R108) and *MtPT5*-overexpressing plants (*35S:MtPT5-1* and *35S:MtPT5-2*). Data represent mean±SE (n=3). Different letters indicate significant difference at p<0.05 (One-way ANOVA, Tukey test). **(B–D)** Phenotypic comparation of wild type (R108) and *MtPT5*-overexpressing plants. **(B)** Four-day-old seedlings were transferred to 1/2 Hoagland and grown for 20 days. Inset is leaves detached from the indicated plants. Bars=1cm. **(C)** Plants grown in soil for 3 months. Bar=5cm. **(D)** Top leaflets detached from **(C)**. Bar=1cm. **(E)** Quantification of leaf areas. The fourth leaf of wild type (R108) and *MtPT5*-overexpressing plants shown in **(B)** were taken for quantification. Data represent mean±SE (n=8). Different letters indicate significant difference at p<0.05 (One-way ANOVA, Tukey test). **(F)** Fresh weight of four expanded leaves detached from (B). Data represent mean±SE (n=8). Different letters indicate significant difference at p<0.05 (One-way ANOVA, Tukey test).

whole plants were also measured. Various plants grown in hydroponic culture for 20 days were harvested. The *MtPT5*-overexpressing plants showed significantly higher Pi contents than wild type (Figure 5B). Taken together, these data indicate that overexpression of *MtPT5* enhances Pi accumulation in *M. truancatula*.

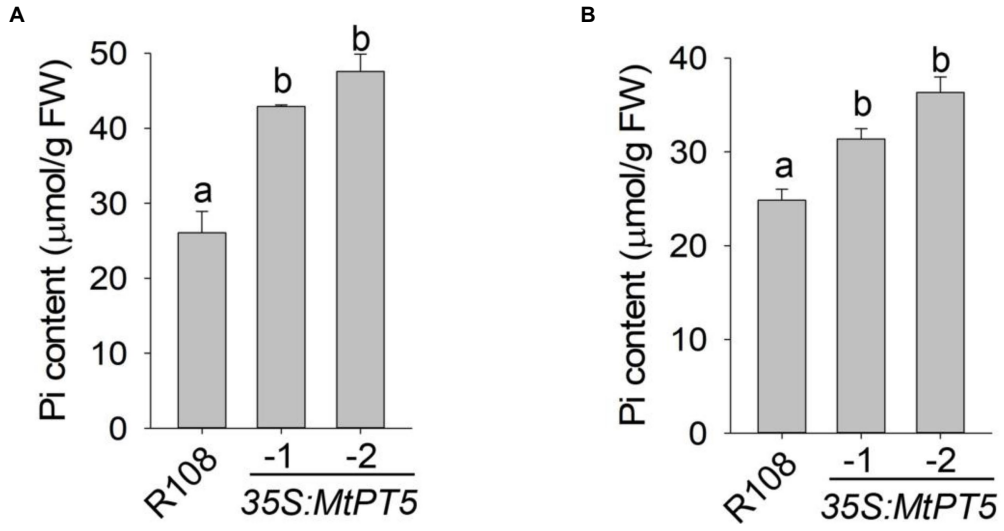


FIGURE 5
Overexpression of *MtPT5* enhances Pi accumulation of *M. truncatula*. **(A)** Pi content in top leaflets of wild type (R108) and *MtPT5*-overexpressing plants grown in soil for 3 months. Data represent mean±SE (n=3). Different letters indicate significant difference at p<0.05 (One-way ANOVA, Tukey test). **(B)** Pi content of wild type (R108) and *35S:MtPT5*-overexpressing plants. Four-day-old seedlings were transferred to 1/2 Hoagland and grown for 20 days, then the whole plants were taken for Pi extraction. Data represent mean±SE (n=5). Different letters indicate significant difference at p<0.05 (One-way ANOVA, Tukey test).

Discussion

Phosphorus (P) is a major determinant of agriculture production. Plants absorb Pi via PHT1 transporters (Harrison et al., 2002), while some of them participate in Pi translocation and remobilization among different organs and tissues (Chang et al., 2019; Wang et al., 2020). It provides opportunities for improving crop performance by studying the functions of PHT1s (Chen and Liao, 2017; Han et al., 2022). Currently, 11 PHT1s have been found in *M. truncatula*. MtPT4 is responsible for Pi acquisition from mycorrhiza and plant root branching (Harrison et al., 2002; Volpe et al., 2016). MtPT6 was reported to promote Pi acquisition in *Arabidopsis* (Cao et al., 2020). Except for MtPT4, the functions of other PHT1s in *M. truncatula* are still unclear. In this study, we uncovered that Pi transporter MtPT5 plays an important role in leaf growth and Pi accumulation in *M. truncatula*.

Analysis of different PHT1s

We found two more PHT1s (Mt4g083960 and Mt5g068140) in *M. truncatula* by searching Ensemblplants.[4] Alignment analysis showed that the 13 PHT1s all contained 12 predicted transmembrane domains, in accordance with the previous report (Pedersen et al., 2013). To choose one member of PHT1 family for further study in *M. truncatula*, phylogenetic tree was firstly constructed using PHT1s from *M. truncatula*, *Arabidopsis*, soybean, maize, and rice. The analysis showed that MtPT5 was closely related to AtPHT1;1 and GmPT7. AtPHT1;1 is an essential Pi transporter in *Arabidopsis*. Under Pi sufficient condition, the mutation of *PHT1;1* leads to about 50% reduction of Pi uptake compared with wild-type plants. The Pi uptake of *pht1;1Δ4Δ* mutant reduces about 75% compared with wild type (Shin et al., 2004). GmPT7 is a nodule-located Pi transporter and responsible for the direct Pi acquisition from soil and Pi translocation from nodules to plant. Overexpression of *GmPT7* improves shoot P content, nitrogen (N) content and soybean yield (Chen et al., 2019). The phylogenetic analysis indicates that MtPT5 probably have essential roles in *M. truncatula* Pi nutrition. The amino acid sequence of MtPT5 shared 84% identity with MtPT1, MtPT2 and MtPT3, whereas MtPT5 displayed an opposite affinity for Pi (Liu et al., 2008). This indicates the multiple functions of different PHT1s in Pi utilization even though PHT1s share high amino acid identities.

Function of MtPT5

MtPT5 was reported to be a membrane-located high-affinity Pi transporter (Liu et al., 2008). To examine the Pi uptake activity of MtPT5 in plants, the coding sequence of *MtPT5* driven by 35S promoter (*35S:MtPT5*) was constructed and introduced into *Arabidopsis* double mutant *pht1;1Δ4Δ*. Phenotypic analysis

4 http://plants.ensembl.org/Medicago_truncatula/Info/Index

showed that *pht1;1Δ4Δ* mutant displayed an arsenate-tolerant phenotype, while the wild type and *35S:MtPT5/pht1;1Δ4Δ* plants displayed arsenate-toxic phenotype. Pi content measurement showed that the Pi content in *pht1;1Δ4Δ* was significantly reduced compared with wild type, in accordance with previously reported (Shin et al., 2004). The Pi contents in *35S:MtPT5/pht1;1Δ4Δ* transgenic lines were rescued to the level of wild type. Taken together, these data demonstrate that MtPT5 has the Pi-transporter activity in plants.

To identify the function of *MtPT5* in *M. truncatula*, two independent *MtPT5*-overexpressing lines (*35S:MtPT5-1* and *35S:MtPT5-2*) were generated with significantly higher *MtPT5* transcript levels. The *MtPT5*-overexpressing lines displayed larger leaves compared with wild type, and the leaf biomasses of the transgenic plants were increased dramatically. The Pi contents of top leaflets and whole plant in *MtPT5*-overexpressing lines were much higher than that in wild-type plants. These data demonstrate that overexpression of *MtPT5* enhances *M. truncatula* leaf growth and Pi accumulation.

Conclusion

Expression analysis showed that *MtPT5* was highly accumulated in shoots, roots and nodules. Previous reports demonstrated that *ZmPT7*, which is expressed in both roots and leaves, participates in Pi acquisition and redistribution in maize (Wang et al., 2020). GmPT7, located to nodules, is responsible for the direct Pi uptake from soil and translocation to fixation zones (Chen et al., 2019). The expression profile of *MtPT5* suggests that it probably have multiple functions in Pi nutrition. In this study, we demonstrate that MtPT5 plays a vital role in Pi accumulation, and overexpression of *MtPT5* promotes the leaf growth of *M. truncatula* dramatically. Leaf size is a vital trait to improve the yield and quality of forage, such as legume alfalfa (*Medicago sativa* L.) (Warman et al., 2011; Zhang et al., 2019). It was reported that about 70% protein of alfalfa is stored in leaves, while the cellulose content in leaves is only 1/3 of that in stems (Yang et al., 2016). Hence, our study provides a clue for elevating alfalfa Pi efficiency and genetic breeding.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

XW and QY planned and designed the research. XW and CW performed the experiments. XW wrote the manuscript. QY supervised this work and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by China Postdoctoral Science Foundation (2021M693441).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1005895/full#supplementary-material

References

Ames, B. N. (1996). Assay of inorganic phosphate, total phosphate and phosphatases. *Methods Enzymol.* 8, 115–118.

Bun-ya, M., Nishimura, M., Harashima, S., and Oshima, Y. (1991). The *PHO84* gene of Saccharomyces cerevisiae encodes an inorganic phosphate transporter. *Mol. Cell. Biol.* 11, 3229–3238.

Cao, Y., Liu, J., Li, Y., Zhang, J., Li, S., An, Y., et al. (2020). Functional analysis of the phosphate transporter gene MtPT6 from *Medicago truncatula*. *Front. Plant Sci.* 11:620377. doi: 10.3389/fpls.2020.620377

Castrillo, G., Sánchez-Bermejo, E., de Lorenzo, L., Crevillén, P., Fraile-Escanciano, A., Tc, M., et al. (2013). WRKY6 transcription factor restricts arsenate uptake and transposon activation in *Arabidopsis*. *Plant Cell* 25, 2944–2957. doi: 10.1105/tpc.113.114009

Catarecha, P., Segura, M. D., Franco-Zorrilla, J. M., García-Ponce, B., Lanza, M., Solano, R., et al. (2007). A mutant of the *Arabidopsis* phosphate transporter PHT1;1 displays enhanced arsenic accumulation. *Plant Cell* 19, 1123–1133. doi: 10.1105/tpc.106.041871

Chang, M. X., Gu, M., Xia, Y. W., Dai, X. L., Dai, C. R., Zhang, J., et al. (2019). OsPHT1;3 mediates uptake, translocation, and remobilization of phosphate under extremely low phosphate regimes. *Plant Physiol.* 179, 656–670. doi: 10.1104/pp.18.01097

Che, J., Yamaji, N., Miyaji, T., Mitani-Ueno, N., Kato, Y., Shen, R. F., et al. (2020). Node-localized transporters of phosphorus essential for seed development in rice. *Plant Cell Physiol.* 61, 1387–1398. doi: 10.1093/pcp/pcaa074

Chen, L., and Liao, H. (2017). Engineering crop nutrient efficiency for sustainable agriculture. *J. Integr. Plant Biol.* 59, 710–735. doi: 10.1111/jipb.12559

Chen, L., Qin, L., Zhou, L., Li, X., Chen, Z., Sun, L., et al. (2019). A nodule-localized phosphate transporter GmPT7 plays an important role in enhancing symbiotic N2 fixation and yield in soybean. *New Phytol.* 221, 2013–2025. doi: 10.1111/nph.15541

Chiou, T. J., and Lin, S. I. (2011). Signaling network in sensing phosphate availability in plants. *Annu. Rev. Plant Biol.* 62, 185–206. doi: 10.1146/annurev-arplant-042110-103849

Clough, S. J., and Bent, A. F. (1998). Floraldip: a simplified method for *agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743. doi: 10.1046/j.1365-313x.1998.00343.x

Cosson, V., Durand, P., d'Erfurth, I., Kondorosi, A., and Ratet, P. (2006). *Medicago truncatula* transformation using leaf explants. *Methods Mol. Biol.* 343, 115–127. doi: 10.1007/978-1-4939-1695-5_4

Dai, C., Dai, X., Qu, H., Men, Q., Liu, J., Yu, L., et al. (2022). The rice phosphate transporter OsPHT1;7 plays a dual role in phosphorus redistribution and anther development. *Plant Physiol.* 188, 2272–2288. doi: 10.1093/plphys/kiac030

Dobre, P., Jurcoane, S., Cristea, S., Matei, F., Moraru, A. C., and Dincă, L. (2014). Influence of N, P chemical fertilizers, row distance and seeding rate on camelina crop. *Agrolife Sci. J.* 3:1

Han, Y., White, P. J., and Cheng, L. (2022). Mechanisms for improving phosphorus utilization efficiency in plants. *Ann. Bot.* 129, 247–258. doi: 10.1093/aob/mcab145

Harrison, M. J., Dewbre, G. R., and Liu, J. (2002). A phosphate transporter from *Medicago truncatula* involved in the acquisition of phosphate released by arbuscular mycorrhizal fungi. *Plant Cell* 14, 2413–2429. doi: 10.1105/tpc.004861

Heuer, S., Gaxiola, R., Schilling, R., Herrera-Estrella, L., Lopez-Arredondo, D., Wissuwa, M., et al. (2017). Improving phosphorus use efficiency: a complex trait with emerging opportunities. *Plant J.* 90, 868–885. doi: 10.1111/tpj.13423

Karandashov, V., and Bucher, M. (2005). Symbiotic phosphate transport in arbuscular mycorrhizas. *Trends Plant Sci.* 10, 22–29. doi: 10.1016/j.tplants.2004.12.003

Liu, F., Chang, X. J., Ye, Y., Xie, W. B., Wu, P., and Lian, X. M. (2011). Comprehensive sequence and whole-life-cycle expression profile analysis of the phosphate transporter gene family in rice. *Mol. Plant* 4, 1105–1122. doi: 10.1093/mp/ssr058

Liu, J., Versaw, W. K., Pumplin, N., Gomez, S. K., Blaylock, L. A., and Harrison, M. J. (2008). Closely related members of the *Medicago truncatula* PHT1 phosphate transporter gene family encode phosphate transporters with distinct biochemical activities. *J. Biol. Chem.* 283, 24673–24681. doi: 10.1074/jbc.M802695200

López-Arredondo, D. L., Leyva-González, M. A., Alatorre-Cobos, F., and Herrera-Estrella, L. (2013). Biotechnology of nutrient uptake and assimilation in plants. *Int. J. Dev. Biol.* 57, 595–610. doi: 10.1387/ijdb.130268lh

López-Arredondo, D. L., Leyva-González, M. A., González-Morales, S. I., LópezBucio, J., and Herrera-Estrella, L. (2014). Phosphate nutrition: improving low-phosphate tolerance in crops. *Annu. Rev. Plant Physiol.* 65, 95–123. doi: 10.1146/annurev-arplant-050213-035949

Loth-Pereda, V., Orsini, E., Courty, P. E., Lota, F., Kohler, A., Diss, L., et al. (2011). Structure and expression profile of the phosphate Pht1 transporter gene family in mycorrhizal *Populus trichocarpa*. *Plant Physiol.* 156, 2141–2154. doi: 10.1104/pp.111.180646

Marschner, H., and Rimmington, G. (1988). Mineral nutrition of higher plants. *Plant Cell Environ.* 11, 147–148.

Muchhal, U. S., Pardo, J. M., and Raghothama, K. G. (1996). Phosphate transporters from the higher plant *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* 93, 10519–10523. doi: 10.1073/pnas.93.19.10519

Mudge, S. R., Rae, A. L., Diatloff, E., and Smith, F. W. (2002). Expression analysis suggests novel roles for members of the Pht1 family of phosphate transporters in *Arabidopsis*. *Plant J.* 31, 341–353. doi: 10.1046/j.1365-313X.2002.01356.x

Pedersen, B. P., Kumar, H., Waight, A. B., Risenmay, A. J., Roe-Zurz, Z., Chau, B. H., et al. (2013). Crystal structure of a eukaryotic phosphate transporter. *Nature* 496, 533–536. doi: 10.1038/nature12042

Péret, B., Clément, M., Nussaume, L., and Desnos, T. (2011). Root developmental adaptation to phosphate starvation: better safe than sorry. *Trends Plant Sci.* 16, 442–450. doi: 10.1016/j.tplants.2011.05.006

Raghothama, K. G. (1999). Phosphate acquisition. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* 50, 665–693. doi: 10.1146/annurev.arplant.50.1.665

Seo, H. M., Jung, Y., Song, S., Kim, Y., Kwon, T., Kim, D. H., et al. (2008). Increased expression of *OsPT1*, a high-affinity phosphate transporter, enhances phosphate acquisition in rice. *Biotechnol. Lett.* 30, 1833–1838. doi: 10.1007/s10529-008-9757-7

Shin, H., Shin, H. S., Dewbre, G. R., and Harrison, M. J. (2004). Phosphate transport in *Arabidopsis*: Pht1;1 and Pht1;4 play a major role in phosphate acquisition from both low-and high-phosphate environments. *Plant J.* 39, 629–642. doi: 10.1111/j.1365-313X.2004.02161.x

Smith, V. H., and Schindler, D. W. (2009). Eutrophication science: where do we go from here? *Trends Ecol. Evol.* 24, 201–207. doi: 10.1016/j.tree.2008.11.009

Veneklaas, E. J., Lambers, H., Bragg, J., Finnegan, P. M., Lovelock, C. E., Plaxton, W. C., et al. (2012). Opportunities for improving phosphorususe efficiency in crop plants. *New Phytol.* 195, 306–320. doi: 10.1111/j.1469-8137.2012.04190.x

Versaw, W. K., and Garcia, L. R. (2017). Intracellular transport and compartmentation of phosphate in plants. *Curr. Opin. Plant Biol.* 39, 25–30. doi: 10.1016/j.pbi.2017.04.015

Vincent, A. G., Schleucher, J., Grobner, G., Vestergren, J., Persson, P., Jansson, M., et al. (2012). Changes in organic phosphorus composition in boreal forest humus soils: The role of iron and aluminium. *Biogeochemistry* 108, 485–499. doi: 10.1007/s10533-011-9612-0

Volpe, V., Giovannetti, M., Sun, X. G., Fiorilli, V., and Bonfante, P. (2016). The phosphate transporters LjPT4 and MtPT4 mediate early root responses to phosphate status in non-mycorrhizal roots. *Plant Cell Environ.* 39, 660–671. doi: 10.1111/pce.12659

Wang, F., Cui, P. J., Tian, Y., Huang, Y., Wang, H. F., Liu, F., et al. (2020). Maize ZmPT7 regulates pi uptake and redistribution which is modulated by phosphorylation. *Plant Biotechnol. J.* 18, 2406–2419. doi: 10.1111/pbi.13414

Wang, H., Xu, Q., Kong, Y. H., Chen, Y., Duan, J. Y., Wu, W. H., et al. (2014). *Arabidopsis* WRKY45 transcription factor activates PHOSPHATE TRANSPORTER1;1 expression in response to phosphate starvation. *Plant Physiol.* 164, 2020–2029. doi: 10.1104/pp.113.235077

Warman, L., Moles, A. T., and Edwards, W. (2011). Not so simple after all: searching for ecological advantages of compound leaves. *Oikos* 120, 813–821. doi: 10.1111/j.1600-0706.2010.19344.x

Yan, X., Chen, X., Ma, C., Cai, Y., and Zhang, F. (2021). What are the key factors affecting maize yield response to and agronomic efficiency of phosphorus fertilizer in China? *Field Crop Res* 270:108221. doi: 10.1016/j.fcr.2021.108221

Yang, Q. C., Kang, J. M., Zhang, T. J., Liu, F. Q., Long, R. C., and Sun, Y. (2016). Distribution, breeding and utilization of alfalfa germplasm resources. *Chin. Sci. Bull.* 61, 261–270.

Zak, D., Kronvang, B., Carstensen, M. V., Hoffmann, C. C., Kjeldgaard, A., Larsen, S. E., et al. (2018). Nitrogen and phosphorus removal from agricultural runoff in integrated buffer zones. *Environ. Sci. Technol.* 52, 6508–6517. doi: 10.1021/acs.est.8b01036

Zhang, F., Chen, X., and Vitousek, P. (2013). An experiment for the world. *Nature* 497, 33–35. doi: 10.1038/497033a

Zhang, C., Li, Z., Zhang, C. Y., Li, M., Lee, Y., and Zhang, G. G. (2019). Extract methods, molecular characteristics, and bioactivities of polysaccharide from alfalfa (*Medicago sativa* L.). *Nutrients* 11:1181. doi: 10.3390/nu11051181

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 September 2022
DOI 10.3389/fpls.2022.976311

OPEN ACCESS

EDITED BY
Jorge Fernando Pereira,
Embrapa Gado de Leite, Brazil

REVIEWED BY
Susilaine Maira Savassa,
Center for Nuclear Energy in Agriculture,
University of São Paulo, Brazil
Muhammad Hamzah Saleem,
University of Chinese Academy of Sciences,
China
Muhammad Naeem,
Agriculture Department,
Government of Punjab,
Pakistan

*CORRESPONDENCE
Fugui Mi
mfguinm@163.com

SPECIALTY SECTION
This article was submitted to
Crop and Product Physiology,
a section of the journal
Frontiers in Plant Science

RECEIVED 23 June 2022
ACCEPTED 25 July 2022
PUBLISHED 06 September 2022

CITATION
Li X, Zhang L, Ren H, Wang X and
Mi F (2022) Zinc toxicity response in
Ceratoides arborescens and identification
of *CaMTP*, a novel zinc transporter.
Front. Plant Sci. 13:976311.
doi: 10.3389/fpls.2022.976311

Zinc toxicity response in *Ceratoides arborescens* and identification of *CaMTP*, a novel zinc transporter

Xingyue Li[1], Lin Zhang[2], Haiyan Ren[1], Xiaoyu Wang[1] and Fugui Mi[1]*

[1]College of Grassland, Resources and Environment, Inner Mongolia Agricultural University, Hohhot, China, [2]M-Grass Ecology and Environment (Group) Co., Ltd., Hohhot, China

Zinc (Zn) is an essential micronutrient for several physiological and biochemical processes. Changes in soil Zn levels can negatively affect plant physiology. Although the mechanism of Zn nutrition has been studied extensively in crops and model plants, there has been little research on steppe plants, particularly live in alkaline soils of arid and semiarid regions. *Ceratoides arborescens* is used in arid and semiarid regions as forage and ecological restoration germplasm, which is studied can enrich the mechanism of Zn nutrition. The plants were exposed to three different Zn treatments, Zn-deficient (-Zn 0mML^{-1}), Zn-normal (Control, 0.015mML^{-1}), and Zn-excess (+Zn, 0.15mML^{-1}), for 3 weeks. Individual biomass, ion concentrations, photosynthetic system, and antioxidant characteristics were measured. High Zn supply significantly decreased plant biomass and induced chlorosis and growth defects and increased Zn concentration but decreased Fe and Ca concentrations, unlike in controls (p<0.05). High Zn supply also reduced plant chlorophyll content, which consequently decreased the photosynthesis rate. Increased concentrations of malondialdehyde and soluble sugar and activities of peroxidase and superoxide dismutase could resist the high-level Zn stress. In contrast, low Zn supply did not affect plant growth performance. We also identified a novel protein through RNA transcriptome analysis, named *CaMTP*, that complemented the sensitivity of a yeast mutant to excessive Zn, which was found to be localized to the endoplasmic reticulum through transient gene expression in Nicotiana benthamiana. The gene *CaMTP* identified to be highly sensitive to Zn stress is a potential candidate for overcoming mineral stress in dicot crop plants.

KEYWORDS

zinc deficiency, zinc excessive, photosynthesis, peroxisome, transport, endoplasmic reticulum

Introduction

Zn, the second most abundant transition metal in living organisms, is vital to crop nutrition as it is required in various enzymatic reactions, metabolic processes, and oxidation reduction reactions (Broadley et al., 2007; Al Jabri et al., 2022). Zn is one of the most limited micronutrients in soils and plants (Amini et al., 2022). With high pH or alkaline stress, many arid and semiarid lands are Zn deficient (McFarland et al., 1990; Scharwies and Dinneny, 2019). Zn deficiency could lead to a lack of cellular integrity, dwarf stems, light chlorosis, and decreases in chlorophyll synthesis (Khatun et al., 2018; Kabir et al., 2021). The deficiency of Zn also causes the accumulation of excessive reactive oxygen species (ROS) that may be due to the lower concentration of the Cu-Zn-SOD enzyme (Marreiro et al., 2017). Another effect of Zn deficiency is reduced plant pigments and interference with the process of photosynthesis, thereby inhibiting plant growth (Sidhu et al., 2020).

Anthropogenic activities, including mining, smelting, and fertilizing with sewage sludge, are associated with Zn buildup in soils and water, resulting in Zn pollution (Broadley et al., 2007; Sinclair and Krämer, 2012). High Zn levels may also induce toxicity affecting most vital processes in plants (Marichali et al., 2016). Common symptoms of Zn toxicity include growth inhibition, repression of root elongation owing to the inhibition of cell proliferation, alteration in water and nutrient uptake, loss of membrane integrity, disruption of redox homeostasis, reduction of chlorophyll content and subsequent photosynthesis, generation of ROS, and the manifestation of oxidative stress (Balafrej et al., 2020). The induction of the antioxidant system has been described as a putative strategy against trace metal toxicity; accumulating evidence indicates that the activities of catalase (CAT), superoxide dismutase (SOD), guaiacol peroxidase (GPX), and ascorbate peroxidase (APX) are increased in Zn-exposed plants (Rout and Das, 2003; Ahmad et al., 2022).

Plants have established a tightly controlled system to balance the uptake, utilization, and storage of Zn ions (Yin et al., 2022). One mechanism involves the production of root exudates to tolerate deficiency or excess (Santi and Schmidt, 2009; Tsednee et al., 2014). Furthermore, enzymes with antioxidant properties play key roles in controlling radicals (superoxide radical, hydrogen radical, and hydroxyl radical) and peroxides at the cellular level and protecting plants from abiotic stresses (Rout and Das, 2003; Kumar Tewari et al., 2008; Kabir et al., 2021). At the molecular level, Zn homeostasis in plants is tightly regulated by Zn sensors and metal chelators involved in Zn acquisition and sequestration (Yin et al., 2022). Zn cannot diffuse across cell membranes; hence specific Zn transporters are required to transport Zn into the cytoplasm (Kaur and Garg, 2021). Some plant membrane transporter families, such as ZRT/IRT-like proteins (ZIP) family, P-type ATPase family, natural resistance-associated macrophage protein (NRAMP) family, and cation diffusion facilitator (CDF) family, have been shown to participate in the Zn^{2+} uptake, transport, and maintaining homeostasis (Krämer et al., 2007). CDF, also referred to as metal tolerance/transport protein (MTP) in plants, is involved in metal transport out of the cytoplasm and is induced by excess Zn (Ricachenevsky et al., 2013). MTP3 is located on the tonoplast (Zhao et al., 2018) and complements the yeast mutant zrc1cot1, which is sensitive to excess Zn levels (Lu et al., 2009), while MTP2 is located on the endoplasmic reticulum (ER) membrane (Sinclair et al., 2018), and MTP5 and MTP12 are located at Golgi in *Arabidopsis thaliana* and cucumber (Migocka et al., 2018).

The magnitude of Zn tolerance in plants depends on different plant species (Mateos-Naranjo et al., 2014), and the physiological range between deficiency and toxicity of Zn is narrow (Glińska et al., 2016). Many species grew in acidic soils since they are rich in metals (Meyer et al., 2016). A study with the establishment of salt-tolerant plants on soils contaminated with soluble salts by petroleum exploration activities in arid and semiarid areas showed that the survival rates of *Ceratoides lunata* and *Kochia prostrata* were 61 and 40%, respectively, but *Atripkx nummukzria* suffered 100% mortality (McFarland et al., 1990). This fact indicates that *Ceratoides* sp. can tolerate salt, but little is known regarding its response to Zn deficiency and toxicity. *Ceratoides arborescens*, a forage endemic to China, was selected in this study because it accumulates mineral elements in adequate capacities, produces copious amounts of high-quality biomass, and is widely used in arid and semiarid regions of China as an excellent forage and ecological restoration grass (Liu and Yi, 2004).

To date, data on the uptake or accumulation of nutrient elements and metabolic profiling of Zn stress in *C. arborescens* remain limited. Therefore, this study investigated the effects of Zn treatment on *C. arborescens* by analyzing plant growth, contents of ions and chlorophylls, antioxidant activities, and the photosynthetic system within a controlled hydroponic culture and identified the transcripts encoding proteins involved in Zn transport through yeast complementation. Our study is essential for bioaugmentation in arid and semiarid regions of dicot crop plants and elucidates the physiological and molecular mechanisms underlying the differential responses to various Zn supply levels in *C. arborescens*.

Materials and methods

Plant materials and cultivation technique

In October 2020, diaspores of *C. arborescens* were collected from the desert steppe of Siziwang Banner, Inner Mongolia, China (E 112.113942°, N 41.996905°). Dry seeds were stored at 4°C before being used in the germination experiment.

Seeds were sterilized with 75% ethyl alcohol for 3 min, followed by double Milli-Q water before being transferred to the germination tray in a growth chamber with a 12/12 h light/dark light period and 100–120 $\mu mol\, m^{-2}\, S^{-1}$ light intensity. Sprouted three-day homogeneous plantlets were transferred to a hydroponic solution of pH 5.7 and cultured for 2 weeks.

The composition of the hydroponic solution (μM L^{-1}) was as follows: KNO_3 (5,000 μM L^{-1}), $CaCl_2$ (1,000 μM L^{-1}), H_3PO_4 (1,000 μM L^{-1}), $MgSO_4.7H_2O$ (1,000 μM L^{-1}), KI (2.5 μM L^{-1}), H_3BO_3 (50 μM L^{-1}), Fe-EDTA (50 μM L^{-1}), $MnSO_4{\cdot}H_2O$ (50 μM L^{-1}), $ZnSO_4{\cdot}7H_2O$ (15 μM L^{-1}), $Na_2MoO_4.2H_2O$ (0.52 μM L^{-1}), $CoCl_2{\cdot}6H_2O$ (0.053 μM L^{-1}), and $CuSO_4{\cdot}5H_2O$ (0.05 μM L^{-1}). Then, the seedlings were transferred to the following treatment solutions (pH 5.7) and cultured for 3 weeks: Zn- excess (+Zn), $ZnSO_4{\cdot}7H_2O$ (150 μM L^{-1}); Zn-normal (control), $ZnSO_4{\cdot}7H_2O$ (15 μM L^{-1}); and Zn-deficient (-Zn), $ZnSO_4{\cdot}7H_2O$ (0 μM L^{-1}) (Wang et al., 2017).

Metal contents analysis

Tissues from the root, stem, and leaf were harvested and dried at 105°C in an oven for 3 days, and then the Zn, Fe, and Ca levels were measured. The dried tissues were weighed and ground to a powder, which was later extracted with 5% HNO_3 for 3 days. The extracts were centrifuged at 12,000 × *g* for 30 min, and the supernatants obtained were then filtered through a 0.45 μm membrane. The levels of Zn, Fe, and Ca were quantified using an Inductive Coupled Plasma Emission Spectrometer (ICP-OES_PQ9000; Analytikjena, German).

Chlorophyll contents and photosynthesis activity parameters analysis

Plant samples, weighing 0.2 g, were grounded with liquid nitrogen and then extracted with 10 ml of 96% ethanol for 24 h in darkness; all these steps were carried out at 4°C. The extract was centrifuged at 10000 × *g* for 10 min, the absorbance of the supernatant was measured at 649 and 665 nm, and chlorophyll concentrations were calculated (van Dijk and Roelofs, 1988). The photosynthetic activity parameters, such as the quantum yield of photosystem I [Y(I)], the quantum yield of photosystem II [Y(II)], and the non-photochemical quenching (NPQ) levels were measured (photosynthesis measurement system GFS-3000; WALZ, German).

Antioxidant enzyme activities and lipid peroxidation analysis

Approximately 0.5 g of fresh root tissues were grounded using liquid nitrogen, followed by homogenization in 1.5 ml of an extraction buffer consisting of 0.05 M pre-cooled phosphate buffer (pH 7.8). The extract was centrifuged at 10,000 × *g* for 20 min at 4°C, and the supernatant was collected and stored at −20°C. SOD activity was determined spectrophotometrically at a wavelength of 560 nm based on the inhibition of nitroblue tetrazolium (NBT) photochemical reduction (Giannopolitis and Ries, 1977). Peroxidase (POD) activity was assessed by determining guaiacol oxidation by H_2O_2 at a wavelength of 470 nm (Giannopolitis and Ries, 1977).

Approximately 0.5 g of fresh roots or shoots were homogenized with 5 ml of 5% (v/v) trichloroacetic acid (TCA) and then centrifuged at 10,000 × *g* for 10 min. MDA concentrations were determined by using the following formula (Hodges et al., 1999):

$$MDA\left(\mu mol/g\ FW\right) = 6.45 \times \left(OD_{532} - OD_{600}\right) - 0.56 \times OD_{450}$$

Approximately 0.5 g of fresh root tissues were grounded in quartz sand and distilled water, followed by homogenization in 10 ml of distilled water and then heated at 80°C for 30 min. The homogenate was filtered, the filtrate volume was made up to 100 ml, and the soluble sugar (SS) levels were measured using resorcinol colorimetry (Williams and Martin, 1967).

Gene expression and bioinformatics analysis

Roots of *C. arborescens* grown in Zn-excess conditions and control hydroponic solutions were collected and subjected to RNA extraction using the RNAiso for Polysaccharide-Rich Plant Tissue kit (Takara, China). Transcriptome analysis was conducted by Beijing Capital Bio Corporation according to the standard procedure of the Illumina HiSeq 2,500 sequencing platform. Transcripts encoding putative Zn^{2+} transporting proteins were identified by searching the transcript annotation tables for the keywords Zinc, Zn^{2+}, Zinc transporting, Zn^{2+} transporting, ZIP, and CDF. *CaMTP* homolog was searched using its amino acid sequence at NCBI Protein Blast.[1] The protein structure was predicated based on its amino acid sequence.[2]

Subcellular localization of *CaMTP*

CaMTP was inserted into the vector pUC57 between the NotI restriction enzyme site on one side and the NotI and KpnI restriction enzyme sites on the other side of the vector. The full-length coding sequence of *CaMTP* without the stop codon was amplified from pUC57 using PrimeStar HS DNA Polymerase (Takara) and the primers "*CaMTP*-pENTR 3C-F" and "*CaMTP*-pENTR 3C-R" (Table 1) containing BamHI and NotI restriction enzyme sites, respectively. The PCR product was ligated into the pEASY-Blunt Simple Cloning Vector (TransGen, Beijing, China) and confirmed *via* sequencing. The *CaMTP* coding sequence was ligated between the BamHI and NotI sites of pENTR 3C (Thermo Fisher, United States). Subsequently, the *CaMTP* coding sequence without the stop codon was ligated downstream of the 35S promoter and upstream of the GFP coding sequence in plant binary expression vector pGWB605 (Nakagawa et al., 2007) using

1 https://blast.ncbi.nlm.nih.gov/Blast.cgi (Accessed July 28, 2022).

2 http://www.ebi.ac.uk/interpro/ (Accessed July 28, 2022).

TABLE 1 Sequence of primers used to insert *CaMTP* into the pENTR 3C and pYUL2 vectors.

Primer name	Primer sequences (5′-3′)
CaMTP-pENTR 3C-F	CGCG GATC CATG TTT ACA GCC CGA CAT GAT CCC ACA C
CaMTP-pENTR 3C-R	TT GC GGCC GCGGAGC CCA ACG TCC AAC GAC AGA CAT
CaMTP-pYUL2-F	CGCG GATC CATG TTT ACA GCC CGA CAT GAT CCC ACA C
CaMTP-pYUL2-R	CCCA AGCT TTTA AGC CCA ACG TCC AAC GAC AGA CAT

Gateway LR Clonase (Thermo Fisher). The pGWB605 vector harboring 35S::*CaMTP*-GFP fragment was introduced into Agrobacterium strain GV3101.

Agrobacterium cells containing pGWB605-35S::*CaMTP*-GFP or individual subcellular marker vectors were co-infiltrated into the lower epidermis of tobacco (N benthamiana) leaves. Confocal microscopy (LSM 710; Zeiss, China) or super-resolution confocal microscopy was used to examine GFP and m-Cherry fluorescence (LSM 880; Zeiss, China).

Yeast strains and vector manipulation

The yeast strains utilized in this study were (1) Zn^{2+} uptake deficient mutant zrt1/zrt2 (ZHY3; MAT; ade6; can1; his3; leu2; trp1; ura3; zrt1::LEU2; zrt2::HIS3) and (2) Zn^{2+} accumulation defective and high [Zn^{2+}] sensitive mutant zrc1/cot1 (BY4741) (Lin et al., 2008).

The yeast expression vector, pFL61 (Minet et al., 1992), was used to transform the zrt1/zrt2Δyeast mutants, with *AtZIP7* serving as a positive control (Milner et al., 2013). *CaMTP* from pUC57 was digested using NotI. The digested product was subsequently ligated into the NotI site of pFL61, downstream of the phosphoglycerate kinase (PGK) promoter. The pFL61 vector harboring the insert in the correct orientation was identified *via* dual digestion of the plasmid with NotI and EcoRI.

The yeast expression vector pYUL2 was used to transform Δzrc1/cot1, in which *Medicago truncatula* MTP3 (*MtMTP3*) served as a positive control. The *CaMTP* coding sequence was ligated between the BamHI and HindIII sites downstream of the alcohol dehydrogenase1 (ADH1) promoter in pYUL2 following amplification with the primers "*CaMTP*-pYUL2-F" and "*CaMTP*-pYUL2-R" (Table 1).

Yeast transformation and phenotypic analysis

Following the Yeastmaker Yeast Transformation System's instructions, competent yeast cells were prepared and transformed (Takara). Following the directions on the Yeast Plasmid Extraction kit, plasmids were extracted and purified from the yeast cells (Solarbio, China). By screening the minimal synthetically defined (SD) base medium with -Ura DO Supplement (SD/-Ura; Takara) and PCR-based verification of the predicted inserts, yeast cells carrying the appropriate plasmid were identified.

Similar-sized yeast colonies were selected from the SD/-Ura selective medium and grown in 5 ml 1× Yeast Extract Peptone Dextrose medium (YPD) for 12–14 h at 250 rpm at 30°C. YPD contains 1% yeast extract, 2% peptone, and 2% dextrose. The pellets were twice rinsed with sterile water after the cultures were centrifuged at 700 g for 5 min. An OD600 of 1 was achieved by dissolving the yeast pellets in sterilized water and adjusting the volume. To create solutions with an OD600 of 0.1, 0.01, 0.001, and 0.0001, the cultures were diluted. A volume of 10 μl of these solutions were spotted onto solid full-nutrient medium or medium used to test Zn^{2+} and Fe^{2+}-dependent growth phenotypes.

The positive and negative controls employed for the Zn^{2+} uptake-defective yeast mutant zrt1zrt2 were pFL61-AtZIP7 and pFL61 empty vector, respectively. The YPD medium is referred to as a full-nutrient medium. The YPD medium was added with 50 M ZnCl2 and 1 mM EDTA to investigate the Zn^{2+}-deficient growth phenotype zrt1zrt2 (pH 6.5). The positive and control samples were pYUL2-MtMTP3 and the empty vector pYUL2, respectively, for assessing the growth phenotype of the Zn^{2+}-storage-defective and excess Zn^{2+}-sensitive yeast mutant zrc1cot1. The Zn^{2+}-excess growth media was the YPD medium supplemented with 7 mM $ZnSO_4$.

Statistical analysis

One-way ANOVA and Tukey's HSD test was used to examine the effects of the Zn treatments on plant dry weight, metal ion concentration, photosynthetic pigment levels, photosynthesis rate, and activity of antioxidant enzymes. Treatment effects were considered to be statistically significant at $p < 0.05$. Statistical analyses were performed in R version 3.6.1 (R Core Team, 2018).

Results

Effects of Zn treatments on plant growth

At the end of the experiment, we found that Zn-excess inhibited the growth of *C. arborescens*, displaying retarded lateral root growth and leaf chlorosis (Figure 1A). Excessive Zn significantly decreased the root and leaf dry weight by 79.8 and 67.9% compared to the control, respectively (Figure 1B). However, Zn-deficient treatment did not affect the organic dry weight of *C. arborescens* (Figure 1B).

Effects of Zn treatments on metal ion concentration

Zn-excess treatment significantly decreased Ca and Fe levels but increased Zn levels in the roots and shoots (Table 2).

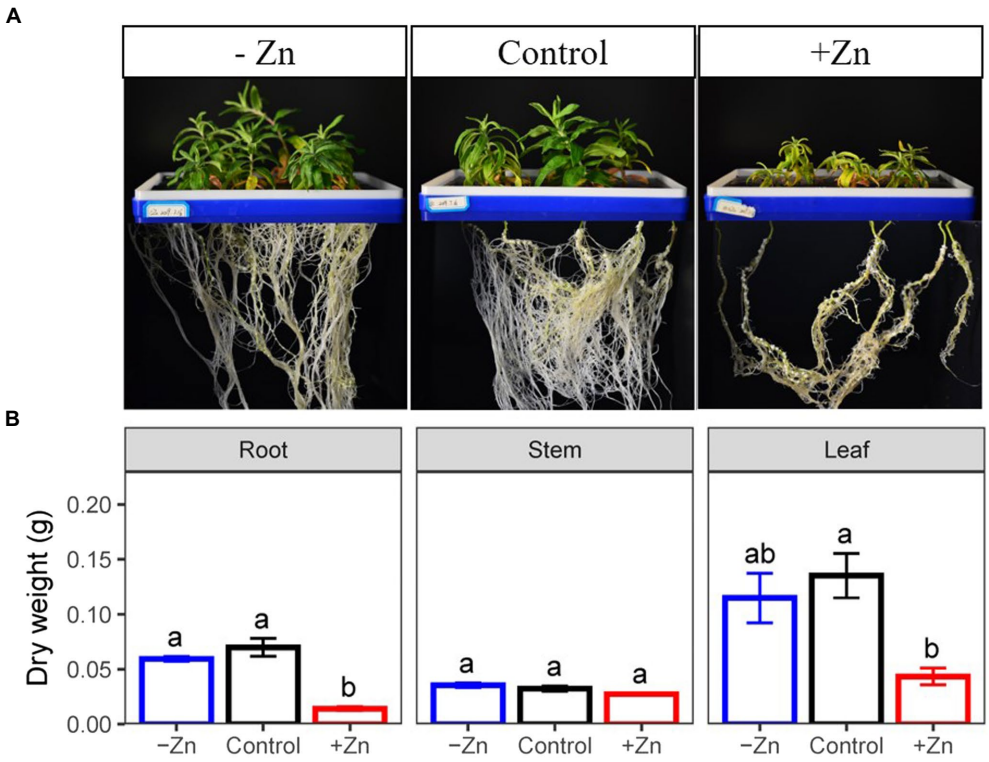


FIGURE 1
(A) Root, stem, and **(B)** leaf dry weight of *Ceratoides arborescens* subjected to Zn-deficient (0mML^{-1}), control (0.015mML^{-1}), and Zn-excess (0.15mML^{-1}) conditions. Different lowercase letters indicate significant differences among treatments (p<0.05).

TABLE 2 Zn, Fe, and Ca concentration in the root, stem, and leaf of *Ceratoides arborescens* subjected to Zn-deficient (0mML^{-1}), control (0.015mML^{-1}), and Zn-excess (0.15mML^{-1}) treatments.

Treatment		−Zn	Control	+Zn
Zn concentration (μg mg^{-1})	Root	0.276 ± 0.038^{b}	0.406 ± 0.118^{b}	2.272 ± 0.074^{a}
	Stem	0.092 ± 0.011^{b}	0.075 ± 0.007^{b}	1.227 ± 0.002^{a}
	Leaf	0.083 ± 0.004^{b}	0.066 ± 0.010^{b}	1.259 ± 0.10^{a}
Fe concentration (μg mg^{-1})	Root	5.084 ± 0.387^{a}	4.543 ± 0.607^{a}	1.158 ± 0.267^{b}
	Stem	0.105 ± 0.017^{a}	0.049 ± 0.002^{b}	0.036 ± 0.007^{b}
	Leaf	0.065 ± 0.004^{a}	0.064 ± 0.001^{a}	0.024 ± 0.001^{b}
Ca concentration (μg mg^{-1})	Root	6.104 ± 0.237^{a}	6.796 ± 0.470^{a}	4.104 ± 0.102^{b}
	Stem	11.062 ± 0.660^{ab}	11.888 ± 0.614^{a}	8.903 ± 0.451^{b}
	Leaf	9.477 ± 1.232^{b}	16.176 ± 0.152^{a}	9.508 ± 0.705^{b}

Different lowercase letters indicate significant differences among treatments (p < 0.05).

Compared to the control, the concentration of Zn increased by 4.6 times for roots, 15 times for stems, and 18 times for leaves under Zn-excess. Meanwhile, the concentration of Fe in the roots, stems, and leaves decreased by 74.5, 26.6, and 62.5%, respectively, and the concentration of Ca in the roots, stems, and leaves decreased by 39.6, 25.1, and 41.2%, respectively, under Zn-excess conditions. However, Zn-deficient treatment does not affect the concentration of Zn in plants. The concentration of Fe in stems increased by 1.14 times, while the concentration of Ca in leaves decreased by 41.4%.

Effects of Zn treatments on chlorophyll contents and photosynthesis activity parameters

Treatment with high levels of Zn significantly influenced the chlorophyll contents and photosynthesis activity parameters. On average, Zn-excess treatment decreased the levels of chlorophyll (Figure 2A), Y(I) (Figure 2B), and Y(II) (Figure 2C) of *C. arborescens* by 30.6, 26.2, and 23.8%, respectively. Zn-excess treatment increased the NPQ level by 1.6 times (Figure 2D). However, Zn-deficient treatment did not affect the levels of chlorophyll, Y(I), Y(II), or NPQ of *C. arborescens* (Figure 2).

Assessment of lipid peroxidation and the activity of antioxidant enzymes

Under excess levels of Zn, the concentrations of (A) MDA and (B) SS and the activities of (C) SOD and (D) POD in the roots of *C. arborescens* increased significantly. Compared to control, MDA concentration, SS concentration, SOD activity, and POD activity in the roots subjected to Zn-excess conditions were increased by 44.3, 50.7, 9.4, and 27.1%, respectively. However, Zn deficiency did not affect the activities of POD and SOD or the concentrations of MDA and SS (Figure 3).

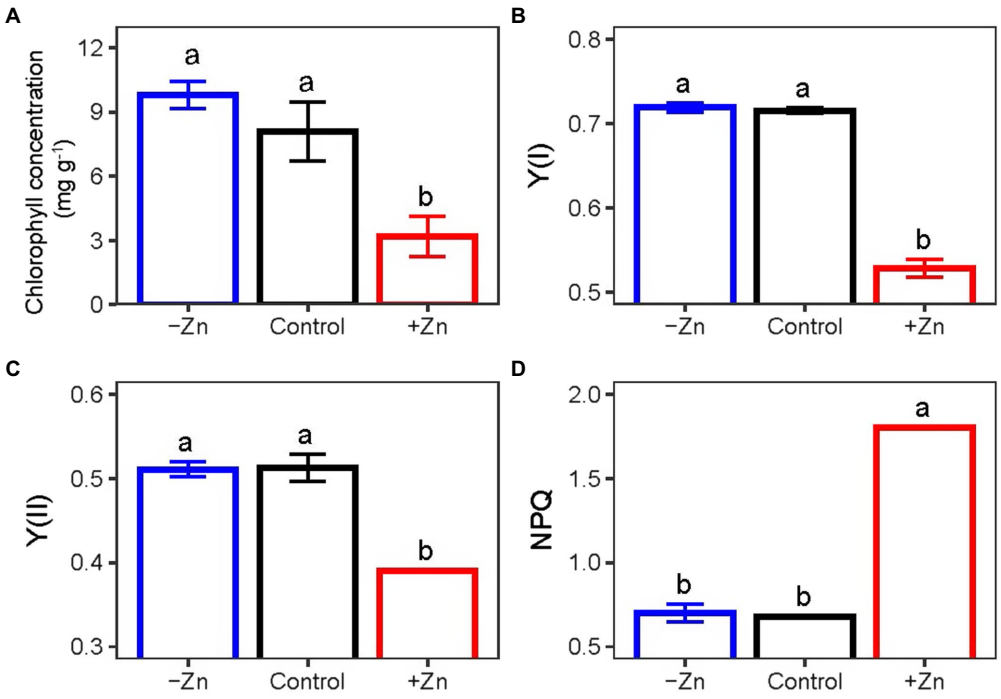


FIGURE 2
(A) The chlorophyll, **(B)** Y(I), **(C)** Y(II), and (**D**) NPQ levels of *C. Arborescens* under Zn-deficient (0mML^{-1}), control (0.015mML^{-1}), and Zn-excess (0.15mML^{-1}) conditions. Different lowercase letters indicate significant differences among treatments (p<0.05).

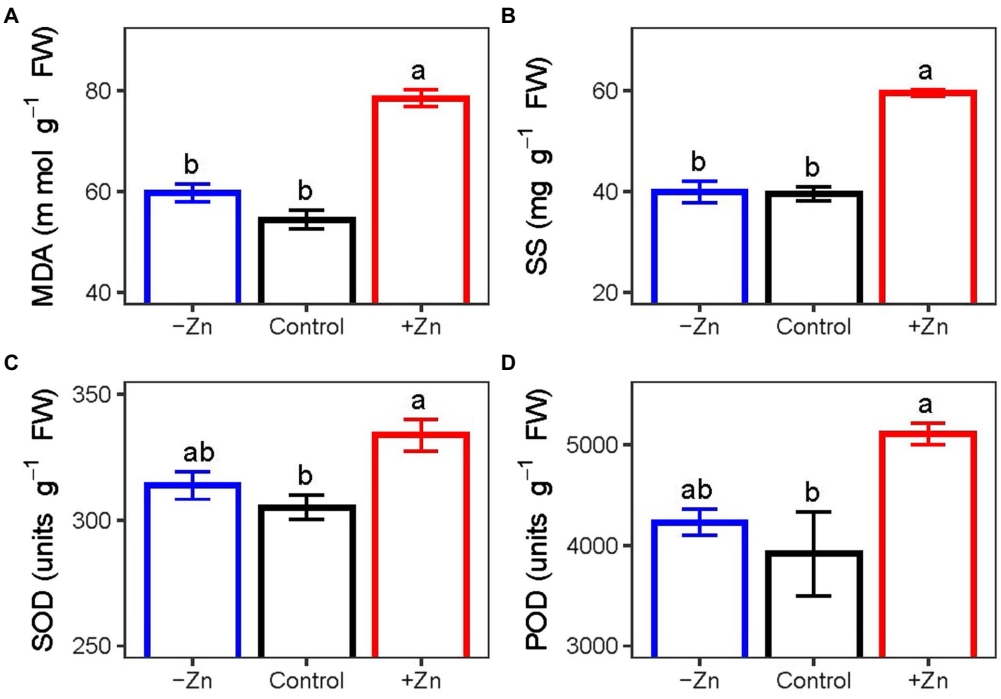


FIGURE 3
The concentrations of (**A**) MDA and (**B**) SS and activities of (**C**) SOD and (**D**) POD of *C. arborescens* subjected to Zn-deficient (0mML^{-1}), control (0.015mML^{-1}), and Zn-excess (0.15mML^{-1}) conditions. Different lowercase letters indicate significant differences among treatments (p<0.05).

Identification of a *Ceratoides arborescens*-specific Zn transport protein-encoding transcript

We analyzed the root transcriptome of *C. arborescens*, and discovered 39 distinct start codon-containing transcripts that may encode Zn-regulated transport proteins (Supplementary Table S1). We selected the full-length transcript DN154722 by assessing its sequence integrity and encoding protein structure prediction. We referred to it as *CaMTP*, while in Gene Ontology, it is referred to as "zinc II ion transport" (Figure 4A). With 80% residue identity and coverage, proteins from various species are regarded as homologs (Yu et al., 2004). Considering this, we could not identify any homolog of *CaMTP* in the NCBI protein database, where "Organism Optional" was *Dicotyledoneae*; instead, five proteins from five different plant species were identified with approximately 86% query coverage and 37% identity with that of *CaMTP*. Three of the five proteins had similar structures to that of *CaMTP* (Figure 4B).

CaMTP localizes in the ER

To determine the subcellular localization of *C. arborescens* proteins in tobacco plant cells, we used the constructs expressing *CaMTP*-GFP fusion proteins. As shown in Figure 5, the *CaMTP*-GFP fusion protein showed strong co-localization with the ER marker ER-mCherry under a confocal laser scanning microscope, indicating that *CaMTP* is located in the ER.

CaMTP complements a Zn^{2+} sequestration-deficient yeast mutant

Although AtZIP7 was utilized as a positive control, it did compliment the Zn^{2+}-dependent growth of the yeast mutant zrt1zrt2, and neither did *CaMTP* (Figure 6A). In the yeast mutant zrc1cot1, which is susceptible to high Zn^{2+} levels, we transiently expressed *CaMTP*. Similar to the positive control, MtMTP3, *CaMTP* expression halted the mutant's development when Zn^{2+} levels were excessive (Figure 6B).

Discussion

Zn is a mineral involved in several physiological and biochemical processes, and its deficiency or excess interferes with the optimal growth of plants (Tiong et al., 2015; Bhantana et al., 2021; Shinozaki and Yoshimoto, 2021; Al Jabri et al., 2022). In our study, high Zn supply has negative effects on the plant, as demonstrated by the decline in the dry weight of leaves and roots as well as leaf chlorosis and abnormally brown roots. These results are consistent with previous findings. For example, growth was restricted when *Moso bamboo* was treated with elevated levels of Zn due to the changes in the root structure (Liu et al., 2014). Growth inhibition and chlorosis in tomato leaves were observed when exposed to excess Zn (Cherif et al., 2010). Compared with these results, *Juncus acutus* could survive and displayed no toxicity symptoms when subjected to Zn at concentrations as high as 100 mM L^{-1}, a concentration considerably greater than those applied in our experiments (Mateos-Naranjo et al., 2014). This

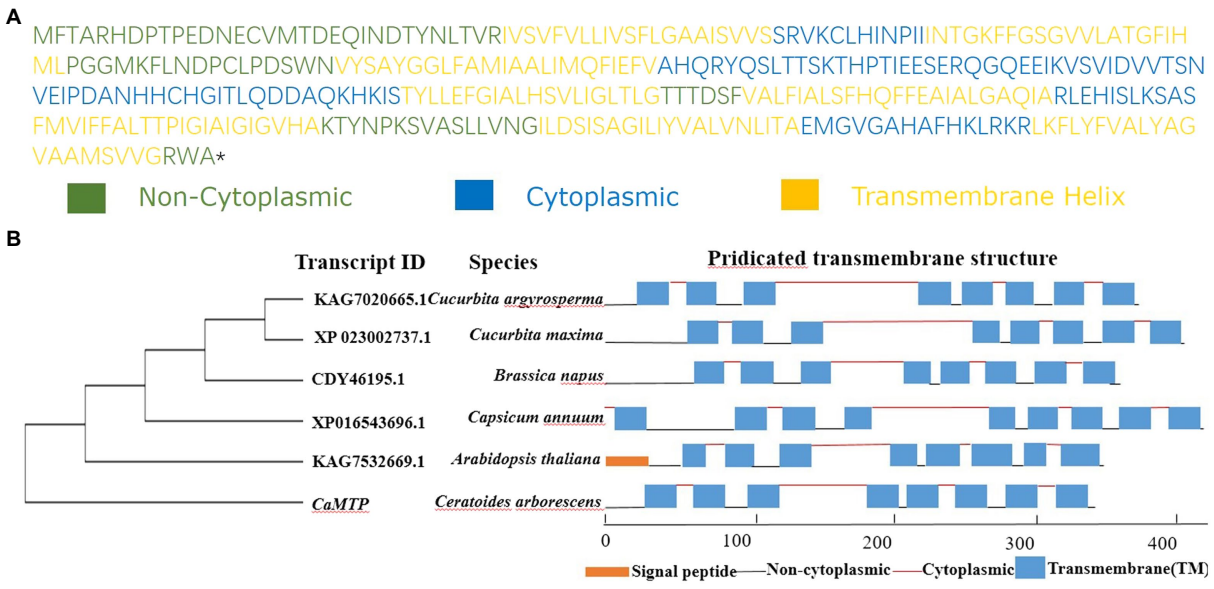


FIGURE 4
(**A**) Protein sequence of *CaMTP*. (**B**) Phylogenetic and structure comparison of *CaMTP* with other similar plant protein sequences. Different protein structure components are illustrated with different colors as indicated.

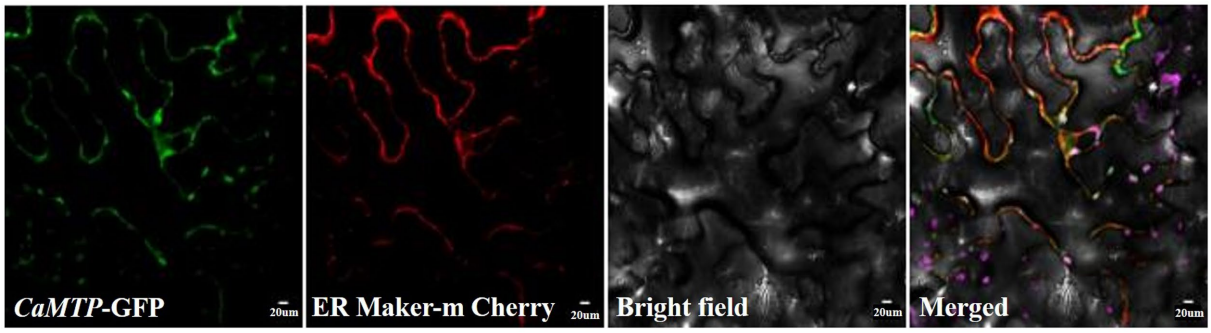


FIGURE 5
CaMTP-GFP and the ER marker ER-mCherry were transiently expressed in tobacco epidermal tissue and observed using conventional confocal microscopy.

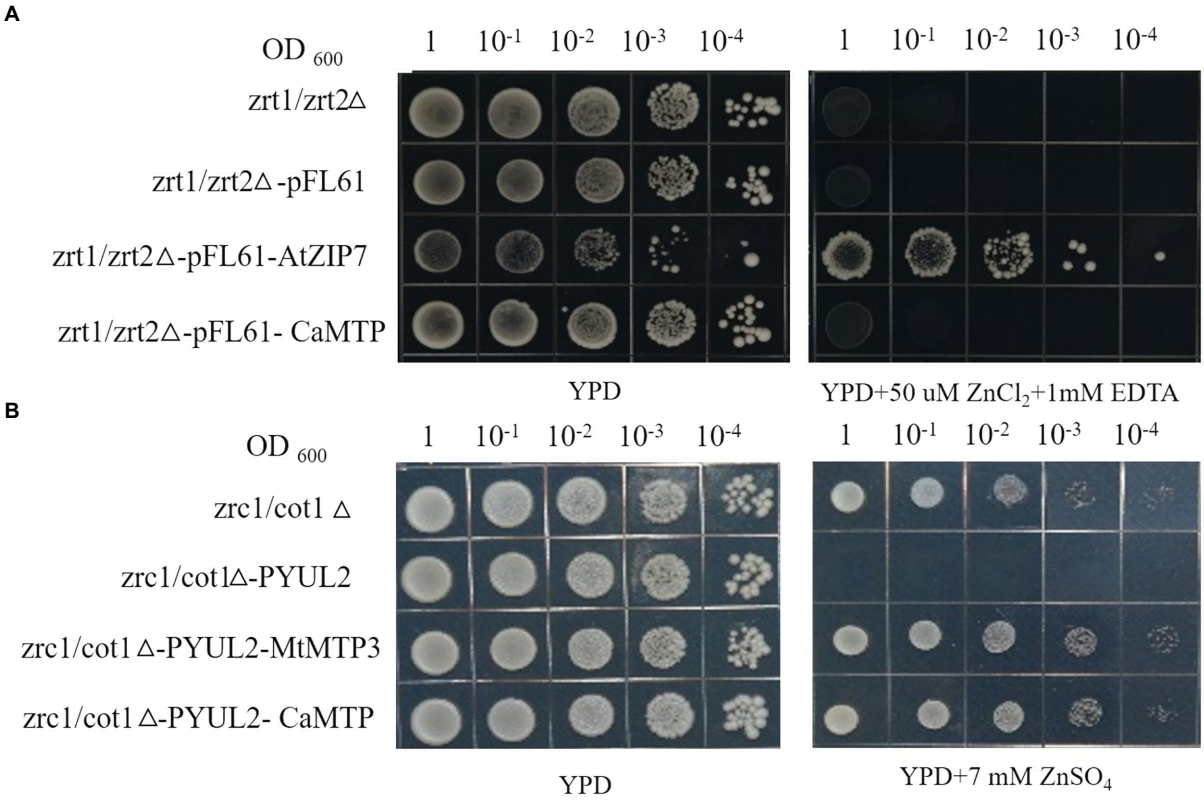


FIGURE 6
Complementation of yeast Zn^{2+} transport mutants with *CaMTP*. (**A**) The yeast cultures were spotted in full-nutrient YPD medium or Zn^{2+}-deficient medium with EDTA. (**B**) The yeast cultures were spotted in full-nutrient YPD medium or Zn^{2+} excess medium with 7mM $ZnSO_4$.

discrepancy in Zn tolerance was associated with the capacity to accumulate Zn in roots (at concentrations up to 2,500 mg kg^{-1}) and largely avoiding its transport to tillers. Although the Zn concentration of *C. arborescens* increased 18 times in the leaves and 15 times in the stems, while only 4.6 times in the roots, but the concentration of zinc in roots is nearly twice that of aboveground parts. Furthermore, the root was the most responsive to Zn in terms of dry weight reduction (79.8%), when compared to that of the leaf (67.9%) and stem (15.6%). Roots are seriously affected by heavy metals may be due to that Zn exposure could induce significant modifications in root architecture either by decreasing root-specific superficial area (Balafrej et al., 2020) or lignification of plant root tips (Glińska et al., 2016), thereby lowering the capacity of water and nutrient absorption, impacting plant growth and thus the plant dry weight negatively (Scheid et al., 2017).

We also found that excessive Zn inhibited photosynthesis by reducing Y(I) (Photosystem I; PSI) and Y(II) (Photosystem II; PSII) since Zn toxicity negatively impacts plant photosynthesis

and thus decreases plant biomass (Sagardoy et al., 2010; Kaur and Garg, 2021). Similar results were reported in the *Datura* species grown at higher Zn concentrations (Vaillant et al., 2005). PSII efficiency declined with the increase in Zn concentration (Dhir et al., 2011). Under metal stress, an imbalance between light absorption and light energy usage results in a surplus of light energy, causing an increase in NPQ and a decrease in the PSII efficiency (Sárvári, 2005). The negative impact of high Zn concentration on the integrity of the plasma membrane and photosynthetic electron transport impeded photosynthesis and reduced membrane permeability (Vijayarengan and Mahalakshmi, 2013). Furthermore, chlorophyll is important for plant photosynthesis (Croft et al., 2017). The reduction of chlorophyll content possibly declined the number of PSI units along with activity (Kudoh and Sonoike, 2002). Zn interferes with chlorophyll biosynthesis either by substituting for the Fe/Mg ions and inhibiting key biosynthetic enzymes, mainly δ-aminolevulinic acid dehydratase and protochlorophyllide reductase, or through the oxidative damage of chloroplasts (Marichali et al., 2016). Excess Zn reduced chlorophyll concentration and caused leaf chlorosis in our study, which was in line with the findings based on the mustard plant (Cherif et al., 2010; Khan and Khan, 2014). Thus, the reduction in chlorophyll concentration in leaves could be linked to a decrease in Fe content (Hu et al., 2017). However, the chlorophyll synthesis in tobacco leaves was not significantly inhibited under Zn stress, as the expression of other important proteins did not significantly change (Zhang et al., 2020).

Moreover, one of the consequences of heavy metals exposure in plants is lower absorption of other mineral elements(Kabata-Pendias, 2000; Kaur and Garg, 2021; Zaheer et al., 2022). In this respect, our assessment of mineral nutrients showed that high Zn caused a nutritional imbalance in *C. arborescens* plants, and Ca and Fe concentrations in roots and leaves were lower than those in the control plants. This fact indicates that high Zn supply impacts the transport of other elements and the uptake of *C. arborescens*, because the transporter of divalent metal cations, such as Zn Fe Mn, generally exhibits broad substrate specificity (Vigani et al., 2018; Kabir et al., 2021). Previous studies have also suggested that Zn transport proteins can transport Fe or Cu ions (Sinclair and Krämer, 2012; Sinclair et al., 2018). The interactions between Zn-Ca and Zn-Fe have also been previously described (Kabata-Pendias, 2000). The reduction of Ca concentration might be involved in the regulation of carbohydrate metabolism under Zn stress (Zhang et al., 2017). The reduction of Fe concentration might be attributed to competing sites being the direct cause of enthalpy among metal ions (Li et al., 2020).

Metal toxicity exacerbates ROS production, which damages lipids, proteins, and nucleic acids and disrupts metabolism (Miller et al., 2008; Zaheer et al., 2020b). However, plants evolved sophisticated antioxidant defense mechanisms to counteract and undo the harm caused by ROS, including activities of various antioxidant enzymes such as SOD and POD (Rout and Das, 2003; Kumar Tewari et al., 2008; Javed et al., 2020). In this study, higher SOD and POD activities were observed in the seedlings of *C. arborescens* under excess Zn. Our findings are concordant with the results of *Spinacia oleracea, Handroanthus impetiginosus* and *Tabebuia roseoalba* (Gai et al., 2017; Zaheer et al., 2020a). Excessive Zn levels can promote MDA production in plants based on enhanced lipid peroxidation *via* excessive generation of free radicals (Cherif et al., 2010). Our results revealed that MDA concentrations in shoots under Zn-excess conditions were 44.3% higher compared to the control, a distinction from the results of the mulberry plant under Zn-deficient conditions where the increase observed in Zn-excess plants was not significant (Kumar Tewari et al., 2008). Additionally, to regulate their response to heavy metal stress, plants could increase the synthesis and storage of osmolytes like soluble proteins, soluble carbohydrates, and free proline (Sharma and Dietz, 2006). Herein, we discovered that roots' SS levels were significantly elevated in response to Zn stress, suggesting a defense mechanism against Zn toxicity. SS functions as signal molecules in sugar sensing and signaling systems in addition to their regular osmoprotective roles of stabilizing cell membranes and preserving turgor (Wu et al., 2011).

Zn deficiency causes leaf bronzing, growth stunting, delayed maturity, and plant yield loss (Broadley et al., 2007). Conversely, Zn deficiency did not affect the growth of *C. arborescens*. These results were corroborated by the plant biomass amount and the unchanged concentration of chlorophyll under Zn deficiency. In *Zea mays* L. cv. PAC (Zn tolerant line), Zn deficiency did not significantly affect shoot height or weight, indicating that PAC can resist Zn deficiency, which was corroborated by the reduction of Zn concentration in both root and shoot in DAC (Zn sensitive line), but the concentration in PAC was maintained under Zn deficiency (Khatun et al., 2018). The decrease in root and shoot Zn concentration in *Ratan* (Kabir et al., 2021) suggests that Zn homeostasis cannot be maintained under Zn deficiency by these genotypes. In this study, we also found that the Zn concentration in both shoots and roots did not change. In contrast, Zn deficiency decreased the contents of Ca in leaves. A previous study also indicated that autophagy resupplies can alleviate the Zn deficiency state and promote optimal growth when environmental Zn is unavailable (Shinozaki and Yoshimoto, 2021). Furthermore, SOD and POD are involved in plant stress tolerance (Kumar Tewari et al., 2008). Herein, we found that the activities of POD and SOD did not change under Zn deficiency, suggesting that *C. arborescens* was not stressed under Zn deficiency. In contrast, increased activities of SOD and POD have been reported in response to Zn deficiency in tomatoes (Kabir et al., 2021), while decreased SOD activities were recorded in *Coffea arabica* (dos Santos et al., 2019). This also reflects that different species respond differently to Zn deficiency. We reasoned that the high Zn content in the seeds (Supplementary Table S2) could compensate for seedling growth in *C. arborescens*; hence it grows well in Zn deficiency.

Identifying novel proteins to increase crop stress tolerance in wild plants, particularly those acclimated to severe settings, is promising (Yang et al., 2018). The rapid advancement of high-throughput sequencing technologies (Bräutigam and Gowik, 2010) and the identification of yeast mutants (Zhao and Eide, 1996; Lin et al., 2008) have tremendously aided in the discovery of new genes and proteins. Our study reports the discovery of a novel protein in

C. arborescens, which showed strong tolerance against Zn stresses. Transcript profiling and comparison of deduced protein sequences identified a transmembrane transcript, *CaMTP*, whose significant homolog was not found in other plants. Contrary to the findings for all previously described members of the ZIP family, *CaMTP* has no passive transport activity for Zn^{2+} (Milner et al., 2013). We transiently expressed *CaMTP* in yeast mutant zrc1cot1Δ, which is sensitive to excess Zn^{2+} (Lin et al., 2008). The expression of *CaMTP* rescued the growth of the mutant under high Zn^{2+} conditions, just like the positive control, *MtMTP3* (Zhao et al., 2018). Members of this protein family typically transport metal out of the cytosol (Krämer et al., 2007; Zhao et al., 2018). We observed that *CaMTP* is located in the ER, similar to MTP2, which may mediate the uptake of cytoplasmic Zn^{2+} into ER (Sinclair et al., 2018). The Arabidopsis CDF family proteins Metal Tolerance Protein 12 (*AtMTP12*) and 5 (AtMTP5) are co-located on the Golgi membrane, and interact with each other (Fujiwara et al., 2015). The cucumber protein *CsMTP5* is also located on the Golgi apparatus (Migocka et al., 2018). This fact indicates that *CaMTP* has characteristics of the CDF family Zn transporter, while the transport function of other metals is not clear.

In conclusion, our study examined the tolerance of Zn and identified a specific Zn transport protein-encoding transcript in *C. arborescens*. The results showed that excess Zn drastically decreased the biomass and induced chlorosis and growth defects in plants. Excess Zn increased the concentration of Zn but decreased that of Fe and Ca in the roots and shoots, along with the content of *C. arborescens*, thereby decreasing photosynthesis activity. The increased MDA and SS levels and POD and SOD activities may be part of a mechanism to resist high-level Zn stress. However, the shortage of Zn supply did not affect the growth of *C. arborescens*, suggesting that this forage could be established in the Zn-deficient area. Further, we identified a novel endoplasmic-localized protein named *CaMTP*, which complemented the increased sensitivity of a yeast mutant to excessive Zn. The Zn-deficiency tolerance mechanisms in plants are not fully known, to further understand Zn homeostasis in plants, the mechanisms related to Zn deficiency in *C. arborescens* should be investigated. Moreover, to further research *CaMTP*, the genetic transformation system of *C. arborescens* should be built.

Data availability statement

The data presented in the study are deposited in the NCBI GenBank (https://www.ncbi.nlm.nih.gov/). The associated BioProject, Bio-Sample and SRA numbers were PRJNA870161, SAMN30353563-SAMN30353571, and SRR21095981-SRR21095989 respectively.

Author contributions

FM designed the study. XL conducted the experiments, analyzed the data, and wrote the manuscript. LZ guided the experiment. HR and XW revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by international cooperation between China and the European Union Center (2017YFE0111000).

Acknowledgments

We would like to thank Zhi Qi for providing the experimental site and equipment, Wanjie Chen for statistical advice, and Editage (www.editage.com) for English language editing.

Conflict of interest

LZ is employed by M-Grass Ecology and Environment (Group) Co., Ltd.

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

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.976311/full#supplementary-material

References

Ahmad, S., Mfarrej, M. F. B., El-Esawi, M. A., Waseem, M., Alatawi, A., Nafees, M., et al. (2022). Chromium-resistant *Staphylococcus aureus* alleviates chromium toxicity by developing synergistic relationships with zinc oxide nanoparticles in wheat. *Ecotoxicol. Environ. Saf.* 230:113142. doi: 10.1016/j.ecoenv.2021.113142

Al Jabri, H., Saleem, M. H., Rizwan, M., Hussain, I., Usman, K., and Alsafran, M. (2022). Zinc oxide nanoparticles and their biosynthesis: overview. *Life* 12, 594. doi: 10.3390/life12040594

Amini, S., Arsova, B., and Hanikenne, M. (2022). The molecular basis of zinc homeostasis in cereals. *Plant Cell Environ.* 45, 1339–1361. doi: 10.1111/pce.14257

Balafrej, H., Bogusz, D., Triqui, Z.-E. A., Guedira, A., Bendaou, N., Smouni, A., et al. (2020). Zinc hyperaccumulation in plants: a review. *Plan. Theory* 9, 562. doi: 10.3390/plants9050562

Bhantana, P., Rana, M. S., Sun, X., Moussa, M. G., Saleem, M. H., Syaifudin, M., et al. (2021). Arbuscular mycorrhizal fungi and its major role in plant growth, zinc

nutrition, phosphorous regulation and phytoremediation. *Symbiosis* 84, 19–37. doi: 10.1007/s13199-021-00756-6

Bräutigam, A., and Gowik, U. (2010). What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. *Plant Biol.* 12, 831–841. doi: 10.1111/j.1438-8677.2010.00373.x

Broadley, M. R., White, P. J., Hammond, J. P., Zelko, I., and Lux, A. (2007). Zinc in plants. *New Phytol.* 173, 677–702. doi: 10.1111/j.1469-8137.2007.01996.x

Cherif, J., Derbel, N., and Nakkach, M., Bergmann, H. von, Jemal, F., and Lakhdar, Z. B. (2010). Analysis of in vivo chlorophyll fluorescence spectra to monitor physiological state of tomato plants growing under zinc stress. *J. Photochem. Photobiol. B* 101, 332–339. doi: 10.1016/j.jphotobiol.2010.08.005

Croft, H., Chen, J. M., Luo, X., Bartlett, P., Chen, B., and Staebler, R. M. (2017). Leaf chlorophyll content as a proxy for leaf photosynthetic capacity. *Global Change Biol.* 23, 3513–3524. doi: 10.1111/gcb.13599

Dhir, B., Sharmila, P., Pardha Saradhi, P., Sharma, S., Kumar, R., and Mehta, D. (2011). Heavy metal induced physiological alterations in *Salvinia natans*. *Ecotoxicol. Environ. Saf.* 74, 1678–1684. doi: 10.1016/j.ecoenv.2011.05.009

dos Santos, J. O., Andrade, C. A., Dázio de Souza, K. R., Santos, M. d. O., Brandão, I. R., Alves, J. D., et al. (2019). Impact of zinc stress on biochemical and biophysical parameters in *Coffea arabica* seedlings. *J. Crop. Sci. Biotechnol.* 22, 253–264. doi: 10.1007/s12892-019-0097-0

Fujiwara, T., Kawachi, M., Sato, Y., Mori, H., Kutsuna, N., Hasezawa, S., et al. (2015). A high molecular mass zinc transporter MTP 12 forms a functional heteromeric complex with MTP 5 in the Golgi in *Arabidopsis thaliana*. *FEBS J.* 282, 1965–1979. doi: 10.1111/febs.13252

Gai, A. P. C., dos Santos, D. S., and Vieira, E. A. (2017). Effects of zinc excess on antioxidant metabolism, mineral content and initial growth of Handroanthus impetiginosus (Mart. ex DC.) Mattos and Tabebuia roseoalba (Ridl.) Sandwith. *Environ. Exp. Bot.* 144, 88–99. doi: 10.1016/j.envexpbot.2017.09.006

Giannopolitis, C. N., and Ries, S. K. (1977). Superoxide dismutases: I. Occurrence in higher plants. *Plant Physiol.* 59, 309–314. doi: 10.1104/pp.59.2.309

Glińska, S., Gapińska, M., Michlewska, S., Skiba, E., and Kubicki, J. (2016). Analysis of *Triticum aestivum* seedling response to the excess of zinc. *Protoplasma* 253, 367–377. doi: 10.1007/s00709-015-0816-3

Hodges, D. M., DeLong, J. M., Forney, C. F., and Prange, R. K. (1999). Improving the thiobarbituric acid- reactive-substances assay for estimating lipid peroxidation in plant tissues containing anthocyanin and other interfering compounds. *Planta* 207, 604–611. doi: 10.1007/s004250050524

Hu, X. Y., Page, M. T., Sumida, A., Tanaka, A., Terry, M. J., and Tanaka, R. (2017). The iron–sulfur cluster biosynthesis protein SUFB is required for chlorophyll synthesis, but not phytochrome signaling. *Plant J.* 89, 1184–1194. doi: 10.1111/tpj.13455

Javed, M. T., Saleem, M. H., Aslam, S., Rehman, M., Iqbal, N., Begum, R., et al. (2020). Elucidating silicon-mediated distinct morpho-physio-biochemical attributes and organic acid exudation patterns of cadmium stressed Ajwain (*Trachyspermum ammi* L.). *Plant Physiol. Biochem.* 157, 23–37. doi: 10.1016/j.plaphy.2020.10.010

Kabata-Pendias, A. (2000). *Trace Elements in Soils and Plants*. Boca Raton, FL: CRC Press, doi: 10.1201/9781420039900

Kabir, A. H., Akther, M. S., Skalicky, M., Das, U., Gohari, G., Brestic, M., et al. (2021). Downregulation of Zn-transporters along with Fe and redox imbalance causes growth and photosynthetic disturbance in Zn-deficient tomato. *Sci. Rep.* 11, 6040–6012. doi: 10.1038/s41598-021-85649-w

Kaur, H., and Garg, N. (2021). Zinc toxicity in plants: a review. *Planta* 253, 129. doi: 10.1007/s00425-021-03642-z

Khan, M. I. R., and Khan, N. A. (2014). Ethylene reverses photosynthetic inhibition by nickel and zinc in mustard through changes in PS II activity, photosynthetic nitrogen use efficiency, and antioxidant metabolism. *Protoplasma* 251, 1007–1019. doi: 10.1007/s00709-014-0610-7

Khatun, M. A., Hossain, M. M., Bari, M. A., Abdullahil, K. M., Parvez, M. S., Alam, M. F., et al. (2018). Zinc deficiency tolerance in maize is associated with the up-regulation of Zn transporter genes and antioxidant activities. *Plant Biol.* 20, 765–770. doi: 10.1111/plb.12837

Krämer, U., Talke, I. N., and Hanikenne, M. (2007). Transition metal transport. *FEBS Lett.* 581, 2263–2272. doi: 10.1016/j.febslet.2007.04.010

Kudoh, H., and Sonoike, K. (2002). Irreversible damage to photosystem I by chilling in the light: cause of the degradation of chlorophyll after returning to normal growth temperature. *Planta* 215, 541–548. doi: 10.1007/s00425-002-0790-9

Kumar Tewari, R., Kumar, P., and Nand Sharma, P. (2008). Morphology and physiology of zinc-stressed mulberry plants. *J. Plant Nutr. Soil Sci.* 171, 286–294. doi: 10.1002/jpln.200700222

Li, G., Li, C., Rengel, Z., Liu, H., and Zhao, P. (2020). Title: excess Zn-induced changes in physiological parameters and expression levels of TaZips in two wheat genotypes. *Environ. Exp. Bot.* 177:104133. doi: 10.1016/j.envexpbot.2020.104133

Lin, H. L., Kumánovics, A., Nelson, J. M., Warner, D. E., Ward, D. M., and Kaplan, J. (2008). A single amino acid change in the yeast vacuolar metal transporters ZRC1 and COT1 alters their substrate specificity. *J. Biol. Chem.* 283, 33865–33873. doi: 10.1074/jbc.M804377200

Liu, D., Chen, J., Mahmood, Q., Li, S., Wu, J., Ye, Z., et al. (2014). Effect of Zn toxicity on root morphology, ultrastructure, and the ability to accumulate Zn in Moso bamboo (*Phyllostachys pubescens*). *Environ. Sci. Pollut. Res.* 21, 13615–13624. doi: 10.1007/s11356-014-3271-3

Liu, H. Y., and Yi, J. (2004). The evaluation on different types *Ceratoides Arborescens* in reproductive characteristics and productive properties of seed. *Seed* 23, 3–6.

Lu, M., Chai, J., and Fu, D. (2009). Structural basis for autoregulation of the zinc transporter YiiP. *Nat. Struct. Mol. Biol.* 16, 1063–1067. doi: 10.1038/nsmb.1662

Marichali, A., Dallali, S., Ouerghemmi, S., Sebei, H., Casabianca, H., and Hosni, K. (2016). Responses of *Nigella sativa* L. to zinc excess: focus on germination, growth, yield and yield components, lipid and terpene metabolism, and total phenolics and antioxidant activities. *J. Agric. Food Chem.* 64, 1664–1675. doi: 10.1021/acs.jafc.6b00274

Marreiro, D., Cruz, K., Morais, J., Beserra, J., Severo, J., and de Oliveira, A. (2017). Zinc and oxidative stress: current mechanisms. *Antioxidants* 6, 24. doi: 10.3390/antiox6020024

Mateos-Naranjo, E., Castellanos, E. M., and Perez-Martin, A. (2014). Zinc tolerance and accumulation in the halophytic species *Juncus acutus*. *Environ. Exp. Bot.* 100, 114–121. doi: 10.1016/j.envexpbot.2013.12.023

McFarland, M. L., Ueckert, D. N., Hartmann, S., and Hons, F. M. (1990). Transplanting shrubs for revegetation of salt-affected soils. *Landsc. Urban Plan.* 19, 377–381. doi: 10.1016/0169-2046(90)90043-2

Meyer, S. T., Castro, S. R., Fernandes, M. M., Soares, A. C., de Souza Freitas, G. A., and Ribeiro, E. (2016). Heavy-metal-contaminated industrial soil: uptake assessment in native plant species from Brazilian Cerrado. *Int. J. Phytoremediation* 18, 832–838. doi: 10.1080/15226514.2016.1146224

Migocka, M., Małas, K., Maciaszczyk-Dziubinska, E., Posyniak, E., Migdal, I., and Szczech, P. (2018). Cucumber Golgi protein CsMTP5 forms a Zn-transporting heterodimer with high molecular mass protein CsMTP12. *Plant Sci.* 277, 196–206. doi: 10.1016/j.plantsci.2018.09.011

Miller, G., Shulaev, V., and Mittler, R. (2008). Reactive oxygen signaling and abiotic stress. *Physiol. Plant.* 133, 481–489. doi: 10.1111/j.1399-3054.2008.01090.x

Milner, M. J., Seamon, J., Craft, E., and Kochian, L. V. (2013). Transport properties of members of the ZIP family in plants and their role in Zn and Mn homeostasis. *J. Exp. Bot.* 64, 369–381. doi: 10.1093/jxb/ers315

Minet, M., Dufour, M.-E., and Lacroute, F. (1992). Complementation of Saccharomyces cerevisiae auxotrophic mutants by Arabidopsis thaliana cDNAs. *Plant J.* 2, 417–422.

Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., et al. (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J. Biosci. Bioeng.* 104, 34–41. doi: 10.1263/jbb.104.34

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Ricachenevsky, F. K., Menguer, P. K., Sperotto, R. A., Williams, L. E., and Fett, J. P. (2013). Roles of plant metal tolerance proteins (MTP) in metal storage and potential use in biofortification strategies. *Front. Plant Sci.* 4:1144. doi: 10.3389/fpls.2013.00144

Rout, G. R., and Das, P. (2003). Effect of metal toxicity on plant growth and metabolism: I. *Zinc Agron.* 23, 3–11. doi: 10.1051/agro:2002073

Sagardoy, R., Vázquez, S., Florez-Sarasa, I. D., Albacete, A., Ribas-Carbó, M., Flexas, J., et al. (2010). Stomatal and mesophyll conductances to CO_2 are the main limitations to photosynthesis in sugar beet (*Beta vulgaris*) plants grown with excess zinc. *New Phytol.* 187, 145–158. doi: 10.1111/j.1469-8137.2010.03241.x

Santi, S., and Schmidt, W. (2009). Dissecting iron deficiency-induced proton extrusion in Arabidopsis roots. *New Phytol.* 183, 1072–1084. doi: 10.1111/j.1469-8137.2009.02908.x

Sárvári, É. (2005). Effects of heavy metals on chlorophyll-protein complexes in higher plants: causes and consequences. *Handb. Photosynth.* 1262, 865–888. doi: 10.1201/9781420027877.ch45

Scharwies, J. D., and Dinneny, J. R. (2019). Water transport, perception, and response in plants. *J. Plant Res.* 132, 311–324. doi: 10.1007/s10265-019-01089-8

Scheid, D. L., Marco, R. D., Grolli, A. L., Silva, R. F. d., Da Ros, C. O., and Andreazza, R. (2017). Growth, tolerance and zinc accumulation in Senna multijuga and *Erythrina crista-galli* seedlings. *Rev. Bras. Eng. Agríc. E Ambient.* 21, 465–470. doi: 10.1590/1807-1929/agriambi.v21n7p465-470

Sharma, S. S., and Dietz, K.-J. (2006). The significance of amino acids and amino acid-derived molecules in plant responses and adaptation to heavy metal stress. *J. Exp. Bot.* 57, 711–726.

Shinozaki, D., and Yoshimoto, K. (2021). Autophagy balances the zinc–iron seesaw caused by Zn-stress. *Trends Plant Sci.* 26, 882–884. doi: 10.1016/j.tplants.2021.06.014

Sidhu, G. P. S., Bali, A. S., Singh, H. P., Batish, D. R., and Kohli, R. K. (2020). Insights into the tolerance and phytoremediation potential of *Coronopus didymus* L. (Sm) grown under zinc stress. *Chemosphere* 244:125350. doi: 10.1016/j.chemosphere.2019.125350

Sinclair, S. A., and Krämer, U. (2012). The zinc homeostasis network of land plants. *Biochim. Biophys. Acta BBA Mol. Cell Res.* 1823, 1553–1567. doi: 10.1016/j.bbamcr.2012.05.016

Sinclair, S. A., Senger, T., Talke, I. N., Cobbett, C. S., Haydon, M. J., and Krämer, U. (2018). Systemic upregulation of MTP2-and HMA2-mediated Zn partitioning to the shoot supplements local Zn deficiency responses. *Plant Cell* 30, 2463–2479. doi: 10.1105/tpc.18.00207

Tiong, J., McDonald, G., Genc, Y., Shirley, N., Langridge, P., and Huang, C. Y. (2015). Increased expression of six ZIP family genes by zinc (Zn) deficiency is associated with enhanced uptake and root-to-shoot translocation of Zn in barley (*Hordeum vulgare*). *New Phytol.* 207, 1097–1109. doi: 10.1111/nph.13413

Tsednee, M., Yang, S.-C., Lee, D.-C., and Yeh, K.-C. (2014). Root-secreted nicotianamine from Arabidopsis halleri facilitates zinc hypertolerance by regulating zinc bioavailability. *Plant Physiol.* 166, 839–852. doi: 10.1104/pp.114.241224

Vaillant, N., Monnet, F., Hitmi, A., Sallanon, H., and Coudret, A. (2005). Comparative study of responses in four Datura species to a zinc stress. *Chemosphere* 59, 1005–1013. doi: 10.1016/j.chemosphere.2004.11.030

van Dijk, H. F., and Roelofs, J. G. (1988). Effects of excessive ammonium deposition on the nutritional status and condition of pine needles. *Physiol. Plant.* 73, 494–501. doi: 10.1111/j.1399-3054.1988.tb05431.x

Vigani, G., Bohic, S., Faoro, F., Vekemans, B., Vincze, L., and Terzano, R. (2018). Cellular fractionation and nanoscopic X-ray fluorescence imaging analyses reveal changes of zinc distribution in leaf cells of Iron-deficient plants. *Front. Plant Sci.* 9:1112. doi: 10.3389/fpls.2018.01112

Vijayarengan, P., and Mahalakshmi, G. (2013). Zinc toxicity in tomato plants. *World Appl. Sci. J.* 24, 649–653.

Wang, Y., Kang, Y., Ma, C., Miao, R., Wu, C., Long, Y., et al. (2017). CNGC2 is a Ca^{2+} influx channel that prevents accumulation of apoplastic Ca^{2+} in the leaf. *Plant Physiol.* 173, 1342–1354. doi: 10.1104/pp.16.01222

Williams, M. W., and Martin, G. C. (1967). Quantitative gas chromatography of sugars and related polyhydroxy compounds in horticultural crops. *HortScience* 2, 68–69.

Wu, X. X., Zhu, W. M., Zhang, H., Ding, H. D., and Zhang, H. J. (2011). Exogenous nitric oxide protects against salt-induced oxidative stress in the leaves from two genotypes of tomato (Lycopersicom esculentum Mill.). *Acta Physiol. Plant.* 33, 1199–1209. doi: 10.1007/s11738-010-0648-x

Yang, J., Wu, C. L., Yu, Y. H., Mao, H. P., Bao, Y. Y., Kang, Y., et al. (2018). A mongolian pine specific endoplasmic reticulum localized CALMODULIN-LIKE calcium binding protein enhances arabidopsis growth. *J. Plant Physiol.* 226, 1–11. doi: 10.1016/j.jplph.2018.04.006

Yin, S., Duan, M., Fang, B., Zhao, G., Leng, X., and Zhang, T. (2022). Zinc homeostasis and regulation: zinc transmembrane transport through transporters. *Crit. Rev. Food Sci. Nutr.* 1–11. doi: 10.1080/10408398.2022.2048292

Yu, H. Y., Luscombe, N. M., Lu, H. X., Zhu, X. W., Xia, Y., Han, J.-D. J., et al. (2004). Annotation transfer between genomes: protein–protein interologs and protein–DNA regulogs. *Genome Res.* 14, 1107–1118. doi: 10.1101/gr.1774904

Zaheer, I. E., Ali, S., Saleem, M. H., Ali, M., Riaz, M., Javed, S., et al. (2020a). Interactive role of zinc and iron lysine on *Spinacia oleracea* L. growth, photosynthesis and antioxidant capacity irrigated with tannery wastewater. *Physiol. Mol. Biol. Plants* 26, 2435–2452. doi: 10.1007/s12298-020-00912-0

Zaheer, I. E., Ali, S., Saleem, M. H., Arslan Ashraf, M., Ali, Q., Abbas, Z., et al. (2020b). Zinc-lysine supplementation mitigates oxidative stress in rapeseed (*Brassica napus* L.) by preventing phytotoxicity of chromium, when irrigated with tannery wastewater. *Plan. Theory* 9:1145. doi: 10.3390/plants9091145

Zaheer, I. E., Ali, S., Saleem, M. H., Yousaf, H. S., Malik, A., Abbas, Z., et al. (2022). Combined application of zinc and iron-lysine and its effects on morpho-physiological traits, antioxidant capacity and chromium uptake in rapeseed (*Brassica napus* L.). *PLoS One* 17:e0262140. doi: 10.1371/journal.pone.0262140

Zhang, H. H., Xu, Z. S., Guo, K. W., Huo, Y. Z., He, G. Q., Sun, H., et al. (2020). Toxic effects of heavy metal cd and Zn on chlorophyll, carotenoid metabolism and photosynthetic function in tobacco leaves revealed by physiological and proteomics analysis. *Ecotoxicol. Environ. Saf.* 202:110856. doi: 10.1016/j.ecoenv.2020.110856

Zhang, Y., Wang, Y., Ding, Z., Wang, H., Song, L., Jia, S., et al. (2017). Zinc stress affects ionome and metabolome in tea plants. *Plant Physiol. Biochem.* 111, 318–328. doi: 10.1016/j.plaphy.2016.12.014

Zhao, H., and Eide, D. (1996). The ZRT2 gene encodes the low affinity zinc transporter in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 271, 23203–23210. doi: 10.1074/jbc.271.38.23203

Zhao, S. B., Li, Y. G., Zhao, B., and Lin, H. (2018). Identification and expression analyses of Mt MTP3, a zinc transporter of CDF family in Mdicago truncatula. *J. Huazhong Agric. Univ.* 37, 52–60. doi: 10.13300/j.cnki.hnlkxb.2018.03.008

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 October 2022
DOI 10.3389/fpls.2022.1027551

OPEN ACCESS

EDITED BY
Jiyu Zhang,
Lanzhou University, China

REVIEWED BY
Wen-ting Peng,
Fujian Agriculture and Forestry University, China
Lu Qin,
Oil Crops Research Institute (CAAS), China
Lingyan Jiang,
Hainan University, China

*CORRESPONDENCE
Zhijian Chen
jchen@scau.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

RECEIVED 25 August 2022
ACCEPTED 20 September 2022
PUBLISHED 06 October 2022

CITATION
Zou X, Huang R, Wang L, Wang G, Miao Y, Rao I, Liu G and Chen Z (2022) SgNramp1, a plasma membrane-localized transporter, involves in manganese uptake in *Stylosanthes guianensis*. *Front. Plant Sci.* 13:1027551. doi: 10.3389/fpls.2022.1027551

SgNramp1, a plasma membrane-localized transporter, involves in manganese uptake in *Stylosanthes guianensis*

Xiaoyan Zou[1,2†], Rui Huang[1†], Linjie Wang[1,2], Guihua Wang[3], Ye Miao[1,2], Idupulapati Rao[4], Guodao Liu[1] and Zhijian Chen[1,2*]

[1]Key Laboratory of Tropical Crops Germplasm Resources Genetic Improvement and Innovation of Hainan Province, Institute of Tropical Crop Genetic Resources, Chinese Academy of Tropical Agricultural Sciences, Haikou, China, [2]College of Tropical Crops, Hainan University, Haikou, China, [3]Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China, [4]Crops for Nutrition and Health, Alliance of Bioversity International and International Center for Tropical Agriculture, Cali, Colombia

Transporters belonging to the natural resistance-associated macrophage protein (Nramp) family play important roles in metal uptake and homeostasis. Although Nramp members have been functionally characterized in plants, the role of Nramp in the important tropical forage legume *Stylosanthes guianensis* (stylo) is largely unknown. This study aimed to determine the responses of *Nramp* genes to metal stresses and investigate its metal transport activity in stylo. Five *SgNramp* genes were identified from stylo. Expression analysis showed that *SgNramp* genes exhibited tissue preferential expressions and diverse responses to metal stresses, especially for manganese (Mn), suggesting the involvement of *SgNramps* in the response of stylo to metal stresses. Of the five *SgNramps*, *SgNramp1* displayed the highest expression in stylo roots. A close correlation between *SgNramp1* expression and root Mn concentration was observed among nine stylo cultivars under Mn limited condition. The higher expression of *SgNramp1* was correlated with a high Mn uptake in stylo. Subsequent subcellular localization analysis showed that SgNramp1 was localized to the plasma membrane. Furthermore, heterologous expression of *SgNramp1* complemented the phenotype of the Mn uptake-defective yeast (*Saccharomyces cerevisiae*) mutant Δ*smf1*. Mn concentration in the yeast cells expressing *SgNramp1* was higher than that of the empty vector control, suggesting the transport activity of SgNramp1 for Mn in yeast. Taken together, this study reveals that SgNramp1 is a plasma membrane–localized transporter responsible for Mn uptake in stylo.

KEYWORDS

nramp, trace metal, metal homeostasis, gene expression, subcellular localization

Introduction

Trace metals, such as manganese (Mn), iron (Fe), zinc (Zn), and copper (Cu), are essential elements for plant growth and productivity. These metal ions involve in multiple physiological and biochemical processes in plants, such as stabilizing the structure of many biomacromolecules and serve as cofactors for key proteins and enzymes (Marschner, 2012). For example, Mn is an essential component in the Mn cluster of the oxygen-evolving complex in photosystem II (PSII) that is required for the electron transport chain in photosynthesis. Mn also acts as a critical cofactor of numerous enzymes, such as superoxide dismutase (SOD), catalase (CAT), decarboxylases, and RNA polymerases (Chen and Li, 2021). Despite its necessity at a low dose, the metal ions presented at excess levels can be harmful to plants, which are similar to the toxic effects of non-essential heavy metals, such as aluminum (Al), cadmium (Cd), arsenic (As), and lead (Pb). The excess amount of trace elements and heavy metals can cause phytotoxicity to plant cells, such as triggering oxidative stress, inhibiting enzyme activity, impeding the structure and function of photosynthetic apparatus, and damaging respiration and energy metabolism, ultimately inhibiting plant growth (Fryzova et al., 2018; Huang et al., 2020; Riyazuddin et al., 2021).

To maintain normal growth, plants have to regulate metal uptake and homeostasis in response to the fluctuated metal status in soils from deficient to toxic levels, which can be achieved *via* a variety of metal transporters, such as members of the natural resistance-associated macrophage protein (Nramp), Zn-regulated transporter/Fe-regulated transporter-like protein (ZRT/IRT), metal tolerance protein (MTP), and heavy metal ATPase (HMA) (Wang et al., 2020; Kaur et al., 2021; Pottier et al., 2022). Among these metal transporters, Nramp, an integral membrane protein, is a critical transporter for divalent metals. Most Nramp proteins include a unique motif and 10–12 conserved transmembrane domains (TMDs) (Li et al., 2021b). Nramp proteins have been characterized to transport a wide range of metal ions across the cellular membrane, including Mn, Fe, Zn, Cd, or Al (Li et al., 2019). To date, a set of Nramp members have been identified in various plants, such as Arabidopsis (Thomine et al., 2000; Li et al., 2022), rice (*Oryza sativa*) (Xia et al., 2010; Sasaki et al., 2012; Mani and Sankaranarayanan, 2018; Chang et al., 2020), *Medicago truncatula* (Tejada-Jiménez et al., 2015), soybean (*Glycine max*) (Kaiser et al., 2003; Qin et al., 2017), barley (*Hordeum vulgare*) (Wu et al., 2016) and tobacco (*Nicotiana tabacum*) (Liu et al., 2022).

For example, six Nramp members are identified in Arabidopsis, and five of them have been functionally characterized (Gao et al., 2018; Li et al., 2022). Among them, AtNramp1 is mainly expressed in roots and is a plasma membrane (PM)–localized transporter responsible for the high-affinity Mn uptake in Arabidopsis (Cailliatte et al., 2010). In addition to Mn, AtNramp1 can also transport Fe and cobalt (Co) *in vivo* (Curie et al., 2000; Cailliatte et al., 2010). AtNramp2 is found to be localized to the trans-Golgi network and is implicated in remobilization of Mn in Golgi, which is required for Arabidopsis root growth under Mn deficiency (Gao et al., 2018). It has been demonstrated that both AtNramp3 and AtNramp4 are induced by Fe deficiency and are essential for vacuolar Fe mobilization during seed germination in Arabidopsis (Thomine et al., 2003; Lanquar et al., 2005). Both AtNramp3 and AtNramp4 can rescue the growth of the yeast (*Saccharomyces cerevisiae*) mutant Δ*smf1*, which is defective in Mn uptake (Thomine et al., 2000); mutation analysis reveals the roles of AtNramp3 and AtNramp4 in intracellular Mn homeostasis in Arabidopsis (Lanquar et al., 2010). AtNramp6 can transport Fe and Cd in yeast cells, and disruption of AtNramp6 reduces lateral root growth in Arabidopsis under Fe deficiency (Cailliatte et al., 2009; Li et al., 2019). Recently, it has been reported that AtNramp6 together with AtNramp1 is cooperatively involved in Mn utilization in Arabidopsis (Li et al., 2022). The transport activities of Nramp homologues for various metal ions are also demonstrated in other plant species, including OsNramps from rice for Mn, Fe, Cd and Al (Xia et al., 2010; Yamaji et al., 2013; Peris-Peris et al., 2017; Li et al., 2021c), MtNramp1 from *M. truncatula* for Fe and Mn (Tejada-Jiménez et al., 2015), HvNramp5 from barley for Mn and Cd (Wu et al., 2016), MhNramp1 from apple (*Malus hupehensis*) for Cd (Zhang et al., 2020), FeNramp5 from buckwheat (*Fagopyrum esculentum*) for Mn and Cd (Yokosho et al., 2021), and NtNramp1 from tobacco for Fe and Cd (Liu et al., 2022).

Although the functions of Nramp homologues have been characterized and analyzed in different plants, the role of the Nramp members in metal transport activity remains to be elucidated in the tropical forage legume, *Stylosanthes guianensis* (stylo). Stylo is an acid soil adapted, pioneer forage legume, which is widely used for multipurpose including land reclamation and restoration, soil quality improvement, and fodder for farm animals (Tang et al., 2009; Marques et al., 2018; Guo et al., 2019). Stylo displays superior level of tolerance to nutrient and metal stresses that are common in acid soils of the tropics, including phosphorus (P) deficiency, excess Mn stress, and Al toxicity (Sun et al., 2014; Chen et al., 2015; Chen et al., 2021). Considering the key role of Nramp in metal ion uptake and homeostasis, this study aimed to determine the responses of *Nramp* genes to metal stress and investigate its roles in metal transport activity. Five *SgNramps* were identified from stylo and their expression patterns were analyzed. A PM-localized transporter SgNramp1 was further demonstrated to be involved in Mn uptake. The results of this study provide a candidate gene for breeding stylo cultivars with high-efficient Mn uptake.

Materials and methods

Identification of *nramp* genes in *S. guianensis*

Five *Nramp* genes were isolated from the previously reported transcriptomic data in stylo (Jiang et al., 2018; Jia et al., 2020; Chen et al., 2021). The full length of each *Nramp* was amplified from the cDNA library stock of stylo roots (Song et al., 2022). These five *Nramp* genes were named from *SgNramp1* to *SgNramp5* under GenBank accession numbers ON338040, ON338041, ON338042, ON338043, and ON338044, respectively. ClustalX and MAGA4 were used to perform multiple alignment and phylogenetic analyses, respectively. TMD was predicted by the HMMER (https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan). The conserved motifs of Nramp proteins were analyzed by MEME (https://meme-suite.org/meme/tools/meme) and Pfam (http://pfam.xfam.org/search/sequence) programs with default parameters.

Plant growth and treatments

In this study, the stylo cultivar "Reyan NO.5" was used for gene expression analysis. Stylo seeds were sown on moistened filter papers at 25°C in the dark for 2–3 days. The germinated seedlings were then transplanted into Hoagland solution as previously described (Song et al., 2022). The macronutrients and micronutrients in the modified Hoagland solution included 2 mM $Ca(NO_3)_2$, 3 mM KNO_3, 0.25 mM KH_2PO_4, 0.5 mM $MgSO_4$, 25 μM $MgCl_2$, 80 μM Fe-Na-EDTA, 5 μM $MnSO_4$, 0.5 μM $ZnSO_4$, 1.5 μM $CuSO_4$, 0.09 μM $(NH_4)_6Mo_7O_{24}$, and 23 μM $Na_2B_4O_7$. The pH of the solution was adjusted to 5.8 every 2 days. Stylo seedlings were grown in a greenhouse at 25–32°C under normal sunlight conditions with a photoperiod of about 13h light. For expression analysis of *SgNramps*, root, stem, and leaf tissues were harvested at 21 days of growth, whereas flower and seed were harvested at 120 days and 140 days, respectively.

To investigate the responses of *SgNramps* to deficiencies of Mn, Fe, Zn, and Cu, the 14-day-old stylo seedlings precultured in half-strength Hoagland solution were transferred to nutrient solution without Mn, Fe, Zn, or Cu application for 7 days. Leaf and root were harvested and used to gene expression analysis. To analyze the effects of excess Mn, Fe, Zn, and Cu treatments on the transcripts of *SgNramps*, the 14-day-old stylo seedlings precultured in half-strength Hoagland solution were transplanted into fresh Hoagland solution supplied with 400 μM $MnSO_4$, 800 μM Fe-Na-EDTA, 20 μM $ZnSO_4$, or 10 μM $CuSO_4$ for excess Mn, Fe, Zn, or Cu treatments, respectively. After 7 days of treatments, leaf and root were harvested and used for gene expression analysis. The stylo seedlings grown in full-strength Hoagland solution for 21 days were set as the control treatment.

In addition, nine stylo cultivars belonging to *S. guianensis*, including "Reyan NO.2", "Reyan NO.5", "Reyan NO.10", "Reyan NO.13", "Reyan NO.18," "Reyan NO.20", "Reyan NO.21", "Reyan NO.22", and "Reyan NO.24", were used to evaluate the genotypic variations in *SgNramp1* expression. Fourteen-day-old stylo seedlings precultured in half-strength Hoagland solution as described above were transplanted into fresh Hoagland solution supplied with 0.1 or 5 μM $MnSO_4$ regarding as low Mn and normal Mn treatments, respectively. After 7 days of Mn treatments, roots were harvested for *SgNramp1* expression and Mn concentration analyses.

Quantitative real-time polymerase chain reaction analysis

Total RNA was extracted using TRIzol reagent (TIANGEN Biotech, China), and then it was used to synthesize first strand cDNA *via* the HiScript III cDNA synthesis kit (Vazyme, Nanjing, China). The synthesized cDNA was further used to perform quantitative real-time polymerase chain reaction (qRT-PCR) analysis using SYBR Green Master mix reagent (Vazyme, Nanjing, China). qRT-PCR reaction was detected by a QuantStudio™ 6 Flex Real-Time equipment (Thermo Fisher Scientific, Massachusetts, USA). Gene specific primers for qRT-PCR analysis are detailed in Supplementary Table S1. The expression of each *SgNramp* was calculated relative to the housekeeping gene *SgEF-1a* (Song et al., 2022). Three biological replicates were included in this experiment.

Analysis of Mn concentration

After samples thoroughly dried in an oven for 7 days, the dried samples were digested with concentrated nitric acid at 140° C. Mn concentration in the digested solution was determined by atomic absorption spectroscopy. The standard reference material (GBW07603) was used to validate Mn determination as previously described (Li et al., 2021a).

Subcellular localization of SgNramp1

The open reading frame (ORF) of *SgNramp1* was amplified by PCR using *SgNramp1-GFP* primers (Supplementary Table S1). The PCR product was cloned into the N-terminus of the green fluorescent protein (GFP) of the *pBWA(V)HS* vector (Li et al., 2021a). The constructs were introduced into *Agrobacterium tumefaciens* strain Gv3101 and were then

transiently expressed in leaves of 5-week-old tobacco (*Nicotiana benthamiana*) seedlings according to Li et al. (2021). The PM marker (OsMCA1) (Kurusu et al., 2012) fused with red fluorescence protein (mKATE) was used to co-localization with SgNramp1 or empty vector. The fluorescence was detected by a confocal laser scanning TCS SP8 microscopy (Leica, Wetzlar, Germany). GFP fluorescence was detected at 500–530 nm, whereas the red fluorescence was detected at 580–630 nm.

Metal transport analysis of SgNramp1 in yeast cells

In this study, the Mn uptake-defective mutant Δ*smf1* (MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; YOL122c::kanMX4), the Fe uptake-defective mutant Δ*fet3fet4* (MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; YMR058w::kanMX4) and the Cd-sensitive mutant Δ*ycf1* (MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; YDR135c::kanMX4) were obtained from the Euroscarf. The ORF of *SgNramp1* was amplified using *SgNramp1-pYES2* primers (Supplementary Table S1). The amplified product was cloned into the yeast expression vector *pYES2* (Invitrogen, Carlsbad, United States). The construct and empty vector were then introduced into the above yeast mutants using the LiOAc/PEG method. The yeast transformants were incubated in a liquid synthetic complete (SC–U/Glu) medium consisting of yeast nitrogen base, glucose, and amino acids without uracil at 30°C for about 6 h. After optical density (OD_{600}) of the yeast cells reached to 0.6, the yeast cells were diluted to an OD_{600} of 0.2 and three 10-fold serial dilutions were prepared. Then, 10 μl of each dilution was spotted onto the induction SC–U plates containing 2% (w) galactose (Gal) and 1% raffinose according to Li et al. (2021a). For Mn transport analysis in Δ*smf1*, the plate was added with 0, 20, and 25 mM Mn-chelator ethylene glycol tetraacetic acid (EGTA). For Fe transport analysis in Δ*fet3fet4*, the plate was added with 0, 20, and 40 μM $FeSO_4$. For assay on the tolerance to Cd in Δ*ycf1*, the plate was added with 0, 40, and 60 μM $CdCl_2$. The yeast cells were incubated at 30°C for 2 d prior to photographing.

To assess yeast growth in liquid medium, the transformed yeast strain Δ*smf1* was precultured in SC–U liquid medium containing 2% (w) Gal and 1% raffinose to the log phase. The precultured yeast cells were diluted to an OD_{600} of 0.05 and were then grown in SC-U liquid medium added with or without 20 mM EGTA. The OD_{600} value of yeast cells was measured from 0-48h. To determine Mn uptake in yeast, the yeast cells with an initial OD_{600} of 0.2 were incubated in SC–U liquid medium supplemented with 20 and 40 μM $MnSO_4$ for 24 h. To analyze kinetics of Mn uptake, the yeast cells with an initial OD_{600} of 0.2 were incubated in SC–U liquid medium supplemented with 0–10 μM $MnSO_4$ for 6 h. After treatments, the yeast cells were harvested and washed with 5 mM $CaCl_2$ solution, followed by washing with deionized water twice. The collected cells were used for Mn determination. The K_m and V_{max} values of Mn uptake in yeast cells expressing *SgNramp1* were estimated by Lineweaver-Burke plots of Mn concentrations at various Mn treatments after subtracting the Mn concentration in yeast cells transformed with *pYES2* empty vector.

Statistical analyses

One-way ANOVA and Student's *t*-test were analyzed by using SPSS13.0 (SPSS Institute, Chicago, United States).

Results

Characterization of *SgNramp* genes in *S. guianensis*

In this study, five *Nramp* genes were cloned in stylo, which were named as *SgNramp1* to *SgNramp5* (Supplementary Table S2). The full-length sequences of *SgNramps* varied from 1,344 to 1,779 bp, and they encoded proteins differing from 447 to 591 amino acid residues in length with molecular weight ranging from 48.3 to 64.5 kDa (Supplementary Table S2). All of the SgNramp members possessed the conserved Nramp domain (PF01566). Among them, four SgNramp (SgNramp1, 2, 3, and 5) proteins included 12 conserved TMDs, whereas SgNramp4 harbored 10 TMDs (Figure 1 and Supplementary Table S2). Furthermore, the SgNramp proteins carried the unique amino acid residues GQSST(/A)ITGTYAGQF(/Y)I(/V)MQ(/G)GFLN(/D) of plant Nramp proteins (Figure 1). The amino acid sequences of five SgNramps shared 31.8–81.2% identities with each other (Supplementary Figure S1). In addition, SgNramp1 to SgNramp5 shared 32.1–77.0% homology identities with AtNramp1 in Arabidopsis, 35.9–76.0% identities with MtNramp1 in *M. truncatula* and 34.1–81.8% homology identities with GmDMT1 in soybean (Supplementary Figure S1).

Phylogenetic analysis was performed using Nramp members from stylo, Arabidopsis, rice, and *M. truncatula*. Results showed that Nramp proteins were divided into two major groups (Groups I and II) (Figure 2). Group I included three SgNramps (SgNramp1, 4 and 5), two Nramp proteins from Arabidopsis, five Nramp members from rice, and three Nramps from *M. truncatula*. SgNramp2 and SgNramp3 clustered with four Nramp members from Arabidopsis, two Nramp members from rice, and four Nramps from *M. truncatula* were classified into Group II (Figure 2). Interestingly, Nramp members in Group I contained five to six conserved motifs, whereas all Nramp members in Group II contained seven conserved motifs. Among them, SgNramp1 and SgNramp5 included six conserved motifs, whereas SgNramp4 only harbored five conserved motifs. SgNramp2 and SgNramp3 contained seven conserved

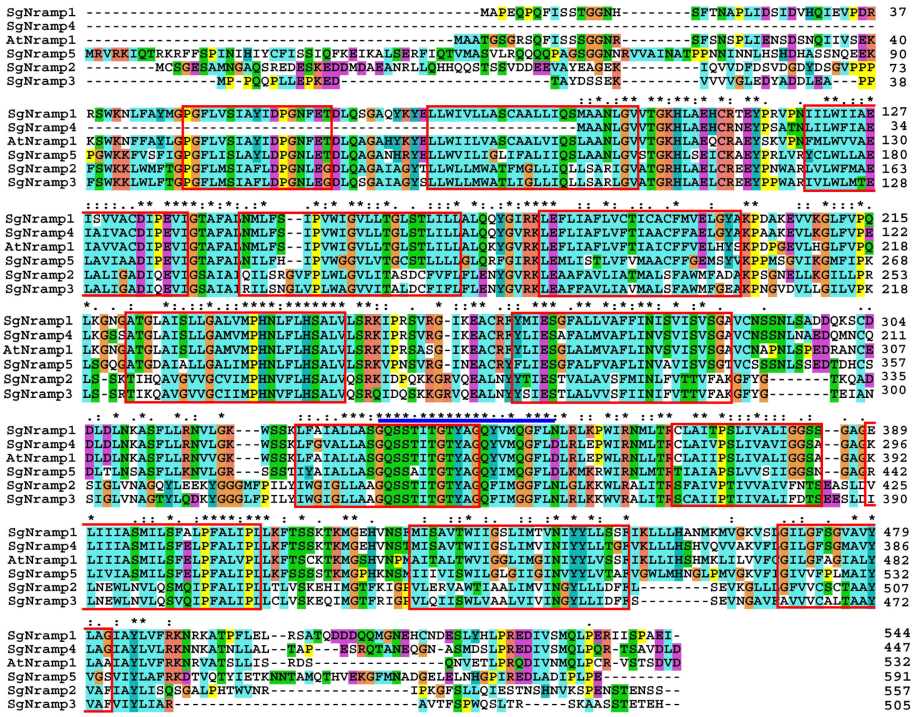


FIGURE 1
Multiple alignment of SgNramps from *S. guianensis* and AtNramp1 from Arabidopsis. The transmembrane domains (TMDs) were boxed with red color. The unique amino acid residues of the Nramp proteins were labeled with blue line above the sequence. The following Nramp proteins were used: SgNramp1 (ON338040), SgNramp2 (ON338041), SgNramp3 (ON338042), SgNramp4 (ON338043), SgNramp5 (ON338044), and AtNramp1 (At1g80830). * in the figure indicates the same amino acid residue included in all Nramp proteins.

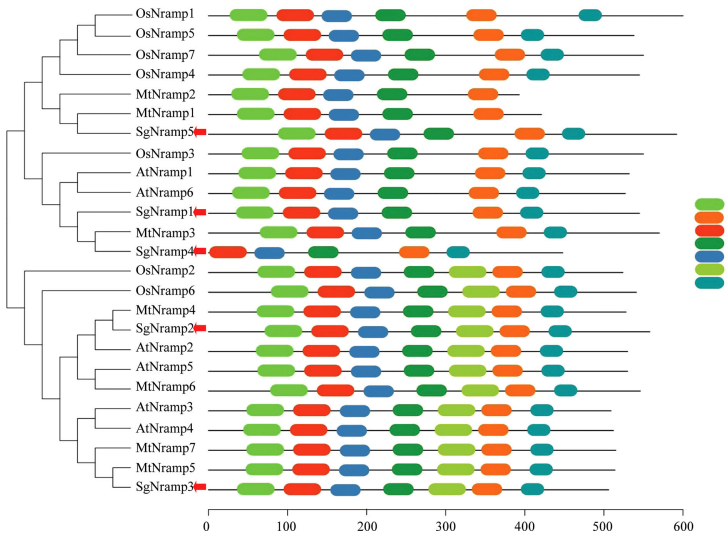


FIGURE 2
Phylogenetic analysis and conserved motifs of Nramp proteins in stylo and other plants. The first two letters of each Nramp member indicate the plant species. At, *Arabidopsis thaliana*; Os, *Oryza sativa*; Mt, *Medicago truncatula*. The red arrows indicate SgNramps from stylo. The sequences of Nramp proteins were obtained from the previous studies (Tejada-Jiménez et al., 2015; Qin et al., 2017; Mani and Sankaranarayanan, 2018). The phylogenetic tree was constructed by MEGA4. Conserved motifs of Nramp proteins were analyzed by MEME. Each motif is presented by a colored box. The sequences of different motifs of Nramp proteins were summarized in Supplementary Table S3.

motifs (Figure 2 and Supplementary Table S3). The conserved motifs included 32–50 amino acids, and putative Nramp domain was predicted in the conserved motifs (Supplementary Table S3).

Expressions of *SgNramps* in different tissues and their responses to metal stresses

The expression patterns of *SgNramps* in tissues of root, stem, leaf, flower, and seed of stylo were measured *via* qRT-PCR. Results showed that *SgNramps* exhibited different expressions in various tissues of stylo (Figure 3). Among them, *SgNramp1* mainly expressed in root and seed and exhibited the lowest expression in leaf. *SgNramp2* displayed preferential expressions in flower. The expressions of *SgNramp3* in stem and flower were higher than those of other tissues, whereas *SgNramp4* showed the higher level of expression in flower and seed compared with the other tissues. *SgNramp5* exhibited a higher expression in stem than in the other tissues (Figure 3).

To investigate the responses of *SgNramps* to trace metal stresses, including deficient and excess Mn, Fe, Zn, and Cu treatments, the transcription profiles of *SgNramps* were measured in leaf and root. Results showed that the transcript of each *SgNramp* gene was regulated by at least one metal stress in leaf and/or root (Figure 4). For example, *SgNramp1* was enhanced, and *SgNramp2*, *SgNramp3* and *SgNramp5* were suppressed by Mn limitation in leaf, whereas *SgNramp2* was enhanced but *SgNramp4* and *SgNramp5* were suppressed under Mn deficient condition in root (Figure 4A). In addition, the number of *SgNramps* regulated by Mn stress with more than twofold (the absolute value of $\log_2 \geq 1$) was higher than those in Fe, Zn, and Cu stresses. For example, a total of 6, 3, and 1 *SgNramp* genes were regulated by deficiencies of Mn, Fe, and Cu in leaf and/or root, respectively, whereas the transcripts of 6, 4, and 3 *SgNramp* genes were regulated by excess Mn, Fe, and Cu treatments in leaf and/or root, respectively (Figures 4A, B). Interestingly, the transcripts of *SgNramp5* were enhanced by more than twofold under limited or excess Fe stress in both leaf and root (Figure 4).

Genotypic differences in *SgNramp1* expression and root Mn concentration

Of the five *SgNramps*, *SgNramp1* displayed high similarity with AtNramp1 that is responsible for Mn uptake in roots of Arabidopsis (Supplementary Figure S1). Furthermore, *SgNramp1* displayed the highest expression in stylo roots compared with the other four *SgNramps* (Figure 3B). Thus, we further assessed whether *SgNramp1* is responsible for genotypic difference in Mn uptake in various stylo cultivars. Results showed that a close correlation (R^2 = 0.64, P < 0.01) was found between *SgNramp1* expression and root Mn concentration in nine stylo cultivars under 0.1 μM $MnSO_4$ condition (low Mn treatment) (Figure 5A).

Among the tested nine stylo cultivars, two cultivars, Reyan NO.18 and Reyan NO.22, exhibited significant genotypic variation in *SgNramp1* expression and root Mn concentration. Furthermore, Mn concentration in root of Reyan NO.18 was significantly higher than that in Reyan NO.22 under both 0.1 and 5 μM $MnSO_4$ treatments, which were considered as low Mn or normal Mn treatments, respectively (Figure 5B). In addition, compared with normal Mn treatment, the expressions of *SgNramp1* were significantly increased by low Mn in Reyan NO.18 but not in Reyan NO.22 (Figure 5C). Therefore,

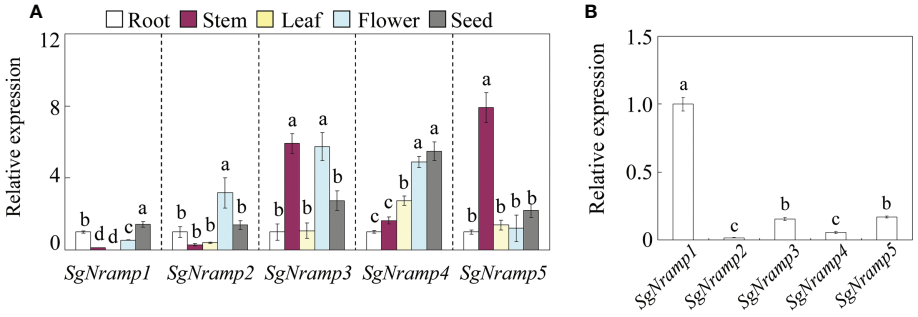


FIGURE 3
Expression analysis of *SgNramps* in various tissues. **(A)** Expressions of *SgNramps* in root, stem, leaf, flower, and seed. The expression of each gene in the roots was set to one, and then the expression level of each gene in other tissues was normalized to its expression in the roots. **(B)** Expressions of *SgNramps* in stylo root. The expression of *SgNramp1* was set to one, and then the expression level of the other *SgNramps* was normalized to *SgNramp1*. Data are means of three biological replicates with standard error (SE). Different letters represent significant differences among various tissue **(A)** or *SgNramp* genes **(B)** at P < 0.05.

10.3389/fpls.2022.1027551

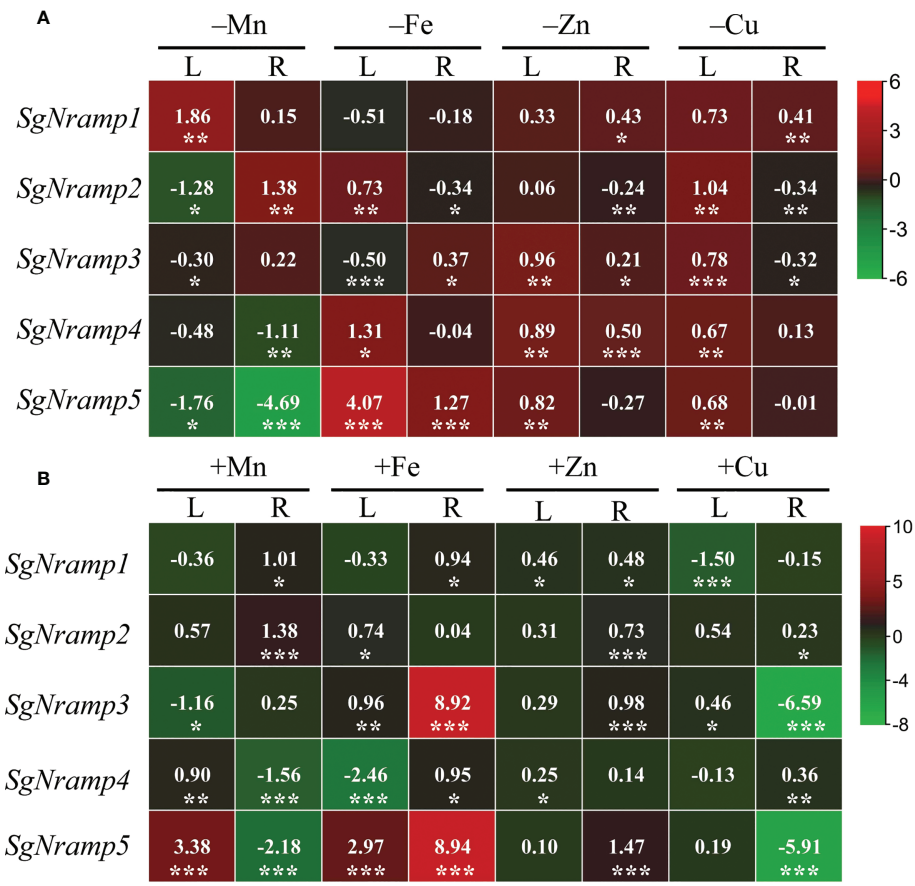


FIGURE 4
Expression analyses of *SgNramps* in leaf and root of stylo under metal stresses. **(A)** Expressions of *SgNramps* in leaf (L) and root (R) of stylo grown in nutrient solution without $MnSO_4$ (−Mn), Fe-Na-EDTA (−Fe), $ZnSO_4$ (−Zn), or $CuSO_4$ (−Cu) applications for 7 days. **(B)** Expressions of *SgNramps* in leaf (L) and root (R) of stylo grown in nutrient solution supplied with 400 μM $MnSO_4$ (+Mn), 800 μM Fe-Na-EDTA (+Fe), 20 μM $ZnSO_4$ (+Zn), or 10 μM $CuSO_4$ (+Cu) for 7 days. Stylo seedlings grown in full-strength Hoagland solution were set as the control treatment. The data include three biological replicates. The values represented the fold change of each *SgNramp* expression between metal stress and the control (CK), which were calculated as $\log_2$(metal stress/CK). The positive and negative values represent up- and down-regulation, respectively. Asterisks below the values indicate significant differences between metal stress and CK. *$P < 0.05$; **$0.001 < P < 0.01$; ***$P < 0.001$.

SgNramp1 is likely to be involved in Mn uptake in stylo and was then selected for further analysis.

Subcellular localization of SgNramp1

SgNramp1 was cloned and fused to the N-terminus of GFP (*35S:SgNramp1-GFP*). Then, subcellular localization analysis was performed in tobacco epidermis cells through transiently expressed *35S:SgNramp1-GFP* and empty vector. Results showed that the fluorescence of GFP alone in epidermis cells was found in the cytoplasm, nucleus, and PM, whereas GPF fluorescence of SgNramp1 was found to be co-localized with that of the PM marker in the epidermis cells (Figure 6), suggesting that SgNramp1 is a PM–localized protein.

Heterologous expression of *SgNramp1* in yeast cells

Subsequently, the coding sequence of *SgNramp1* was cloned into the *pYES2* vector and transformed into various yeast strains. As shown in Figure 7, no differences in the growth were observed in the Mn uptake-defective yeast mutants Δ*smf1* expressing either the *pYES2* empty vector or *SgNramp1* gene in the control medium without EGTA supplement. Under Mn limited condition by added with 20 or 25 mM EGTA, *SgNramp1* expression rescued the growth of the yeast strain Δ*smf1* compared with the empty vector control, although the growth of both strains was inhibited by low Mn stress (Figure 7). No significant differences in growth performance were observed between the Fe uptake-defective yeast mutants Δ*fet3fet4* transformed with either the *pYES2* empty vector or *SgNramp1*

10.3389/fpls.2022.1027551

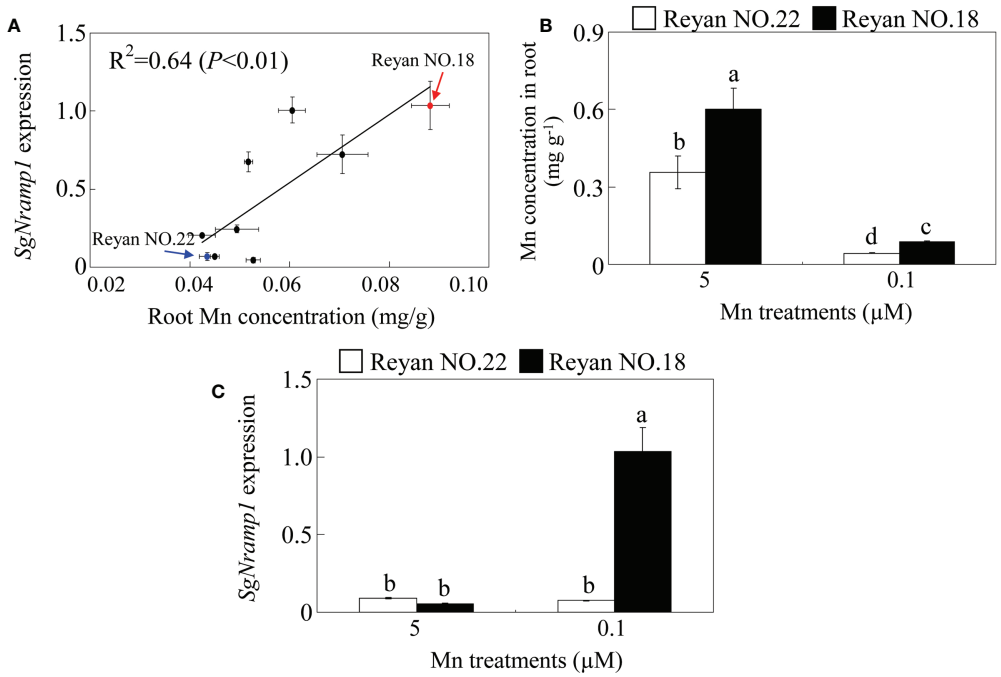


FIGURE 5
Genotypic variation analysis of *SgNramp1*. **(A)** Analysis of correlation between *SgNramp1* expression and root Mn concentration in various stylo cultivars under 0.1 μM $MnSO_4$ treatments. The red and blue arrows indicate Reyan NO.18 and Reyan NO.22 cultivars, respectively. **(B)** Mn concentration in roots of Reyan NO.18 and Reyan NO.22 under 0.1 or 5 μM $MnSO_4$ treatments. **(C)** Expression of *SgNramp1* in roots of Reyan NO.18 and Reyan NO.22 under 0.1 or 5 μM $MnSO_4$ treatments. Fourteen-day-old stylo seedlings were subjected to 0.1 or 5 μM $MnSO_4$ treatments for 7 days that were considered as low Mn and normal Mn treatments, respectively. Stylo roots were used for *SgNramp1* expression analysis and Mn determination. Data are means of three biological replicates with SE. Different letters represent significant differences at $P < 0.05$.

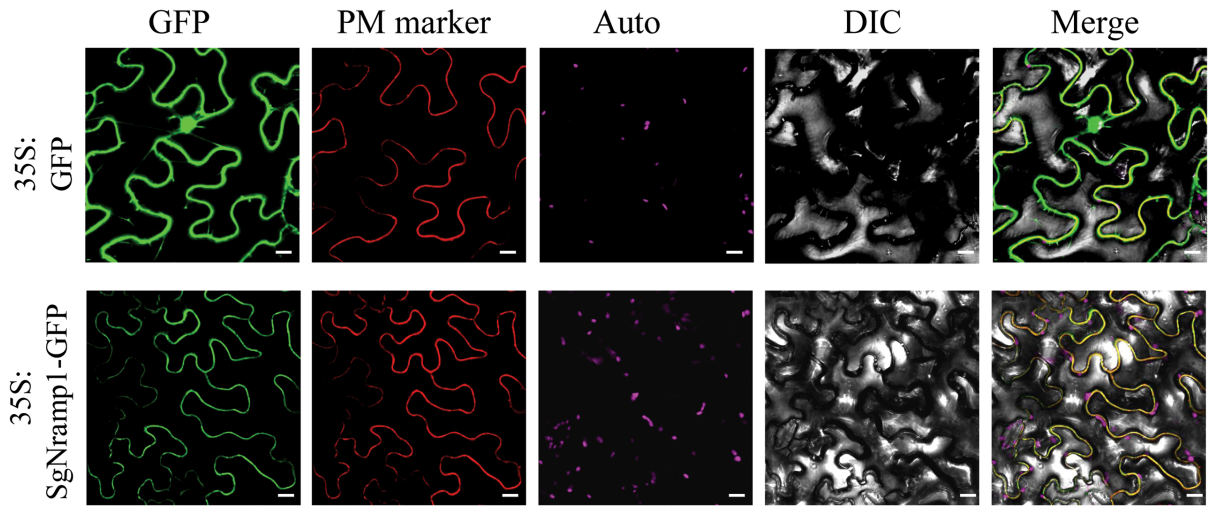


FIGURE 6
Subcellular localization of SgNramp1. The *35:SgNramp1-GFP* construct and *35:GFP* empty vector were transiently expressed in tobacco epidermis cells. SgNramp1 was co-expressed with the plasma membrane (PM) marker, which was fused with red fluorescence protein (mKATE). Signals of GFP fusion protein, PM marker, chlorophyll autofluorescence (Auto), bright-field images (DIC), and the merged images (Merge) were shown from left to right. Scale bar is 20 μm.

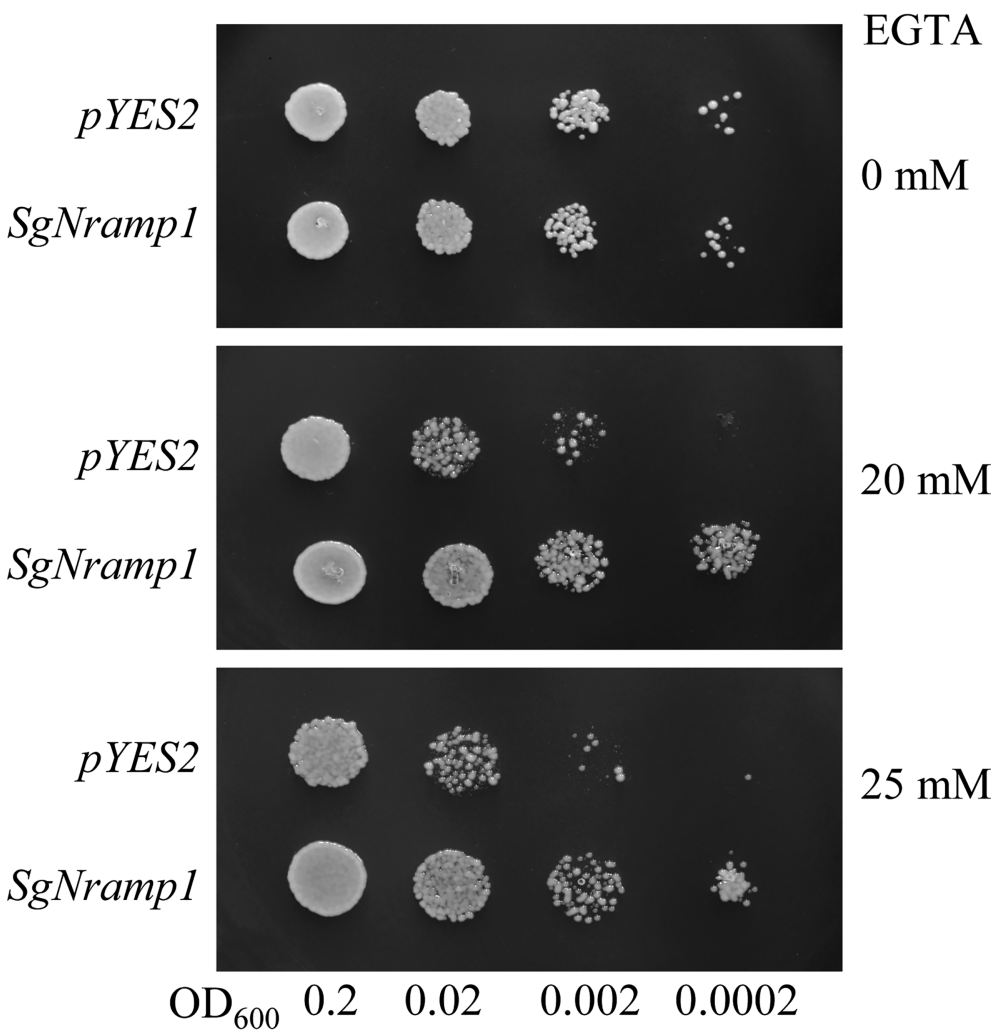


FIGURE 7
Manganese transport analysis of SgNramp1 in yeast. The yeast mutant Δ*smf1* expressing either the *pYES2* empty vector or *SgNramp1* was spotted on SC-U/Gal solid medium containing 0, 20, or 25 mM Mn-chelator EGTA. The photo shows the growth of yeast cells at 30°C for 2 days.

in the medium with 0, 20, or 40 μM Fe supplements (Supplementary Figure S2A). In addition, the Cd-sensitive yeast strains Δ*ycf1* expressing either the *pYES2* empty vector or *SgNramp1* exhibited the same sensitivity to Cd stress in the medium added with 20 or 60 μM Cd (Supplementary Figure S2B). These results suggest that *SgNramp1* can transport Mn but not Fe and Cd.

To further investigate the differences in yeast growth and Mn uptake, the yeast strains Δ*smf1* expressing the *pYES2* empty vector or *SgNramp1* were grown in a liquid SC-U/Gal medium added with or without 20 mM EGTA. The growth curve showed the similar growth status of the two strains under without EGTA condition (Figure 8A). In the presence of 20 mM EGTA, the growth of yeast strain expressing *SgNramp1* was better than the yeast strain transformed with the empty vector control. The OD_{600} values of yeast cells expressing *SgNramp1* were 52.3 and 69.8% higher than that of the yeast strain transformed with the empty vector control at 24 h and 48 h of incubation, respectively (Figure 8A). Furthermore, Mn concentration in the yeast cells expressing *SgNramp1* was 1.4- and 1.9-fold higher than those in the yeast cells transformed with the empty vector control supplemented with 20 and 40 μM $MnSO_4$, respectively (Figure 8B). In addition, kinetic analysis of Mn uptake in yeast showed that the K_m value was 3.8 ± 0.40 μM and the V_{max} value was 1.13 ± 0.09 μg g^{-1} min^{-1} (Figure 8C). Thus, SgNramp1 is involved in Mn uptake in yeast cells.

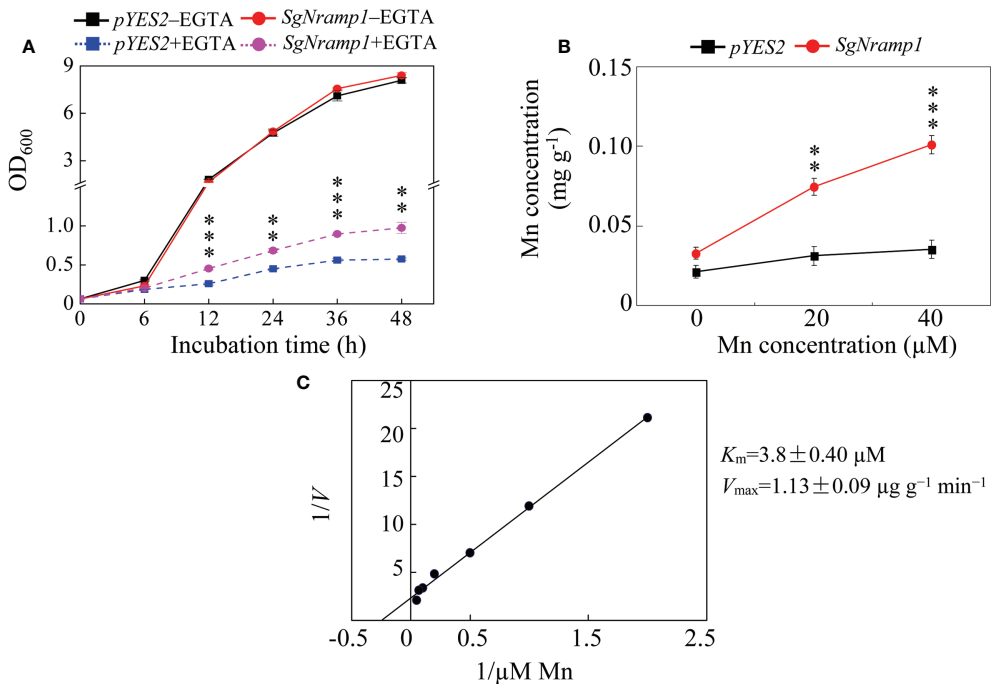


FIGURE 8
Growth and Mn uptake in yeast mutant Δ*smf1* expressing *SgNramp1*. **(A)** The yeast mutant Δ*smf1* expressing either the *pYES2* empty vector or *SgNramp1* was grown in SC-U/Gal liquid medium with or without 20 mM EGTA for 0h to 48h. **(B)** Mn concentrations in the yeast cells grown in SC-U/Gal liquid medium containing 0, 20, or 40 μM $MnSO_4$ for 24h. **(C)** The kinetics of Mn uptake in yeast cells. The kinetics parameters are shown next to the figure. Data are means of three biological replicates with SE. Asterisks indicate significant differences between the *pYES2* empty vector and *SgNramp1*. **0.001 < *P* < 0.01; ***P < 0.001.

Discussion

Numerous studies demonstrated that Nramp proteins participate in the uptake, translocation, and distribution of metal ions, regulating metal homeostasis in plants (Sasaki et al., 2012; Li et al., 2021c). Although Nramp homologues have been reported in plants, Nramp members and their functional roles in the important tropical legume, stylo, have not been elucidated so far. In this study, we identified five *SgNramp* genes in stylo. All of them possessed the conserved Nramp domain (PF01566) and carried the unique amino acid residues of Nramp proteins (Figure 1), which have been widely identified in AtNramps from Arabidopsis, OsNramps from rice, GmNramps from soybean, PvNramps from bean (*Phaseolus vulgaris*), and CsNramps from tea (*Camellia sinensis*) (Thomine et al., 2000; Yamaji et al., 2013; Qin et al., 2017; Ishida et al., 2018; Li et al., 2021b). Furthermore, Nramp proteins have been reported to include 10–12 TMDs. For example, there are 12 TMDs in AtNramp1/2, 10 TMDs in OsNramp5, 11 TMDs in HvNramp5, and 12 TMDs in FeNramp5 (Curie et al., 2000; Sasaki et al., 2012; Wu et al., 2016; Yokosho et al., 2021). Consistently, all of the SgNramp proteins contained 12 conserved TMDs, except SgNramp4, which harbored 10 TMDs (Figure 1 and Supplementary Table S2). These conserved structures are essential for the functions of Nramp proteins.

Phylogenetic analysis showed that SgNramp1 and SgNramp4, together with AtNramp1 and AtNramp6 from Arabidopsis, MtNramp3 from *M. truncatula*, and OsNramp3 from rice were grouped into the same cluster in Group I (Figure 2). AtNramp1 is a PM-localized transporter that participates in Mn uptake in Arabidopsis (Cailliatte et al., 2010). AtNramp6 is able to transport Fe and regulates lateral root growth in Arabidopsis under low Fe stress (Cailliatte et al., 2009; Li et al., 2019). Furthermore, AtNramp6 is suggested to cooperate with AtNramp1 in controlling Mn homeostasis in Arabidopsis (Li et al., 2022). OsNramp3 is characterized to be an important node-based transporter essential for preferential Mn distribution, mediating Mn remobilization between young leaves and old tissues of rice in response to variable Mn levels (Yamaji et al., 2013; Yang et al., 2013). As the closest homologue of SgNramp5 in Group I, MtNramp1 is implicated in Fe transport in the yeast cells, and it participates in apoplastic Fe uptake by

rhizobia-infected cells in nodules of *M. truncatula* (Tejada-Jiménez et al., 2015). In Group II, SgNramp2 exhibited high similarity to AtNramp2, which is localized to the trans-Golgi network and is proposed to involve in Mn transport in Golgi, regulating root growth of Arabidopsis under low Mn stress (Gao et al., 2018). SgNramp3 was clustered with AtNramp3/4 and MtNramp5/7 in the same subgroup (Figure 2). In addition to participate in vacuolar Fe mobilization during seed germination, AtNramp3 and AtNramp4 are demonstrated to be responsible for intracellular Mn homeostasis in Arabidopsis (Thomine et al., 2003; Lanquar et al., 2005; Lanquar et al., 2010). Therefore, stylo SgNramps might possess the potential role in regulating metal homeostasis in stylo.

Nramp proteins have been characterized to function in the regulation of metal uptake and homeostasis, and thus they may exhibit preferential expressions in various plant tissues. For example, a set of *Nramp* homologues are observed to be mainly expressed in roots, such as *AtNramp1* from Arabidopsis (Cailliatte et al., 2010), *OsNramp4/5* from rice (Xia et al., 2010; Sasaki et al., 2012), *MtNramp1* from *M. truncatula* (Tejada-Jiménez et al., 2015), *HvNramp5* from barley (Wu et al., 2016), *TtNramp6* from wheat (*Triticum turgidum*) (Wang et al., 2019a), *AhNramp1* from peanut (*Arachis hypogaea*) (Wang et al., 2019b), *FeNramp5* from buckwheat (Yokosho et al., 2021), and *NtNramp1* from tobacco (Liu et al., 2022). The root-specific expression of *Nramps* is beneficial for metal uptake. Similarly, variations in the abundances of *SgNramps* were observed in different tissues of stylo, such as *SgNramp1* mainly expressed in root and seed (Figure 3), suggesting the key role of *SgNramp* members in regulating metal homeostasis in specific tissues.

In addition, *SgNramp* genes exhibited diverse responses to Mn, Fe, Zn, and Cu stresses as well as heavy metal toxicity in stylo (Figure 4 and Supplementary Figure S3). In Arabidopsis, the transcripts of *AtNramp1* in roots are enhanced by both low Fe and Mn levels, whereas both *AtNramp3* and *AtNramp4* abundances are increased in roots and shoots by Fe starvation but not Mn deficiency (Curie et al., 2000; Thomine et al., 2000; Cailliatte et al., 2010). In rice, the expression of *OsNramp5* gene is up-regulated in roots by Zn and Fe deficiencies but not low Mn and Cu conditions (Xia et al., 2010; Sasaki et al., 2012). On the other hand, although some of the *Nramp* homologues are not response to metal stresses, they are able to transport metal ions. For example, *AtNramp2* and *AtNramp6* genes are not regulated by deficiencies of Mn, Fe, and Zn, but *AtNramp2* is found to be involved in remobilization of Mn in Golgi for Arabidopsis root growth (Gao et al., 2018), and *AtNramp6* is shown to participate in regulation of Fe and Mn homeostasis (Li et al., 2019; Li et al., 2022). Although *OsNramp3* is unaffected by deficiencies of Zn, Fe, Mn, and Cu at transcriptional level, its encoding protein is rapidly degraded after high Mn exposure, regulating Mn distribution in young leaves and old tissues (Yamaji et al., 2013). Thus, changes in the abundances of *SgNramp* genes might potentially contribute to regulate metal homeostasis in various tissues of stylo for dealing with the fluctuated metal levels.

In this study, we found that the transcript levels of *SgNramp1* are responsible for the genotypic differences in root Mn levels in stylo, as reflected by a close correlation between *SgNramp1* expression and Mn concentration in various stylo cultivars (Figure 5). Furthermore, a higher level of Mn was observed in the stylo cultivar Reyan NO.18 compared with Reyan NO.22, which was correlated with higher *SgNramp1* expression in Reyan NO.18 under low Mn condition (Figure 5). Similarly, differences in Mn efficiency have been observed among barley genotypes, which has been attributed to differences in Mn uptake capability; by analyzing two barley genotypes contrasting in Mn acquisition efficiency and gene expression, the role in Mn uptake has been suggested for HvIRT1, which exhibits higher expression in a Mn-efficient barley genotype than a Mn-inefficient genotype and has the ability to restore Mn uptake of Δ*smf1* mutant (Pedas et al., 2005, Pedas et al., 2008).

The metal transport activity of SgNramp1, which was localized to the PM, was further analyzed in yeast cells. Heterologous expression of *SgNramp1* in yeast mutants showed that *SgNramp1* can restore the growth of the yeast strain Δ*smf1* through increasing Mn uptake (Figures 7 and 8). It has been demonstrated that a homologue of *SgNramp1*, *AtNramp1*, encoding a PM-localized transporter, can restore the growth of yeast mutant Δ*smf1* under Mn limitation (Curie et al., 2000; Thomine et al., 2000). Furthermore, loss function of *AtNramp1* disrupts the growth and Mn concentration in Arabidopsis under low Mn condition, demonstrating that AtNramp1 is the transporter responsible for Mn uptake in Arabidopsis (Cailliatte et al., 2010). In addition, broad transport substrates of Nramp homologues were observed in other plants, such *M. truncatula*, rice, barley, and buckwheat (Ishimaru et al., 2012; Tejada-Jiménez et al., 2015; Wu et al., 2016; Yokosho et al., 2021). For example, OsNramp5 in rice transports Mn, Cd, and Fe (Ishimaru et al., 2012), whereas HvNramp5 in barley and FeNramp5 in buckwheat only transport Mn and Cd but not Fe (Wu et al., 2016; Yokosho et al., 2021). Thus, it is reasonable to assume that *SgNramp1* might contribute to Mn acquisition in stylo. However, the function of *SgNramp1* may be differed in various stylo cultivars. Further research work is needed, such as generation of transgenic stylo plants overexpressing or suppressing *SgNramp1*, to verify its exact function conducted in both nutrient solutions and soil conditions with different Mn status and availability.

In summary, we identified five *SgNramp* genes in stylo. Changes observed in the abundances of *SgNramps* suggest the involvement of these genes in regulating metal homeostasis in tissues of different plant parts of stylo during metal stress conditions. Moreover, SgNramp1 is a PM-localized transporter and it is responsible for Mn uptake. Increases of *SgNramp1* expression may contribute to enhance the acquisition of Mn in stylo. Therefore, *SgNramp1* might be a good candidate for breeding future stylo cultivars with efficient Mn uptake through gene-editing technology.

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

ZC designed the research. XZ, RH, LW and GW performed the experiments and analyzed the data. YM and GL prepared the plant material for this work. XZ, RH and ZC wrote the manuscript. IR, GL and ZC discussed and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by the National Natural Science Foundation of China (32271756, 31861143013), the National Natural Science Foundation of Hainan (321RC645), and the China Agriculture Research System of MOF and MARA (CARS-34).

Acknowledgments

The authors appreciate Yufen Xing for technical help.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1027551/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Homology identity analysis of SgNramp proteins. **(A)** Homology identity among SgNramp proteins. **(B)** Homology identity of SgNramps with AtNramp1 (accession no. At1g80830) from Arabidopsis. **(C)** Homology identity of SgNramps with MtNramp1 (accession no. Medtr3g088460) from *M. truncatula*. **(D)** Homology identity of SgNramps with GmDMT1 (accession no. AY169405) from soybean. The homology identity (%) was analyzed by Clustal.

SUPPLEMENTARY FIGURE 2
Transport activity of SgNramp1 in yeast mutants Δ*fet3fet4* and Δ*ycf1*. **(A)** The yeast mutant Δ*fet3fet4* defective in Fe uptake expressing either the *pYES2* empty vector or *SgNramp1* was grown in SC-U/Gal medium containing 0, 20, or 40 μM Fe. **(B)** The yeast mutant Δ*ycf1* sensitive in Cd stress expressing either the *pYES2* empty vector or *SgNramp1* was grown in SC-U/Gal medium containing 0, 20, or 60 μM Cd. The photo shows the growth of yeast cells at 30°C for 2 days.

SUPPLEMENTARY FIGURE 3
Expression of *SgNramps* in roots of stylo under heavy metal treatments. Fourteen-day-old stylo seedlings were separately subjected to 0 (CK), 100 μM $AlCl_3$, 40 μM $CdCl_2$, or 20 μM $LaCl_3$ treatments for 2 days. Roots were harvested for gene expression analysis. Data are means of three biological replicates with standard error (SE). Asterisks indicate significant differences between control and treatments. *$0.01 < P < 0.05$; **$0.001 < P < 0.01$; ***$P < 0.001$.

References

Cailliatte, R., Lapeyre, B., Briat, J. F., Mari, S., and Curie, C. (2009). The NRAMP6 metal transporter contributes to cadmium toxicity. *Biochem. J.* 422, 217–228. doi: 10.1042/BJ20090655

Cailliatte, R., Schikora, A., Briat, J. F., Mari, S., and Curie, C. (2010). High-affinity manganese uptake by the metal transporter NRAMP1 is essential for *Arabidopsis* growth in low manganese conditions. *Plant Cell* 22, 904–917. doi: 10.1105/tpc.109.073023

Chang, J. D., Huang, S., Yamaji, N., Zhang, W. W., Ma, J. F., and Zhao, F. J. (2020). OsNRAMP1 transporter contributes to cadmium and manganese uptake in rice. *Plant Cell Environ.* 43, 2476–2491. doi: 10.1111/pce.13843

Chen, Z. J., and Li, J. F. (2021). "Mechanism of manganese uptake and homeostasis in plant cell," in *Cation transporters in plants, ed s. upadhyay*, vol. 13. (Elsevier, Netherlands: Academic Press), 227–246. doi: 10.1016/B978-0-323-85790-1.00025-7

Chen, Z. J., Song, J. L., Li, X. Y., Arango, J., Cardoso, J. A., Rao, I., et al. (2021). Physiological responses and transcriptomic changes reveal the mechanisms underlying adaptation of *Stylosanthes guianensis* to phosphorus deficiency. *BMC Plant Biol.* 21, 466. doi: 10.1186/s12870-021-03249-2

Chen, Z. J., Sun, L. L., Liu, P. D., Liu, G. D., Tian, J., and Liao, H. (2015). Malate synthesis and secretion mediated by a manganese-enhanced malate dehydrogenase confers superior manganese tolerance in *Stylosanthes guianensis*. *Plant Physiol.* 167, 176–188. doi: 10.1104/pp.114.251017

Curie, C., Alonso, J. M., Le Jean, M., Ecker, J. R., and Briat, J. F. (2000). Involvement of NRAMP1 from *Arabidopsis thaliana* in iron transport. *Biochem. J.* 347, 749–755. doi: 10.1042/0264-6021:3470749

Fryzova, R., Pohanka, M., Martinkova, P., Cihlarova, H., Brtnicky, M., Hladky, J., et al. (2018). Oxidative stress and heavy metals in plants. *Rev. Environ. Contam. Toxicol.* 245, 129–156. doi: 10.1007/398_2017_7

Gao, H. L., Xie, W. X., Yang, C. H., Xu, J. Y., Li, J. J., Wang, H., et al. (2018). NRAMP2, a trans-golgi network-localized manganese transporter, is required for *Arabidopsis* root growth under manganese deficiency. *New Phytol.* 217, 179–193. doi: 10.1111/nph.14783

Guo, P. F., Liu, P. D., Lei, J., Chen, C. H., Qiu, H., Liu, G. D., et al. (2019). Improvement of plant regeneration and *Agrobacterium*-mediated genetic transformation of *Stylosanthes guianensis*. *Trop. Grassl-Forrajes.* 7, 480–492. doi: 10.17138/tgft(7)480-492

Huang, S., Wang, P., Yamaji, N., and Ma, J. F. (2020). Plant nutrition for human nutrition: hints from rice research and future perspectives. *Mol. Plant* 13, 825–835. doi: 10.1016/j.molp.2020.05.007

Ishida, J. K., Caldas, D. G. G., Oliveira, L. R., Frederici, G. C., Leite, L. M. P., and Mui, T. S. (2018). Genome-wide characterization of the *NRAMP* gene family in *Phaseolus vulgaris* provides insights into functional implications during common bean development. *Genet. Mol. Biol.* 41, 820–833. doi: 10.1590/1678-4685-GMB-2017-0272

Ishimaru, Y., Takahashi, R., Bashir, K., Shimo, H., Senoura, T., Sugimoto, K., et al. (2012). Characterizing the role of rice NRAMP5 in manganese, iron and cadmium transport. *Sci. Rep.* 2, 286. doi: 10.1038/srep00286

Jia, Y. D., Li, X. Y., Liu, Q., Hu, X., Li, J. F., Dong, R. S., et al. (2020). Physiological and transcriptomic analyses reveal the roles of secondary metabolism in the adaptive responses of *Stylosanthes* to manganese toxicity. *BMC Genomics* 21, 861. doi: 10.1186/s12864-020-07279-2

Jiang, C. D., Liu, L. S., Li, X. F., Han, R. R., Wei, Y. M., and Yu, Y. X. (2018). Insights into aluminum-tolerance pathways in *Stylosanthes* as revealed by RNA-seq analysis. *Sci. Rep.* 8, 6072. doi: 10.1038/s41598-018-24536-3

Kaiser, B. N., Moreau, S., Castelli, J., Thomson, R., Lambert, A., Bogliolo, S., et al. (2003). The soybean NRAMP homologue, GmDMT1, is a symbiotic divalent metal transporter capable of ferrous iron transport. *Plant J.* 35, 295–304. doi: 10.1046/j.1365-313x.2003.01802.x

Kaur, R., Das, S., Bansal, S., Singh, G., Sardar, S., Dhar, H., et al. (2021). Heavy metal stress in rice: uptake, transport, signaling, and tolerance mechanisms. *Physiol. Plant* 173, 430–448. doi: 10.1111/ppl.13491

Kurusu, T., Nishikawa, D., Yamazaki, Y., Gotoh, M., Nakano, M., Hamada, H., et al. (2012). Plasma membrane protein OsMCA1 is involved in regulation of hypo-osmotic shock-induced Ca^{2+} influx and modulates generation of reactive oxygen species in cultured rice cells. *BMC Plant Biol.* 12, 11. doi: 10.1186/1471-2229-12-11

Lanquar, V., Lelievre, F., Bolte, S., Hames, C., Alcon, C., Neumann, D., et al. (2005). Mobilization of vacuolar iron by AtNRAMP3 and AtNRAMP4 is essential for seed germination on low iron. *EMBO J.* 24, 4041–4051. doi: 10.1038/sj.emboj.7600864

Lanquar, V., Ramos, M. S., Lelievre, F., Barbier-Brygoo, H., Krieger-Liszkay, A., Kramer, U., et al. (2010). Export of vacuolar manganese by AtNRAMP3 and AtNRAMP4 is required for optimal photosynthesis and growth under manganese deficiency. *Plant Physiol.* 152, 1986–1999. doi: 10.1104/pp.109.150946

Li, J. F., Dong, R. S., Jia, Y. D., Huang, J., Zou, X. Y., An, N., et al. (2021a). Characterization of metal tolerance proteins and functional analysis of GmMTP8.1 involved in manganese tolerance in soybean. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.683813

Li, J. Q., Duan, Y., Han, Z. L., Shang, X. W., Zhang, K. X., Zou, Z. W., et al. (2021b). Genome-wide identification and expression analysis of the *NRAMP* family genes in tea plant (*Camellia sinensis*). *Plants* 10, 1055. doi: 10.3390/plants10061055

Li, Y., Li, J. J., Yu, Y. H., Dai, X., Gong, C. Y., Gu, D. F., et al. (2021c). The tonoplast-localized transporter OsNRAMP2 is involved in iron homeostasis and affects seed germination in rice. *J. Exp. Bot.* 72, 4839–4852. doi: 10.1093/jxb/erab159

Liu, W. H., Huo, C. H., He, L. S., Ji, X., Yu, T., Yuan, J. W., et al. (2022). The NtNRAMP1 transporter is involved in cadmium and iron transport in tobacco (*Nicotiana tabacum*). *Plant Physiol. Biochem.* 173, 59–67. doi: 10.1016/j.plaphy.2022.01.024

Li, J. Y., Wang, Y. R., Zheng, L., Li, Y., Zhou, X. L., Li, J. J., et al. (2019). The intracellular transporter AtNRAMP6 is involved in fe homeostasis in *Arabidopsis*. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01124

Li, L., Zhu, Z. Z., Liao, Y. H., Yang, C. H., Fan, N., Zhang, J., et al. (2022). NRAMP6 and NRAMP1 cooperatively regulate root growth and manganese translocation under manganese deficiency in arabidopsis. *Plant J* 110, 1564–1577. doi: 10.1111/tpj.15754

Mani, A., and Sankaranarayanan, K. (2018). In silico analysis of natural resistance-associated macrophage protein (NRAMP) family of transporters in rice. *Protein J.* 37, 237–247.

Marques, A., Moraes, L., Aparecida Dos Santos, M., Costa, I., Costa, L., Nunes, T., et al. (2018). Origin and parental genome characterization of the allotetraploid *Stylosanthes scabra* Vogel (Papilionoideae, leguminosae), an important legume pasture crop. *Ann. Bot.* 122, 1143–1159. doi: 10.1093/aob/mcy113

Marschner, H. (2012). *Marschner's mineral nutrition of higher plants*. *3rd ed.* Ed. H. Marschner (Netherlands, Academic Press: Elsevier).

Pedas, P., Hebbern, C. A., Schjoerring, J. K., Holm, P. E., and Husted, S. (2005). Differential capacity for high-affinity manganese uptake contributes to differences between barley genotypes in tolerance to low manganese availability. *Plant Physiol.* 139, 1411–1420. doi: 10.1104/pp.105.067561

Pedas, P., Ytting, C. K., Fuglsang, A. T., Jahn, T. P., Schjoerring, J. K., and Husted, S. (2008). Manganese efficiency in barley: Identification and characterization of the metal ion transporter HvIRT1. *Plant Physiol.* 148, 455–466. doi: 10.1104/pp.108.118851

Peris-Peris, C., Serra-Cardona, A., Sánchez-Sanuy, F., Campo, S., Ariño, J., and San Segundo, B. (2017). Two NRAMP6 isoforms function as iron and manganese transporters and contribute to disease resistance in rice. *Mol. Plant Microbe In.* 30, 385–398. doi: 10.1094/MPMI-01-17-0005-R

Pottier, M., Le Thi, V. A., Primard-Brisset, C., Marion, J., Bianchi, M., Victor, C., et al. (2022). Duplication of *NRAMP3* gene in poplars generated two homologous transporters with distinct functions. *Mol. Biol. Evol.* 39, msac129. doi: 10.1093/molbev/msac129

Qin, L., Han, P., Chen, L., Walk, T. C., Li, Y., Hu, X., et al. (2017). Genome-wide identification and expression analysis of NRAMP family genes in soybean (*Glycine max* l.). *Front. Plant Sci.* 8, 1436 doi: 10.3389/fpls.2017.01436

Riyazuddin, R., Nisha, N., Ejaz, B., Khan, M. I. R., Kumar, M., Ramteke, P. W., et al. (2021). A comprehensive review on the heavy metal toxicity and sequestration in plants. *Biomolecules* 12, 43. doi: 10.3390/biom12010043

Sasaki, A., Yamaji, N., Yokosho, K., and Ma, J. F. (2012). Nramp5 is a major transporter responsible for manganese and cadmium uptake in rice. *Plant Cell* 24, 2155–2167. doi: 10.1105/tpc.112.096925

Song, J. L., Zou, X. Y., Liu, P. D., Cardoso, J. A., Schultze-Kraftc, R., Liu, G. D., et al. (2022). Differential expressions and enzymatic properties of malate dehydrogenases in the response of *Stylosanthes guianensis* to nutrient and metal stresses. *Plant Physiol. Biochem.* 170, 325–337. doi: 10.1016/j.plaphy.2021.12.012

Sun, L. L., Liang, C. Y., Chen, Z. J., Liu, P. D., Tian, J., Liu, G. D., et al. (2014). Superior Al tolerance of *Stylosanthes* is mainly achieved by malate synthesis through an Al-enhanced malic enzyme, SgME1. *New Phytol.* 202, 209–219. doi: 10.1111/nph.12629

Tang, Y. Q., Wu, Z. Y., Liu, G. D., and Yi, K. X. (2009). Research advances in germplasm resources of *Stylosanthes*. *Chin. Bull. Bot.* 44, 752–762. doi: 10.3969/j.issn.1674-3466.2009.06.014

Tejada-Jiménez, M., Castro-Rodríguez, R., Kryvoruchko, I., Lucas, M. M., Udvardi, M., Imperial, J., et al. (2015). *Medicago truncatula* natural resistance-associated macrophage protein1 is required for iron uptake by rhizobia-infected nodule cells. *Plant Physiol.* 168, 258–272. doi: 10.1104/pp.114.254672

Thomine, S., Lelievre, F., Debarbieux, E., Schroeder, J. I., and Barbier-Brygoo, H. (2003). AtNRAMP3, a multispecific vacuolar metal transporter involved in plant responses to iron deficiency. *Plant J.* 34, 685–695. doi: 10.1046/j.1365-313x.2003.01760.x

Thomine, S., Wang, R., Ward, J. M., Crawford, N. M., and Schroeder, J. I. (2000). Cadmium and iron transport by members of a plant metal transporter family in *Arabidopsis* with homology to *Nramp* genes. *Proc. Natl. Acad. Sci. U.S.A.* 97, 4991–4996. doi: 10.1073/pnas.97.9.4991

Wang, C., Chen, X., Yao, Q., Long, D., Fan, X. , Kang, H. Y., et al. (2019a). Overexpression of *TtNRAMP6* enhances the accumulation of cd in *Arabidopsis*. *Gene* 696, 225–232. doi: 10.1016/j.gene.2019.02.008

Wang, N. Q., Qiu, W., Dai, J., Guo, X. T., Lu, Q. F., Wang, T. Q., et al. (2019b). AhNRAMP1 enhances manganese and zinc uptake in plants. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00415

Wang, P. T., Yamaji, N., Inoue, K., Mochida, K., and Ma, J. F. (2020). Plastic transport systems of rice for mineral elements in response to diverse soil environmental changes. new. *Phytol.* 226, 156–169. doi: 10.1111/nph.16335

Wu, D. Z., Yamaji, N., Yamane, M., Kashino-Fujii, M., Sato, K., and Ma, J. F. (2016). The HvNramp5 transporter mediates uptake of cadmium and manganese, but not iron. p*l*ant. *Physiol.* 172, 1899–1910. doi: 10.1104/pp.16.01189

Xia, J. X., Yamaji, N., Kasai, T., and Ma, J. F. (2010). Plasma membrane-localized transporter for aluminum in rice. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18381–18385. doi: 10.1073/pnas.1004949107

Yamaji, N., Sasaki, A., Xia, J. X., Yokosho, K. , and Ma, J. F. (2013). A node based switch for preferential distribution of manganese in rice. *Nat. Commun.* 4, 2442. doi: 10.1038/ncomms3442

Yang, M., Zhang, W., Dong, H. X., Zhang, Y. Y., Lv, K., Wang, D. J., et al. (2013). OsNRAMP3 is a vascular bundles-specific manganese transporter that is responsible for manganese distribution in rice. *Plos One* 8, e83990. doi: 10.1371/journal.pone.0083990

Yokosho, K., Yamaji, N., and Ma, J. F. (2021). Buckwheat FeNramp5 mediates high manganese uptake in roots. *Plant Cell Physiol.* 62, 600–609. doi: 10.1093/pcp/pcaa153

Zhang, W. W., Yue, S. Q., Song, J. F., Xun, M., Han, M. Y., and Yang, H. Q. (2020). *MhNRAMP1* from *Malus hupehensis* exacerbates cell death by accelerating cd uptake in tobacco and apple calli. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00957

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 17 October 2022
DOI 10.3389/fpls.2022.996672

OPEN ACCESS

EDITED BY
Jorge Fernando Pereira,
Embrapa Gado de Leite, Brazil

REVIEWED BY
Meki Muktar,
International Livestock Research Institute, Ethiopia
Zhipeng Liu,
Lanzhou University, China

*CORRESPONDENCE
Junmei Kang
kangjunmei@caas.cn
Zhen Wang
wangzhen@caas.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

RECEIVED 18 July 2022
ACCEPTED 26 September 2022
PUBLISHED 17 October 2022

CITATION
Jiang X, Yu A, Zhang F, Yang T, Wang C, Gao T, Yang Q, Yu L-X, Wang Z and Kang J (2022) Identification of QTL and candidate genes associated with biomass yield and Feed Quality in response to water deficit in alfalfa (*Medicago sativa* L.) using linkage mapping and RNA-Seq. *Front. Plant Sci.* 13:996672. doi: 10.3389/fpls.2022.996672

Identification of QTL and candidate genes associated with biomass yield and Feed Quality in response to water deficit in alfalfa (*Medicago sativa* L.) using linkage mapping and RNA-Seq

Xueqian Jiang[1†], Andong Yu[1†], Fan Zhang[1], Tianhui Yang[2], Chuan Wang[2], Ting Gao[2], Qingchuan Yang[1], Long-Xi Yu[3], Zhen Wang[1*] and Junmei Kang[1*]

[1]Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China, [2]Institute of Animal Science, Ningxia Academy of Agricultural and Forestry Sciences, Ningxia, China, [3]Plant Germplasm Introduction and Testing Research, United States Department of Agriculture-Agricultural Research Service, Prosser, WA, United States

Biomass yield and Feed Quality are the most important traits in alfalfa (*Medicago sativa* L.), which directly affect its economic value. Drought stress is one of the main limiting factors affecting alfalfa production worldwide. However, the genetic and especially the molecular mechanisms for drought tolerance in alfalfa are poorly understood. In this study, linkage mapping was performed in an F1 population by combining 12 phenotypic data (biomass yield, plant height, and 10 Feed Quality-related traits). A total of 48 significant QTLs were identified on the high-density genetic linkage maps that were constructed in our previous study. Among them, nine main QTLs, which explained more than 10% phenotypic variance, were detected for biomass yield (one), plant height (one), CP (two), ASH (one), P (two), K(one), and Mg (one). A total of 31 candidate genes were identified in the nine main QTL intervals based on the RNA-seq analysis under the drought condition. Blast-P was further performed to screen candidate genes controlling drought tolerance, and 22 functional protein candidates were finally identified. The results of the present study will be useful for improving drought tolerance of alfalfa varieties by marker-assisted selection (MAS), and provide promising candidates for further gene cloning and mechanism study.

KEYWORDS

alfalfa, biomass yield, Feed Quality, drought, linkage mapping, RNA-Seq

Introduction

Abiotic stresses, such as drought, cold, and salinity, are misfortunes for agriculture, which seriously limit crop productivity. In particular, drought is an increasing worldwide threat (Gupta et al., 2020). Economic losses in crop production were almost $30 billion in the past decade due to drought stress (Gupta et al., 2020). By 2050, water demand for agriculture could be double and economic losses in crop production due to water scarcity could be higher owing to climate change (Gleick, 2000; Tang, 2019; Gupta et al., 2020).

Alfalfa, the "Queen of Forages", has a capacity to produce high yields and high-quality forage. The dry matter yield of alfalfa ranges from 12 to 19 t ha^{-1} (Nešić et al., 2005). On the other hand, alfalfa is a rich in protein, with over 20% crude protein content in alfalfa hay (Long et al., 2022). Furthermore, alfalfa is well recognized for its high concentration of macroelements (such as N, P, and Ca), microelements (such as Fe, Zn, Cu, and Mn), and vitamins (such as A, E, and K), all of which are beneficial to animal health (Radović et al., 2009). All those characteristics confirm that alfalfa has a superior value in feeding animals. However, alfalfa production and yield stability are severely affected by drought (Ashrafi et al., 2018; Ines et al., 2022). It has been reported that severe drought causes severe economic losses (Nadeem et al., 2019). Slama et al. (2011) studied eight cultivars of *Medicago sativa* and found that the biomass production was reduced by 55–75% for plants subjected to a water deficit condition (Slama et al., 2011). As a result, developing alfalfa cultivars with improved water use efficiency (WUE) and drought tolerance is critical for sustainable alfalfa production in water-limited areas.

Traditional breeding programs to develop new alfalfa varieties are time-consuming and costly. Incorporating marker-assisted selection (MAS) into breeding programs holds one of the promises to meet the demand for alfalfa production. Although, MAS has been widely used in several important crops, such as maize (Samayoa et al., 2019), wheat (Kumar et al., 2018a), and soybean (Kim et al., 2020). It is rarely employed for the commercial development of improved alfalfa varieties. Identifying and developing genetic loci robustly associated with alfalfa drought tolerance is the first step to developing drought resistant varieties by MAS. Quantitative trait loci (QTL) mapping and genome-wide association studies (GWAS) have been used to identify QTLs that influence complex quantitative traits, such as drought resistance. To date, some QTLs/SNPs for drought resistance have been detected by linkage or association mapping. For instance, Ray et al. (2015) identified 10 and 15 QTL associated with increased or reduced forage yield during drought stress in two backcross (BC1) mapping populations. Santantonio et al. (2019) performed linkage mapping of forage yield, WUE, carbon and nitrogen metabolism in the same populations under drought conditions (Santantonio et al., 2019). A diversity panel of 198 alfalfa accessions was used to identify SNPs associated with drought tolerance using association mapping (Zhang et al., 2015; Yu, 2017; Lin et al., 2020). In a greenhouse, Zhang et al. (2015) identified 19 and 15 loci associated with drought resistance index and relative leaf water content, respectively. In the field, SNPs associated with biomass yield and 26 Feed Quality-related traits under water deficit have been identified in the same panel (Yu, 2017; Lin et al., 2020).

Functional genes that contribute to drought tolerance were identified using homology-based cloning, including *MsMYB2L* (Song et al., 2019), *MsZIP* (Li et al., 2013), *MsHSP17.7* (Li et al., 2016), *MsZEP* (Zhang et al., 2016), *MsHSP70* (Li et al., 2017b), *MsCML46* (Du et al., 2021), *MsVDAC* (Yang et al., 2021b), *MsWRKY11* (Wen et al., 2021). However, little is known about genetic factors that contribute to drought tolerance in alfalfa. In the present study, we evaluated biomass yield, plant height, and 10 Feed Quality traits in an F1 mapping population under water deficit. Linkage mapping and RNA-sequencing (RNA-seq) analyses were performed to identify the QTLs associated with drought tolerance, and candidate genes by screening the differentially expressed genes (DEGs) within the QTL intervals. The detected QTLs are valuable resources for alfalfa genetic improvement by MAS, and further investigation of candidate genes can provide insights into genetic factors of alfalfa resistance to drought and other abiotic stresses.

Materials and methods

Mapping population development, genotyping, and genetic linkage maps

The F1 population consisting of 150 progenies was used for evaluating and identifying loci associated with drought tolerance. Mapping population development, genotyping, and construction of genetic linkage maps were provided in our previous study (Jiang et al., 2022). Briefly, the mapping population was developed by crossing two tetraploid alfalfa plants, Cangzhou (CF000735, paternal parent, P1) and Zhongmu NO.1 (CF0032020, maternal parent, P2). The two parents vary in drought tolerance capacity with P1 better than P2. In greenhouse, both parents were imposed to water stress at 40% field capacity for 4 weeks, while control pots maintained 100% field capacity throughout the period of regrowth cycle. The biomass yield of plants under drought stress decreased significantly by 27.4% and 47.3%, as compared to control plants for P1 and P2, respectively. According to the method of Li et al. (2014), alleles segregating with a ratio of less than 2:1 were considered as single-dose allele (SDA) markers. A total of 7,252 and 7,404 high-quality SDA markers were obtained for P1

and P2 parents, respectively; which were then used to construct high-density linkage maps by JoinMap 4.0 software (Van Ooijen, 2006).

Field experiments and phenotyping

In 2016, cloned plants of the F1 population and two parents were transplanted to the field of the Chinese Academy of Agricultural Sciences Research Station at Langfang, Hebei Province (39.59°N, 116.59°E). The field experiment used a randomized complete block design with three replications, where one cloned plant of 150 progenies and two parents were planted in each replication. At the field site, the rainfall during the three alfalfa first growth cycles of 2018, 2019, and 2020 was approximately 24.7, 35.6, and 77.8mm, which was far less than the water requirement for normal growth of alfalfa (Cole et al., 1970; Dobrenz et al., 1971). Drought treatment was applied to the plants by withholding water before the first cut. Thus, three forage regrowth cycles (the first cut of 2018, 2019 and 2020) experienced significant water stress. In the early flowering stage, we measured biomass yield (BY), plant height (PH), and ten quality-related traits, including: the content of crude protein (CP), neutral detergent fiber (NDF), acid detergent fiber (ADF), lignin, dry matter (DM), ASH, K, Ca, Mg, and P. Plant height was the length of the longest stem, and biomass was the plant's fresh weight when the stubble height is 4 ~ 5 cm. The BY and PH of individual plants were measured when the first flower appeared. The Feed Quality was measured using near-infrared reflectance spectroscopy (Foss NIRS 1650), and the details were described by Yang et al. (2021a). Statistical analysis of 12 traits was estimated by the R package psych. The best linear unbiased estimation (BLUE) and broad-sense heritability (H^2) of the 12 traits were estimated by the ANOV function in the IciMapping software (Meng et al., 2015).

Linkage mapping

Combining BLUE values of each trait and the genetic linkage maps, QTL in response to water deficit were identified in the mapping population using the Inclusive Composite Interval Mapping with an additive effect (ICIM-ADD) in QTL IciMapping software (Meng et al., 2015). The QTL with a LOD value ≥ 3.0 was selected as significant QTL. QTLs were named as: q + phenotype + linkage group no., or qFT + linkage group no. + an ordered number to designate QTL in a single linkage group. For example, *qBY6.4-1* indicates the first QTL associated with biomass yield under a water deficit condition in the Chr6.4 linkage group.

RNA-seq analysis

In brief, RNA-seq sequences were filtered using fastp (Chen et al., 2018), and mapped to the XingJiangDaYe reference genome (Chen et al., 2020b) using hisat2 (Kim et al., 2019). Samtools was used to generate and sort BAM files (Li et al., 2009). FeatureCounts v2.0.1 was used to generate read counts for each sample (Liao et al., 2014). The FPKM (Fragments Per Kilobase Million) was utilized to normalize and estimate gene expression values. The $|\log2\ (\text{FoldChange})| \geq 2$ and $P \leq 0.01$ were used as thresholds to assess the significance of gene expression difference. TBtools was used to do a Gene Ontology (GO) enrichment analysis of the differentially expressed genes. (Chen et al., 2020a). Three RNA-sequencing datasets were used for the discovery of differentially expressed candidate genes associated with drought tolerance within the QTL region. The datasets were submitted to NCBI and signed for project numbers as follow: The first dataset, PRJNA525327 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA525327/), was an RNA-seq transcriptome profiling of two alfalfa genotypes (drought-sensitive and drought-resistant) root tissue under PEG-induced drought stress. The second one, PRJNA765383 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA765383/), was an RNA-seq of two genotypes (drought-sensitive and drought-resistant) of alfalfa leaf under drought conditions. The third RNA-seq project (PRJNA450305, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA450305/) was alfalfa seedlings (Zhongmu No.1) treated with 400 mM mannitol under different treatment time points (0, 12, and 24 h).

Prediction of candidate genes

The flanking markers of main QTLs, which explained more than 10% of the phenotypic variance, were used to obtain the physical location of QTLs on the XingJiangDaYe reference genome. The genes located on the flanking markers and within the physical intervals were extracted for further analysis. The first two RAN-seq datasets resulted in four comparative groups (G1~G4). The extracted genes, which were differentially expressed in two or more comparative groups, were used to further identify if they were differentially expressed in different treatment time points in the third RNA-seq project. If a candidate is identified by combining the information on linkage mapping and RNA-seq analysis, then, it will be annotated based on BLSAT search in Ensembl (https://ensembl.gramene.org/).

Quantitative real-time PCR (qRT-PCR) was performed to confirm whether the expression of candidate genes was induced by drought stress. Alfalfa seeds (Zhongmu No.1) were germinated in the MS medium. After seven days germination,

the seedlings with similar growth were transferred to a hydroponic pot filled with 1/2 MS nutrient solution (pH = 5.8). Ten days after transfer, PEG-6000 (20%) was used to simulate drought treatment. Total RNA for each sample was isolated from three palnts under a time course of drought treatments (0, 1, 3, 6, 12, and 24 h). Using Ms-actin gene as the reference gene, qRT-PCR was implemented in triplicate for each treatment with the 7500 Real-Time PCR System (Applied Biosystem, CA, USA). The relative gene expression level was calculated by the 2-△△Ct method.

Results

Phenotypic data analysis

The BLUE values of the 12 traits including yield, plant height and 10 quality-related traits were used for statistical analysis. Two parents appeared to have substantial variations in these measured traits except for ADF and DM (Table S1). For example, the BY of the P2 parent was significantly higher than that of the P1 parent. A frequency distribution histogram based on the F1 population revealed a nearly normal distribution for 12 traits, indicating that the traits measured were quantitative traits (Figure S1). As shown in Figure S2, a positive correlation was observed between BY and PH in the F1 population with a correlation coefficient of 0.69 (Figure S2), indicating that the genotypes with higher PH tended to have higher yields. For quality related traits, CP was positively correlated with lignin (0.34, $P < 0.001$), Ca (0.48, $P < 0.001$), P (0.74, $P < 0.001$), and K (0.65, $P < 0.001$), while negatively correlated with Mg (-0.27, $P < 0.01$), ADF (−0.44, $P < 0.001$), and NDF (-0.37, $P < 0.001$). For lignin, it was positively correlated with ADF (0.55, $P < 0.001$) and NDF (0.57, $P < 0.001$) (Figure S2). The correlations we observed were consistent with those in the previous reports (Yang et al., 2021a).

The H^2 of the 12 traits in the F1 population have a high variation ranging from 0.28 to 0.61 (Table S1). BY and PH had higher heritability of 0.50 and 0.61, respectively, suggesting that high proportion of the variability of these traits were contributed by genetic factor. Among the 10 Feed Quality-related traits, DM and ADF had the lowest heritability, at 0.28 and 0.32, respectively, indicating that a high proportion of the variability in these traits came from environmental factors, with a smaller contribution from genetic differences.

QTL analysis

Under water deficit, BY, PH, and ten quality-related traits of the alfalfa population were evaluated during three consecutive years from 2018 to 2020. Using a threshold of LOD value higher than 3, a total of 48 QTLs were identified with phenotypic variance explanations ranging from 2.93% to 16.03% (Table 1). For two yield related traits, BY and PH, we identified 11 QTLs in two parents with six for BY and five for PH. We detected 37 QTLs related to ten quality traits with 8, 1, 3, 7, 1, 9, 3, and 5 for content of CP, lignin, NDF, ASH, Ca, P, K, and Mg, respectively, while, no QTL was detected for ADF and DM (Table 1).

Among these significant QTLs, nine showed more than 10% of the phenotypic variation explained (PVE) individually, with four in P1 and five in P2 (Table 1; Figure 1). PVE by four main QTLs, qPH7.3, qCP3.3, qASH5.4, and qP4.2, ranged from 10.70 to 16.03% in the P1 parent, with the highest PVE on qPH7.3 and the lowest PVE on qASH5.4 (Table 1; Figure 1A). In the P2 parent, five main QTLs of *qBY6.4-2*, *qCP3.4*, *qP7.1-2*, *qK8.2*, and *qMg4.3* explained 10.13, 10.77, 10.72, 14.00, and 13.43% of the phenotypic variation, respectively (Table 1; Figure 1B). For BY, one main QTL, *qBY6.4-2*, was located on Chr6.4, while one PH-related main QTL, *qPH7.3*, was located on Chr7.3 with a LOD score of 7.03. For ten Feed Quality traits, seven main QTL were identified, including: two for CP, one for ash, two for P, one for K, and one for Mg. In contrast, no main QTL was detected for lignin, ADF, NDF, DM, and Ca (Table 1).

Among the nine main QTLs, three QTLs from P1 and three from P2 were also independently identified in single environment (Table S2). They were considered as consistent QTLs. Specifically, *qCP3.3*, *qCP3.4*, and *qK8.2* were also identified in 2019, 2020, and 2020, respectively. *qBY6.4-2* and *qPH7.3* were detected in two years. The last one, *qP4.2*, was identified in three grown environments (2018, 2019, and 2020) at a position of 81.5 ~ 82.5 cM.

Analysis of DEGs

In order to elucidate the drought-induced transcripts in alfalfa, we performed transcriptome analysis in leaf and root. Four groups (G1 ~ G4) were constructed by comparing the same genotype of root and leaf tissues under different conditions (control and drought) (Figure 2A; Tables S3 , S4). In root, the number of DEGs (drought vs. control) was approximately equal in the two comparative groups. There were 2,600 genes that were up-regulated in G1, while 2,671 genes were up-regulated in G2. Meanwhile, 5,524 and 4,897 down-regulated genes were identified in G1 and G2, respectively (Figure 2A). In contrast, compared to the 1,606 up- and 2,438 down-regulated DEGs identified in the G4, 5,357 up- and 4,109 down-regulated DEGs were identified in G3, indicating that the drought responsive transcriptomic program in leaf had bigger difference than those in root in different alfalfa genotypes (Figure 2A). In drought-resistant genotypes, drought condition upregulated 1,606 and 2,671 genes, and downregulated 2,438 and 4,897 genes in the leaf (G4) and root (G2) respectively. Not surprisingly, drought conditions resulted in more common DEGs in root, with 4,986 in root and 1,973 in leaf (Figure 2B).

10.3389/fpls.2022.996672

TABLE 1 QTLs detected for 12 traits using ICIM-ADD in the F1 population.

Parent	QTL	LG	Position/cM	LeftMarker	RightMarker	LOD	PVE (%)
P1	*qBY1.1*	chr1.1	8.5~9.5	chr1.1:1188379	chr1.1:25288609	6.95	6.62
P1	*qBY1.4*	chr1.4	45.5~46.5	chr1.4:32618614	chr1.4:38011048	4.70	5.78
P1	*qBY2.1*	chr2.1	91.5~92.5	chr2.1:40634643	chr2.1:51966907	3.68	2.93
P1	*qBY3.3*	chr3.3	34.5~35.5	chr3.3:39725093	chr3.3:51651121	6.60	6.22
P1	*qBY6.4-1*	chr6.4	67.5~68.5	chr6.4:19754452	chr6.4:38317952	6.01	8.85
P2	***qBY6.4-2***	chr6.4	9.5~10.5	chr6.4:11679540	chr6.4:13292372	3.16	10.13
P1	*qPH6.1*	chr6.1	68.5~69.5	chr6.1:62112875	chr6.1:59533246	3.83	9.03
P1	***qPH7.3***	chr7.3	83.5~84.5	chr7.3:63690328	chr7.3:63690347	7.30	16.03
P2	*qPH3.1*	chr3.1	56.5~57.5	chr3.1:35511599	chr3.1:24556189	3.63	8.16
P2	*qPH5.4*	chr5.4	46.5~47.5	chr5.4:15100953	chr5.4:24519386	3.73	7.38
P2	*qPH7.1*	chr7.1	96.5~97.5	chr7.1:70529293	chr7.1:62826189	4.26	8.83
P1	*qCP1.4-1*	chr1.4	4.5~5.5	chr1.4:4781072	chr1.4:18320659	6.73	6.61
P1	*qCP1.4-2*	chr1.4	56.5~57.5	chr1.4:40757995	chr1.4:68168405	4.48	7.30
P1	*qCP2.3*	chr2.3	28.5~29.5	chr2.3:25484901	chr2.3:25484816	4.33	4.37
P1	***qCP3.3***	chr3.3	77.5~78.5	chr3.3:82448563	chr3.3:85819420	8.73	11.13
P1	*qCP4.2*	chr4.2	81.5~82.5	chr4.2:83498816	chr4.2:86046636	6.25	6.47
P1	*qCP5.3*	chr5.3	66.5~67.5	chr5.3:47283990	chr5.3:56958651	6.29	7.40
P1	*qCP6.3*	chr6.3	45.5~46.5	chr6.3:24707110	chr6.3:13214112	3.61	5.88
P2	***qCP3.4***	chr3.4	78.5~79.5	chr3.4:51603347	chr3.4:51603328	4.53	10.77
P1	*qlignin6.4*	chr6.4	38.5~39.5	chr6.4:13303817	chr6.4:15215403	4.63	4.08
P2	*qNDF5.1*	chr5.1	46.5~47.5	chr5.1:32987072	chr5.1:48353467	4.62	6.04
P2	*qNDF6.4*	chr6.4	105.5~111.5	chr6.4:61366096	chr6.4:60044123	3.43	5.96
P2	*qNDF8.4*	chr8.4	53.5~54.5	chr8.4:54287851	chr8.4:58806497	5.26	9.14
P1	***qASH5.4***	chr5.4	61.5~62.5	chr5.4:42741251	chr5.4:37142807	3.79	10.70
P1	*qASH8.1*	chr8.1	65.5~66.5	chr8.1:57462848	chr8.1:64832251	3.43	6.12
P2	*qASH1.2*	chr1.2	33.5~34.5	chr1.2:29902008	chr1.2:30572221	3.97	4.00
P2	*qASH1.4*	chr1.4	117.5~118.5	chr1.4:66910982	chr1.4:82180569	8.63	9.27
P2	*qASH4.4*	chr4.4	89.5~90.5	chr4.4:66438303	chr4.4:60823700	7.07	7.33
P2	*qASH5.3*	chr5.3	46.5~48.5	chr5.3:9887846	chr5.3:12938236	3.39	4.24
P2	*qASH8.2*	chr8.2	94.5~95.5	chr8.2:76204355	chr8.2:83554323	6.27	5.86
P1	*qCa4.3*	chr4.3	4.5~7.5	chr4.3:405639	chr4.3:3983154	3.53	9.29
P1	*qP2.4*	chr2.4	59.5~60.5	chr2.4:35238369	chr2.4:33943447	3.87	7.38
P1	*qP4.1*	chr4.1	22.5~23.5	chr4.1:35458256	chr4.1:28488544	4.55	5.71
P1	***qP4.2***	chr4.2	81.5~82.5	chr4.2:83498816	chr4.2:86046636	10.45	14.01
P1	*qP5.2*	chr5.2	56.5~57.5	chr5.2:70405584	chr5.2:79796969	4.42	6.07
P1	*qP7.4*	chr7.4	64.5~65.5	chr7.4:56305064	chr7.4:74538855	3.29	5.38
P2	*qP2.2*	chr2.2	91.5~92.5	chr2.2:46344535	chr2.2:68020962	3.31	4.80
P2	*qP3.4*	chr3.4	78.5~79.5	chr3.4:51603347	chr3.4:51603328	5.66	9.32
P2	*qP7.1-1*	chr7.1	5.5~7.5	chr7.1:15149480	chr7.1:15158721	3.95	7.53
P2	***qP7.1-2***	chr7.1	34.5~35.5	chr7.1:18573847	chr7.1:21230294	5.55	10.72
P2	*qK1.1*	chr1.1	28.5~29.5	chr1.1:29260583	chr1.1:26628171	5.32	8.08
P2	*qK5.3*	chr5.3	148.5~152.5	chr5.3:79503684	chr5.3:79503581	4.49	5.83
P2	***qK8.2***	chr8.2	18.5~19.5	chr8.2:28815850	chr8.2:26672846	8.01	14.00
P2	*qMg1.1*	chr1.1	45.5~46.5	chr1.1:53165176	chr1.1:46575248	3.11	3.52
P2	*qMg1.4*	chr1.4	14.5~17.5	chr1.4:4062805	chr1.4:4058912	4.17	2.98
P2	*qMg3.4*	chr3.4	85.5~86.5	chr3.4:68267075	chr3.4:73873783	7.09	6.98
P2	***qMg4.3***	chr4.3	118.5~119.5	chr4.3:85323781	chr4.3:80402482	12.24	13.43
P2	*qMg7.3*	chr7.3	13.5~14.5	chr7.3:24844583	chr7.3:4955198	5.68	4.89

A total of 48 QTLs were mapped by best linear unbiased estimation (BLUE) values on P1 and P2 linkage maps. Main QTLs (PVE > 10%) were bold. LG, linkage group; Position/cM, 1-LOD support interval; Leftmarker, the marker on the left of the LOD peak; Rightmarker, the marker on the right of the LOD peak; LOD, the logarithm of the odds; PVE, the percentage of the phenotypic variation explained by QTL.

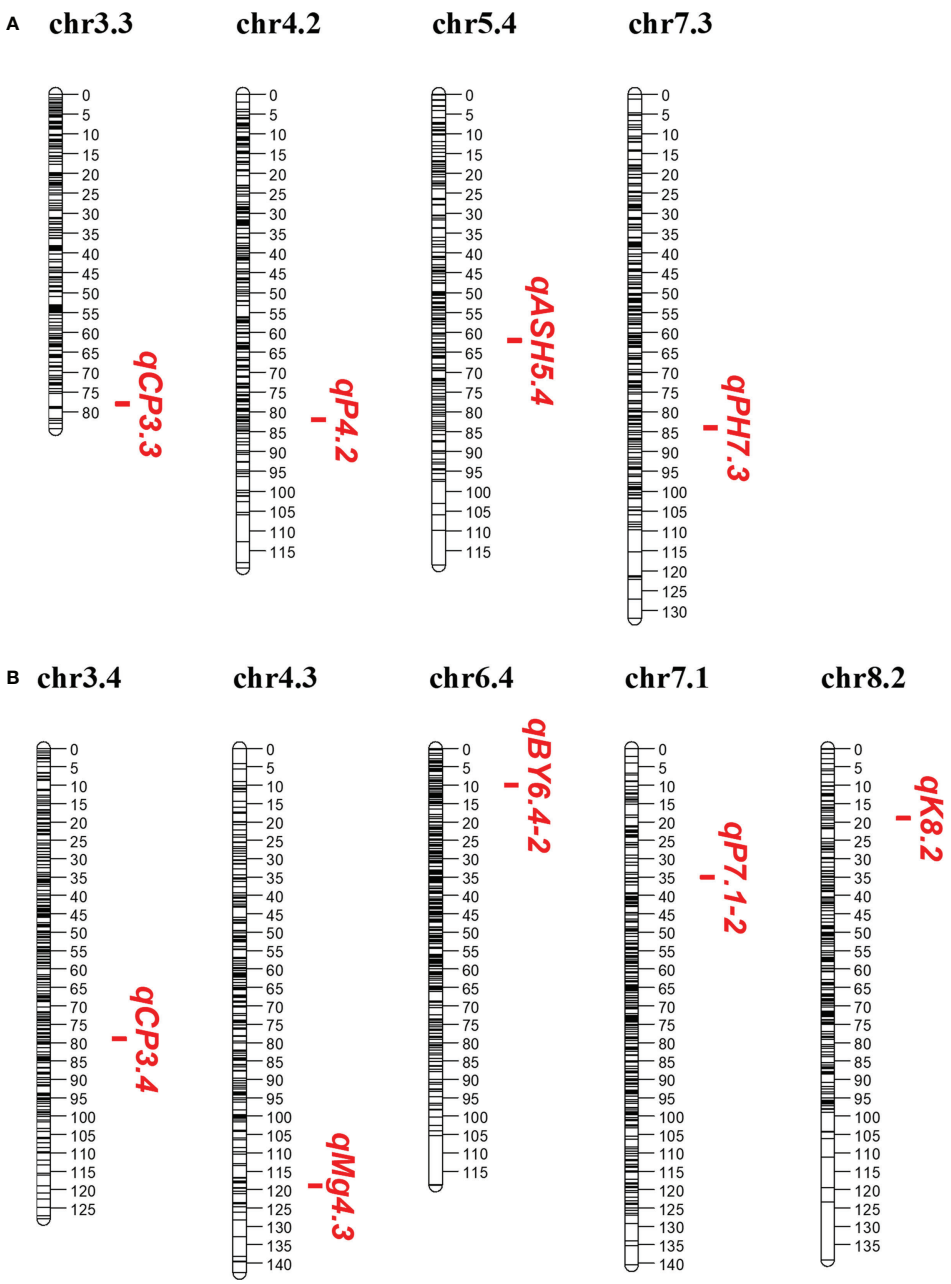


FIGURE 1
Main QTLs (PVE > 10%) detected in this study. **(A, B)** Main QTLs in the P1 and P2 parent, respectively.

Through GO enrichment analysis, the common DEGs in alfalfa root and leaf were then classified into the biological process (BP), cellular component (CC), and molecular function (MF) (Figures 2C, D; Tables S5 , S6). The common DEGs in root were enriched in the BP terms "response to acid chemical", "monocarboxylic acid biosynthetic process", "monocarboxylic acid biosynthetic process", "oxoacid metabolic process", "organic acid metabolic process", "organic hydroxy compound biosynthetic process" and so on (Figure 2C; Table S5). These GO terms were also enriched in leaf (Table S6), indicating that these common genes were possibly involved in responding to drought stress. Subsequent GO enrichment analysis of 274 common DEGs in all four comparative groups resulted that these DEGs had a role in "response to water deprivation", "response to osmotic stress" and "response to abscisic acid" and other responses (Figure 3, Table S7). Based on their functions, these genes likely play crucial roles in drought tolerance.

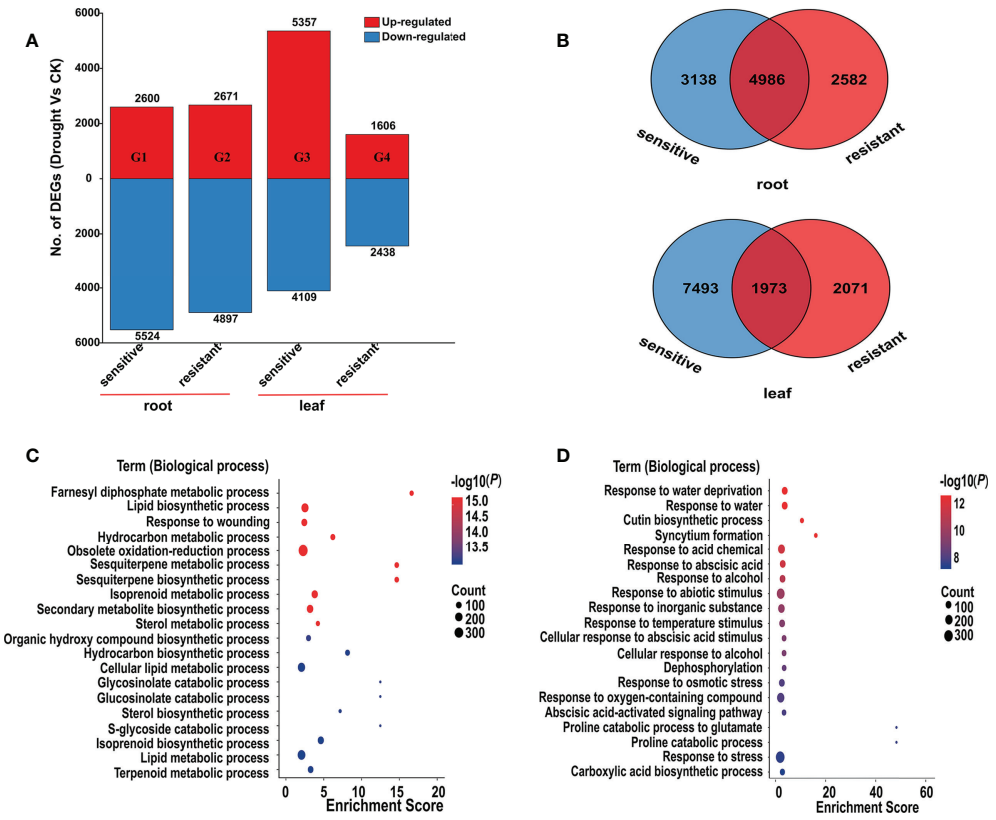


FIGURE 2
Differentially expressed genes (DEGs) in root and leaf. **(A)**, Number of upregulated genes (red) and downregulated genes (blue) in four comparative groups (G1 ~ G4). You can see the specific information of these four comparative groups in the Table S3. **(B)**, Venn diagram of DEGs among root and leaf. **(C, D)**, Bubble chart of GO enrichment analysis of the DEGs in root **(C)** and leaf **(D)**. The ordinate represents different GO terms (biological progress) and the abscissa represents enrichment Score. Circle size represents the gene number while circle color represents the value of –log10 (*P*). The *P* value was corrected by Benjamini & Hochberg (BH) method.

Integration of DEGs with QTLs

Linkage mapping revealed nine major QTL (PVE > 10%), which were selected as genetic regions to identify candidate genes responsive to drought tolerance. We first located their physical positions on the XingJiangDaYe reference genome. For instance, two physical intervals of *qPH7.3* and *qCP3.4* were less than 1Mb. We therefore screened candidate genes within 1cM upstream and downstream of these two QTL flanking markers. In this way, a total of 1,455 genes were extracted from the nine main QTL intervals (Table 2). Among them, three genes (*MS.gene013861*, *MS.gene065379*, and *MS.gene034021*) were located on flanking markers. *chr4.3:85323781*, the left flanking markers of *qMg4.3*, was lined with *MS.gene034021*, which was differentially expressed in G1 and G2 (Table S8).

To detect the candidate genes within the nine intervals, we combined the linkage mapping and RNA-seq analysis results. For G1, a total of 82 DEGs located in the nine main QTL regions, and the number of DEGs within each QTL ranged from 2 (*qBY6.4-2*/*qPH7.3*) to 24 (*qCP3.3*) (Table 2; Figure 4A). Also, 83, 80, and 37 DEGs located in the nine main QTL regions for G2, G3, and G4, respectively (Table 2; Figures 4A–D). Additionally, there were 52 common DEGs in roots (G1 and G2) and 22 common DEGs located in the nine QTL regions in leaves (G3 and G4) (Table 2; Figures 4E, F). Interestingly, the abovementioned DEGs in leaves had different expression patterns. For example, *MS.gene08151* was down-regulated in drought condition in a drought-resistant genotype, while it was up-regulated in a drought-sensitive genotype (Figure 4F). However, the same phenomenon was not observed in root.

To obtain more evidence to determine the reliable candidate genes, we further analysis the expression pattern of these genes in drought conditions at different time points (0h, 12h, and 24h). Among them, 31 candidates within QTL regions were differentially expressed at different time points (Figure S3). Notably, five candidate genes were differentially expressed across four comparative groups and at different treatment time, of which two (*MS.gene068413* and *MS.gene068423*) were

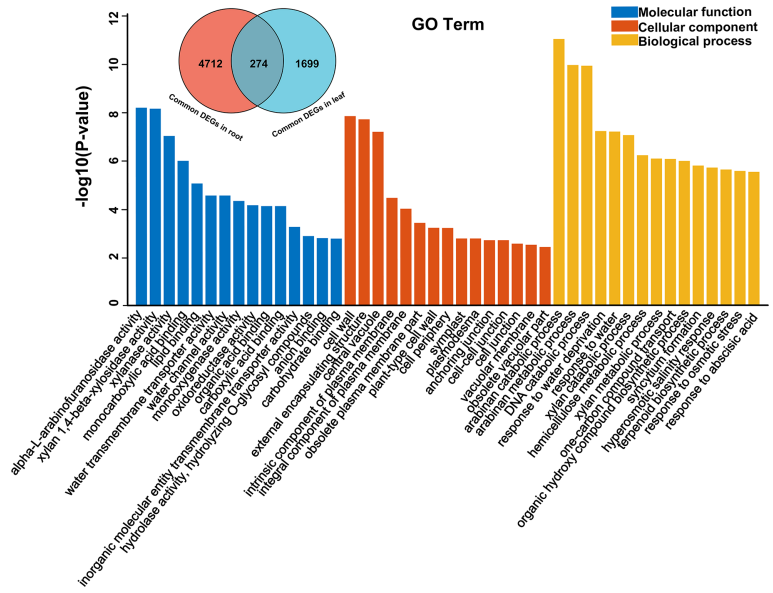


FIGURE 3
GO enrichment bar chart of 274 common DEGs among root and leaf.

differentially expressed in *qP4.2*, two (*MS.gene037560* and *MS.gene037580*) in *qMg4.3*, and one (*MS.gene027542*) in *qASH5.4* (Table 2; Figures 4A–D).

Furthermore, functional annotation of the 31 identified genes revealed that 22 of them may be involved in drought stress regulation (Table 3). We investigated the expression level of all 22 candidate genes by qRT-PCR analysis, and found that all candidate genes were significantly induced under drought stress compared with untreated control (Figure 5). The results revealed that 13 candidate genes were consistent with the transcriptome results under drought stress with eight up-regulated genes (*MS.gene012838*, *MS.gene38684*, *MS.gene012839*, *MS.gene068423*, *MS.gene068413*, *MS.gene037560*, *MS.gene037580*, *MS.gene067823*) and five down-regulated genes (*MS.gene73180*, *MS.gene33495*, *MS.gene33688*, *MS.gene33704*, *MS.gene97835*) (Table S8; Figure 5). The expression patterns of the rest nine genes were different between the RNA-seq and qPT-PCR. For example, the expression level of *MS.gene99452* was down-regulated under

TABLE 2 Number of genes, DEGs within nine main QTL (PVE > 10%) regions.

QTL	Genes	DEGs						
		Root			Leaf			[c]Common
		G1	G2	[a]Common	G3	G4	[b]Common	
qCP3.3	280	24	21	15	15	6	3	0
qCP3.4	97	3	4	3	5	2	1	0
qP4.2	138	8	14	8	8	2	2	2
qMg4.3	273	15	16	9	19	12	8	2
qASH5.4	149	10	6	5	11	5	4	1
qBY6.4-2	60	2	4	1	4	1	0	0
qP7.1-2	223	11	11	6	19	7	3	0
qPH7.3	75	2	3	2	3	0	0	0
qK8.2	160	7	4	3	12	2	1	0
Total	1455	82	83	52	96	37	22	5

[a]Common DEGs among G1 and G2.
[b]Common DEGs among G3 and G4.
[c]Common DEGs among four comparative groups (G1 ~ G4).

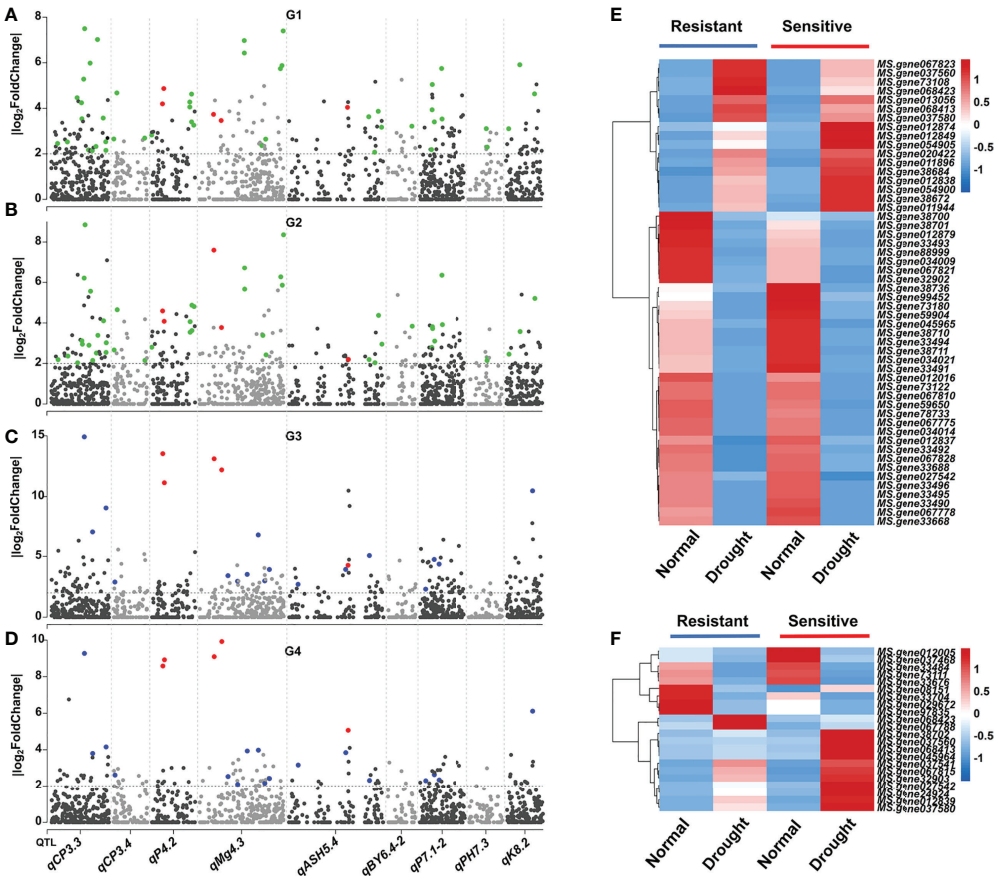


FIGURE 4
Combined transcriptome analysis and linkage mapping revealing the potential causal genes for drought response in alfalfa. **(A–D)** Distribution of DEGs within nine main QTL regions in four comparative groups, respectively. Green, blue, and red colors represent common DEGs among root, leaf, and four comparative groups, respectively. **(E)**, Heatmap of 52 common DEGs among root. **(F)** Heatmap of 22 common DEGs among leaf.

drought stress in RNA-seq (Table S8). However, it was up-regulated in qRT-PCR with the highest expression level at 6 h after drought treatment (Figure 5). Differences in expression patterns may be due to differences in drought treatment conditions and the inherent genetic variation between the materials used for different experiments. These candidate genes identified in our study should be considered as putative candidates, further investigation should be performed.

Discussion

Analyzing the correlation coefficient in alfalfa forage yield and quality during drought stress can provide useful insight into phenotypic relationships between these traits in water-limited environments. To this end, we collected phenotypic data of two yield-related and 10 quality-related traits in the F1 mapping population for three years through drought-stressed growth cycle. P content was positively correlated with BY (0.34) and CP content (0.74), suggesting that applying P fertilizer may be beneficial to yield and CP content under drought conditions. For CP, it was negatively correlated (-0.37) with NDF, while the two traits had no significant correlation in a previous study (Yang et al., 2021a). Our results provided useful information to explore the relationship between alfalfa yield and quality during drought stress.

Achieving a high yield and high nutritional value is a main goal in alfalfa breeding. Forage yield and quality are easily influenced by the environment, such as drought stress. Previous studies suggested that biomass yield and Feed Quality under drought conditions may involve different mechanisms compared to non-stress conditions (Yu, 2017; Lin et al., 2020). It is therefore crucial to identify QTL associated with forage yield and quality under water deficit for genetic gain of alfalfa. In this study, we collected 12 traits with the heritability ranging from 0.28 to 0.61 and then performed linkage mapping. However, 48 QTL were detected for only ten traits. There were no QTL mapped for ADF and DM, probably due to the low heritability

TABLE 3 Potential candidate genes identified in this study.

QTL	GeneID	Position				BLAST-P	E-value	%ID
		Chr.	Start_Pos	End_Pos	Stand	protein_coding description		
qCP3.3	MS.gene012838	chr3.3	85656764	85658555	+	homeobox associated leucine zipper protein	4.6E-50	91.0
qCP3.3	MS.gene38684	chr3.3	84651967	84654396	+	Serine/Threonine kinase, plant-type protein	0	96.2
qCP3.3	*MS.gene38710*	chr3.3	84266981	84269413	+	ABA response element-binding factor	3.1E-114	99.5
qCP3.3	*MS.gene012839*	chr3.3	85634143	85642942	–	branched-chain amino acid aminotransferase	3.3E-126	97.9
qCP3.4	*MS.gene73180*	chr3.4	52454644	52454955	+	Lipid transfer protein	1.3E-51	94.9
qP4.2	*MS.gene068423*	chr4.2	84224298	84225247	–	phosphatidylethanolamine-binding protein	3.5E-111	95.3
qP4.2	*MS.gene068413*	chr4.2	84132136	84133091	–	phosphatidylethanolamine-binding protein	8.6E-113	96.4
qP4.2	*MS.gene99452*	chr4.2	83500593	83501671	+	PAR1 protein	3E-126	95.4
qP4.2	*MS.gene33495*	chr4.2	85748885	85750285	+	glycoside hydrolase family 1 protein	8.1E-119	95.5
qP4.2	*MS.gene33494*	chr4.2	85849862	85850722	+	glycoside hydrolase family 1 protein	2.4E-31	87.5
qP4.2	*MS.gene33492*	chr4.2	85977904	85980624	+	glycoside hydrolase family 1 protein	9.6E-144	91.3
qMg4.3	*MS.gene33688*	chr4.3	84072885	84075965	–	LRR receptor-like kinase family protein	0	95.5
qMg4.3	*MS.gene33668*	chr4.3	84295620	84297870	–	WRKY transcription factor	1.2E-154	97.8
qMg4.3	*MS.gene037560*	chr4.3	81683850	81684764	–	phosphatidylethanolamine-binding protein	5.2E-112	95.9
qMg4.3	*MS.gene037580*	chr4.3	81240143	81241108	–	UPF0098 protein CPn_0877	2E-118	95.8
qMg4.3	*MS.gene33704*	chr4.3	83826342	83826998	+	abscisic acid receptor	1.2E-123	95.3
qASH5.4	*MS.gene027542*	chr5.4	40643863	40645210	+	DUF1262 family protein	0	77.3
qASH5.4	*MS.gene029672*	chr5.4	37713550	37715701	–	myo-inositol oxygenase	0	97.8
qASH5.4	*MS.gene97835*	chr5.4	40503630	40505071	–	DUF538 family protein	1.2E-21	97.6
qP7.1-2	*MS.gene067823*	chr7.1	19308600	19310009	–	spermidine hydroxycinnamoyl transferase	0	93.2
qP7.1-2	*MS.gene08151*	chr7.1	18898187	18900692	–	plasma membrane H+-ATPase	4E-39	78.8
qP7.1-2	*MS.gene067788*	chr7.1	19687484	19688272	+	DUF4283 domain protein	1.4E-44	63.5

and no significant differences were observed between the two traits in the two parents. MAS has been one of the most efficient breeding methods due to the advantages of reducing the time and labor required in field tests (Collard and Mackill, 2008; Kumar et al., 2018b). Moreover, quantitative genetics theory predicted that the efficiency of MAS was inversely related to the heritability of the traits (Knapp, 1998; Oladosu et al., 2019). Water scarcity tolerance was a complex trait controlled by polygenes with low to medium heritability (Elena, 2022). Hence, utilizing the QTLs identified in this study, particularly those quality related QTLs of low to moderate heritability, can significantly reduce the time and resources required for breeding efforts.

Transcriptome analysis is an effective method for the identification of differentially expressed stress-responsive genes (Sathik et al., 2018; Liu et al., 2020). Ascertaining the DEGs that change upon water stress in alfalfa is of crucial in understanding the genetic base of drought tolerance. In alfalfa, although drought-responsive DEGs have been reported in previous studies (Ma et al., 2021; Singer et al., 2021), there were some limitations due to the lack of a reference genome. The release of a high-quality alfalfa reference genome has given new power to transcriptional analysis (Chen et al., 2020b). In this study, we aligned the sequencing data to the XingJiangDaYe reference genome and finally obtained a total of 274 DEGs through root and leaf. Through GO enrichment analysis, the most significant enrichment terms were "arabinan catabolic process" and "arabinan metabolic process", which played a role in maintaining the flexibility of the plant cell wall during water deficit (Moore et al., 2008a; Moore et al., 2008b). In addition, terms related to cell wall biosynthesis such as "xylan catabolic process", "hemicellulose metabolic process", and "xylan metabolic process" were also enriched in these DEGs, indicating that changes in the expression of cell wall-related genes were a vital and integral component of alfalfa's response to water stress. Understanding how the wall adapts to loss of water should provide new insight into crop improvement (Lenk et al., 2019; Ganie and Ahammed, 2021).

Alfalfa is a self-incompatible plant, and the traditional QTL fine mapping and map-based cloning are limited by constructing near-isogenic lines. It has been reported that integration of conventional linkage mapping and RNA-seq can rapidly identify candidates associated with complex traits of interest to replace the fine-mapping process (Park et al., 2019; Derakhshani et al., 2020). In this study, we finally identified 22 drought responsive genes by integrating QTL and DEGs. Plants have evolved complex networks of drought stress response. Abscisic acid (ABA) regulates a number of physiological responses in drought-stressed plants, ensuring a

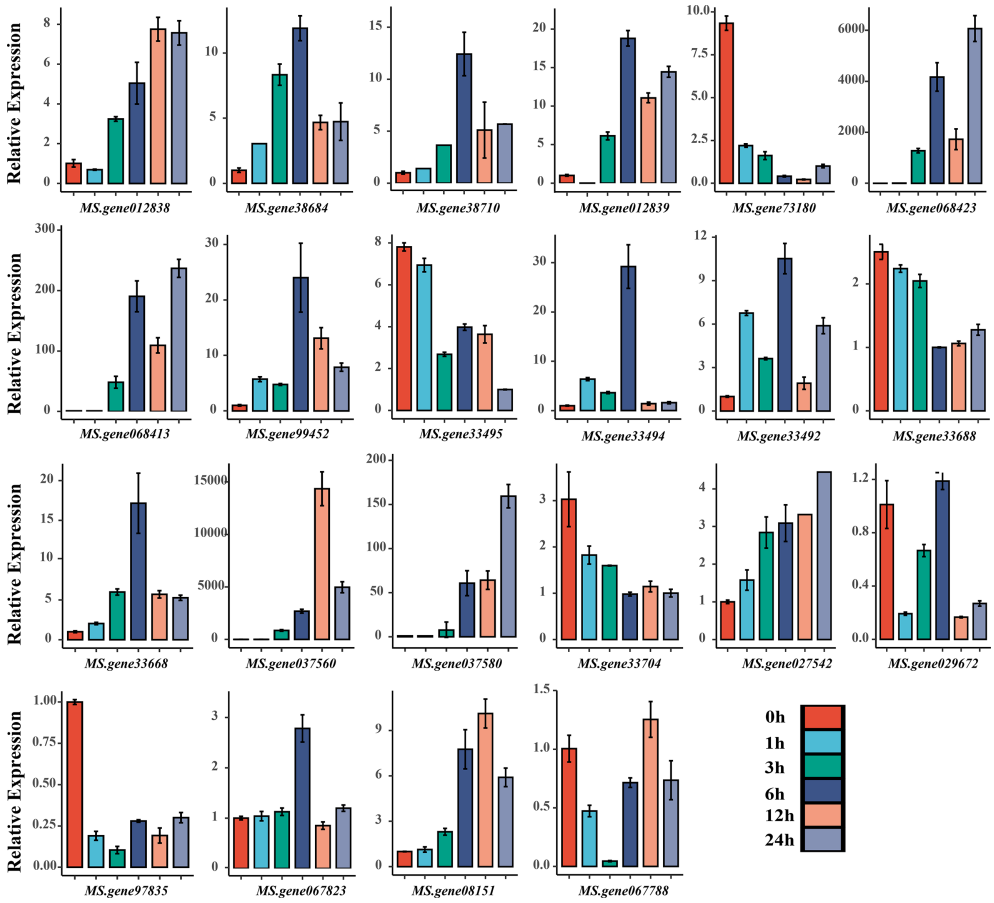


FIGURE 5
qRT-PCR analysis of the 22 candidate genes under drought stress.

balance of optimal development and stress tolerance (Soma et al., 2021). *MS.gene38710* was annotated as an ABA-responsive element binding factor that can be phosphorylated by upstream genes to regulate the expression of drought-responsive genes (Feng et al., 2019; Zhang et al., 2020). In plants, the phosphatidylethanolamine-binding protein (PEBP) family has been identified to have a crucial role in the regulation of plant growth and developmental processes. Meanwhile, omics data suggested that the PEBP family was also involved in the drought stress response (Manoharan et al., 2016; Li et al., 2017a; Schneider et al., 2019). *OsMFT1*, a member of the PEBP family, has been shown to improve rice drought tolerance by interacting with two key drought-related transcription factors, *OsbZIP66* and *OsMYB26* (Chen et al., 2021). In our study, three PEBP family members (*MS.gene068423*, *MS.gene068413*, and *MS.gene037560*) were identified in two main QTLs, suggesting that they may play an important role in alfalfa drought response. In addition, some differential genes located in the main QTL interval should also be focused on in future research. For example, *MS.gene034021* was annotated as Zein-binding protein, which has not been reported to be related to plant drought resistance. However, it was differentially expressed in G1 and G2 with the high $\log_2$FoldChange value of -7.40 and -8.36, respectively. Further molecular investigation is needed to illustrate the function of these genes in drought response.

Conclusion

In the present study, nine main QTLs associated with biomass yield, plant height, and the content of CP, ASH, K, Mg, and P under water deficit condition were identified in the F1 population. The integration of linkage mapping with RNA-seq analysis under water stress, 22 DEGs, excavated from QTL-regions, were identified as potential candidates. The closely linked markers and candidate genes identified in the present study will provide a tool for the MAS breeding program, and new insight for further revealing the molecular mechanism of drought tolerance in alfalfa.

Data availability statement

RNA-seq raw data used in this study was downloaded from the SRA database in NCBI with the Bioproject accession numbers of PRJNA525327, PRJNA765383, and PRJNA450305.

Author contributions

QY, JK and ZW conceived and designed the experiments and developed the mapping population. XJ, TY, CW, and Ting Gao collected phenotypic data. XJ, FZ, and AY performed data analysis. XJ and AY wrote the manuscript. L-XY, ZW, and JK revised and finalized the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (32071868) and the earmarked fund for China Agriculture Research System (CARS-35-04) and the key research project of Ningxia province for alfalfa breeding program (2019NYY203) and Agricultural Science and Technology Innovation Program (ASTIP-IAS14).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.996672/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Frequency distributions of best linear unbiased estimation (BLUE) values of 12 phenotypic data in the F_1 population.

SUPPLEMENTARY FIGURE 2
Phenotypic correlations of 12 traits in F_1 population. Numbers indicate the magnitude of the correlation coefficient. Negatively correlated variables are red, whereas positively correlated variables are blue, with the intensity depending on the magnitude of the correlation. Asterisks indicate the significance level, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

SUPPLEMENTARY FIGURE 3
Hierarchical clustering analysis of DEGs in 0h, 6h, and 24h after drought stress. Red, blue and white elements in the matrix indicate up-regulated, down-regulated, and no change genes, respectively.

SUPPLEMENTARY TABLE 1
Statistical analysis of 12 traits (BLUE values) in two parents and F1 population.

SUPPLEMENTARY TABLE 2
Consistent QTLs detected in this study. The main QTLs, which were independently identified in a single environment, were considered as consistent QTLs.

SUPPLEMENTARY TABLE 3
Information of four comparative groups **(G1-G4)**.

SUPPLEMENTARY TABLE 4
The list of differentially expressed genes (DEGs) in four comparative groups. The DEGs threshold was set at $P \leq 0.01$ and $|\log2 \text{ fold change (FC)}| > 2$.

SUPPLEMENTARY TABLE 5
GO enrichment analysis of the 4,986 common DEGs in root. The 4,986 DEGs were identified in both G1 and G2.

SUPPLEMENTARY TABLE 6
GO enrichment analysis of the 1,973 common DEGs in leaf. The 1,973 DEGs were identified in both G3 and G4.

SUPPLEMENTARY TABLE 7
GO enrichment analysis of the 274 common DEGs among root and leaf. The 274 DEGs were identified in all four comparative groups.

SUPPLEMENTARY TABLE 8
List of the position information, the log2FoldChange value, and the annotation information of 1,455 genes within nine main QTL regions.

References

Ashrafi, M., Azimi-Moqadam, M.-R., Moradi, P., MohseniFard, E., Shekari, F., and Kompany-Zareh, M. (2018). Effect of drought stress on metabolite adjustments in drought tolerant and sensitive thyme. *Plant Physiol. Biochem.* 132, 391–399. doi: 10.1016/j.plaphy.2018.09.009

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

Chen, Y., Shen, J., Zhang, L., Qi, H., Yang, L., Wang, H., et al. (2021). Nuclear translocation of OsMFT1 that is impeded by OsFTIP1 promotes drought tolerance in rice. *Mol. Plant* 14, 1297–1311. doi: 10.1016/j.molp.2021.05.001

Chen, H., Zeng, Y., Yang, Y., Huang, L., Tang, B., Zhang, H., et al. (2020b). Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. *Nat. Commun.* 11, 2494. doi: 10.1038/s41467-020-16338-x

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

Cole, D., Dobrenz, A., Massengale, M., and Wright, L. N. (1970). Water requirement and its association with growth components and protein content of alfalfa (*Medicago sativa* l.) 1. *Crop Sci.* 10, 237–240. doi: 10.2135/cropsci1970.0011183X001000030008x

Collard, B. C., and Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. B: Biol. Sci.* 363, 557–572. doi: 10.1098/rstb.2007.2170

Derakhshani, B., Jafary, H., Maleki Zanjani, B., Hasanpur, K., Mishina, K., Tanaka, T., et al. (2020). Combined QTL mapping and RNA-seq profiling reveals candidate genes associated with cadmium tolerance in barley. *PloS One* 15, e0230820. doi: 10.1371/journal.pone.0230820

Dobrenz, A., Cole, D., and Massengale, M. (1971). Yield components and leaf characteristics associated with the water requirement of alfalfa 1. *Crop Sci.* 11, 124–125. doi: 10.2135/cropsci1971.0011183X001100010045x

Du, B., Chen, N., Song, L., Wang, D., Cai, H., Yao, L., et al. (2021). Alfalfa (*Medicago sativa* l.) MsCML46 gene encoding calmodulin-like protein confers tolerance to abiotic stress in tobacco. *Plant Cell Rep.* 40, 1907–1922. doi: 10.1007/s00299-021-02757-7

Elena, B. (2022). Improving drought tolerance: Can comparative transcriptomics support strategic rice breeding? *Plant Stress* 3, 100058. doi: 10.1016/j.stress.2022.100058

Feng, R. J., Ren, M. Y., Lu, L. F., Peng, M., Guan, X., Zhou, D. B., et al. (2019). Involvement of abscisic acid-responsive element-binding factors in cassava (Manihot esculenta) dehydration stress response. *Sci. Rep.* 9, 12661. doi: 10.1038/s41598-019-49083-3

Ganie, S. A., and Ahammed, G. J. (2021). Dynamics of cell wall structure and related genomic resources for drought tolerance in rice. *Plant Cell Rep.* 40, 437–459. doi: 10.1007/s00299-020-02649-2

Gleick, P. H. (2000). *The world's water 2000-2001: the biennial report on freshwater resources* (Washington, DC: Island Press).

Gupta, A., Rico-Medina, A., and Caño-Delgado, A. I. (2020). The physiology of plant responses to drought. *Science* 368, 266–269. doi: 10.1126/science.aaz7614

Ines, S., Talbi, O., Nasreddine, Y., Rouached, A., Gharred, J., Jdey, A., et al. (2022). Drought tolerance traits in medicago species: A review. *Arid Land Res. Manage.* 36, 67–83. doi: 10.1080/15324982.2021.1936289

Jiang, X., Yang, T., Zhang, F., Yang, X., Yang, C., He, F., et al. (2022). RAD-Seq-Based high-density linkage maps construction and quantitative trait loci mapping of flowering time trait in alfalfa (*Medicago sativa* l.). *Front. Plant Sci.* 13, 899681. doi: 10.3389/fpls.2022.899681

Kim, J.-M., Kim, K.-H., Jung, J., Kang, B. K., Lee, J., Ha, B.-K., et al. (2020). Validation of marker-assisted selection in soybean breeding program for pod shattering resistance. *Euphytica* 216, 1–9. doi: 10.1007/s10681-020-02703-w

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

Knapp, S. J. (1998). Marker-assisted selection as a strategy for increasing the probability of selecting superior genotypes. *Crop Sci.* 38, 1164–1174. doi: 10.2135/cropsci1998.0011183X003800050009x

Kumar, A., Jain, S., Elias, E. M., Ibrahim, M., and Sharma, L. K. (2018a). "An overview of QTL identification and marker-assisted selection for grain protein content in wheat," in *Eco-friendly agro-biological techniques for enhancing crop productivity*, 245–274. (Singapore: Springer). doi: 10.1007/978-981-10-6934-5_11

Kumar, A., Sandhu, N., Dixit, S., Yadav, S., Swamy, B., and Shamsudin, N. A. A. (2018b). Marker-assisted selection strategy to pyramid two or more QTLs for quantitative trait-grain yield under drought. *Rice* 11, 1–16. doi: 10.1186/s12284-018-0227-0

Lenk, I., Fisher, L. H., Vickers, M., Akinyemi, A., Didion, T., Swain, M., et al. (2019). Transcriptional and metabolomic analyses indicate that cell wall properties are associated with drought tolerance in brachypodium distachyon. *Int. J. Mol. Sci.* 20, 1758. doi: 10.3390/ijms20071758

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

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

Li, J. X., Hou, X. J., Zhu, J., Zhou, J. J., Huang, H. B., Yue, J. Q., et al. (2017a). Identification of genes associated with lemon floral transition and flower development during floral inductive water deficits: A hypothetical model. *Front. Plant Sci.* 8, 1013. doi: 10.3389/fpls.2017.01013

Li, Z., Long, R., Zhang, T., Wang, Z., Zhang, F., Yang, Q., et al. (2017b). Molecular cloning and functional analysis of the drought tolerance gene MsHSP70 from alfalfa (*Medicago sativa* l.). *J. Plant Res.* 130, 387–396. doi: 10.1007/s10265-017-0905-9

Li, Z.-y., Long, R.-c., Zhang, T.-j., Yang, Q.-c., and Kang, J.-m. (2016). Molecular cloning and characterization of the MsHSP17. 7 gene from *Medicago sativa* l. *Mol. Biol. Rep.* 43, 815–826. doi: 10.1007/s11033-016-4008-9

Li, X., Wei, Y., Acharya, A., Jiang, Q., Kang, J., and Brummer, E. C. (2014). A saturated genetic linkage map of autotetraploid alfalfa (Medicago sativa l.) developed using genotyping-by-Sequencing is highly syntenous with the medicago truncatula genome. *G3 Genesgenetics* 4, 1971–1979. doi: 10.1534/g3.114.012245

Lin, S., Medina, C. A., Boge, B., Hu, J., Fransen, S., Norberg, S., et al. (2020). Identification of genetic loci associated with forage quality in response to water deficit in autotetraploid alfalfa (*Medicago sativa* l.). *BMC Plant Biol.* 20, 303. doi: 10.1186/s12870-020-02520-2

Li, Y., Sun, Y., Yang, Q., Fang, F., Kang, J., and Zhang, T. (2013). Isolation and characterization of a gene from *Medicago sativa* l., encoding a bZIP transcription factor. *Mol. Biol. Rep.* 40, 1227–1239. doi: 10.1007/s11033-012-2165-z

Liu, J., Deng, J. L., and Tian, Y. (2020). Transcriptome sequencing of the apricot (Prunus armeniaca l.) and identification of differentially expressed genes involved in drought stress. *Phytochemistry* 171, 112226. doi: 10.1016/j.phytochem.2019.112226

Long, R., Zhang, F., Zhang, Z., Li, M., Chen, L., Wang, X., et al. (2022). Genome assembly of alfalfa cultivar zhongmu-4 and identification of SNPs associated with agronomic traits. *Genomics Proteomics Bioinf* 20, 14–28. doi: 10.1016/j.gpb.2022.01.002

Manoharan, R. K., Han, J. S., Vijayakumar, H., Subramani, B., Thamilarasan, S. K., Park, J. I., et al. (2016). Molecular and functional characterization of FLOWERING LOCUS T homologs in allium cepa. *Molecules* 21, 217. doi: 10.3390/molecules21020217

Ma, Q., Xu, X., Wang, W., Zhao, L., Ma, D., and Xie, Y. J. P. P. (2021). Comparative analysis of alfalfa (*Medicago sativa* l.) seedling transcriptomes reveals genotype-specific drought tolerance mechanisms. *Plant Physiol. Biochem.* 166, 203–214. doi: 10.1016/j.plaphy.2021.05.008

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

Moore, J. P., Farrant, J. M., and Driouich, A. (2008a). A role for pectin-associated arabinans in maintaining the flexibility of the plant cell wall during water deficit stress. *Plant Signaling Behav.* 3, 102–104. doi: 10.4161/psb.3.2.4959

Moore, J. P., Vicré-Gibouin, M., Farrant, J. M., and Driouich, A. (2008b). Adaptations of higher plant cell walls to water loss: drought vs desiccation. *Physiologia Plantarum* 134, 237–245. doi: 10.1111/j.1399-3054.2008.01134.x

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, 2541. doi: 10.3390/ijms20102541

Nešić, Z., Tomić, Z., Žujović, M., and Krnjaja, V. (2005). Production characteristics of domestic alfalfa (*Medicago sativa* l.) cultivars in agroecological conditions of srem district. *Biotechnol. Anim. Husbandry* 21, 169–174. doi: 10.2298/BAH0502169N

Oladosu, Y., Rafii, M. Y., Samuel, C., Fatai, A., Magaji, U., Kareem, I., et al. (2019). Drought resistance in rice from conventional to molecular breeding: A review. *Int. J. Mol. Sci.* 20, 3519. doi: 10.3390/ijms20143519

Park, M., Lee, J.-H., Han, K., Jang, S., Han, J., Lim, J.-H., et al. (2019). A major QTL and candidate genes for capsaicinoid biosynthesis in the pericarp of capsicum chinense revealed using QTL-seq and RNA-seq. *Theor. Appl. Genet.* 132, 515–529. doi: 10.1007/s00122-018-3238-8

Radović, J., Sokolović, D., and Marković, J. (2009). ALFALFA-MOST IMPORTANT PERENNIAL FORAGE LEGUME IN ANIMAL HUSBANDRY. *Biotechnol. Anim. Husbandry* 25, 465–475. doi: 10.2298/BAH0906465R

Ray, I. M., Han Y, E. L., Meenach, C. D., Santantonio, N., Sledge, M. K., Pierce, C. A., et al. (2015). Identification of quantitative trait loci for alfalfa forage biomass productivity during drought stress. *Crop Sci.* 55, 2012–2033. doi: 10.2135/cropsci2014.12.0840

Samayoa, L. F., Butrón, A., and Malvar, R. A. (2019). Usefulness of marker-assisted selection to improve maize for increased resistance to sesamia nonagrioides attack with no detrimental effect on yield. *Ann. Appl. Biol.* 174, 219–222. doi: 10.1111/aab.12480

Santantonio, N., Pierce, C. A., Steiner, R. L., and Ray, I. M. (2019). Genetic mapping of water use efficiency and carbon and nitrogen metabolism in drought-stressed alfalfa. *Crop Sci.* 59, 92–106. doi: 10.2135/cropsci2018.05.0307

Sathik, M., Luke, L. P., Rajamani, A., Kuruvilla, L., Sumesh, K., and Thomas, M. (2018). *De novo* transcriptome analysis of abiotic stress-responsive transcripts of hevea brasiliensis. *Mol. Breed.* 38, 1–17. doi: 10.1007/s11032-018-0782-5

Schneider, S., Turetschek, R., Wedeking, R., Wimmer, M. A., and Wienkoop, S. (2019). A protein-linger strategy keeps the plant on-hold after rehydration of drought-stressed beta vulgaris. *Front. Plant Sci.* 10, 381. doi: 10.3389/fpls.2019.00381

Singer, S. D., Subedi, U., Lehmann, M., Burton Hughes, K., Feyissa, B. A., Hannoufa, A., et al. (2021). Identification of differential drought response mechanisms in medicago sativa subsp. sativa and falcata through comparative assessments at the physiological, biochemical, and transcriptional levels. *Plants* 10, 2107. doi: 10.3390/plants10102107

Slama, I., Tayachi, S., Jdey, A., Rouached, A., and Abdelly, C. (2011). Differential response to water deficit stress in alfalfa (*Medicago sativa*) cultivars: Growth, water relations, osmolyte accumulation and lipid peroxidation. *Afr. J. Biotechnol.* 10, 16250–16259. doi: 10.5897/AJB11.1202

Soma, F., Takahashi, F., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2021). Cellular phosphorylation signaling and gene expression in drought stress responses: ABA-dependent and ABA-independent regulatory systems. *Plants (Basel Switzerland)* 10, 756. doi: 10.3390/plants10040756

Song, Y., Lv, J., Qiu, N., Bai, Y., Yang, N., and Dong, W. (2019). The constitutive expression of alfalfa MsMYB2L enhances salinity and drought tolerance of arabidopsis thaliana. *Plant Physiol. Biochem.* 141, 300–305. doi: 10.1016/j.plaphy.2019.06.007

Tang, QJSCES (2019). Global change hydrology: Terrestrial water cycle and global change. *Science China. Earth Sciences* 63, 459–462. doi: 10.1007/s11430-019-9559-9

Van Ooijen, J. (2006). *JoinMap® 4, software for the calculation of genetic linkage maps in experimental populations* Vol. 33 (Wageningen: Kyazma BV).

Wen, W., Wang, R., Su, L., Lv, A., Zhou, P., and An, Y. (2021). MsWRKY11, activated by MsWRKY22, functions in drought tolerance and modulates lignin biosynthesis in alfalfa (*Medicago sativa* l.). *Environ. Exp. Bot.* 184, 104373. doi: 10.1016/j.envexpbot.2021.104373

Yang, C., Zhang, F., Jiang, X., Yang, X., He, F., Wang, Z., et al. (2021a). Identification of genetic loci associated with crude protein content and fiber composition in alfalfa (*Medicago sativa* l.) using QTL mapping. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.608940

Yang, M., Duan, X., Wang, Z., Yin, H., Zang, J., Zhu, K., et al. (2021b). Overexpression of a voltage-dependent anion-selective channel (VDAC) protein-encoding gene, MsVDAC, from *Medicago sativa* confers cold and drought tolerance to transgenic tobacco. *Genes* 12, 1706. doi: 10.3390/genes12111706

Yu, L. X. (2017). Identification of single-nucleotide polymorphic loci associated with biomass yield under water deficit in alfalfa (*Medicago sativa* l.) using genome-wide sequencing and association mapping. *Front. Plant Sci.* 8, 1152. doi: 10.3389/fpls.2017.01152

Zhang, H., Liu, D., Yang, B., Liu, W. Z., Mu, B., Song, H., et al. (2020). Arabidopsis CPK6 positively regulates ABA signaling and drought tolerance through phosphorylating ABA-responsive element-binding factors. *J. Exp. Bot.* 71, 188–203. doi: 10.1093/jxb/erz432

Zhang, Z., Wang, Y., Chang, L., Zhang, T., An, J., Liu, Y., et al. (2016). MsZEP, a novel zeaxanthin epoxidase gene from alfalfa (*Medicago sativa*), confers drought and salt tolerance in transgenic tobacco. *Plant Cell Rep.* 35, 439–453. doi: 10.1007/s00299-015-1895-5

Zhang, T., Yu, L. X., Zheng, P., Li, Y., Rivera, M., Main, D., et al. (2015). Identification of loci associated with drought resistance traits in heterozygous autotetraploid alfalfa (*Medicago sativa* l.) using genome-wide association studies with genotyping by sequencing. *PloS One* 10, e0138931. doi: 10.1371/journal.pone.0138931

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 30 November 2022
DOI 10.3389/fpls.2022.995031

OPEN ACCESS

EDITED BY
Jorge Fernando Pereira,
Embrapa Gado de Leite, Brazil

REVIEWED BY
Nisha Singh,
Gujarat Biotechnology University, India
Poliana Fernanda Giachetto,
EMBRAPA Agricultural Informatics, Brazil

*CORRESPONDENCE
Ke Jin
jinke@caas.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

RECEIVED 15 July 2022
ACCEPTED 09 September 2022
PUBLISHED 30 November 2022

CITATION
Liu Y, Fan W, Cheng Q, Zhang L, Cai T, Shi Q, Wang Z, Chang C, Yin Q, Jiang X and Jin K (2022) Multi-omics analyses reveal new insights into nutritional quality changes of alfalfa leaves during the flowering period. *Front. Plant Sci.* 13:995031. doi: 10.3389/fpls.2022.995031

Multi-omics analyses reveal new insights into nutritional quality changes of alfalfa leaves during the flowering period

Yinghao Liu[1†], Wenqiang Fan[2†], Qiming Cheng[3], Lianyi Zhang[4], Ting Cai[4], Quan Shi[4], Zuo Wang[4], Chun Chang[1], Qiang Yin[1], Xiaowei Jiang[1] and Ke Jin[1*]

[1]Key Laboratory of Forage Cultivation, Processing and High Efficient Utilization of Ministry of Agriculture and Rural Affairs, Institute of Grassland Research of Chinese Academy of Agricultural Sciences, Hohhot, China, [2]Key Laboratory of Grassland Resources, Ministry of Education, Inner Mongolia Agricultural University, Hohhot, China, [3]College of Animal Science, Guizhou University, Guiyang, China, [4]Key Laboratory of Efficient Utilization of Forage, Inner Mongolia Agricultural and Animal Husbandry Technology Extension Center, Hohhot, China

High-quality alfalfa is an indispensable resource for animal husbandry and sustainable development. Its nutritional quality changes dramatically during its life cycle and, at present, no molecular mechanisms for nutrient metabolic variation in alfalfa leaves at different growth stages have been clearly reported. We have used correlation and network analyses of the alfalfa leaf metabolome, proteome, and transcriptome to explore chlorophyll, flavonoid, and amino acid content at two development stages: budding stage (BS) and full-bloom stage (FBS). A high correlation between the expression of biosynthetic genes and their metabolites revealed significant reductions in metabolite content as the plant matured from BS to FBS. L-Glutamate, the first molecule of chlorophyll biosynthesis, decreased, and the expression of *HemA*, which controls the transformation of glutamyl-tRNA to glutamate 1-semialdehyde, was down-regulated, leading to a reduction in leaf chlorophyll content. Flavonoids also decreased, driven at least in part by increased expression of the gene encoding CYP75B1: flavonoid 3′-monooxygenase, which catalyzes the hydroxylation of dihydroflavonols and flavonols, resulting in degradation of flavonoids. Expression of *NITRILASE 2 (NIT2)* and *Methyltransferase B* (*metB)*, which regulate amino acid metabolism and influence the expression of genes of the glycolysis-TCA pathway, were down-regulated, causing amino acid content in alfalfa leaves to decrease at FBS. This study provides new insights into the complex regulatory network governing the content and decrease of chlorophyll, amino acids, flavonoids, and other nutrients in alfalfa leaves during maturation. These results further provide a theoretical basis for the generation of alfalfa varieties exhibiting higher nutritional quality, high-yield cultivation, and a timely harvest.

KEYWORDS

alfalfa leaves, metabolomics, proteomics, transcriptomics, chlorophyll, amino acids, flavonoids

Introduction

Alfalfa (*Medicago sativa* L.) is a high-yield perennial forage legume with excellent nutrition, digestibility, and palatability. It is a good source of feed for high-yielding dairy cows as it is rich in protein, amino acids, chlorophyll, flavonoids, and other nutritious substances (Popovic et al., 2001). Metabolic processes occurring in the alfalfa leaf, such as photosynthesis, respiration, and transpiration, are closely related to plant growth, nutrition, yield, quality, and resistance (Dong et al., 2021). Therefore, study of alfalfa leaves at different growth stages may provide information to ensure high yield and nutrient quality.

To date, only changes in alfalfa leaf protein metabolism have been studied at different developmental stages (Fan et al., 2018); there are few reports on how chlorophyll, amino acid, and flavonoid metabolism change with development. Chlorophyll is indispensable for absorption and transduction of light energy during photosynthesis (Gao et al., 2018). Chlorophyll biosynthesis is a tightly regulated process, since its derivatives can produce highly toxic compounds upon illumination if not bound to specific proteins, such as HEMA1 and chlorophyll synthase ChlG (Stenbaek and Jensen, 1995). Overexpression in tobacco of HEMA1, which exhibits glutamyl-tRNA reductase activity, has been shown to increase chlorophyll content and the rate of photosynthesis (Zhang et al., 2010; Schmied et al., 2011). ChlG regulates the chlorophyll biosynthetic pathway by modulating stable assembly of chlorophyll-binding proteins and other thylakoid membrane components (Shalygo et al., 2009). Mutations in genes involved in the chlorophyll synthetic pathway can cause alterations in chlorophyll content and leaf color. For example, in rice, mutation in the gene encoding protoporphyrin IX methyltransferase, which catalyzes formation of protoporphyrin IX mono-methyl ester, results a yellow-green plant leaf phenotype (Wang et al., 2017).

Flavonoids are widely distributed plant secondary metabolites. These polyphenolic compounds are synthesized *via* the phenylpropanoid pathway and classified into ten major group: flavones, isoflavones, flavanones, flavonols, dihydroflavonols, flavan-3-ols (F3Os), anthocyanins, anthocyanidins, leucoanthocyanidins (flavan-3,4-diols), and polymeric proanthocyanidins (PAs) (Dantas et al., 2020). The most common alfalfa flavonoids are glycosides of the flavone aglycones apigenin, luteolin, tricin, and chrysoeriol (Goawska et al., 2010). In animals, the flavonoids in alfalfa mainly have biological activities such as anti-oxidation, anti-aging, anti-tumor and enhancing body immunity (Jing et al., 2015). Thus, alfalfa flavonoids can significantly improve the production performance and antioxidant capacity in animals, such as cattle (Zhan et al., 2017) and broiler chickens (Ouyang et al., 2016).

A third group of important metabolites are amino acids, which are the main circulating form of plant nitrogen. Amino acid metabolism through the tricarboxylic acid (TCA) cycle generates energy and essential elements for protein synthesis, carbohydrate metabolism, endogenous hormone regulation, and energy storage (Less and Galili, 2009). Although legumes have a high content of organic nitrogen, they, like other plants, contains low levels of some essential amino acids — lysine (Lys), methionine (Met), threonine (Thr), and tryptophan (Trp) — constraining their nutritional quality for animals, including humans (Galili et al., 2016). Cysteine and methionine can also be limiting, particularly for ruminants; in cattle, methionine increases both yield and protein content in milk (Bird and Moir, 1972; Tong et al., 2014).

Plant leaf nutrient absorption is one of the important processes in nutrient utilization and circulation. Previous studies have shown that leaf nutrients are affected by temporal and spatial environmental factors (temperature, precipitation, habitat, etc.), and their responses often show different patterns within and between species. For some species (e.g., alfalfa), the uptake and utilization of nutrients vary with growth (growth period, stubble, age change), leading to changes in plant nutrient reuptake. So questions arise, such as what is the performance of nutrient reuptake in the leaves of perennial forage legumes at different growth stages? And what is its mechanism of regulation?

To date, few studies have linked gene and protein expression with molecular mechanisms that regulate metabolite accumulation and nutritional quality in alfalfa leaves during the flowering period. Here, we have combined transcriptomics, proteomics, and broad metabolomics analyses to explore accumulation of primary and secondary metabolites in alfalfa leaves at different developmental stages, focusing on chlorophyll, flavonoid, and amino acid metabolism. Our results will provide new insights and avenues to improve alfalfa leaf nutritional quality, particularly during flowering.

Materials and methods

Plant material and sampling

Alfalfa plants (*Medicago sativa* L. var. WL319HQ) were planted in 2019 and assessed during the 2020 growing season in the experimental field of the Inner Mongolia Agricultural University (Hohhot, Inner Mongolia, China). Leaf samples were collected at two different development stages, namely the budding stage (BS, lower buds appeared in >80% of the branches) and full-bloom stage (FBS, 80% flowering). On June 1, 2020 (BS), 60 plants at a uniform growth stage were selected. Leaves from half of these plants (30) were collected and mixed to generate the BS samples. The remaining 30 plants were allowed to grow naturally for 2 weeks (15 June), when leaves were harvested to generate the FBS samples. Leaves from 10 plants were combined at each stage to generate the three biological replicates used for subsequent analyses.

Leaf sampling occurred in the same way for BS and FBS stages. All leaves from the above ground part of the plant were harvested and mixed to provide an accurate assessment of whole-plant nutritional content as delivered for animal forage. Leaves were harvested at 6pm, when light intensity and temperature are relatively low; at this time, plant metabolism is relatively slow, so this time was chosen to minimize individual plant differences in metabolism. Samples were immediately flash-frozen in liquid nitrogen and stored at -80°C until analysis.

Chemical composition analysis

Amino acid content was determined according to the national standard of the People's Republic of China, GB/T 18246-2000 "Determination of Amino acids in Feed" (Dong et al., 2018). Flavonoid content was determined by colorimetry (Guo et al., 2020); and chlorophyll content by spectrophotometry (Harmut, 1987).

Untargeted metabolomics analysis

Freeze-dried samples were homogenized into powder. Metabolites were extracted from approximately 80 mg of sample in 1.2 mL 70% (w/v) aqueous methanol overnight at 4°C. After centrifugation to remove undissolved residue, the supernatant was absorbed and filtered, and stored at -80°C before the analysis. Quality control (QC) samples were prepared by pooling 10 µl of each sample.

UPLC-MS/MS was performed as described by Wang et al. (2019), with a minor modification that the gradient was kept at 40% B for only 2.9 min before final increase to 85% and column re-calibration. to monitor 10 candidate ions per cycle.

Qualitative analysis was performed by searching a self-compiled database (Shanghai Applied Protein Technology Co., Ltd.). Significantly different expressed metabolites (DEMs) were determined by Variable Importance in Projection (VIP) ≥1 and fold-change >1.5 or <0.67 (Wang et al., 2019).

Tandem mass tag-based proteomic analysis

The freeze-dried samples were ground into a fine powder. Sample lysis, protein extraction, and filter-aided sample preparation (FASP Digestion) were performed as described previously (Wisniewski et al., 2009).

Samples (100 µg peptide) were labeled using a tandem mass tag (TMT) Isobaric Label Reagent Set (Thermo Fisher Scientific Inc., Waltham, MA, USA). A Pierce high pH reversed-phase fractionation kit (Thermo Fisher Scientific Inc., Waltham, MA, USA) was used to fractionate TMT-labeled digest samples into nine fractions according to manufacturer's instructions. LC-MS/MS analysis were performed as described by Ren et al. (2020).

The MS raw data for each sample were searched, identified, and quantified using the MASCOT engine (Matrix Science, London, UK; version 2.2) embedded into Proteome Discoverer 1.4 software. Differentially expressed proteins (DEPs) were selected using fold-change >1.2 or <0.83 and P-value < 0.05 (Jing et al., 2020).

Transcriptomic analysis

RNA was extracted, and the RNA-sequencing (RNA-seq) library constructed and sequenced as described by Xu et al. (2022). Six libraries were analyzed (three biological replicates at two developmental stages), using the alfalfa genome from: https://figshare.com/articles/dataset/genome_fasta_sequence_and_annotation_files/12327602. Differentially expressed genes (DEGs) were selected with DESeq2 software using p-adjust<0.05 and $|\log_2$ fold change$|\geq 1$ as described by Xu et al., 2022.

Bioinformatics analysis

Bioinformatic analyses of DEGs, DEPs and DEMs were performed with Python programming. At p-adjust<0.05, it was considered that Gene Ontology (GO, http://www.geneontology.org/) gene function and KEGG pathways (http://www.genome.jp/kegg/) were significantly enriched. DEGs in the protein–protein interaction (PPI) network were predicted by STRING (https://string-db.org/), and the drawn network diagram and network node degree were analyzed by Cytoscape 3.5.1. Transcription factors (TFs) were identified using PlantTFDB 4.0 (http://planttfdb.cbi.pku.edu.cn/).

Statistical analysis

Chemical composition data were analyzed by one-way ANOVA.

For metabolomics analyses, multivariate statistical analysis, SIMCA-P (version 14.1, Umetrics, Umea, Sweden) was employed. After Pareto scaling, principal component analysis (PCA) and orthogonal partial least-squares discriminant analyses (OPLS-DA) were performed. The variable importance in projection (VIP) values were generated from the OPLS-DA model.

Analysis of GO and KEGG enrichment was performed using Fisher's exact test and Benjamini- Hochberg correction for multiple testing was further applied to adjust derived p-values.

The coexpression network analyses was screened by Pearson correlation analysis.

Further comparisons were performed as described in figure legends. All statistical analyses were plotted as described by Wang et al. (2021).

Results

Physiological analysis

The chlorophyll, amino acid, and flavonoid content of alfalfa leaves was determined and quantified at different developmental stages (Figure 1). For all analyzed compounds, the content was significantly higher ($P < 0.05$) in the budding stage (BS) than in the full-bloom stage (FBS). The total chlorophyll content was 12.57 mg/g at BS and 8.29 mg/g at FBS, respectively (Figure 1A). Consistently, the contents of Chl_a and Chl_b at BS (7.45 mg/g and 4.91 mg/g, respectively) were higher than at FBS (2.82 and 1.75, respectively) (Figures 1B, C). The total amino acid (TAA) content in alfalfa leaves was 23.51% at BS, higher than the 20.49% at FBS (Figure 1D). When looking at the essential amino acid (EAA, Figure 1E) and nonessential amino acid (NEAA, Figure 1F) content, BS leaves contained 8.7% EAA and 13.3% NEAA compared with 6.6% EAA and 12.6% NEAA in FBS leaves. Flavonoid analysis showed that the total flavonoid (Figure 1G) content of alfalfa leaves was with 3.27 mg/g at BS, higher than 2.19 mg/g at FBS. This trend was also observed for specific flavonoids, including apigenin (Figure 1H, 1.5 mg/g *vs.* 0.46 mg/g) and luteolin (Figure 1I, 0.85mg/g *vs.* 0.27 mg/g).

Untargeted metabolomic analysis

To explore how leaf nutritional quality changes during alfalfa development, the metabolomic composition of BS and FBS leaves was compared (Supplementary Figure 1A). Repeatability and reliability of the MS data was verified *via* QC sample test curves (Supplementary Figure 1B). 56 metabolites clearly separated into two groups (BS and FBS) by Principal Component Analysis (PCA, Supplementary Figure 1C), and further delineated using OPLS-DA to identify differentially abundant metabolites (Supplementary Figure 1D). That R^2 Y and Q^2 Y scores were both ≥0.8 confirms the correctness of the differences identified in the two growth stages in alfalfa leaf metabolism.

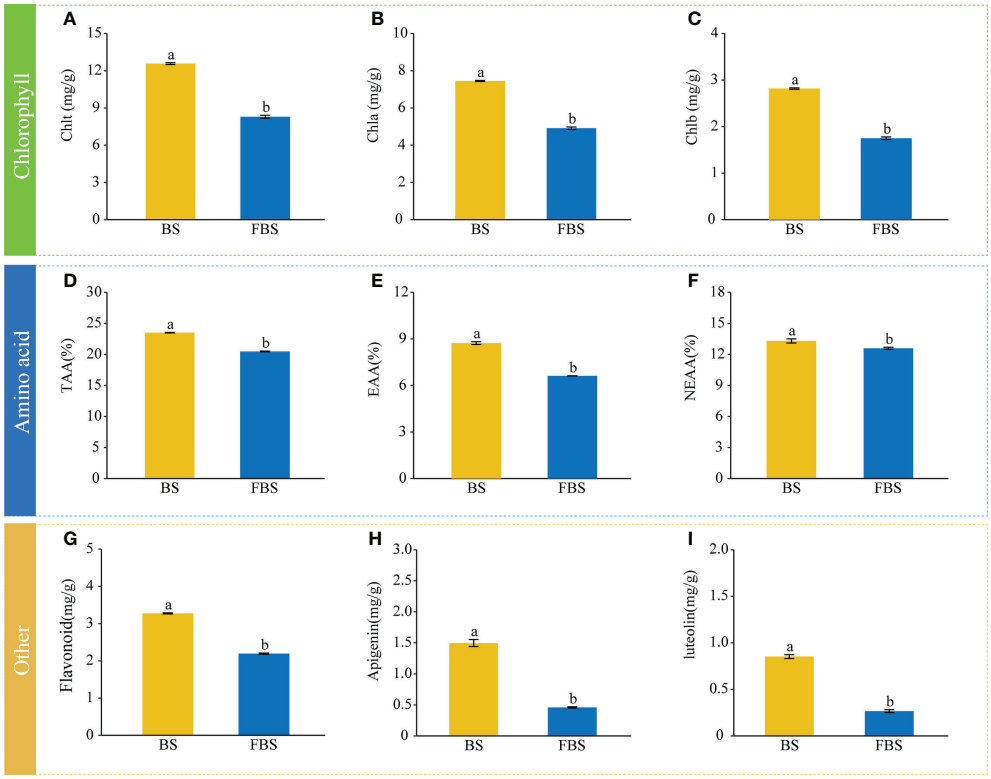


FIGURE 1
Abundances of metabolites involved in chlorophyll, amino acid, and flavonoid biosynthesis during budding stage (BS) and full-bloom stage (FBS). **(A)** total chlorophyll Chl_t, **(B)** chlorophyll a Chl_a, **(C)** chlorophyll b Chl_b, **(D)** total amino acids TAA, **(E)** essential amino acids EAA, **(F)** non-essential amino acids NEAA, **(G)** flavonoids, **(H)** apigenin, and **(I)** luteolin. Letters indicate significant differences ($P<0.05$, Duncan's method).

Metabolite abundance was quantified based on mass spectrum signal intensity (Figure 2A). At FBS, the signal intensity of Chl, amino acids, and flavonoids decreased by 23.46%, 12.36%, and 14.29%, respectively, compared with levels during BS; and 24 and 25 upregulated and downregulated metabolites, respectively, were identified (Figure 2B). Differentially expressed metabolites (DEMs) were used for KEGG enrichment analysis. In Figure 2C, the left semi-circle shows DEMs, while the right semi-circle shows enriched pathways, with lines linking DEMs with pathways. DEMs were enriched in *Metabolic pathways* and *Biosynthesis of secondary metabolites*, especially pathways involving L-glutamate, L-glutamine, L-tryptophan, and L-phenylalanine (Figure 2C).

TMT-based proteomic analysis

The quality and quantity of plant metabolites depends on the presence of the metabolizing proteins, so we analyzed differential protein expression between the two developmental stages. A total of 337 proteins was identified (Figure 3A); 38 proteins showed significant changes in expression from BS to FBS, including 15 upregulated and 23 downregulated proteins (Figure 3B).

The proteomics KEGG analysis for metabolism and signaling pathways is presented in Figure 3C, and the number of proteins per pathway in Figure 3D. In total, 10 pathways were upregulated and six down-regulated. Of the upregulated pathways, the amino sugar and nucleotide sugar metabolism pathway was the most enriched. In line with the metabolite analysis, photosynthesis and flavonoid metabolism pathways were downregulated. To further characterize the differential protein expression, a GO analysis was performed. Enrichment could be found in 15 categories of biological processes (BP), nine in molecular functions (MF), and six in cellular components (CC) (Figure 3E).

Transcriptomic analysis

After filtering low-quality reads, 245,922,094 clean reads were obtained from our six sequenced libraries (Supplementary Table 1). Clean reads from each sample were aligned to the alfalfa reference genome, and 96,267 expressed genes were detected.

Our transcriptomic analysis revealed that 7,537 genes were upregulated, and 6,843 genes were downregulated at FBS compared to BS (Figure 4A). To further analyze the function of proteins encoded by DEGs, a GO analysis was performed, and 15 categories were enriched in biological processes (BP), nine in molecular functions (MF), and six in cellular components (CC) (Figure 4B). Similar to the metabolomic and proteomic analyses,

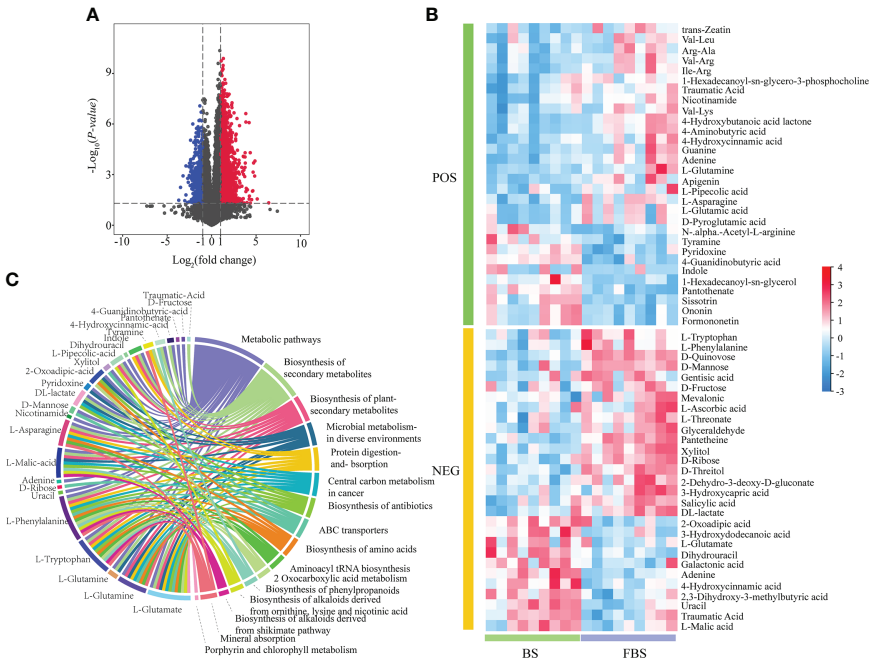


FIGURE 2
(A) Volcano map of differentially expressed metabolites. **(B)** Heat map of differentially expressed metabolites. **(C)** Circle diagram of GO enrichment analysis of differentially expressed metabolites.

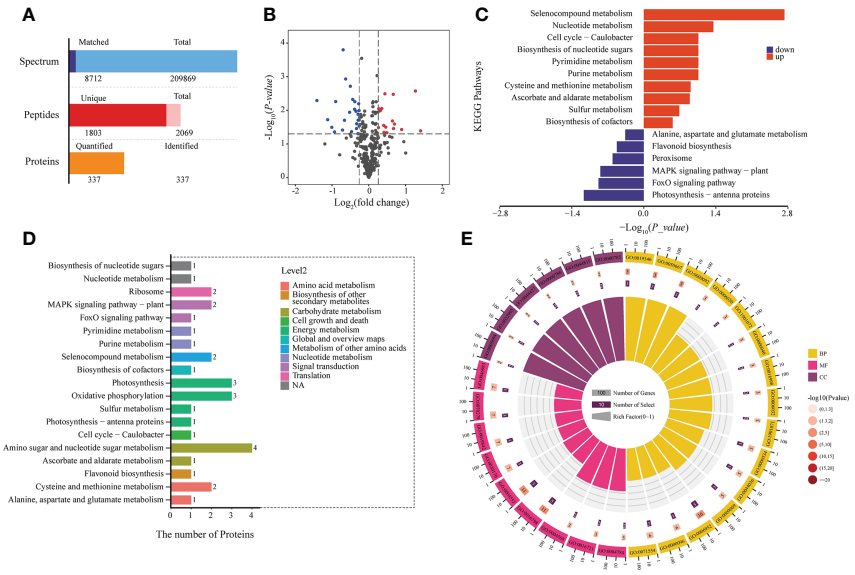


FIGURE 3
(A) Histogram of protein identification and quantitative results. **(B)** Volcano map of differentially expressed proteins. **(C)** KEGG enrichment of up-regulated proteins. **(D)** KEGG results of all differentially expressed proteins. **(E)** GO enrichment of all differentially expressed proteins.

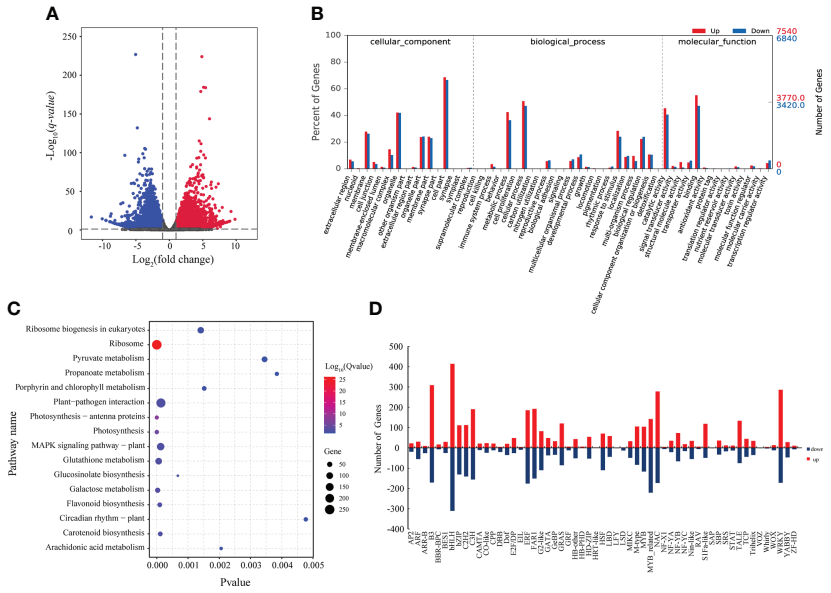


FIGURE 4
(A) Volcano map of DEGs. **(B)** GO enrichment analysis of DEGs. **(C)** KEGG enrichment of DEGs. **(D)** Statistical map of differentially expressed transcription factor families.

the KEGG pathway analysis of the transcriptome indicated that 13 genes were differentially expressed involved in Ribosome biogenesis in eukaryotes, two in Ribosome, and one in Photosynthesis (Figure 4C).

A total of 7,228 DEGs encoding 26 families of transcription factors were identified (Figure 4D), including the large transcription factor families bHLH, MYB, and YABBY. These results indicate that these transcription

factors participate in regulating changes in gene expression between BS and FBS.

Expression patterns of gene and proteins within chlorophyll, amino acids, and flavonoids synthetic pathways

Based on our results describing changes in metabolite content, and transcription and translation of key genes, we have proposed some transcriptional regulatory relationships between candidate transcription factors and genes in the synthetic pathways of chlorophyll, amino acids, and flavonoids in alfalfa leaves (Figures 5–7). The *Hem* family genes regulate and catalyze many steps in the chlorophyll metabolism pathway, especially *HemA*, *HemL*, *HemE*, *HemY*, and *HemH*. From BS to FBS, some members of the *Hem* family are upregulated whereas other chlorophyll biosynthetic genes were downregulated, such as *ChlH*, *POR*, *DVR*, *CAO*, *ChlG*, and *NOL* (Figure 5).

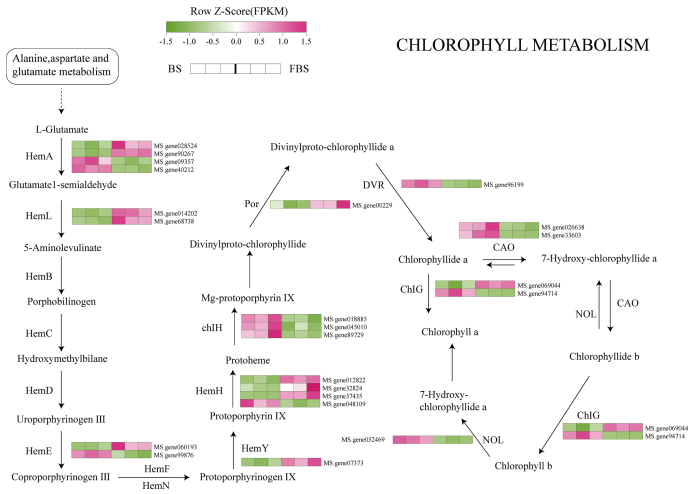


FIGURE 5
Chlorophyll biosynthesis pathway highlighting differentially expressed genes.

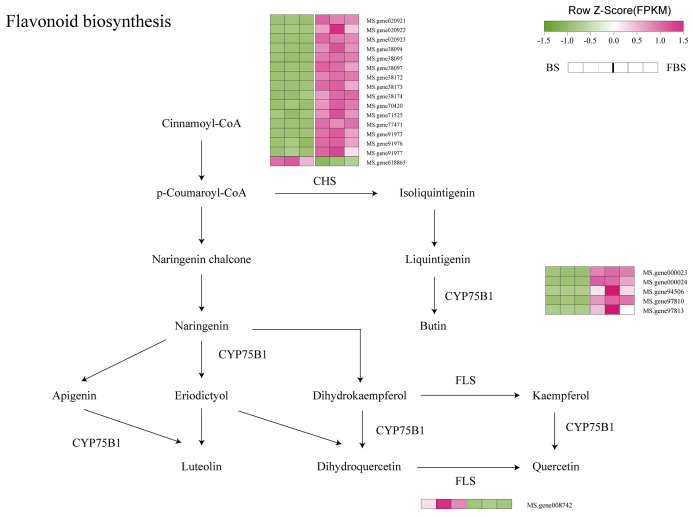


FIGURE 6
Flavonoid biosynthesis pathway highlighting differentially expressed genes.

CYP75B1 family genes participate in the flavonoid metabolic pathway. Indeed, among the up-regulated genes from BS to FBS are several genes belonging to this family, as well as many of the chalcone synthase gene family (Figure 6).

For the amino acid metabolism pathway, we found that *NIT2* and *metB* genes were down-regulated at FBS compared with BS, whereas *Glutamate dehydrogenase 1 (Glud1)* family genes were upregulated (Figure 7).

Coexpression network analysis

The genes, proteins, and metabolites of the chlorophyll, flavonoids, and amino acids coexpression network reveals 78 nodes and 327 edges (Figure 8). Seven encoded Hub proteins and three Hub metabolites were identified according to their margin >20. In this analysis, the interaction between the carbon-nitrogen hydrolase family protein (fragment) and DL-lactate was

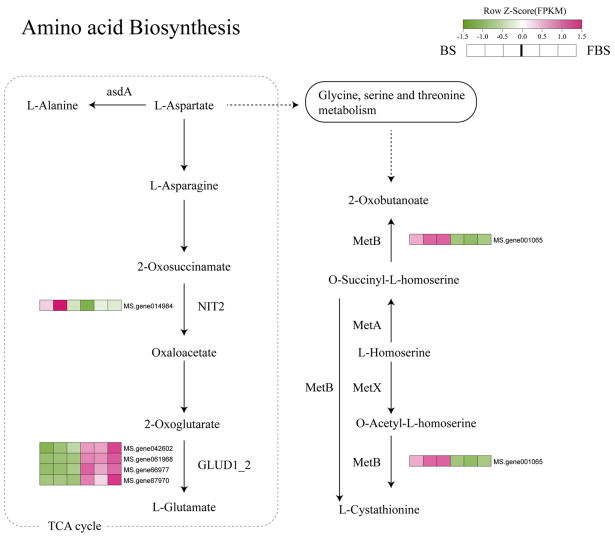


FIGURE 7
Amino acid biosynthesis pathway highlighting differentially expressed genes.

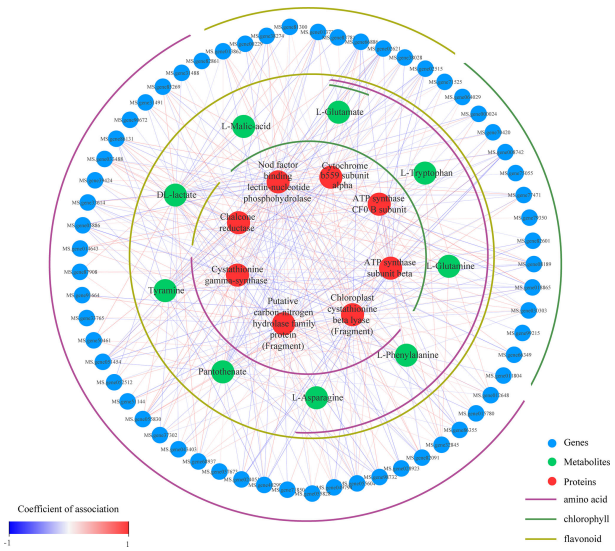


FIGURE 8
Network diagram of alfalfa metabolomic, proteomic, and transcriptomic analysis.

the strongest. Interestingly, there were relatively more negative correlation nodes between the metabolism of these hubs and proteins.

Discussion

Molecular regulation of chlorophyll biosynthesis in alfalfa leaves during flowering

Photosynthesis requires chlorophyll, which absorbs and converts solar energy for cell electron transfer and energy conversion (Zhang et al., 2021). The content of chlorophyll in alfalfa leaves affects not only yield and appearance, but also its nutritional quality (Saric-Krsmanovic et al., 2018). Our study indicates that there are some differences in metabolite accumulation between BS and FBS, especially chlorophyll that decreases significantly. Studies have shown that alfalfa forage protein significantly correlates with chlorophyll A and chlorophyll A/B, with chlorophyll A/B being the main factor affecting protein synthesis, followed by chlorophyll A (Liu, 2018). Therefore, the decrease in chlorophyll content may be one of the reasons for the decrease of protein content at different developmental stages.

The content of chlorophyll at BS was higher than at FBS. During early vegetative growth, alfalfa accumulates more nutrients and chlorophyll whereas during reproductive growth, the plants consume more nutrients than produced by photosynthesis while the chlorophyll content gradually declines. In contrast, Yu et al. (2017) found that chlorophyll content increased with leaf age. Chen et al. (2010) found that chlorophyll content differed during rice growth, peaking during heading and filling stages, and being low at seedling and tillering stages. Our divergent results may be due to differences between species, varieties, or environmental factors; or due to the fact that our study focused only on two developmental stages, close together but relevant for forage nutrition.

In our previous research, we found a direct correlation between chlorophyll and protein (Fan et al., 2018), but the specific regulatory mechanism was not then explained. Chlorophyll synthesis is triggered by the conversion of glutamate to 5-aminolevulinic acid (ALA). L-Glutamate is a precursor for chlorophyll synthesis, and an increased L-glutamate content in plants is beneficial for chlorophyll synthesis. After spraying glutamate on citrus leaves, the chlorophyll content was significantly higher in the treatment group (Wu et al., 2018). In another experiment, L-glutamate was added to the sugar-free medium of small plants (Gou et al., 2020), which increased the dry weight, chlorophyll content, and net photosynthetic rate in these plants. In alfalfa leaves, our metabolomic results revealed that L-glutamate content at FBS was significantly lower than at BS, in line with our previous experimental results.

ALA is a key precursor of chlorophyll biosynthesis catalyzed by glutamyl-tRNA synthetase, glutamyl-tRNA reductase (Glu TR), and glutamine-1-hemicaldehyde-aminotransferase. Glu TR, encoded by *HemA*, is the central controller of chlorophyll biosynthesis (Zeng et al., 2020; Luo et al., 2021). Other involved catalytic enzymes are bile pigment deaminase, urinary porphyrin synthase, urinary porphyrin iii decarboxylase, fecal porphyrin oxidation decarboxylase, and porphyrin oxidase (Zhang et al., 2010), encoded by *HemC*, *HemD*, *HemE*, *HemF*, and *HemG*, respectively. *HemA* is a gene family made up of multiple members, the number of which may vary from plant to plant. The model plant *Arabidopsis thaliana* has five *HemA* members, while cucumber has only two genes(Zhang et al., 2010). Here, we found 10 genes, including *HemA*, *HemB*, *HemC*, *HemD*, *HemE*, *HemF*, *HemH*, *HemY*, *HemL*, and *HemN*, five of which (*HemA*, *HemL*, *HemE*, *HemY*, and *HemH*) play decisive roles in the metabolism of L-glutamate to protoheme. To the best of our knowledge, this study has clarified the metabolic regulation pathway of *Hem* family genes in alfalfa for the first time.

A reduction in expression of *Hem* genes can lead to a decrease, or even loss, of enzyme activity and a corresponding decrease in chlorophyll production. Indeed, when *HemA* of *S. cerevisiae* was ectopically expressed in tobacco, transgenic plants oversynthesized ALA under light conditions and exhibited significantly enhanced leaf photosynthesis (Zhang et al., 2021). Li et al. (2020) showed that, in Arabidopsis, *HemA1* was induced by light, and its expression level changed with light quality (red, far-red, blue, and white light). Our results are consistent with previous studies, showing that the decrease in L-glutamate and the down-regulation of *HemA* expression may result in the inhibition of chlorophyll biosynthesis precursors, and ultimately lead to the decrease in leaf chlorophyll content.

Molecular regulation of flavonoid biosynthesis in alfalfa leaves during flowering

The structurally diverse flavonoid metabolites contribute to various physiological plant processes, including biotic and abiotic stress response, nodule organogenesis (in legumes), fertility, pigmentation, and developmental regulation (Wang et al., 2020; Marco et al., 2021). Flavonoids also regulate root growth *via* reactive oxygen species (ROS) (Maloney et al., 2014) and the stomatal opening of guard cells (Watkins et al., 2017). Moderate salt stress in crops and medicinal plants can induce

flavonoid accumulation to improve product quality (Lim et al., 2012). Flavonoids also help to resist infection by fungi and bacteria (Zhou et al., 2016).

Our study indicates that the flavonoid content in alfalfa leaves at FBS was significantly lower than at BS. This study provides the first report on how flavonoid content changes in alfalfa leaves during plant maturation, and indicates that flavonoid content is closely related to, and changes during, plant development. Similarly, in stevia, the accumulation of flavonoids is closely related to the growth stage, and its content show a dynamic trend of increasing (first 70 days), decreasing (day 70–94), and the remaining stable (Zhou et al., 2016).

Flavonoids in alfalfa mainly comprise apigenin, digloflavone, kainic acid, quercetin, and myricetin, among which the first two are the most common (Lui et al., 2020).

At present, *CYP75* family genes have attracted much attention, but there is no relevant report about *CYP75* in alfalfa. Relevant studies have shown that flavonoid synthesis in grapes is regulated by *CYP75* family genes, and the flavonoid content at the budding stage is significantly higher than that at flowering (Castellarin et al., 2006). Consistent with these results, this study has shown that *CYP75B1* plays a key role in the transformation of cinnamoyl-CoA into luteolin, apigenin and butin, especially in the transformation of liquintigenin to butin. Five genes (*Ms. Gene000023*, *Ms. Gene000024*, *Ms. Gene94506*, *Ms. Gene97810*, and *Ms. Gene97813*) were up-regulated, indicating that these genes play key roles in the transformation of these metabolites. However, their specific mechanisms of regulation require further elucidation.

Molecular regulation of amino acid biosynthesis in alfalfa leaves during flowering

Alfalfa has a high total amino acid content, one of the important indices of plant quality (Liu et al., 2022). Studies have shown that the amino acid content is positively correlated with the crude protein content during growth, and that, with time, amino acid content decreases significantly (Liu, 2018). This observation is consistent with results from our proteomics analyses.

Amino acids are essential molecules that form proteins and contribute to the nitrogen cycle (Yu et al., 2020), and are also the precursors of many metabolites (Dinkeloo et al., 2018) such as the phytohormones auxin and ethylene (Du et al., 2022). Amino acid content and composition varies between tissues and cell types, and across different developmental stages (Kishor et al., 2015). To our knowledge, amino acid content, biosynthesis, and regulation in alfalfa leaves during development have not been investigated.

Increases in amino acid content may be due to increased amino acid biosynthesis or protein degradation; decreased glycolysis; or the metabolism of other nitrogen-containing compounds, including chlorophyll, purines, nucleotides, and alkaloids (Batista-Silva et al., 2019; Zhao et al., 2019). Here, we found that most DEGs between BS and FBS were involved in the glycolysis-TCA pathway, which suggests that decreased amino acid content in mature alfalfa leaves may be due to alterations in carbohydrate metabolism.

One member of the *NIT2* family was significantly down-regulated. Wang et al. (2010) described a similar phenomenon in chlorotic leaves of *Malus domestica* (apple), in which amino acid content was significantly decreased due to depressed glycolysis and TCA cycle activity. While *NIT2* is downregulated in the glutamate pathway, our results also showed that *MetB* family genes, participating in the metabolism of glycine to 2-oxobutanoate and the transformation and synthesis of L-cystathionine to 2-oxobutanoate, were significantly down-regulated. At present, there is no report on *MetB* in amino acid metabolism and synthesis, and its gene function needs to be further studied.

The glycolysis-TCA cycle synthesizes fundamental metabolites, such as amino acids and the reducing agent NADH, and provides energy for plant growth and development (Xiong et al., 2021). Phosphoenolpyruvate (PEP), an important intermediate in the glycolysis-TCA cycle, is also a precursor for amino acid biosynthesis, particularly aromatic amino acids (Li et al., 2020). Higher expression of glycolysis-TCA–related enzymes could increase amino acid synthesis (Katsuki et al., 2020). For example, wild ginseng has a higher amino acid content and higher expression of glycolysis-TCA–related enzymes than cultivated ginseng (Sun et al., 2016). TCA cycle intermediates have also been closely correlated with amino acid pools in tobacco (Zhao et al., 2016).

Results from this study clearly show that both glycolysis and the TCA cycle are downregulated in alfalfa leaves at budding compared with flowering, indicated by lower expression of several genes encoding key enzymes (e.g., *NIT2*, *metB*) and the decreased abundance of carbohydrate metabolism intermediates. These data provide the first description of amino acid regulation during alfalfa leaf development and provide the basis for further study to explore the precise roles of enzymes such as NIT2 and metB for use as precision tools to improve alfalfa leaf quality.

Conclusion

We have analyzed the differences in chlorophyll, amino acid, and flavonoid content in alfalfa leaves at transcriptional, translational, and metabolic levels between two different developmental stages, the budding and the flowering stages. We found that as alfalfa matures, the content of chlorophyll, amino acids, and flavonoids significantly decreases. We identified ten *Hem* family genes present in alfalfa, which are involved in the L-glutamate regulatory and chlorophyll

biosynthetic pathways. From BS to FBS, we observed a decrease in L-glutamate content and *HemA* expression in alfalfa leaves, which may inhibit the production of chlorophyll synthesis precursor and eventually lead to the observed decrease in leaf chlorophyll content. Expression of the key gene family *CYP75B1* decreased, resulting in a decline in flavonoids, such as apigenin and luteolin. In mature alfalfa leave, the amino acid synthesis genes *NIT2* and *metB* decreased, whereas *Glud1* increased, expression. Most DEGs were involved in the glycolysis-TCA pathway, causing a reduction in amino acid content and therefore nutritional quality. We have described some of the molecular mechanisms underpinning the nutritional value of alfalfa, providing a theoretical basis to guide future improvements in high-quality alfalfa hay production.

Data availability statement

The data presented in the study are deposited in the NCBI repository, accession number PRJNA644634.

Author contributions

YL and WF conducted the experiments, analyzed the data, and wrote the manuscript. QC, LZ, TC, XJ, and QY provided technical support, and revised the manuscript critically. QS and ZW provided the experimental materials. KJ conceived the experiment and manuscript revision. All authors contributed to the article and approved the final manuscript for submission.

Funding

This work was supported by the Fundamental Research Funds of Chinese Academy of Agriculture (Grant No. 1610332022014), Inner Mongolia Autonomous Region science and technology planning project (Grant No. 2022YFHH0046), Inner Mongolia Natural Science Foundation project (Grant No2022LHQN3003) and Guizhou University Cultivation Project (Guida Renji Hezi [2020]9).

Acknowledgments

We would like to thank all colleagues and friends who have contributed to this study.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.995031/full#supplementary-material

References

Batista-Silva, W., Heinemann, B., Rugen, N., Nunes-Nesi, A., Araújo, W. L., Braun, H. P., et al. (2019). The role of amino acid metabolism during abiotic stress release. *Plant Cell Environ.* 42 (5), 1630–1644. doi: 10.1111/pce.13518

Bird, P. R., and Moir, R. J. (1972). Sulphur metabolism and excretion studies in ruminants. 8. methionine degradation and utilization in sheep when infused into the rumen or abomasum. *Aust. J. Biol. Sci.* 25, 835–848. doi: 10.1071/bi9731429

Castellarin, S. D., Di Gaspero, G., Marconi, R., Nonis, A., Peterlunger, E., Paillard, S., et al. (2006). Colour variation in red grapevines (Vitis vinifera l.): genomic organisation, expression of flavonoid 3'-hydroxylase, flavonoid 3',5'-hydroxylase genes and related metabolite profiling of red cyanidin-/blue delphinidin-based anthocyanins in berry skin. *BMC Genomics* 7 (1), 12. doi: 10.1186/1471-2164-7-12

Chen, X. L., Chen, C., and Zhou, L. (2010). Determination and correlativity analysis of chlorophyll content at different developmental stages in rice. *Modern Agric. Sci. Technology.* 17), 42–44,52. doi: 10.3969/j.issn.1007-5739.2010.17.017

Dantas, R., Rodrigo, S., Constantin, P., Barbosa, R., Anderson, L., Soares, R., et al. (2020). Biosynthesis and metabolic actions of simple phenolic acids in plants. *Phytochem. Rev.* 19, 865–906. doi: 10.1007/s11101-020-09689-2

Dinkeloo, K., Boyd, S., and Pilot, G. (2018). Update on amino acid transporter functions and on possible amino acid sensing mechanisms in plants. *Semin. Cell Dev. Biol.* 74, 105.133. doi: 10.1016/j.semcdb.2017.07.010

Dong, S., Liu, T., and Dong, M. (2018). Determination of 18 kinds of amino acids in fresh tea leaves by hplc coupled with pre-column derivatization. *Asian Agric. Res.* 10 (02), 59–62+67.

Dong, S. W., Sang, L. J., Xie, H. L., Chai, M. F., and Wang, Z. Y. (2021). Comparative transcriptome analysis of salt stress-induced leaf senescence in medicago truncatula. *Front. Plant Science.* 12. doi: 10.3389/fpls.2021.666660

Du, Z. K., Lin, W. D., Zhu, J. X., and Li, J. M. (2022). Amino acids profiling and transcriptomic data integration demonstrates the dynamic regulation of amino acids synthesis in the leaves of cyclocarya paliurus. *PeerJ.* 10, e13689. doi: 10.7717/peerj.13689

Fan, W. Q., Ge, G. T., Liu, Y. H., Wang, W., Liu, L. Y., and Jia, Y. S. (2018). Proteomics integrated with metabolomics: analysis of the internal causes of nutrient changes in alfalfa at different growth stages. *BMC Plant Biol.* 18 (1), 78. doi: 10.1186/s12870-018-1291-8

Galili, G., Amir, R., and Fernie, A. R. (2016). The regulation of essential amino acid synthesis and accumulation in plants. *Annu. Rev. Plant Biol.* 67, 8.1–8.26. doi: 10.1146/annurev-arplant-043015-112213

Gao, J., Wang, H., Yuan, Q., and Yue, F. (2018). Structure and function of the photosystem supercomplexes. *Front. Plant Science.* 9. doi: 10.3389/fpls.2018.00357

Goawska, S., Ukasik, I., Kapusta, T., and Janda, B. (2010). Analysis of flavonoids content in alfalfa. *Ecol. Chem. Engineering* 17 (2–3), 261–267.

Gou, T. Y., Yang, L., Hu, W. X., Chen, X. H., Zhu, Y. X., Guo, J., et al. (2020). Silicon improves the growth of cucumber under excess nitrate stress by enhancing nitrogen assimilation and chlorophyll synthesis. *Plant Physiol. Biochem.* 152, 53–56. doi: 10.1016/j.plaphy.2020.04.031

Guo, H. M., Zhu, H. S., and Guo, J. X.. (2020). Comparative study on extraction and detection methods of total flavonoids from Alfalfa at different growth stages. *Chinese Journal of Grassland* 42 (03), 141–146. doi: 10.16742/j.zgcdxb.20190270

Harmut, A. (1987). Chlorophylls and carotenoids: Pigments of photosynthetic membranes. *Methods Enzymology*. 148, 350–383. doi: 10.1016/0076-6879(87)48036-1

Jing, D., Chen, W., Hu, R., Zhang, Y., and Liang, G. (2020). An integrative analysis of transcriptome, proteome and hormones reveals key differentially expressed genes and metabolic pathways involved in flower development in loquat. *Int. J. Mol. Sci.* 21 (14), 5107. doi: 10.3390/ijms21145107

Jing, C. L., Dong, X. F., and Tong, J. M. (2015). Optimization of ultrasonic-assisted extraction of flavonoid compounds and antioxidants from alfalfa using response surface method. *Molecules* 20 (9), 15550–15571. doi: 10.3390/molecules200915550

Katsuki, M., Daisuke, S., and Shunsuke, K. (2020). Optimal ratio of carbon flux between glycolysis and the pentose phosphate pathway for amino acid accumulation in corynebacterium glutamicum. *ACS Synthetic Biol.* 9 (7), 1615–1622. doi: 10.1021/acssynbio.0c00181

Kishor, P. B. K., HimaKumari, P., Sunita, M. S. L., and Sreenivasulu, N. (2015). Role of proline in cell wall synthesis and plant development and its implications in plant ontogeny. *Front. Plant Science.* 6. doi: 10.3389/fpls.2015.00544

Less, H., and Galili, G. (2009). Coordinations between gene modules control the operation of plant amino acid metabolic networks. *BMC Syst. Biol.* 3, 14. doi: 10.1186/1752-0509-3-14

Li, N., Liu, Y., Liang, Z., Lou, Y., and Wang, G. (2020). Influence of fuel oil on platymonas helgolandica: an acute toxicity evaluation to amino acids. *Environ. Pollution.* 271, 116226. doi: 10.1016/j.envpol.2020.116226

Lim, J., Park, K., and Kim, B. (2012). Effect of salinity stress on phenolic compounds and carotenoids in buckwheat sprout. *Food Chem.* 135 (3), 1065–1070. doi: 10.1016G/j.foodchem.2012.05.068

Liu, L. Y. (2018). *Effects of environmental factors on nutrients in alfalfa drying process and field regulation strategies* (Inner Mongolia Agricultural University).

Liu, L. Y., Jia, Y. S., and Wang, Z. J. (2022). Study on main environmental factors affecting natural drying of alfalfa. *J. Grass Industry.* 31 (2), 121–132. doi: 10.11686/cyxb2020529

Lui, A. C. W., Lam, P. Y., Kwun, H. C., Wang, L. X., Tobimatsu, Y., and Lo, C. (2020). Convergent recruitment of 5'-hydroxylase activities by CYP75B flavonoid b-ring hydroxylases for tricin biosynthesis in medicago legumes. *New Phytol* 228 (1), 268–284. doi: 10.1111/nph.16498

Luo, J., Wang, H., Chen, S., and Ren, S. (2021). CmNAC73 mediates the formation of chrysanthemum green by directly activating the expression of chlorophyll biosynthesis genes HEMA1 and CRD1. *Genes* 12 (5), 704. doi: 10.3390/genes12050704

Maloney, G., Dinapoli, K., and Muday, G. (2014). The anthocyanin reduced tomato mutant demonstrates the role of flavonols in tomato lateral root and root hair development. *Plant Physiol.* 166 (2), 614–631. doi: 10.1104/pp.114.240507

Marco, A., Mariana, B., Natália, G., Maycon, A., Cleverson, R., Jaquelyne, P., et al. (2021). Physiological impact of flavonoids on nodulation and ureide metabolism in legume plants. *Plant Physiol. Biochem.* 166 (8), 512–521. doi: 10.1016/j.plaphy.2021.06.007

Ouyang, K. H., Xu, M. S., Jiang, Y., and Wang, W. J. (2016). Effects of alfalfa flavonoids on broiler performance, meat quality and gene expression. *Can. J. Anim. Science.* 96 (3), 332–341. doi: 10.1139/CJAS-2015-0132

Popovic, S., Stjepanovic, M., Grljusic, S., Cupic, T., and Tucak, M. (2001). Protein and fiber contents in alfalfa leaves and stems. *Quality in Lucerne and Medics for Animal Production* (Eds.: I. Delgado and J. Lloveras 45, 215–218.

Ren, Y., Pan, J., Zhang, Z., Zhao, J., and Hu, G. (2020). Identification of an up-accumulated polyamine oxidase 2 in pollen of self-incompatible 'wuzishatangju' mandarin using comparative proteomic analysis. *Scientia Horticulturae.* 266, 109279. doi: 10.1016/j.scienta.2020.109279

Saric-Krsmanovic, M., Bozic, D., Radivojevic, L., Gajic Umiljendic, J., and Vrbnicanin, S. (2018). Impact of field dodder (cuscuta campestris yunk.) on chlorophyll fluorescence and chlorophyll content of alfalfa and sugar beet plants. *Russian J. Plant、Physiology.* 65 (5), 726–731. doi: 10.1134/S102144371805014X

Schmied, J., Hedtke, B., and Grimm, B. (2011). Overexpression of hema1 encoding glutamyl-trna reductase. *J. Plant Physiol.* 168 (12), 1372–1379. doi: 10.1016/j.jplph.2010.12.010

Shalygo, N., Czarnecki, O., Peter, E., and Grimm, B. (2009). Expression of chlorophyll synthase is also involved in feedback-control of chlorophyll biosynthesis. *Plant Mol. Biol.* 71 (4), 425–436. doi: 10.1007/s11103-009-9532-8

Stenbaek, A., and Jensen, P. E. (1995). Chlorophyll biosynthesis. *Phytochemistry* 7 (7), 1039–1057. doi: 10.2307/3870056

Sun, H., Liu, F. B., and Sun, L. W. (2016). Proteomic analysis of amino acid metabolism differences between wild and cultivated panax ginseng. *J. Ginseng Res.* 40 (2), 113–120. doi: 10.1016/j.jgr.2015.06.001

Tong, Z. Y., Xie, C., Ma, L., Liu, L. P., Jin, Y. S., Dong, J. L., et al. (2014). Co-Expression of bacterial aspartate kinase and adenylylsulfate reductase genes substantially increases sulfur amino acid levels in transgenic alfalfa (medicago sativa l.). *PloS One* 9 (2), e88310. doi: 10.1371/journal.pone.0088310

Wang, F., Ge, S. F., Xu, X. X., Xin, Y. X., and Zhang, D. X. (2021). Multiomics analysis reveals new insights into the apple fruit quality decline under high nitrogen conditions. *J. Agric. Food Chem.* 69 (19), 5559–5572. doi: 10.1021/acs.jafc.1c01548

Wang, Z. H., Hong, X., Hu, K. K., Wang, Y., Wang, X. X., Du, S. Y., et al. (2017). Impaired magnesium protoporphyrin ix methyltransferase (chlm) impedes chlorophyll synthesis and plant growth in rice. *Front. Plant Sci.* 8, 1694. doi: 10.3389/fpls.2017.01694

Wang, H. J., Liu, S. H., Wang, T. L., Liu, H. W., Xu, X. H., Chen, K. S., et al. (2020). The moss flavone synthase I positively regulates the tolerance of plants to drought stress and UV-b radiation. *Plant Science.* 298, 110591. doi: 10.1016/j.plantsci.2020.110591

Wang, H., Ma, F., and Cheng, L. (2010). Metabolism of organic acids, nitrogen and amino acids in chlorotic leaves of 'Honeycrisp' apple (Malus domestica borkh) with excessive accumulation of carbohydrates. *Planta.* 232, 511–522. doi: 10.1007/s00425-010-1194-x

Wang, Z. Q., Shi, H. R., Yu, S. F., Zhou, W. L., Li, J., Liu, S. H., et al. (2019). Comprehensive transcriptomics, proteomics, and metabolomics analyses of the mechanisms regulating tiller production in low-tillering wheat. *Theor. Appl. Genet.* 132, 2181–2193. doi: 10.1007/s00122-019-03345-w

Watkins, J., Chapman, J., and Muday, G. (2017). Abscisic acid- induced reactive oxygen species are modulated by flavonols to control stomata aperture. *Plant Physiol.* 175 (4), 1807–1825. doi: 10.1104/pp.17.01010

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

Wu, Y., Xin, J., Liao, W., Hu, L., Dawuda, M. M., Zhao, X., et al. (2018). 5-aminolevulinic acid (ala) alleviated salinity stress in cucumber seedlings by enhancing chlorophyll synthesis pathway. *Front. Plant Science.* 9. doi: 10.3389/fpls.2018.00635

Xiong, Y., Qu., Y., and Han, H. (2021). Unraveling physiological and metabolomic responses involved in phlox subulata l. tolerance to drought stress. *Plant Mol. Biol. Rep.* 39, 98–111. doi: 10.1007/s11105-020-01238-7

Xu, B., Wu, R. L., Shi, F. L., Gao, C., and Wang, J. (2022). Transcriptome profiling of flower buds of male-sterile lines provides new insights into male sterility mechanism in alfalfa. *BMC Plant Biology* 22 (1), 1–13. doi: 10.1186/s12870-022-03581-1

Yu, W. X., Yang, Q. P., Huang, S. M., Guo, Z. W., and Yu, D. Q. (2017). Analysis on the main influencing factors of chlorophyll value in leaves of phyllostachys prominens. *J. Zhejiang Forestry Sci. Technology.* 2), 36–39. doi: 10.3969/j.issn.1001-3776.2014.02.009

Zeng, Z. Q., Lin, T. Z., Zhao, J. Y., and Zheng, T. H. (2020). The OsHemA gene encoding glutamyl-TrNA reductase (GluTR) is critical for chlorophyll biosynthesis in rice (oryza sativa). *J. Integr. Agriculture.* 19 (3), 12. doi: 10.1016/s2095-3119(19)62710-3

Zhang, J. L., Li, X. G., Xu, X. H., Chen, H. P., Li, Y. L., and Guy, R. D. (2021). Leaf morphology, photosynthesis and pigments change with age and sunlight in savin juniper. *Plant Biol.* 23 (6), 1097–1108. doi: 10.1111/plb.13256

Zhang, C. H., Zhang, H. Y., and Yu, X. Z. (2010). Effects of selenium on photosynthetic characteristics and chlorophyll fluorescence parameters of tomato seedling leaves under low temperature stress. *Chin. Agric. Sci. Bulletin.* 26 (5), 152–157.

Zhan, J. S., Liu, M. M., Su, X. S., Zhan, K., Zhang, C. G., and Zhao, G. Q. (2017). Effects of alfalfa flavonoids on the production performance, immune system, and ruminal fermentation of dairy cows. *Asian-Australasian J. Anim. Sci.* 30 (10), 1416–1424. doi: 10.5713/ajas.16.0579

Zhao, Y., Xu, F. L., Liu, J., Guan, F. C., Quan, H., and Meng, F. J. (2019). The adaptation strategies of herpetospermum pedunculosum (Ser.) baill at altitude gradient of the Tibetan plateau by physiological and metabolomic methods. *BMC Genomics* 20, 451. doi: 10.1186/s12864-019-5778-y

Zhao, J., Zhao, Y., Hu, C., Zhao, C., Zhang, J., and Li, L. (2016). Metabolic profiling with gas chromatography-mass spectrometry and capillary electrophoresis-mass spectrometry reveals the carbon-nitrogen status of tobacco leaves across different planting areas. *J. Proteome Res.* 15, 468–476 468. doi: 10.1021/acs.jproteome.5b00807

Zhou, X., He, J., and Wang, T. (2016). Combined effects of blue and ultraviolet lights on the accumulation of flavonoids in tartary buckwheat sprouts. *Polish J. Food Nutr. Sci.* 66 (2), 93–98. doi: 10.1515/pjfns-2015-0042

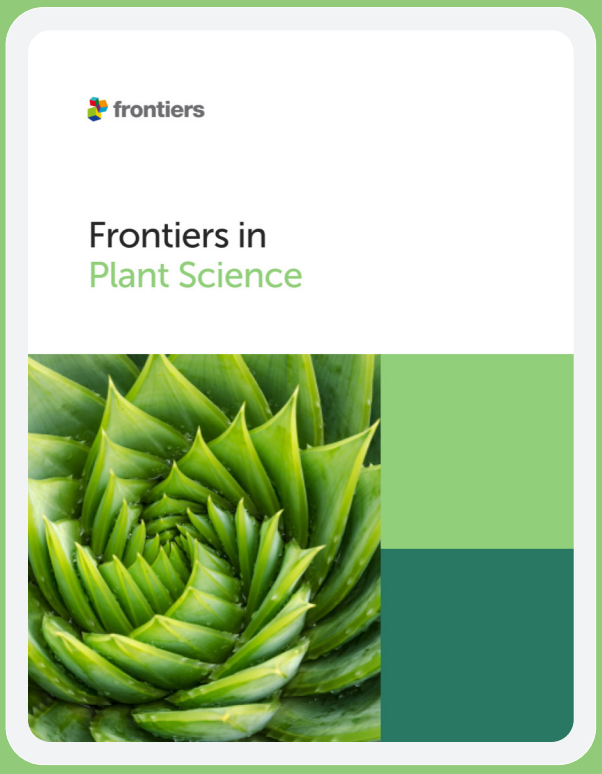
Frontiers in
Plant Science

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