frontiers | Research Topics

Transcriptional/post-transcriptional regulations in agricultural species after stresses

Edited by
Suxu Tan, Rani Alex and Turgay Unver

Published in
Frontiers in Genetics
Frontiers in Veterinary Science

ISSN 1664-8714
ISBN 978-2-83251-345-3
DOI 10.3389/978-2-83251-345-3

About Frontiers

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

Frontiers journal series

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

Dedication to quality

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

What are Frontiers Research Topics?

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

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

Transcriptional/post-transcriptional regulations in agricultural species after stresses

Topic editors

Suxu Tan — Qingdao University, China

Rani Alex — Indian Council of Agricultural Research (ICAR), India

Turgay Unver — FicusBio, Türkiye

Citation

Tan, S., Alex, R., Unver, T., eds. (2023). *Transcriptional/post-transcriptional regulations in agricultural species after stresses*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-345-3

Table of contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 04 January 2023
DOI 10.3389/fgene.2022.1127832

OPEN ACCESS

EDITED AND REVIEWED BY
Martino Cassandro,
University of Padua, Italy

*CORRESPONDENCE
Suxu Tan,
tansuxu@qdu.edu.cn

SPECIALTY SECTION
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

RECEIVED 20 December 2022
ACCEPTED 20 December 2022
PUBLISHED 04 January 2023

CITATION
Tan S, Alex R and Unver T (2023), Editorial: Transcriptional and post-transcriptional regulations in agricultural species after stresses.
Front. Genet. 13:1127832.
doi: 10.3389/fgene.2022.1127832

Editorial: Transcriptional and post-transcriptional regulations in agricultural species after stresses

Suxu Tan[1]*, Rani Alex[2] and Turgay Unver[3,4]

[1]Institute of Aquatic Biotechnology, College of Life Sciences, Qingdao University, Qingdao, China, [2]ICAR-National Dairy Research Institute, Karnal, Haryana, India, [3]Ficus Biotechnology, Ankara, Türkiye, [4]Faculty of Engineering, Ostim Technical University, Ankara, Türkiye

KEYWORDS

agriculture, transcriptiome, post-transcriptional regulation, stress, gene

Editorial on the Research Topic
Transcriptional and post-transcriptional regulations in agricultural species after stresses

According to the announcement of United Nations (https://www.un.org), the world population hit 8 billion on 15 November 2022, which could reach a whopping 10 billion people by 2050. Agriculture is the very basic foundation of the modern human society and civilization. In a broad sense, agriculture comprises planting, animal husbandry, aquaculture, and forestry. Agricultural species provide food, feed, fiber, fuel, etc., on which we rely to nourish and sustain ourselves and the next-generation. Agricultural species, however, face various stresses during development, including but not limited to bacteria, viruses, drought, temperature, feed, and chemicals. Each of these adverse conditions may compromise physiological and cellular functions, as well as activate the reprogramming of transcription to maintain the homeostasis of an individual. Understanding the molecular mechanisms and regulations underlying these processes could help select breeding populations, genetically modify organisms, facilitate diagnosis, and improve welfare, ultimately benefiting human society and the whole world.

Referring to the central dogma of molecular biology (Crick, 1970), phenotype can be determined and regulated at multiple layers including the DNA, RNA, and protein level. In response to environmental cues and changing conditions, the RNA-level transcriptional and post-transcriptional regulations come into effect swiftly as pioneers. This Research Topic focuses on the transcriptional and post-transcriptional regulations in agricultural species responding/adapting to various stresses, aiming to reveal the underlying molecular mechanisms and provide insights into their applications in practice.

This Research Topic covers a wide spectrum of agricultural species, including livestock (cattle, sheep, pig, and chicken), fish (Atlantic salmon), cereals and legumes (wheat, maize, soybean, and pea), vegetable (eggplant), fruit (pear), and ornamental plants (red plum, lily, and lotus), in response to various stresses and conditions, such as heat, cold, drought, hypoxia, bacteria, chemicals, weaning, and mastitis. Most studies in this Research Topic examined the whole transcriptome to reveal the underlying molecular mechanisms following stresses. For example, in the context of global warming, Niu et al. employed transcriptome sequencing to study the mechanism by which high temperature affect grain abortion of maize, by comparing heat-resistant and heat-sensitive variety under a 7-day heat stress treatment after pollination. They unveiled that heat stress mainly resulted in reduced carbohydrate availability for grain development, leading to reduced kernel number.

In addition to the transcriptome profiling, Nayan et al. utilized MeDIP-Seq to investigate the genome-wide methylation specific to pathogen-caused subclinical mastitis in water buffalo and its consequential effect on the gene expression landscape. Furthermore, the authors built a Buffalo Subclinical Mastitis Methylome-Transcriptome database (BSCM2TDb) to catalogue the results of this study, which could help buffalo breeders in breed improvement and disease management programs.

Some articles focus on the studies of gene families, instead of the whole transcriptome. For instance, Wang et al. identified 719 putative Type 2C protein phosphatase (PP2C) genes in eight Rosaceae species (Chinese white pear, European pear, Japanese apricot, apple, peach, strawberry, sweet cherry, and black raspberry), and further studied the evolution of PP2C genes. Moreover, the authors performed gene expression analyses specifically in Chinese white pear and identified candidate genes that participated in stress responses, including heat, cold, drought, NaCl, and abscisic acid.

Researchers of this Research Topic also studied post-transcriptional regulations, such as the profiles of lncRNA, miRNA, and circRNA in Tibetan sheep, hybrid hens, and Duroc pigs, respectively. For example, Lu et al. revealed the regulatory mechanisms of the lncRNAs and mRNAs in the adaptation of Tibetan sheep to hypoxia, by comparing lung tissues from high-altitude Tibetan sheep and low-altitude Hu sheep. Functional enrichment and interaction network analysis results showed that lncRNA and mRNA may adapt to hypoxia *via* lipid metabolism, facilitating further analyses of plateau adaptability.

This Research Topic produced vast and valuable data and information, identified crucial biological pathways and candidate genes, and increased our knowledge of the underlying mechanisms of stress responses in agricultural species. These laid a solid foundation to move forward to use genomic breeding approaches, including marker-assisted selection, genomic selection and editing, to speed breeding of resistant populations and subspecies. In addition, the enriched transcriptomics data of this Research Topic, integrated with other omics data of genomics, proteomics, metabolomics, and phenomics, i.e., multi-omics strategies with systems biology, could enhance our understanding and accelerate the breeding improvement (Yang et al., 2021). Furthermore, organic and precision agricultural systems coupled with artificial intelligence have the potential to achieve a balance between productivity, nutrition, economic development, social welfare, and environmental impact (Gebbers and Adamchuk, 2010; Reganold and Wachter, 2016).

We would like to thank all authors and reviewers who contribute to this Research Topic. Besides, we want to pay tribute to agricultural practitioners for their hard work, especially the sacrifices of our father and mother, grandfather and grandmother, extending to our ancestors, who tried their best to feed us well and make our lives better, with the help of agriculture. Similarly, we need to do more to ensure our children have enough food in the future and make them proud of us. To achieve the daunting task, concerted efforts of farmers, researchers, engineers, and policymakers are much-needed.

Author contributions

ST wrote the original version of the manuscript. RA and TU revised the manuscript 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

References

Crick, F. (1970). Central dogma of molecular biology. *Nature* 227, 561–563. doi:10.1038/227561a0

Gebbers, R., and Adamchuk, V. I. (2010). Precision agriculture and food security. *Science* 327, 828–831. doi:10.1126/science.1183899

Reganold, J. P., and Wachter, J. M. (2016). Organic agriculture in the twenty-first century. *Nat. Plants* 2, 15221. doi:10.1038/nplants.2015.221

Yang, Y., Saand, M. A., Huang, L., Abdelaal, W. B., Zhang, J., Wu, Y., et al. (2021). Applications of multi-omics technologies for crop improvement. *Front. Plant Sci.* 12, 563953. doi:10.3389/fpls.2021.563953

frontiers in Genetics

ORIGINAL RESEARCH
published: 01 October 2021
doi: 10.3389/fgene.2021.762135

Genome-Wide Identification and Low-Temperature Expression Analysis of bHLH Genes in *Prunus mume*

Aiqin Ding†, Anqi Ding†, Ping Li, Jia Wang, Tangren Cheng, Fei Bao and Qixiang Zhang**

Beijing Key Laboratory of Ornamental Plants Germplasm Innovation and Molecular Breeding, Beijing Laboratory of Urban and Rural Ecological Environment, Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants of Ministry of Education, Engineering Research Center of Landscape Environment of Ministry of Education, National Engineering Research Center for Floriculture, Beijing Forestry University, Beijing, China

Prunus mume is an illustrious ornamental woody plant with colorful flowers, delicate fragrances, and graceful tree forms. Low temperature limits its geographical distribution. The basic helix-loop-helix (bHLH) proteins exist in most eukaryotes as a transcription factor superfamily, which play a crucial role in metabolism, physiology, development, and response to various stresses of higher organisms. However, the characteristics of the bHLH gene family and low-temperature response remain unknown in *P. mume*. In the present study, we distinguished 95 *PmbHLH* genes in the *P. mume* whole-genome and analyzed their features. *PmbHLHs* were divided into 23 subfamilies and one orphan by phylogenetic analysis. Similar gene structures and conserved motifs appeared in the same subfamily. These genes were situated in eight chromosomes and scaffolds. Gene duplication events performed a close relationship to *P. mume*, *P. persica,* and *P. avium*. Tandem duplications probably promoted the expansion of *PmbHLHs*. According to predicted binding activities, the PmbHLHs were defined as the Non-DNA-binding proteins and DNA-binding proteins. Furthermore, *PmbHLHs* exhibited tissue-specific and low-temperature induced expression patterns. By analyzing transcriptome data, 10 *PmbHLHs* which are responsive to low-temperature stress were selected. The qRT-PCR results showed that the ten *PmbHLH* genes could respond to low-temperature stress at different degrees. There were differences in multiple variations among different varieties. This study provides a basis to research the evolution and low-temperature tolerance of *PmbHLHs*, and might enhance breeding programs of *P. mume* by improving low-temperature tolerance.

Keywords: *Prunus mume*, basic helix-loop-helix gene family, genome-wide analysis, expression pattern, low temperature stress 3

OPEN ACCESS

Edited by:
Suxu Tan,
Michigan State University,
United States

Reviewed by:
Yong Wang,
Jiangsu University, China
Chengsong Zhu,
University of Texas Southwestern Medical Center, United States

****Correspondence:***
Fei Bao
baofei@bjfu.edu.cn
Qixiang Zhang
zqxbjfu@126.com

†These authors share first authorship

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *21 August 2021*
Accepted: *16 September 2021*
Published: *01 October 2021*

Citation:
Ding A, Ding A, Li P, Wang J, Cheng T, Bao F and Zhang Q (2021) Genome-Wide Identification and Low-Temperature Expression Analysis of bHLH Genes in Prunus mume. *Front. Genet. 12:762135. doi: 10.3389/fgene.2021.762135*

INTRODUCTION

Plants are subject to various unsuitable environmental stresses when they grow in a natural environment. Low-temperature stress is a severe natural disaster, which divides into chilling stress and freezing stress. Chilling stress mainly affects the process of photosynthesis and respiratory metabolism of plants, resulting in the disorder of plant cell function to make them

grow abnormally and even causing growth stagnation (Pearce, 1988). Freezing stress freezes plant cells, then causes mechanical damage to the plant cell membrane and eventually might cause plant death (Mccully et al., 2004; Knight et al., 2009). To adapt and resist low-temperature stress, plants have evolved a set of complex and fine regulation mechanisms. Transcription factors are vital in plant signal regulatory networks. When plants suffer from low-temperature stress, transcription factors can activate low-temperature responsive genes by binding *cis*-acting elements on gene promoters. Thus, they regulate signal transduction pathways to improve low-temperature tolerance in plants.

The basic helix–loop–helix (bHLH) proteins, which belong to superfamily transcription factors are widely spread in plants, animals, and fungi (Ledent and Vervoort, 2001). The bHLH superfamily contains two highly conserved domains: the basic region and helix-loop-helix (HLH) region (Atchley et al., 1999). The basic region is composed of 15–20 amino acids and is located at the N-terminal of the bHLH domain, which can recognize and bind DNA (Atchley and Fitch, 1997). In the plant bHLH domain, 50% of the basic region contains a highly conserved His5-Glu9-Arg13 sequence, which can bind to E-box (5′-CANNTG-3′) element. This is necessary for bHLH to bind to DNA (Pires and Dolan, 2010). In the C-terminus of the bHLH domain, the HLH region is composed of about 40 amino acids. This region is characterized by two α-helices connected by a loop with variable length (Murre et al., 1989). The HLH domain can promote interaction between proteins to form homodimers or heterodimers and interact with E-box elements in the genes promoter region (Massari and Murre, 2000; Huq and Quail, 2002). Therefore, the biological functions of most bHLH transcription factors involve forming dimers.

The bHLH transcription factors are related to plant growth and development (Groszmann et al., 2008; Ding et al., 2009), floral organ formation (Buti et al., 2020), secondary metabolism (Nemesio-Gorriz et al., 2017; Wang et al., 2019), and stress resistance (Chinnusamy, 2003; Zhou et al., 2009; Seo et al., 2011; Gao et al., 2020). Under low-temperature stress, many studies have proved that bHLHs are involved in regulation. For example, *ICE1* (INDUCER OF CBF EXPRESSION 1), which belongs to the bHLH transcription factor family, could activate *CBF3* and *COR* genes in response to low temperature in *Arabidopsis thaliana*. Meanwhile, other bHLH transcription factors could also regulate cold tolerance in plants. In rice seedlings, cold stress specifically induced *OsbHLH1* gene expression (Wang et al., 2003). Apple *MdCIbHLH1* played a role in cold tolerance in a CBF dependent manner (Feng et al., 2012).

Prunus mume Sieb. et Zucc, a crucial woody plant with excellent ornamental characteristics for various colors, delicate fragrances, ample flower shapes, and abundant tree forms has been widely used for plant landscaping. *P. mume* originated in *the* Yangtze River Basin and Southwest China. The cultivar was distributed in Northern China and East Asia, with domestication taking place over a long time (Zhang et al., 2018). However, low temperature is still a limiting factor for the northward distribution of *P. mume*. Therefore, it is essential to enhance cold resistance to expand distribution. The bHLH transcription factors have been proved to be involved in resisting low-temperature stress. Nevertheless, the identification of bHLH genes has still not been conducted in *P. mume*. In the present study, we identified 95 *PmbHLH* genes and performed a comprehensive bioinformatics analysis based on *P. mume* genome-wide. We analyzed gene identification, phylogenetic tree, DNA binding activity, gene structure, conserved motifs, protein interaction, chromosomal distribution, synteny analysis, and tissue-specific expression. Combining RNA sequencing data and qRT-PCR analysis, the expression patterns of *PmbHLHs* were estimated under low-temperature stress. Overall, our results could form the foundation for researching the biological function of *PmbHLHs* and enrich the low-temperature resistance gene bank in woody plants.

MATERIAL AND METHODS

Genome-Wide Identification of *PmbHLHs*

We applied the genome project (http://prunusmumegenome.bjfu.edu.cn) to download the whole genome data of *P. mume* (Zhang et al., 2012). The sequences of *A. thaliana* bHLHs (AtbHLHs) were obtained from Pires and Dolan (Pires and Dolan, 2010). The Pfam database (http://pfam.xfam.org, PF00010) was used to obtain the Hidden Markov Model (HMM) profile of the HLH domain and was searched bHLH proteins of *P. mume* with HMMER3 software (http://hmmer.janelia.org) (Sun et al., 2015). To ensure credibility, the E-value cut-off was set at 10^{-5}. SMART online software (http://smart.embl-heidelberg.de/) (Schultz et al., 1998) and NCBI CD-search (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer et al., 2015) were used to confirm bHLH domains in presumptive PmbHLH proteins of *P. mume*.

The WoLF PSORT program (https://wolfpsort.hgc.jp/) was applied to predict the subcellular localization of *PmbHLHs* (Horton et al., 2007). The CDS length, molecular weights (MWs), theoretical isoelectric points (pI), amino-acid sequences (aa), the total number of positively charged residues (Arg + Lys), total number of negatively charged residues (Asp + Glu), grand averages of hydropathicity (GRAVYs), and instability index and aliphatic index of all predicted *PmbHLHs* were calculated using ExPASy (https://web.expasy.org/protparam/) (Artimo et al., 2012).

Phylogenetic Analysis and Multiple Alignment

RAxML version eight software with maximum likelihood (ML) method was used to construct phylogenetic trees (Stamatakis, 2014). The optimal JTT (Jones-Taylor-Thornton) model of amino acid substitution was applied to construct ML phylogenetic trees. The calculated relationships of the phylogenetic tree were supported by performing 1,000 iterations of bootstrap test and visualized by iTOL (https://itol.embl.de/) (Letunic and Bork, 2006). Alignments of PmbHLHs domains were conducted by Muscle (https://www.ebi.ac.uk/Tools/msa/muscle/) (Edgar, 2004).

Gene Structure and Conserved Domain

TBtools (Chen et al., 2020) and NCBI Batch CD-Search (https://www.ncbi.nlm.nih.gov/) (Marchler-Bauer et al., 2015) were applied to analyze and visualize gene structure and conserved domains. We applied the online MEME program to analyze motif structures of PmbHLH proteins (30 motifs were set as the maximum number), others using default parameters (Bailey et al., 2009). Jalview software (Clamp et al., 2004) and Weblogo3 (http://weblogo.berkeley.edu/logo.cgi) were used to visualize and analyze conserved domains.

Chromosomal Distribution, Gene Duplication, and Synteny

We used the GDR database (https://www.rosaceae.org/) to retrieve genomes of *Prunus avium* and *Prunus persica* (Verde et al., 2013; Shirasawa et al., 2017). The *P. mume* genome database provided chromosomal distribution information of *bHLH* genes. Chromosomal location map, as well as duplication events of *bHLH* genes and mutation rates of Ka (nonsynonymous) versus Ks (synonymous) were predicted through TBtools (Chen et al., 2020). Syntenic relationship of *bHLH* genes was analyzed by MCScanX (Multiple Collinearity Scan toolkit) in *P. mume*, *P. avium,* and *P. persica* (Wang et al., 2012). We used the R circlize package to visualize the relationship between three varieties (Zhuo et al., 2018). The *PmbHLH* genes divergence time (T) was calculated through the equation: $T = dS/(2\lambda \times 106)$ Mya, where $\lambda = 1.5 \times 10^{-8}$ s for dicots (Lynch and Conery, 2000).

Gene Expression Analysis

The transcriptome data of *PmbHLHs* in five tissues (bud, fruit, leaf, stem, and root) were downloaded from the NCBI Sequence Read Archive (accession number: SRP014885). Heat map of five tissues and low-temperature stress were drawn by TBtools (Chen et al., 2020) and iTOL (https://itol.embl.de/) based on FPKM values (Letunic and Bork, 2006).

Plant Materials and Treatments

Annual grafted seedlings of true *mume* 'Beijing Yudie' and apricot mei 'Danfenghou' were used for low-temperature treatment. These plant materials were exposed to the same soil humidity (60–70%), air humidity (65–70%), illumination time (12 h/12 h), and light intensity (150 μmol•m^{-2}•s^{-1}). Stems of *P. mume* were treated for chilling treatment (4°C for 8 h and 5 days) and freezing treatment (−5°C for 1 h).

To further research the regulation process of *PmbHLHs* under low-temperature, we carried out detailed multi-stage chilling treatment and freezing treatment.

1) Chilling treatment: in September, we took annual branches of 'Beijing Yudie" and put branches in a 4°C low-temperature incubator and undertook sampling at 0/2/4/8/16/24/48/72 h. Stem segments without bud points were used as sampling materials.
2) Freezing treatment: in November, when the lowest ambient temperature reaches 5°C, we took branches of 'Beijing Yudie', then put them in a refrigerator at 4°C. The second day, they were put in a low-temperature incubator with step-by-step cooling (1 h/°C, keeping it from 0°C to −10°C). Stem segments were sampled at 0/2/4/6/8/10/24/48 h. The above samples were stored at −80°C for RNA extraction. There were three biological replicates.

RNA Extraction and qRT-PCR Analysis

Total RNA of low-temperature treatments was isolated by RNA Extraction Kit (Takara, Beijing, China) based on references. First-strand cDNA synthesis was reversed with DNase-treated RNA (1 μg) through PrimeScript^{TM}RT Reagent Kit with gDNA Eraser (Takara, Dalian, China). The template used cDNA (2 μL) in a 10 μL qRT-PCR by TB Green II Premix Ex Taq (Takara, Dalian, China). We applied the $2^{\Delta\Delta Ct}$, method to calculate relative expression levels and internal control used protein phosphatase 2A (PP2A)gene of *P. mume* (Wang et al., 2014). **Supplementary Table S10** showed ten selected genes and specific primers. Each qRT-PCR was repeated at least three times.

Promoter *Cis*-Acting Elements and Protein Interaction Analysis

Promoter *cis*-acting regulatory elements were analyzed by PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al., 2002). The AraNet V2 tool (Lee et al., 2015) was used to construct the protein interaction network based on homologous proteins of PmbHLHs in *Arabidopsis*. The protein interaction network was visualized by Cytoscape (Shannon, 2003) and STRING software (http://string-db.org/).

RESULTS

Identification of *PmbHLHs* Genes in *P. Mume*

A total of 95 non-redundant *PmbHLHs* were discovered based on HMMER software from *P. mume* genome. According to their location position, these genes were named *PmbHLH01* to *PmbHLH95*. The length of PmbHLH proteins ranged between 91 (PmbHLH36) and 700 (PmbHLH45) amino acids and most of these genes (74%) had lengths of 200–400 aa. The presumptive isoelectric points (pI) ranged from 4.57 (PmbHLH14) to 10.1 (PmbHLH43). About molecular weight values, the smallest was 10.27 kDa (PmbHLH67) and the largest was 78.26 kDa (PmbHLH45). The instability index varied from 36.78 to 92.34, while only one (PmbHLH66) was considered a stable protein. Predicted GRAVY values were all negative, representing all genes that possessed hydrophilic characteristics. The aliphatic index, (Asp + Glu) value and (Arg + Lys) value of PmbHLHs showed diversity features. Analysis of the gene structure of 95 PmbHLH proteins showed that coding genes of most PmbHLH proteins (92.63%) have introns. Subcellular localization of PmbHLH

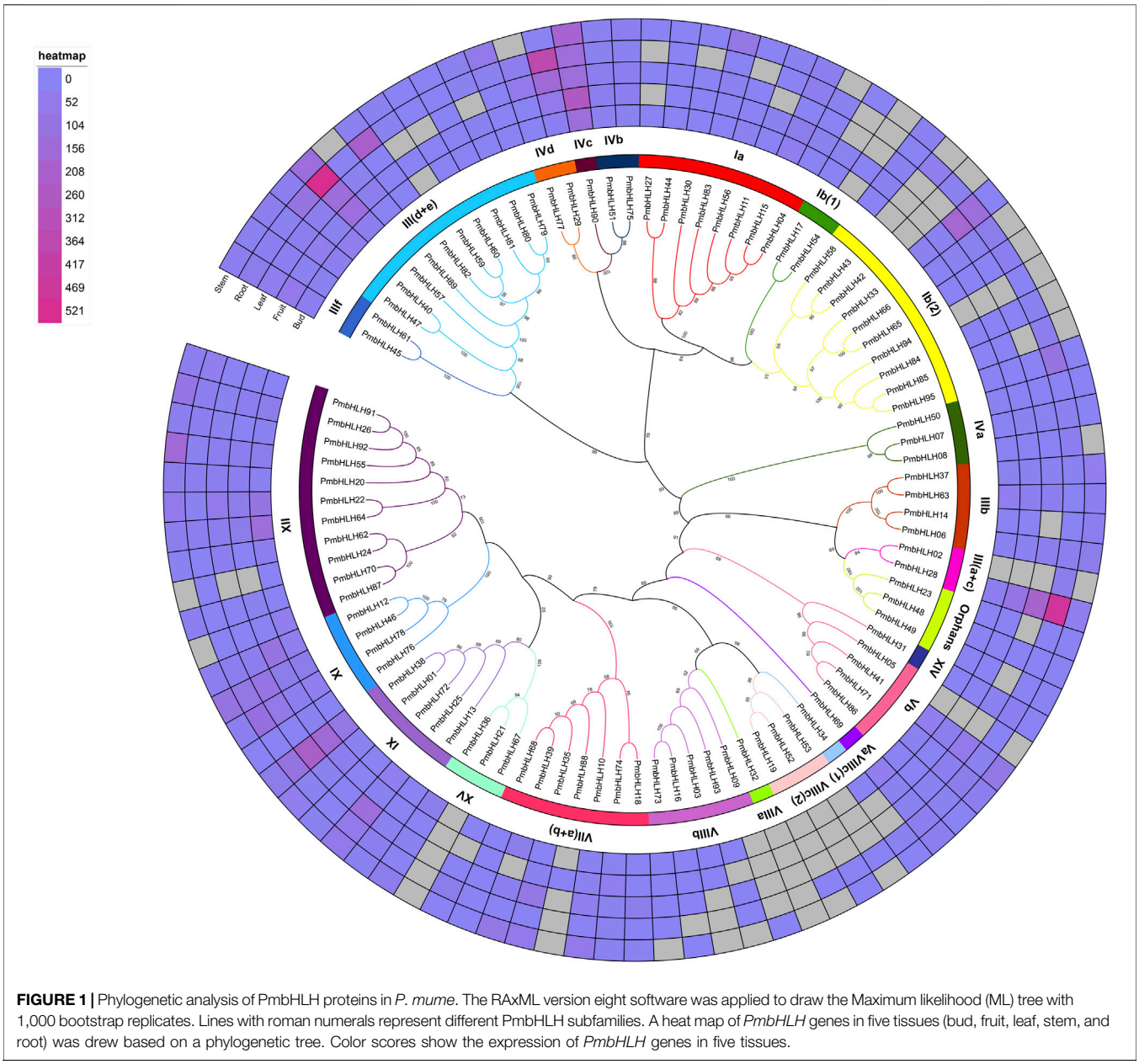


FIGURE 1 | Phylogenetic analysis of PmbHLH proteins in *P. mume*. The RAxML version eight software was applied to draw the Maximum likelihood (ML) tree with 1,000 bootstrap replicates. Lines with roman numerals represent different PmbHLH subfamilies. A heat map of *PmbHLH* genes in five tissues (bud, fruit, leaf, stem, and root) was drew based on a phylogenetic tree. Color scores show the expression of *PmbHLH* genes in five tissues.

proteins were mostly located in the nucleus (**Supplementary Table S1**).

Phylogenetic Analysis, Multiple Sequence Alignment, and DNA-Binding Activity

To research the evolutionary relationship of PmbHLHs, a phylogenetic tree with ML method was established by full-length amino acid sequences of 95 PmbHLHs (**Figure 1**). To clearly describe the classification and potential functions of PmbHLHs, we performed another ML phylogenetic tree with 95 PmbHLH proteins and 162 AtbHLHs (**Supplementary Figure S1** and **Supplementary Table S2**). Based on verified AtbHLHs (Pires and Dolan, 2010), 95 PmbHLH proteins were divided into 23 subfamilies and one orphan (**Figure 1**). XII was the largest subfamily, consisting of 11 members. Five subfamilies [IVc, Va, VIIIa, VIIIc (1), XIV] contained only one bHLH protein. This indicated that PmbHLHs were distributed unevenly in different subfamilies. To date, we know very little about the biological functions of PmbHLHs except for PmICE1 (Cao et al., 2014). However, plenty of bHLH proteins have been confirmed in *Arabidopsis*. Through *Arabidopsis* ortholog analysis, PmbHLHs might play an important role in abiotic stress, hormonal regulation, development, and so on (**Supplementary Table S2**).

The bHLH domain analysis showed that PmbHLH domains consisted of four conserved regions, including the basic region, two helix regions, and a loop region (**Figure 2** and **Supplementary Figure S2**). Additionally, 23 amino acid residues were conserved (>50% consensus ratio) in their bHLH domains. Among residues,

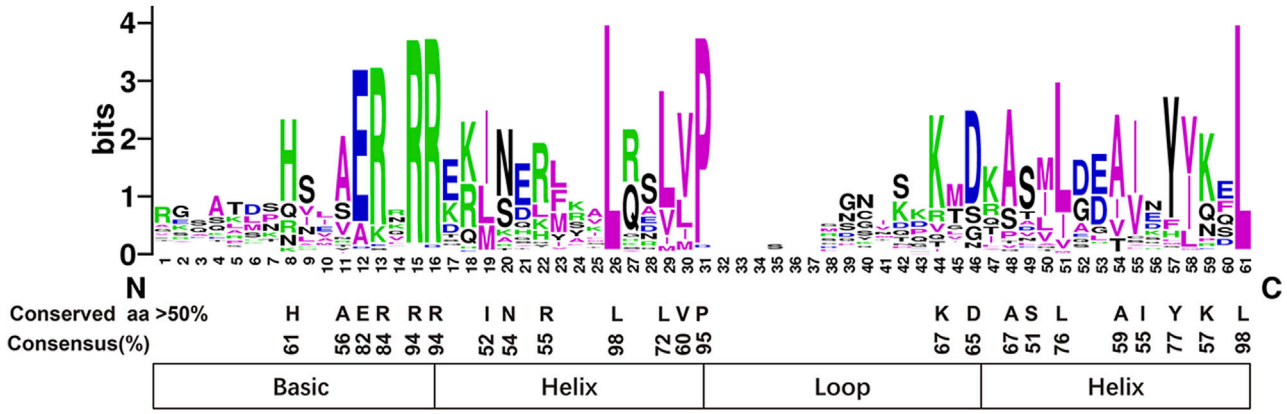


FIGURE 2 | All PmbHLH proteins shown highly conserved in the domain. The overall height of every stack means sequence conservation in the corresponding position. Capital letters represent the conservation of amino acids exceed 50% among 95 PmbHLH domains.

TABLE 1 | Predicted DNA-binding categories based on the bHLH domain.

Predicted activity	Predicted Motif	Number of PmbHLHs	Number of AtbHLHs (Toledo-Ortiz et al.)
DNA binding	—	—	—
E-box	—	—	—
G-box	bHLH	24 (25.26%)	89 (60.54%)
Non-G-box	bHLH	13 (13.68%)	20 (13.61%)
Non-E-box	bHLH	3 (3.16%)	11 (7.48%)
Total	—	40 (42.11%)	120 (81.63%)
Non-DNA binding	HLH	55 (57.89%)	27 (18.37%)

the Arg-15, Arg-16, Leu-26, Pro-31, Leu-61 were highly conserved with a greater than 90% consensus ratio. Among the 23 conserved amino acid residues, the basic region found six conserved residues (His-8, Ala-11, Glu-12, Arg-13, Arg-15, Arg-16), first helix region found seven conserved residues (Ile-19, Asn-20, Arg-22, Leu-26, Leu-29, Val-30, Pro-31), loop region found two conserved residues (Lys-44, Asp-45), second helix region found eight conserved residues (Ala-48, Ser-49, Leu-51, Ala-54, Ile-55, Tyr-57, Lys-59, Leu-61).

The DNA binding activity of target genes was decided by the bHLH domain in the basic region. According to the classification standard in *A. thaliana* (Toledo-Ortiz et al., 2003), PmbHLHs were defined as Non-DNA-binding proteins and DNA-binding proteins (**Figure 2** and **Table 1**). In addition, DNA-binding proteins were divided into E-box-binding proteins (including G-box-binding proteins) and non-E-box binding proteins according to the existence of Glu-12 and Arg-15 (positions were corresponding to positions 13 and 16 in *Arabidopsis*). His/Lys-8, Glu-12. Arg-16 (positions were corresponding to positions 9, 13, and 17 in *Arabidopsis*) are responsible for the binding of the G-box. Based on the conservation of these residues, three PmbHLHs were classified to non-E-box-binding proteins for missing Glu-12/Arg-15 residues and 37 PmbHLHs were classed to E-box-binding proteins. Among 37 E-box-binding proteins, 24 as G-box-binding proteins, while 13 proteins missed the G-box-binding site. Furthermore, 55 of the 95 PmbHLHs were classed in non-DNA-binding proteins due to less than six amino acid residues in the basic region (**Table 1** and **Supplementary Table S3**).

Gene Structure and Conserved Motif

Further to the analysis features of *PmbHLHs*, we investigated intron/exon patterns according to the phylogenetic tree with PmbHLHs sequences (**Figure 3A**). Analysis of genomic DNA sequences showed that number of introns changing from zero to ten (**Figure 3B**). Most of them usually had one to eight introns, except *PmbHLH69* and *PmbHLH92*. Different subfamilies had different intron/exon patterns, while the same classes were similar. Seven genes of 95 *PmbHLHs* were intron-less, five of these were in subfamily VIIIb. Seven subfamilies [Ib (1), IIIf, IVb, Vb, VIIIb, IX, XV] had a concentrated number of exons/introns.

We used MEME online tool to predict thirty conserved motifs of 95 PmbHLH proteins (**Figure 3C** and **Supplementary Table S4**). The number of PmbHLHs motifs was distinctive, ranging from 1 to 16. Most of PmbHLHs shared three to six motifs. Each PmbHLH protein contained motif one and motif two except for PmbHLH31, PmbHLH58, and PmbHLH69. Different subfamilies had unique motif combinations. For instance,

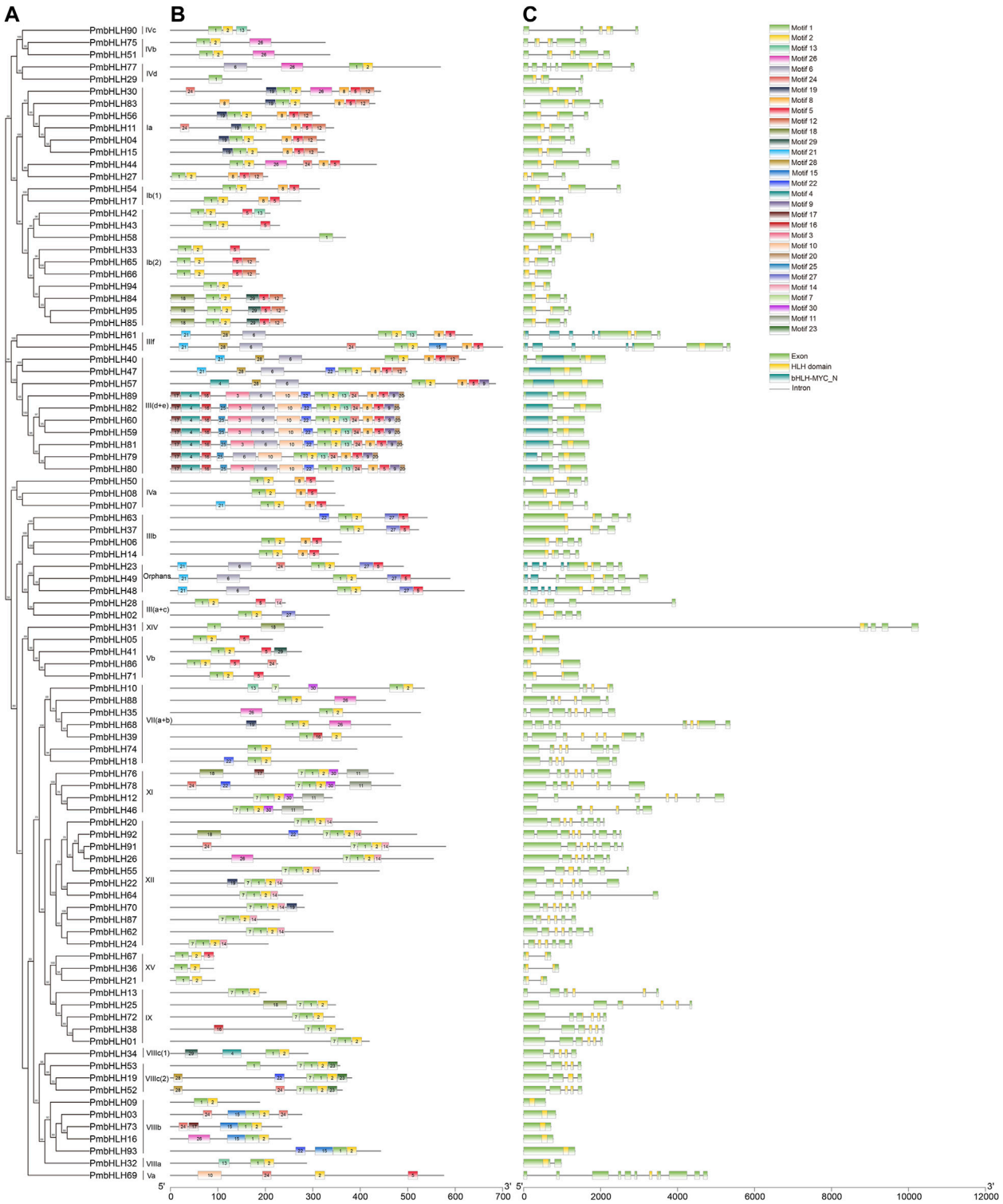


FIGURE 3 | Phylogenetic relationship, motifs analysis, and gene structure in PmbHLH. **(A)** The RAxML version eight software was applied to construct a phylogenetic tree based on PmbHLH proteins. **(B)** Motif composition analysis of PmbHLHs by MEME. Different colorful rectangles with numbers 1–30 represent different motifs. **(C)** Exon-intron structure of PmbHLHs. Grey lines represent introns and green boxes represent exons. Yellow rectangles indicate the position of PmbHLHs conserved domain.

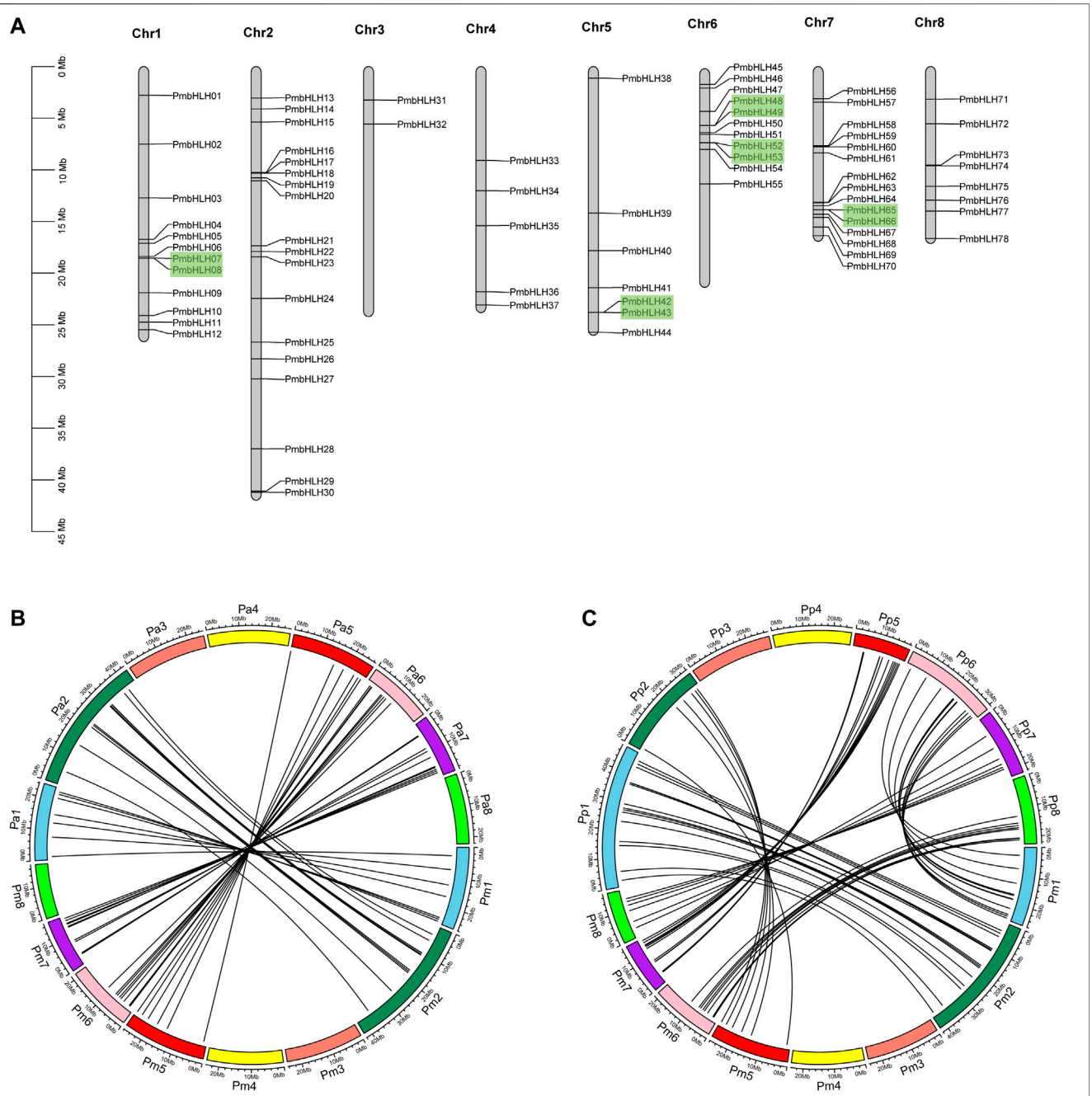


FIGURE 4 | Chromosomal mapping of PmbHLHs and synteny analysis in P. mume (Pm),P. avium (Pa) and P. persica (Pp). **(A)** Chromosomal distribution of *PmbHLH* genes. 78 *PmbHLH* genes were unevenly mapped on eight chromosomes. Tandem duplicated gene pairs are displayed with green blocks. **(B)** Synteny analysis in *P. mume* and *P. avium*. The same color blocks represent the same chromosome label in *P. mume* and *P. avium*. Black lines mean orthologous bHLH in *P. mume* and *P. avium*. **(C)** Synteny analysis in *P. persica* and *P. mume*. The same color blocks represent the same chromosome label in *P. persica* and *P. mume*. Black lines mean orthologous bHLH in *P. persica* and *P. mume*.

subfamily IVb contained motifs 1, 2, and 26, and subfamily Ib (1) contained motifs 1, 2, 5, and 8. Some motifs were unique and existed in only one subfamily. Subfamily III (d + e) contained the most motifs and motif 9, 20, and 25 were specific to it. Motif 11, 23, and 30 were respectively observed in subfamily XI, VIII (2), and XI. These conserved motifs may play special functions. The same subfamily had similar motifs, implying these PmbHLHs might have similar functions. In contrast, the kinds of motifs showed little difference in the same subfamily. For example, PmbHLH07 had motif 21 except common motifs 1, 2, 5, and eight in subfamily IVa.

Chromosomal Distribution and Synteny Analysis

According to genome annotation information, 78 *PmbHLH* genes were mapped on eight chromosomes, while 17 *PmbHLH* genes were localized on unassembled genomic scaffolds (**Figure 4A** and **Supplementary Figure S3**). The distribution of *PmbHLH* genes on each chromosome was irregular. 18 *PmbHLH* genes (18.95%) were present on chromosome 2, which was the maximum number, whereas only two *PmbHLH* genes (2.11%) were located on chromosome 3. The proportion of *PmbHLH* genes on half of the chromosomes accounted for more than 10% but less than 20%.

Duplication events contained genome duplication, tandem duplication, segmental duplication, and transposon duplication, leading to plant evolution (Qiao et al., 2019). Duplication played an important role in *PmbHLH* gene expansion. Among 95 *PmbHLHs*, nine pairs of *PmbHLH* genes were described as tandem duplication, mapping on chromosome 1, 5, 6, 7, and scaffolds (**Figure 4A** and **Supplementary Figure S3**). The largest number of tandem duplications were distributed on Chromosome 6. By contrast, segmental duplication was absent in the *PmbHLH* gene family, which signified segmental duplication was not involved in gene expansion. The selection pressure of gene duplications was estimated by mutation rates of Ka (nonsynonymous) versus Ks (synonymous). We calculated the Ka/Ks of the *PmbHLH* gene family (**Supplementary Table S5**). Results showed that most of the Ka/Ks values of tandem duplication were <1 and changed from 0.270 to 0.949, implying a purifying selection during *PmbHLH* genes expansion. However, *PmbHLH82-PmbHLH83* and *PmbHLH85-PmbHLH86* only existed Ka value, which means that two gene pairs evolved by natural selection. The divergence time of *PmbHLHs* tandem duplication ranged from 3.67 to 64.06 Mya.

To further explore the evolutionary mechanism of the *PmbHLH* gene family, we constructed a syntenic map of *P. mume* associated with *P. avium* and *P. persica*. 41 syntenic orthologous gene pairs were distinguished between *P. avium* and *P. mume*, 59 pairs between *P. persica* and *P. mume* (**Figure 4B** and **Figure 4C**). This indicates that the *P. avium*, *P. persica,* and *P. mume* had a close relationship. Interestingly, we found that one *PmbHLH* gene only corresponds to one gene in these syntenic orthologous gene pairs. For further evolutionary studies, the divergence time of *bHLH* gene pairs was calculated in three varieties (**Supplementary Table S6**). Divergence time started 86.11 Mya to 0.67 Mya between *P. persica* and *P. mume*, and 1-2 Mya occurred in most duplicated events. In *P. avium* and *P. mume*, it began 60.81 Mya to 0.14 Mya and 0.5-2 Mya occurred in most of the duplicated events, which may indicate that the speciation time of these orthologous pairs was shorter in the two varieties. Additionally, higher syntenic genes appeared on chromosomes 2, 6, and 7.

Expression Profile of *PmbHLHs*

Based on the FPKM values from RNA sequencing data, we investigated the expression pattern of *PmbHLHs* among different tissues and different low-temperature treatments. The heatmap of **Figure 1** and **Supplementary Figure S5** show that the *PmbHLH* gene family presented clear tissue-specific expression. Among 95 *PmbHLHs*, 53 genes were expressed in five tissues (bud, fruit, leaf, stem, and root), implying that these *PmbHLHs* may participate in the development and growth process of tissues. While, *PmbHLH21* and *PmbHLH32* (they belonged to subfamily XV and VIIa, respectively) lacked expression in all detected tissues. We discovered that nine *PmbHLHs* were expressed in only one tissue. They included two genes (*PmbHLH54* and *PmbHLH88*) only in fruit, seven genes (*PmbHLH02*, *PmbHLH05*, *PmbHLH19*, *PmbHLH42*, *PmbHLH43*, *PmbHLH53* and *PmbHLH94*) only in root. This probably indicated that the nine genes have special functions in fruit or root. Moreover, most of remaining *PmbHLHs* were expressed in three or four detected tissues (**Supplementary Table S7**). According to the phylogenetic tree analysis, *PmbHLHs* from the same subfamily had a similar expression pattern (**Figure 1**). For instance, IX subfamily (included *PmbHLH01*, *PmbHLH13*, *PmbHLH25*, *PmbHLH38* and *PmbHLH72*) exhibited expression in all detected tissues. Furthermore, Ib (2) subfamily genes (included *PmbHLH42*, *PmbHLH43*, *PmbHLH65*, *PmbHLH66*, *PmbHLH84*, *PmbHLH85*, *PmbHLH94,* and *PmbHLH95*) were highly expressed in root, while *PmbHLH33* and *PmbHLH58* had a high expression level in leaf and stem respectively, implying that the functions of genes in the same family gradually differ in the evolution process.

Previous studies have proved that the cold resistance of apricot mei is stronger than that of true *mume*. Therefore, annual plants of apricot mei 'Danfenghou' and true *mume* 'Beijing Yudie' were treated with low temperatures. Based on transcriptome data, this study analyzed differences in the expression level of the *PmbHLHs* gene family during these three periods. FPKM value greater than one is an effective expression. 62 genes that were effectively expressed after low-temperature stress were detected. The FPKM value of these genes was plotted as a heat map.

As shown in **Figure 5**, according to different expression patterns in different periods, *PmbHLHs* genes in the two varieties were clustered into four groups (**Supplementary Table S8**). *PmbHLH* genes of the 'Danfenghou' group Ⅰ were highly expressed during freezing treatments and had a similar expression pattern to the 'Beijing Yudie' group Ⅲ. There are six genes (*PmbHLH64*, *PmbHLH25*, *PmbHLH10*, *PmbHLH78*, *PmbHLH31,* and *PmbHLH47*) in the two varieties with similar expression patterns, which can respond to freezing stress. The expression level of *PmbHLHs* in the 'Danfenghou' group Ⅱ and 'Beijing Yudie' group Ⅰ gradually decreased with the extension of treatment time. The expression profiles of *PmbHLHs* in the 'Danfenghou' group Ⅲ and 'Beijing Yudie' group Ⅱ were similar. 14 genes increased expression levels when two varieties were treated at 4°C for 8 h. They might be involved in the perception and transport of cold signals. The expression levels of 'Danfenghou' group Ⅱ and 'Beijing Yudie' group Ⅰ both increased at 4°C for 5 days, but there were no overlapping genes in the two cultivars. This indicated that the function of *PmbHLH* genes in the two cultivars after prolonged low-temperature acclimation has a difference.

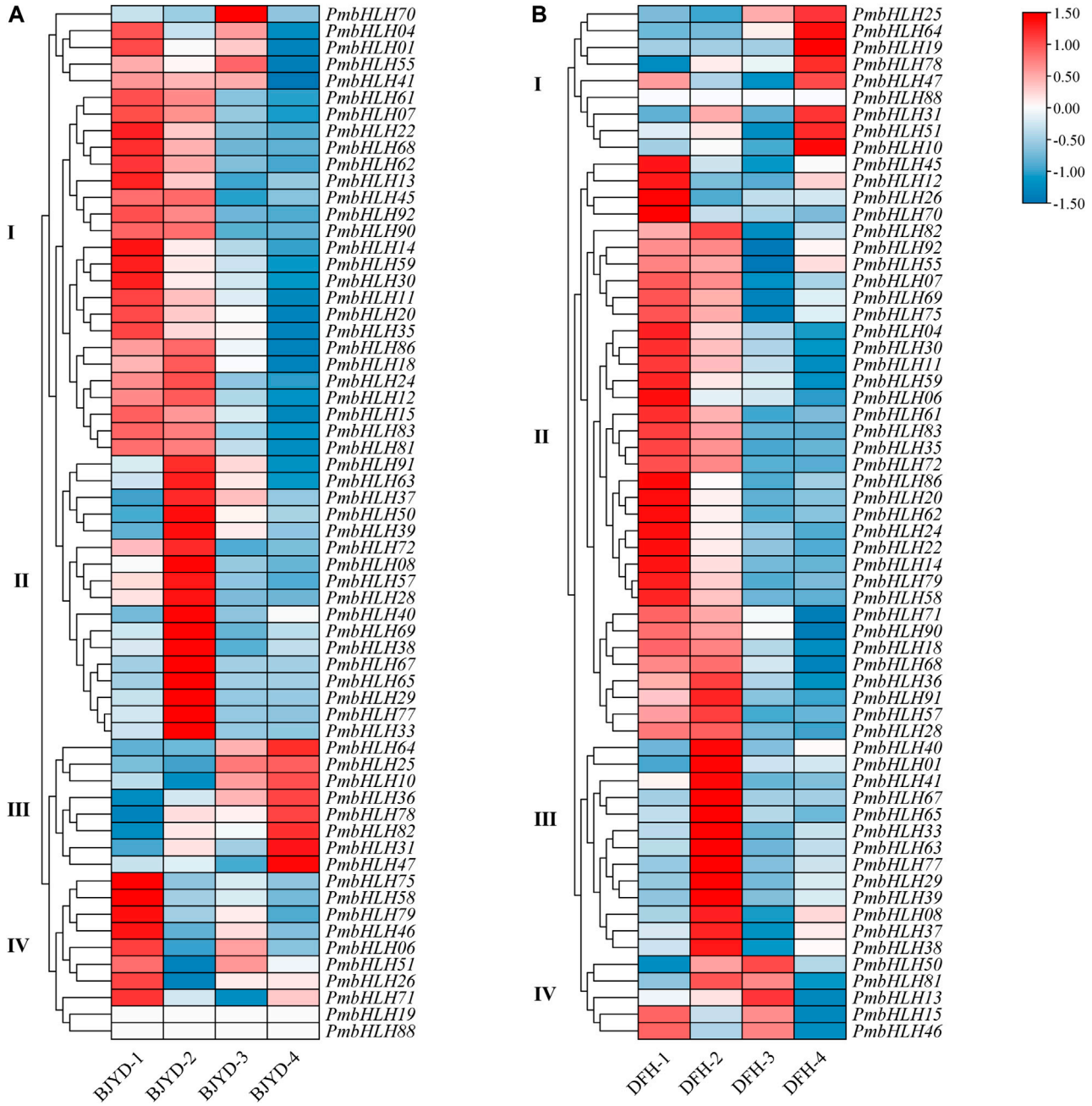


FIGURE 5 | Hierarchical clustering of the expression profile of PmbHLHs in 'Beijing Yudie' **(A)** and 'Danfenghou' **(B)** under low-temperature treatment. "1" represents stems that were treated at 25°C, "2" represents stems that were treated at 4°C for 8 h, "3" represents stems that were treated at 4°C for 5 days, "4" represents stems that were treated at −5°C for 1 h. Heat maps were generated with FPKM values. Colorful scale means relative expression level and is shown at the top. Red represents high expression and blue represents low expression. Hierarchical clustering groups are displayed by roman numerals.

According to FPKM value and fold log2 change value, 10 *PmbHLH* genes were screened out. Compared with control material that has not been treated with low temperature, the transcriptional expression levels of six *PmbHLHs* genes (*PmbHLH25*, *PmbHLH28*, *PmbHLH38*, *PmbHLH40*, *PmbHLH57,* and *PmbHLH78*) increased by 2–5 times at chilling treatment (4°C for 8 h and 5 days), which may be involved in the response of *P. mume* to chilling treatment. The expression levels of *PmbHLH4*, *PmbHLH6*, *PmbHLH26,* and *PmbHLH46* gradually decreased with the extension of 4°C treatment time. Compared with the control, the highest reduction factor reached 5 times. They might negatively regulate downstream low-temperature response genes or proteins to participate in the chilling stress response of *P.*

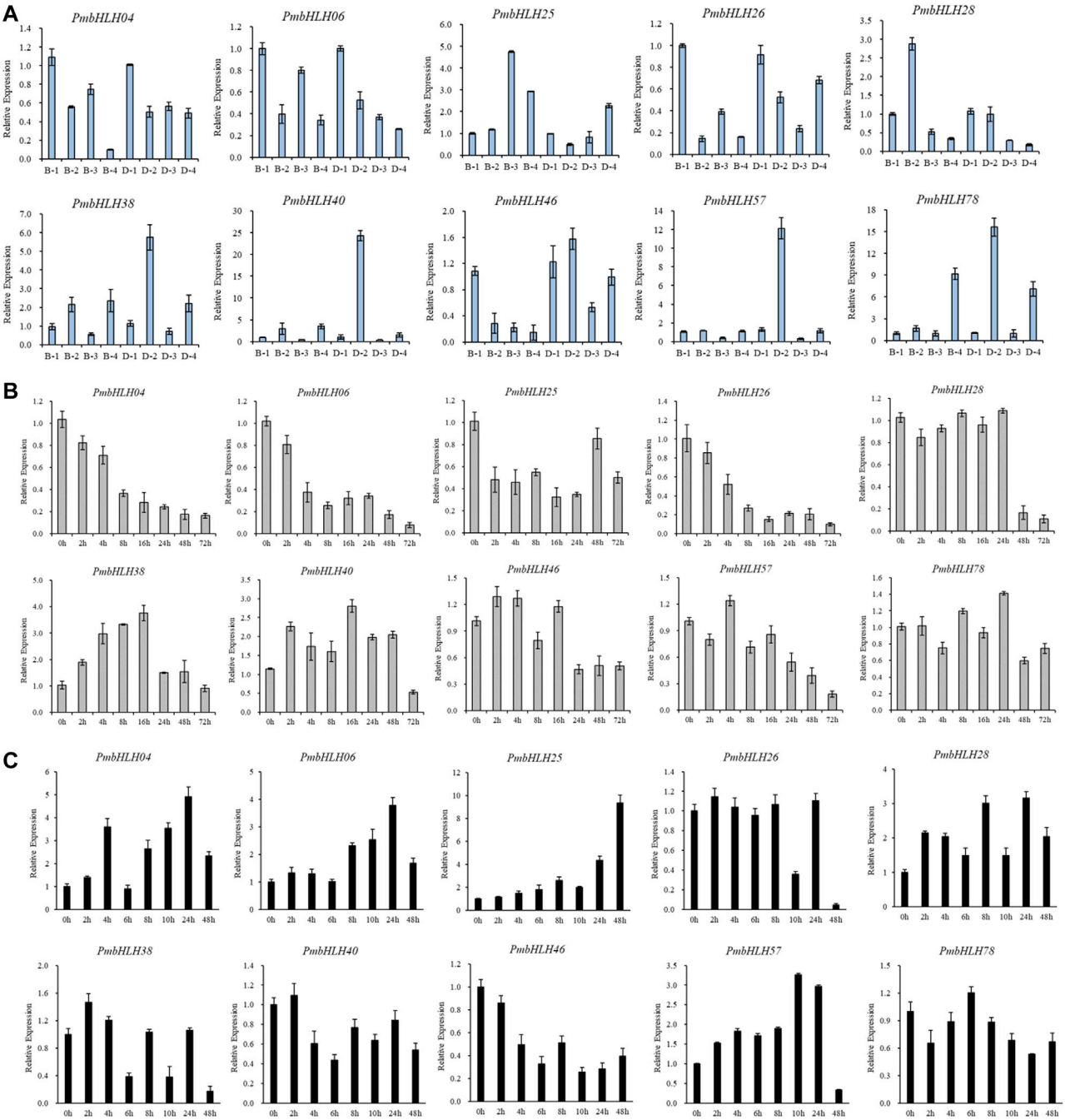


FIGURE 6 | Expression patterns of 10 candidate PmbHLHs in 'Beijing Yudie' and 'Danfenghou' under low-temperature treatments by qRT-PCR. **(A)** Expression patterns under cold treatment (4°C for 8 h and 5 days) and freezing treatment (–5°C for 1 h). "1" represented stems that were treated at 25°C, "2" represented stems that were treated at 4°C for 8 h, "3" represented stems that were treated at 4°C for 5 days, "4" represented stems that were treated at –5°C for 1 h. **(B)** Expression patterns exposed to cold treatment for different times (0/2/4/8/16/24/48/72 h). **(C)** Expression patterns exposed to freezing treatment for different times (0/2/4/6/8/10/24/48 h). The standard deviation of three independent replicates was represented by Error bars.

mume. We used transcriptome to further analyze the expression patterns of *PmbHLHs* genes in different varieties. The trend of expression levels with these 10 differential genes showed the same in strong cold-resistant 'Danfenghou' and weaker cold-resistant 'Beijing Yudie'. However, there were differences in multiple variations among different varieties. For example, the expression of the *PmbHLH28* gene in 'Beijing Yudie' was 4.8 times higher than that in the control, but only 1.1 times higher in the 'Danfenghou' group.

Expression Analysis of *PmbHLHs* Under Low-Temperature Stress

To further investigate *PmbHLHs* function in low temperature comprehensively, the 10 genes were detected by qRT-PCR experiments in which the stems of *P. mume* were treated for chilling treatment (4°C for 8 h and 5 days) and freezing treatment (−5°C for 1 h). The expression of *PmbHLH* genes was distinct in two varieties of *P. mume* (**Figure 6A**). In true *mume* 'Beijing Yudie', six genes were up-regulated in varying degrees under low-temperature stress. Remaining genes were down-regulated. Among up-regulated genes, *PmbHLH25*, *PmbHLH40*, *PmbHLH46,* and *PmbHLH57* were highly expressed under 4°C and −5°C treatment. The greatest expression of *PmbHLH25* and *PmbHLH40* was found in 4°C treatments only. In addition, low temperature induced most of the genes to change significantly in apricot mei 'Danfenghou'. *PmbHLH46* and *PmbHLH25* were up-regulated at 4°C treatment and at −5°C treatment, respectively. *PmbHLH40* and *PmbHLH57* were highly expressed under 4°C and −5°C treatment. In two varieties, four genes (*PmbHLH25*, *PmbHLH38*, *PmbHLH40,* and *PmbHLH78*) could be induced to high expression in true *mume* 'Beijing Yudie' and apricot mei 'Danfenghou'. In addition, the expression of four *PmbHLHs* (*PmbHLH38*, *PmbHLH40*, *PmbHLH57,* and *PmbHLH78*) in apricot mei 'Danfenghou' higher than in true *mume* 'Beijing Yudie', further showing that apricot mei 'Danfenghou' was more resistant than true *mume* 'Beijing Yudie'.

To explore the regulation process of *PmbHLHs*, we tested the expression pattern of 10 genes by qRT-PCR at chilling treatment (0/2/4/8/16/24/48/72 h) and freezing treatment (0/2/4/6/8/10/24/48 h) for different periods using true *mume* 'Beijing Yudie'. As shown in **Figure 6B**, the expression level of *PmbHLH04*, *PmbHLH25*, *PmbHLH26*, *PmbHLH46,* and *PmbHLH57* were reduced with chilling treatment. However, *PmbHLH38* and *PmbHLH40* were induced to high expression and up-regulation expression peaked on treating with 16 h. Under freezing treatment, the expression level of *PmbHLH26*, *PmbHLH40,* and *PmbHLH78* changed slightly. *PmbHLH38* and *PmbHLH40* presented a trend of decreasing. After prolonging the freezing treatment time, *PmbHLH04*, *PmbHLH06*, *PmbHLH25*, *PmbHLH28,* and *PmbHLH57* were induced to high expression. Especially, *PmbHLH25* had the largest expression level increase (approximately 10-fold) at 48 h after being exposed to freezing conditions (**Figure 6C**). These results imply that *PmbHLH* genes may play a role in resisting low-temperature stress.

Promoter *Cis*-Acting Elements and Protein Interaction Analysis

The promoter *cis*-elements (1500 bp) were analyzed through PlantCARE software and the predicted regulation mechanisms of *PmbHLHs*. The result showed that *PmbHLH* genes were abundant in abiotic and biotic elements (LTR, MBS, WRE3, WUN-motif, ARE), light-responsive elements (AAAC-motif, ATCT-motif, LAMP-element, Box 4, I-box, and G-box), plant growth and development-related elements (MSA-like, GCN4_motif, and RY-element). In addition, hormone-responsive elements (ABRE, TGACG-motif, G-box, MYC, P-box, and TATC-box) exhibited a wide range of positions in the promoter. This means that *PmbHLHs* may extensively participate in various physiological biochemistry pathways of *P. mume* (**Supplementary Figure S4**).

To further predict the functions of PmbHLHs, we constructed the interaction network using AraNet V2 based on homologous proteins of *Arabidopsis* (Lee et al., 2015). A total of 56 PmbHLH proteins had orthologs in *Arabidopsis* and predicted about 640 interaction protein pairs (**Figure 7** and **Supplementary Table S9**). The interaction network showed that PmbHLHs might interact with MYB, bHLH, NAC, bZIP, HB, WRKY, ERF, and so on, implying PmbHLHs might exert functions by interacting with other genes. The above results predicted that PmbHLHs might be involved in the response of *P. mume* to low temperature. To further research ten PmbHLHs, we constructed and analyzed the interaction network of candidate ten genes. Among the ten PmbHLHs, PmbHLH04, PmbHLH26, and PmbHLH38 were absent in homologous proteins of *Arabidopsis*. The three candidate genes might be novel and potential proteins in response to low temperature. The other seven candidate PmbHLHs formed an interactive network centered on PmbHLH40, which showed high homology to JAM2. Moreover, they might strongly interact with MYB124, TIFY, NAC, WRKY, and IAA to drive function when suffered from low-temperature stress in *P. mume*. **Supplementary Figure S6** shows that PmbHLH25 (AKS2) could directly interact with PmbHLH28 (AT5G57150). PmbHLH06 (bHLH93) could directly interact with PmbHLH38 (FBH4). This indicated that these PmbHLH might work by forming dimers under low-temperature stress. Overall, interaction networks could provide a crucial reference for investigating the regulation mechanism of PmbHLHs.

DISCUSSION

The bHLH transcription factor family plays a positive or negative role in the physiological and biochemical processes of the development of plant trichome, root hair, photomorphogenesis, light signal transmission, and development of plant tissues and organs. In addition, the transcription factor family has contributed to resisting adverse environmental factors in plants, such as drought resistance, salt tolerance, cold tolerance, plant iron deficiency stress, and so on (Toledo-Ortiz et al., 2003). A great many *bHLH* genes have been distinguished in the plant kingdom, including *A. thaliana* (162) (Pires and Dolan, 2010), *Hibiscus hamabo* (162) (Ni et al., 2021), *Helianthus annuus* (183) (Li et al., 2021), *Capsicum annuum* (107) (Liu et al., 2021), *Sorghum bicolor* (174) (Fan et al., 2021), *Osmanthus fragrans* (206) (Li et al., 2020), *Camellia sinensis* (134) (Liu et al., 2021), *Juglans regia* (102) (Zhao et al., 2021), *Pyrus bretschneideri* (197) (Dong et al., 2021). The number of bHLH transcription factors changed widely among various plants. However, the characteristic of bHLH genes remains unknown in *P. mume*. In our research, we used *P. mume* genome to distinguish 95 *PmbHLH* genes. The number of *PmbHLHs* was

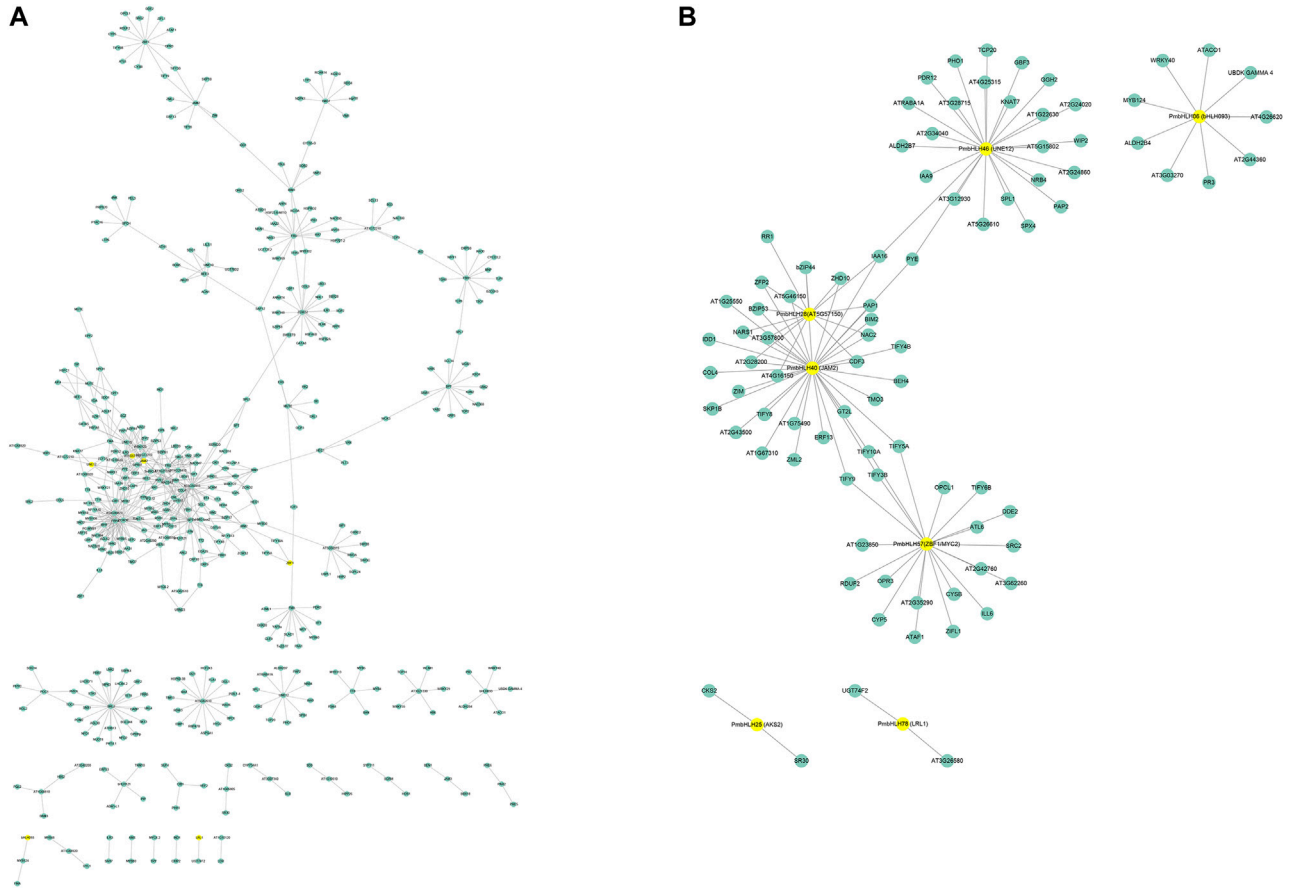


FIGURE 7 | The interaction network for bHLHs in *P. mume* based on orthologs in *Arabidopsis*. **(A)** The whole interaction network for PmbHLHs. **(B)** Protein interaction network of seven candidate PmbHLHs. Yellow circles represent PmbHLHs and blue circles represent other genes.

the same as bHLH in *P. persica*, which further proved that the genetic relationship of *P. persica* and *P. mume* was closer.

PmbHLHs were inhomogeneous on eight chromosomes and scaffolds. These results showed that the bHLH gene family existed specific evolution patterns in different varieties. In a previous study, it was suggested that the expansion of bHLH might derive from gene duplication during evolution with a high proportion of segmental duplications and tandem duplications (Cannon et al., 2004; Kavas et al., 2016). In *P. mume*, nine pairs of *PmbHLH* genes were described as tandem duplication (18.95%) and absent in segmental duplications. Tandem duplication possibly promoted *PmbHLHs* expansion. The rate of gene duplication in this study was lower than in other plants, suggesting gene duplication was the secondary formation in *PmbHLH* gene expansion or *PmbHLHs*, and that loss-functions or redundancies may be lost during evolution. This conclusion was similar to the *PmWRKY* genes family (Bao et al., 2019).

The bHLH domain consists of the basic region, two helix regions, and a loop region (Massari and Murre, 2000). DNA binding activity was determined by the basic region of the bHLH domain and the HLH region was essential in homodimer or heterodimer formation (Carretero-Paulet et al., 2010). In the present study, 23 amino acid residues in the PmbHLHs domains were conserved (>50% consensus ratio). Among them, residues Arg-15, Arg-16, Leu-26, Pro-31, Leu-61 (corresponding to Arg-16, Arg-17, Leu-27, Pro-32, Leu-61 in AtbHLHs) showed highly conserved with greater than 90% consensus ratio. Particularly, residues Leu-26 and Leu-61 were more conservative (**Figure 2**). DNA binding activity in the basic region was decided by greater than five basic amino acid residues (Toledo-Ortiz et al., 2003). According to the classification standard, PmbHLHs were defined as the Non-DNA-binding proteins and DNA-binding proteins (**Figure 2**). Furthermore, Glu-13 could specifically identify E-box and the position of Glu-13 was stabilized through Arg-16. And the existence of His/Lys-9, Glu-13, and Arg-17 could discern G-box binding motif. Therefore, DNA-binding proteins were divided into non-G-box binding proteins, G-box-binding proteins, and non-E-box binding proteins based on these factors. In addition, Leu-27 and Leu-61 played an important role in protein interaction in the helix region (Carretero-Paulet et al., 2010). 98% PmbHLHs had Leu-26 and Leu-61 (positions were equivalent to positions 27 and 61 in *Arabidopsis*), implying that PmbHLHs possessed dimerization capacity.

Expression profiles of genes could present their functions. Therefore, expression patterns of *PmbHLHs* in five tissues

were analyzed. The expression pattern of *PmbHLHs* represented clear tissue-specific expression based on transcriptome data (**Figure 1**). Among them, *PmbHLH17* and *PmbHLH54* had a high expression level in fruit, suggesting that they might associate with embryonic development. *RGE1* had *PmbHLH17* and *PmbHLH54* homologous proteins in *Arabidopsis*, which could result in the retarded growth of embryos (Kondou et al., 2008). Meanwhile, *PmbHLH27*, *PmbHLH30,* and *PmbHLH44* were expressed in leaf and stem. The homolog *FAMA*, *SPCH,* and *MUTE* could control meristem differentiation during stomatal development (Ohashi-Ito and Bergmann, 2006; Pillitteri et al., 2007; Marcos et al., 2017). In subfamily IIIb, *PmbHLH37* was expressed highly in stem, homolog, and *AtICE* was implicated in cold acclimation response and freezing tolerance (Chinnusamy, 2003). Additionally, *PmbHLH02* were expressed in the root and the expression of *AtFIT* was up-regulated when it suffered from iron deficiency stress in *Arabidopsis* roots (Ling et al., 2002). These results contributed to predicting the functional regions and further understanding the functions of *PmbHLHs*.

Low temperature is an important environmental factor affecting plant yield and distribution. When plants suffer from adversity stress at low temperatures, they could feel low-temperature signals to produce a series of physiological and biochemical reactions and regulate gene expression (Zhu, 2016). Previous reports have demonstrated that the *bHLH* transcription factor had a crucial influence in resisting low-temperature stress (Nakamura et al., 2011). In the present study, we distinguished 10 *PmbHLH* genes that might possess low-temperature resistance based on transcriptome data. By analyzing qRT-PCR, the expression of *PmbHLH* genes was distinct in two varieties of *P. mume*. Among up-regulated *PmbHLHs*, four genes could be induced to express in true *mume* 'Beijing Yudie' and apricot mei 'Danfenghou'. Simultaneously, the expression of four *PmbHLHs* in apricot mei 'Danfenghou' was higher than in true *mume* 'Beijing Yudie', further showing that apricot mei 'Danfenghou' was more resistant than true *mume* 'Beijing Yudie'. The remaining four genes were down-regulated when they suffer low-temperature stress. To further research the regulation mode of the 10 *PmbHLHs*, fine regulation detection was carried out. The result showed that *PmbHLHs* were more strongly expressed in −5°C treatment than 4°C treatment. In the meantime, the expression of *PmbHLHs* increased with the prolongation of −5°C treatment time, suggesting *PmbHLHs* could respond to deep freezing. Especially *PmbHLH25,* which possibly has great potential in breeding frost resistance. It was noteworthy that *AtAKS2* (ABA-responsive kinase substrates 2) was *PmbHLH25* homologous proteins in *Arabidopsis*. ABA could induce phosphorylation of *AKS* to change stomatal opening or close, then enable plants to adapt to changing environmental conditions (Takahashi et al., 2013). We speculated that *PmbHLH25* might play an essential role in the process of low-temperature signal perception and transduction.

CONCLUSION

In our study, we identified 95 *PmbHLHs* from *P. mume* genome and comprehensively analyzed their characterization, including gene identification, gene structure, conserved motifs, DNA binding activity, and chromosomal distribution, protein interaction, synteny analysis, and expression profiling. According to the phylogenetic tree, 95 PmbHLHs were classified into 23 subfamilies. The protein interaction and synteny analysis further expounded the potential functions and evolutionary mechanisms of *PmbHLHs*. Moreover, tissue-specific expression revealed that *PmbHLHs* might widely participate in the development of tissues. Combining with the qRT-PCR date, *PmbHLH04, PmbHLH06, PmbHLH25, PmbHLH28, PmbHLH38, PmbHLH40,* and *PmbHLH57* may play a major role in resisting low-temperature stress. These results provide a molecular basis and valuable insights for further studying the functions of *PmbHLHs* in regulating low temperature in *P. mume*.

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

QZ conceived and designed the experiments. FB revised the manuscript. AD and AD performed data analysis and experiments. AD drafted the manuscript. PL, JW, and TC contributed reagents, materials, and analysis tools. All authors have read and approved the final manuscript.

FUNDING

This research was funded by the National Key R and D Program of China (2019YFD1001500), the open funds of the State Key Laboratory of Plant Physiology and Biochemistry (Grant/Award Number: SKLPPBKF 2005) and Special Fund for Beijing Common Construction Project.

ACKNOWLEDGMENTS

We greatly appreciate the Frontiers editors and reviewers for handling our manuscript and providing critical suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.762135/full#supplementary-material

REFERENCES

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., et al. (2012). ExPASy: SIB Bioinformatics Resource portal. *Nucleic Acids Res.* 40, W597–W603. doi:10.1093/nar/gks400

Atchley, W. R., and Fitch, W. M. (1997). A Natural Classification of the Basic helix-loop-helix Class of Transcription Factors. *Proc. Natl. Acad. Sci.* 94, 5172–5176. doi:10.1073/pnas.94.10.5172

Atchley, W. R., Terhalle, W., and Dress, A. (1999). Positional Dependence, Cliques, and Predictive Motifs in the bHLH Protein Domain. *J. Mol. Evol.* 48, 501–516. doi:10.1007/pl00006494

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

Bao, F., Ding, A., Cheng, T., Wang, J., and Zhang, Q. (2019). Genome-Wide Analysis of Members of the *WRKY* Gene Family and Their Cold Stress Response in *Prunus Mume*. *Genes* 10, 911. doi:10.3390/genes10110911

Buti, S., Hayes, S., and Pierik, R. (2020). The bHLH Network Underlying Plant Shade-avoidance. *Physiol. Plantarum* 169, 312–324. doi:10.1111/ppl.13074

Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., and May, G. (2004). The Roles of Segmental and Tandem Gene Duplication in the Evolution of Large Gene Families in *Arabidopsis thaliana*. *Bmc Plant Biol.* 4, 10. doi:10.1186/1471-2229-4-10

Cao, N., Zhang, Q. X., Hao, R. J., Xu, Z. D., Wang, T., and Yang, W. R. (2014). Molecular Cloning and Expression Analysis of Cold-Resistant Transcription Factor PmICE1 from *Prunus Mume*. *J. Northeast. For. Univ* 42, 21–25. doi:10.13759/j.cnki.dlxb.2014.04.005

Carretero-Paulet, L., Galstyan, A., Roig-Villanova, I., Martínez-García, J. F., Bilbao-Castro, J. R., and Robertson, D. L. (2010). Genome-Wide Classification and Evolutionary Analysis of the bHLH Family of Transcription Factors in Arabidopsis, Poplar, Rice, Moss, and Algae. *Plant Physiol.* 153, 1398–1412. doi:10.1104/pp.110.153593

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

Chinnusamy, V. (2003). ICE1: A Regulator of Cold-Induced Transcriptome and Freezing Tolerance in *Arabidopsis*. *Gene Dev.* 17, 1043–1054. doi:10.1101/gad.1077503

Clamp, M., Cuff, J., Searle, S. M., and Barton, G. J. (2004). The Jalview Java Alignment Editor. *Bioinformatics* 20, 426–427. doi:10.1093/bioinformatics/btg430

de Marcos, A., Houbaert, A., Triviño, M., Delgado, D., Martín-Trillo, M., Russinova, E., et al. (2017). A Mutation in the bHLH Domain of the SPCH Transcription Factor Uncovers a BR-dependent Mechanism for Stomatal Development. *Plant Physiol.* 174, 823–842. doi:10.1104/pp.17.00615

Ding, W., Yu, Z., Tong, Y., Huang, W., Chen, H., and Wu, P. (2009). A Transcription Factor with a bHLH Domain Regulates Root Hair Development in rice. *Cell Res* 19, 1309–1311. doi:10.1038/cr.2009.109

Dong, H., Chen, Q., Dai, Y., Hu, W., Zhang, S., and Huang, X. (2021). Genome-wide Identification of PbrbHLH Family Genes, and Expression Analysis in Response to Drought and Cold Stresses in Pear (*Pyrus Bretschneideri*). *BMC Plant Biol.* 21, 86. doi:10.1186/s12870-021-02862-5

Edgar, R. C. (2004). MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res.* 32, 1792–1797. doi:10.1093/nar/gkh340

Fan, Y., Yang, H., Lai, D., He, A., Xue, G., Feng, L., et al. (2021). Genome-wide Identification and Expression Analysis of the bHLH Transcription Factor Family and its Response to Abiotic Stress in Sorghum [*Sorghum Bicolor* (L.) Moench]. *BMC Genomics* 22, 415. doi:10.1186/s12864-021-07652-9

Feng, X.-M., Zhao, Q., Zhao, L.-L., Qiao, Y., Xie, X.-B., Li, H.-F., et al. (2012). The Cold-Induced Basic helix-loop-helix Transcription Factor Gene MdCIbHLH1encodes an ICE-like Protein in Apple. *Bmc Plant Biol.* 12, 22. doi:10.1186/1471-2229-12-22

Gao, F., Robe, K., Bettembourg, M., Navarro, N., Rofidal, V., Santoni, V., et al. (2020). The Transcription Factor *bHLH121* Interacts with bHLH105 (ILR3) and its Closest Homologs to Regulate Iron Homeostasis in *Arabidopsis*. *Plant Cell* 32, 508–524. doi:10.1105/tpc.19.00541

Groszmann, M., Paicu, T., and Smyth, D. R. (2008). Functional Domains of *SPATULA*, a bHLH Transcription Factor Involved in Carpel and Fruit Development in *Arabidopsis*. *Plant J.* 55, 40–52. doi:10.1111/j.1365-313X.2008.03469.x

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: Protein Localization Predictor. *Nucleic Acids Res.* 35, W585–W587. doi:10.1093/nar/gkm259

Huq, E., and Quail, P. H. (2002). PIF4, a Phytochrome-Interacting bHLH Factor, Functions as a Negative Regulator of Phytochrome B Signaling in *Arabidopsis*. *Embo J.* 21, 2441–2450. doi:10.1093/emboj/21.10.2441

Kavas, M., Baloğlu, M. C., Atabay, E. S., Ziplar, U. T., Daşgan, H. Y., and Ünver, T. (2016). Genome-wide Characterization and Expression Analysis of Common Bean bHLH Transcription Factors in Response to Excess Salt Concentration. *Mol. Genet. Genomics* 291, 129–143. doi:10.1007/s00438-015-1095-6

Kim, Y. S., Lee, M., Lee, J.-H., Lee, H.-J., and Park, C.-M. (2015). The Unified ICE-CBF Pathway Provides a Transcriptional Feedback Control of Freezing Tolerance during Cold Acclimation in *Arabidopsis*. *Plant Mol. Biol.* 89, 187–201. doi:10.1007/s11103-015-0365-3

Knight, H., Mugford, S. G., Ülker, B., Gao, D., Thorlby, G., and Knight, M. R. (2009). Identification of *SFR6*, a Key Component in Cold Acclimation Acting post-translationally on CBF Function. *Plant J.* 58, 97–108. doi:10.1111/j.1365-313X.2008.03763.x

Kondou, Y., Nakazawa, M., Kawashima, M., Ichikawa, T., Yoshizumi, T., Suzuki, K., et al. (2008). RETARDED GROWTH of EMBRYO1, a New Basic Helix-Loop-Helix Protein, Expresses in Endosperm to Control Embryo Growth. *Plant Physiol.* 147, 1924–1935. doi:10.1104/pp.108.118364

Kurbidaeva, A., Ezhova, T., and Novokreshchenova, M. (2014). *Arabidopsis thaliana* ICE 2 Gene: Phylogeny, Structural Evolution and Functional Diversification from ICE1. *Plant Sci.* 229, 10–22. doi:10.1016/j.plantsci.2014.08.011

Ledent, V., and Vervoort, M. (2001). The Basic helix-loop-helix Protein Family: Comparative Genomics and Phylogenetic Analysis. *Genome Res.* 11, 754–770. doi:10.1101/gr.177001

Lee, T., Yang, S., Kim, E., Ko, Y., Hwang, S., Shin, J., et al. (2015). AraNet V2: an Improved Database of Co-functional Gene Networks for the Study of *Arabidopsis thaliana* and 27 Other Nonmodel Plant Species. *Nucleic Acids Res.* 43, D996–D1002. doi:10.1093/nar/gku1053

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a Database of Plant Cis-Acting Regulatory Elements and a portal to Tools for In Silico Analysis of Promoter Sequences. *Nucleic Acids Res.* 30, 325–327. doi:10.1093/nar/30.1.325

Letunic, I., and Bork, P. (2006). Interactive Tree of Life (iTOL): an Online Tool for Phylogenetic Tree Display and Annotation. *Bioinformatics* 23, 127–128. doi:10.1093/bioinformatics/btl529

Li, J., Li, X., Han, P., Liu, H., Gong, J., Zhou, W., et al. (2021). Genome-wide Investigation of bHLH Genes and Expression Analysis under Different Biotic and Abiotic Stresses in *Helianthus Annuus* L. *Int. J. Biol. Macromolecules* 189, 72–83. doi:10.1016/j.ijbiomac.2021.08.072

Li, Y., Li, L., Ding, W., Li, H., Shi, T., Yang, X., et al. (2020). Genome-wide Identification of *Osmanthus Fragrans* bHLH Transcription Factors and Their Expression Analysis in Response to Abiotic Stress. *Environ. Exp. Bot.* 172, 103990. doi:10.1016/j.envexpbot.2020.103990

Ling, H.-Q., Bauer, P., Bereczky, Z., Keller, B., and Ganal, M. (2002). The Tomato Fer Gene Encoding a bHLH Protein Controls Iron-Uptake Responses in Roots. *Proc. Natl. Acad. Sci.* 99, 13938–13943. doi:10.1073/pnas.212448699

Liu, R., Song, J., Liu, S., Chen, C., Zhang, S., Wang, J., et al. (2021). Genome-wide Identification of the Capsicum bHLH Transcription Factor Family: Discovery of a Candidate Regulator Involved in the Regulation of Species-specific Bioactive Metabolites. *BMC Plant Biol.* 21, 262. doi:10.1186/s12870-021-03004-7

Liu, R., Wang, Y., Tang, S., Cai, J., Liu, S., Zheng, P., et al. (2021). Genome-wide Identification of the tea Plant bHLH Transcription Factor Family and Discovery of Candidate Regulators of Trichome Formation. *Sci. Rep.* 11, 10764. doi:10.1038/s41598-021-90205-7

Lynch, M., and Conery, J. S. (2000). The Evolutionary Fate and Consequences of Duplicate Genes. *Science* 290, 1151–1155. doi:10.1126/science.290.5494.1151

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., et al. (2015). CDD: NCBI's Conserved Domain Database. *Nucleic Acids Res.* 43, D222–D226. doi:10.1093/nar/gku1221

Massari, M. E., and Murre, C. (2000). Helix-Loop-Helix Proteins: Regulators of Transcription in Eucaryotic Organisms. *Mol. Cel Biol* 20, 429–440. doi:10.1128/MCB.20.2.429-440.2000

Mccully, M. E., Canny, M. J., and Huang, C. X. (2004). The Management of Extracellular Ice by Petioles of Frost-Resistant Herbaceous Plants. *Ann. Bot-london* 94, 665–674. doi:10.1093/aob/mch191

Murre, C., Mccaw, P. S., and Baltimore, D. (1989). A New DNA Binding and Dimerization Motif in Immunoglobulin Enhancer Binding, Daughterless, MyoD, and Myc Proteins. *Cell* 56, 777–783. doi:10.1016/0092-8674(89)

Nakamura, J., Yuasa, T., Huong, T. T., Harano, K., Tanaka, S., Iwata, T., et al. (2011). Rice Homologs of Inducer of CBF Expression (*OsICE*) Are Involved in Cold Acclimation. *Plant Biotechnol.* 28, 303–309. doi:10.5511/plantbiotechnology.11.0421a

Nakata, M., Mitsuda, N., Herde, M., Koo, A. J. K., Moreno, J. E., Suzuki, K., et al. (2013). A bHLH-type Transcription Factor, ABA-INDUCIBLE BHLH-TYPE TRANSCRIPTION FACTOR/JA-ASSOCIATED MYC2-LIKE1, Acts as a Repressor to Negatively Regulate Jasmonate Signaling in Arabidopsis. *The Plant Cell* 25, 1641–1656. doi:10.1105/tpc.113.111112

Nemesio-Gorriz, M., Blair, P. B., Dalman, K., Hammerbacher, A., Arnerup, J., Stenlid, J., et al. (2017). Identification of norway spruce MYB-bHLH-WDR Transcription Factor Complex Members Linked to Regulation of the Flavonoid Pathway. *Front. Plant Sci.* 8, 305–319. doi:10.3389/fpls.2017.00305

Ni, L., Wang, Z., Fu, Z., Liu, D., Yin, Y., Li, H., et al. (2021). Genome-wide analysis of basic helix-loop-helix family genes and expression analysis in response to drought and salt stresses in *Hibiscus hamabo* Sieb. et Zucc. *Int. J. Mol. Sci.* 22, 8748. doi:10.3390/ijms22168748

Ohashi-Ito, K., and Bergmann, D. C. (2006). Arabidopsis FAMA Controls the Final Proliferation/Differentiation Switch during Stomatal Development. *The Plant Cell* 18, 2493–2505. doi:10.1105/tpc.106.046136

Pearce, R. S. (1988). Extracellular Ice and Cell Shape in Frost-Stressed Cereal Leaves: a Low-Temperature Scanning-Electron-Microscopy Study. *Planta* 175, 313–324. doi:10.1007/BF00396336

Pillitteri, L. J., Sloan, D. B., Bogenschutz, N. L., and Torii, K. U. (2007). Termination of Asymmetric Cell Division and Differentiation of Stomata. *Nature* 445, 501–505. doi:10.1038/nature05467

Pires, N., and Dolan, L. (2010). Origin and Diversification of Basic-Helix-Loop-Helix Proteins in Plants. *Mol. Biol. Evol.* 27, 862–874. doi:10.1093/molbev/msp288

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., et al. (2019). Gene Duplication and Evolution in Recurring Polyploidization-Diploidization Cycles in Plants. *Genome Biol.* 20, 38. doi:10.1186/s13059-019-1650-2

Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998). SMART, a Simple Modular Architecture Research Tool: Identification of Signaling Domains. *Proc. Natl. Acad. Sci.* 95, 5857–5864. doi:10.1073/pnas.95.11.5857

Seo, J.-S., Joo, J., Kim, M.-J., Kim, Y.-K., Nahm, B. H., Song, S. I., et al. (2011). *OsbHLH148*, a Basic helix-loop-helix Protein, Interacts with OsJAZ Proteins in a Jasmonate Signaling Pathway Leading to Drought Tolerance in rice. *Plant J.* 65, 907–921. doi:10.1111/j.1365-313X.2010.04477.x

Shannon, P. (2003). Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303

Shirasawa, K., Isuzugawa, K., Ikenaga, M., Saito, Y., Yamamoto, T., Hirakawa, H., et al. (2017). The Genome Sequence of Sweet Cherry (*Prunus Avium*) for Use in Genomics-Assisted Breeding. *Dna Res.* 24, 499–508. doi:10.1093/dnares/dsx020

Stamatakis, A. (2014). RAxML Version 8: a Tool for Phylogenetic Analysis and post-analysis of Large Phylogenies. *Bioinformatics* 30, 1312–1313. doi:10.1093/bioinformatics/btu033

Sun, H., Fan, H.-J., and Ling, H.-Q. (2015). Genome-wide Identification and Characterization of the bHLH Gene Family in Tomato. *Bmc Genomics* 16, 9. doi:10.1186/s12864-014-1209-2

Takahashi, Y., Ebisu, Y., Kinoshita, T., Doi, M., Okuma, E., Murata, Y., et al. (2013). bHLH Transcription Factors that Facilitate K+ Uptake during Stomatal Opening Are Repressed by Abscisic Acid through Phosphorylation. *Sci. Signaling* 6, ra48. doi:10.1126/scisignal.2003760

Toledo-Ortiz, G., Huq, E., and Quail, P. H. (2003). The *Arabidopsis* Basic/Helix-Loop-Helix Transcription Factor Family[w]. *The Plant Cell* 15, 1749–1770. doi:10.1105/tpc.013839

Verde, I., Abbott, A. G., Abbott, A. G., Scalabrin, S., Jung, S., Shu, S., et al. (2013). The High-Quality Draft Genome of Peach (*Prunus Persica*) Identifies Unique Patterns of Genetic Diversity, Domestication and Genome Evolution. *Nat. Genet.* 45, 487–494. doi:10.1038/ng.2586

Wang, H., Li, Y., Pan, J., Lou, D., Hu, Y., and Yu, D. (2017). The bHLH Transcription Factors MYC2, MYC3, and MYC4 Are Required for Jasmonate-Mediated Inhibition of Flowering in *Arabidopsis*. *Mol. Plant* 10, 1461–1464. doi:10.1016/j.molp.2017.08.007

Wang, L., Tang, W., Hu, Y., Zhang, Y., Sun, J., Guo, X., et al. (2019). A MYB/bHLH Complex Regulates Tissue-specific Anthocyanin Biosynthesis in the Inner Pericarp of Red-centered Kiwifruit Actinidia Chinensis Cv. Hongyang. *Plant J.* 99, 359–378. doi:10.1111/tpj.14330

Wang, T., Lu, J., Xu, Z., Yang, W., Wang, J., Cheng, T., et al. (2014). Selection of Suitable Reference Genes for miRNA Expression Normalization by qRT-PCR during Flower Development and Different Genotypes of *Prunus Mume*. *Scientia Horticulturae* 169, 130–137. doi:10.1016/j.scienta.2014.02.006

Wang, Y.-J., Zhang, Z.-G., He, X.-J., Zhou, H.-L., Wen, Y.-X., Dai, J.-X., et al. (2003). A rice Transcription Factor *OsbHLH1* Is Involved in Cold Stress Response. *Theor. Appl. Genet.* 107, 1402–1409. doi:10.1007/s00122-003-1378-x

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: A Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity. *Nucleic Acids Res.* 40, e49. doi:10.1093/nar/gkr1293

Zhang, Q., Chen, W., Sun, L., Zhao, F., Huang, B., Yang, W., et al. (2012). The Genome of *Prunus Mume*. *Nat. Commun.* 3. doi:10.1038/ncomms2290

Zhang, Q., Zhang, H., Sun, L., Fan, G., Ye, M., Jiang, L., et al. (2018). The Genetic Architecture of floral Traits in the Woody Plant *Prunus Mume*. *Nat. Commun.* 9, 1318. doi:10.1038/s41467-018-04093-z

Zhao, W., Liu, Y., Li, L., Meng, H., Yang, Y., Dong, Z., et al. (2021). Genome-Wide Identification and Characterization of bHLH Transcription Factors Related to Anthocyanin Biosynthesis in Red walnut (*Juglans Regia* L.). *Front. Genet.* 12, 632509. doi:10.3389/fgene.2021.632509

Zhou, J., Li, F., Wang, J.-l., Ma, Y., Chong, K., and Xu, Y.-y. (2009). Basic helix-loop-helix Transcription Factor from Wild rice (*OrbHLH2*) Improves Tolerance to Salt- and Osmotic Stress in *Arabidopsis*. *J. Plant Physiol.* 166, 1296–1306. doi:10.1016/j.jplph.2009.02.007

Zhu, J.-K. (2016). Abiotic Stress Signaling and Responses in Plants. *Cell* 167, 313–324. doi:10.1016/j.cell.2016.08.029

Zhuo, X., Zheng, T., Zhang, Z., Zhang, Y., Jiang, L., Ahmad, S., et al. (2018). Genome-Wide Analysis of the NAC Transcription Factor Gene Family Reveals Differential Expression Patterns and Cold-Stress Responses in the Woody Plant *Prunus Mume*. *Genes* 9, 494. doi:10.3390/genes9100494

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 08 October 2021
doi: 10.3389/fgene.2021.728166

Heat Stress After Pollination Reduces Kernel Number in Maize by Insufficient Assimilates

Shiduo Niu, Xiong Du, Dejie Wei, Shanshan Liu, Qian Tang, Dahong Bian, Yarong Zhang, Yanhong Cui* and Zhen Gao**

College of Agronomy, Hebei Agricultural University/ State Key Laboratory of North China Crop Improvement and Regulation/ Key Laboratory of Crop Growth Regulation of Hebei Province, Baoding, China

Global warming has increased the occurrence of high temperature stress in plants, including maize, resulting in decreased the grain number and yield. Previous studies indicate that heat stress mainly damages the pollen grains and thus lowered maize grain number. Other field studies have shown that heat stress after pollination results in kernel abortion. However, the mechanism by which high temperature affect grain abortion following pollination remains unclear. Hence, this study investigated the field grown heat-resistant maize variety "Zhengdan 958" (ZD958) and heat-sensitive variety "Xianyu 335" (XY335) under a seven-day heat stress treatment (HT) after pollination. Under HT, the grain numbers of XY335 and ZD958 were reduced by 10.9% (p = 0.006) and 5.3% (p = 0.129), respectively. The RNA sequencing analysis showed a higher number of differentially expressed genes (DEGs) between HT and the control in XY335 compared to ZD958. Ribulose diphosphate carboxylase (RuBPCase) genes were downregulated by heat stress, and RuBPCase activity was significantly lowered by 14.1% (p = 0.020) in XY335 and 5.3% (p = 0.436) in ZD958 in comparison to CK. The soluble sugar and starch contents in the grains of XY335 were obviously reduced by 26.1 and 58.5%, respectively, with no distinct change observed in ZD958. Heat stress also inhibited the synthesis of grain starch, as shown by the low activities of metabolism-related enzymes. Under HT, the expression of trehalose metabolism genes in XY335 were upregulated, and these genes may be involved in kernel abortion at high temperature. In conclusion, this study revealed that post-pollination heat stress in maize mainly resulted in reduced carbohydrate availability for grain development, though the heat-resistant ZD958 was nevertheless able to maintain growth.

Keywords: heat stress, maize, kernel abortion, assimilates, trehalose

OPEN ACCESS

Edited by:
Suxu Tan,
Michigan State University,
United States

Reviewed by:
Yanxin Zhao,
Beijing Academy of Agricultural and Forestry Sciences, China
Xiao-Gui Liang,
China Agricultural University, China

****Correspondence:***
Xiong Du
duxiong2002@163.com
Yanhong Cui
cyh@hebau.edu.cn
Zhen Gao
gaozhenvision@163.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *21 June 2021*
Accepted: *02 September 2021*
Published: *08 October 2021*

Citation:
Niu S, Du X, Wei D, Liu S, Tang Q, Bian D, Zhang Y, Cui Y and Gao Z (2021) Heat Stress After Pollination Reduces Kernel Number in Maize by Insufficient Assimilates. Front. Genet. 12:728166. doi: 10.3389/fgene.2021.728166

INTRODUCTION

The rising levels of carbon dioxide in the atmosphere causes a greenhouse effect that results in increased temperatures and climatic changes (Xuan et al., 2020). Rising temperature is a global issue because of its impact on crop growth and yield (Wang et al., 2020a). Maize is more sensitive to heat stress (one of the most important abiotic stresses) than wheat and rice (Zhao et al., 2017; Zhang et al., 2019). Simulation result indicated that a 10% reduction in maize yield was shown for each 1°C increase in global temperature (Zhang and Zhao, 2017; Dong et al., 2021). Furthermore, previous

studies have underscored the effects of high temperature on maize growth and development (Obaid et al., 2016; Zhang and Zhao, 2017; Wang et al., 2020b). For example, maize plants exhibited various effects of high temperature at distinct phenological periods (Lizaso et al., 2018). Maize tassels (male flowers) growing at the top of the plant were found to be vulnerable to low-level heat stress and this affected pollen viability (Dong et al., 2021; Wang et al., 2021). At the flowering stage, high temperatures inhibited anther dehiscence, pollen viability, and pollen germination, which caused kernel abortion and maize yield reduction (Carberry et al., 1989; Bakhtavar et al., 2015; Hatfield and Prueger, 2015; Li and Howell, 2021; Wang et al., 2021). Additionally, high temperatures caused a delay in the anthesis-silking interval (ASI) of maize, resulting in reduced kernel number, although plenty of pollen was still present (Wang et al., 2019). Moreover, in widely planted smaller tassel maize varieties, extended ASI distinctly decreased yield (Wang et al., 2019). Pollen sterility under heat stress has been intensely studied recently, but the manner in which high temperatures after pollination causes grain abortion remains unclear.

Leaf photosynthesis is fairly sensitive to high temperatures (Berry and Bjorkman, 1980), resulting in decrease in the net photosynthetic rate (Ben-Asher et al., 2008). Impaired photosynthesis affects biological carbon fixation (Gustin et al., 2018), thus restraining the synthesis of glucose and starch in the kernels and influencing the activities of related enzymes (Yang et al., 2016; Fahad et al., 2017; Basu et al., 2019). RNA sequencing (RNA-seq) analysis has shown that high temperatures downregulate starch synthesis genes involved in carbon metabolism (Bita and Gerats, 2013). Additionally, other stresses also have been shown to reduce assimilate availability, leading to kernel abortion (Puteh et al., 2014; Pan et al., 2015; Shen et al., 2020) and the inhibition of grain filling (Edreira and Otegui, 2012). Inversely, under abiotic stress, trehalose can increase sugar transport into the grains and improve crop grain number or grain size (Griffiths et al., 2016). Specifically, a gene in the maize ear expressing trehalose phosphate phosphatase causes a significantly reduced kernel abortion rate under drought (Nuccio et al., 2015). However, the effects of changes in sugar and trehalose-6-phosphate signaling synthesis genes on kernel abortion under short-term heat stress in field maize remains unclear.

Previous studies have shown that damage to the pollen grains due to high-temperature stress is the main limiting factor to kernel setting (Liu, 2014; Wang et al., 2019). Short-term heat stress after pollination was found to have less influence on maize kernel abortion (Wang et al., 2021). However, we hypothesized that post-pollination heat stress might result in kernel abortion in a heat-sensitive maize variety. Hence, the objectives of this study were to 1) assess the effects of heat stress on the change in kernel number after pollination using heat-sensitive and heat-resistant maize varieties and 2) determine sugar metabolism in the maize kernels under heat stress.

MATERIALS AND METHODS

Experimental Site

The field experiment was conducted in 2019 at the Shenzhou Dryland Farming Experimental Station of the Hebei Academy of Agriculture and Forestry Sciences (Hebei Province, China, 37.91N, 115.71E). **Supplementary Figure S1A** shows the climatic conditions during the growing season of maize. The soil in the experiment was classified as loam fluvo-aquic with 12.53 g kg^{-1} organic matter, 65.8 mg kg^{-1} total nitrogen, 121.9 mg kg^{-1} available potassium, and 15.3 mg kg^{-1} available phosphorus.

Experimental Design and Field Management

This study used the heat-sensitive "Xianyu 335" (XY335)and heat-insensitive maize varietiy: "Zhengdan 958" (ZD958) (Wang et al., 2020a). Both are common maize varieties in China. The maize seeds were manually sown at a density of 75,000 plants per hectare on June 16, 2019. The row spacing was 60 cm and the plant spacing was 22 cm and there were six rows in each greenhouse. After sowing, irrigation water was supplied using the surface flood method. The fertilizer application was done before sowing at a rate of 750 kg per hectare compound fertilizer with a 25:8:12 ratio of N: P_2O_5: K_2O, while top-dressing was done at V12 using 138 kg N ha^{-1} (urea) of fertilizer. Weeds, pests, drought, and diseases were well controlled.

Randomized complete blocks were used in this study, with three replicates per treatment. The silking period was recorded when the silks of 50% of maize plants had reached 2 cm (Abendroth et al., 2011). Five days after silking, artificial unified pollination was conducted following Shen et al. (2020). Six simple greenhouses were then constructed to enclose the maize plants that would undergo heat treatment (HT), with each variety planted in a separate greenhouse. The control maize plants were grown under natural conditions. Each greenhouse measured 5 m in length, 3.5 m in width, and 3.5 m in height. Polyethylene film (0.8 mm thickness) was used as a barrier with 1.2 m openings on the sides for adequate gas exchange (**Supplementary Figure S2**). The HT treatment was conducted from 8:00 to 18:00 for 7 days.

Temperature and humidity recorders (L95-2 Saiouhuachuang Technological Corporation, Beijing, China) were installed in the center of each greenhouse to record data every 10 min and were placed 1.5 m above the ground. The average temperature and maximum temperature in the shed during daily treatment are shown in **Supplementary Figure S1B**. The simple greenhouses were removed after the HT treatment.

Sampling and Measurements

Four days after HT, the net photosynthetic rate (Pn) of the ear leaves (representative source organ) was measured with a portable photosynthetic apparatus system (LI-3400 Li-Cor, USA) under a natural field environment. Each measurement was taken at the center of the ear-leaf.

Light intensity was measured at noon above the canopy three times in each plot by using a LI-250A Light Meter (Li-Cor, USA). The polyethylene film used for the HT allowed a penetration of 95.4% of the incoming solar radiation. There was no significant difference in light intensity between HT and the control (CK) (**Supplementary Figure S1C**).

At 5 days after the HT treatment (10 days after silking), sampling of the leaves and kernels was performed. Each leaf sample (>1 g) was taken from the bottom of the ear leaf at about

20 cm. The fresh kernel samples (>20 grains) were taken as shown in **Supplementary Figure S3**. Samples of the kernels and leaves were immediately frozen in liquid nitrogen. The samples were then refrigerated at −80°C for enzyme activities determination and transcriptome sequencing. At 10 days after silking, three ears per plot were gathered, and two rows of grains were sampled for dry weight measurements. The dry weight was determined by drying the kernels at 80°C until a constant weight. The ear growth rate of the maize was then calculated based on the dry weight of the ear (cob + kernel). At physiological maturity, the maize ears were hand-harvested in each treatment to record the row number and kernel number per row.

Assay of Soluble Sugar and Starch Levels

The soluble sugar content of the kernels was measured using the anthrone colorimetric method. Briefly, 0.5 g of dried powdered of kernels was mixed with 6 ml of water and then heated at 100°C for 30 min. The samples were cooled to room temperature and centrifuged at 3,500 rpm for 15 min to obtain the supernatant as a soluble sugar solution. The supernatant was transferred into another test tube and the last procedure was repeated. The insoluble sediment was diluted with 10 ml of 3 mol L^{-1} HCl and then heated for 45 min at 100°C. Following which it was centrifuged at 35,000 rpm min^{-1} for starch determination. The product was collected and neutralized with 10 ml of 3 mol L^{-1} NaOH. The measurement of soluble sugar and starch levels referred to Hanft and Jones (1986).

Measurement of Photosynthetic Enzyme and Starch Synthase Activities

The fresh kernel samples (0.5 g) were ground into a fine powder and extracted with 450 μL Phosphate Buffer Solution (PBS) at pH 7.2–7.4. Afterwards, the prepared samples were centrifuged at 4,000 rpm min^{-1} for 15 min to separate the supernatants, which were then assayed using an enzyme-linked immunosorbent assay (ELISA) kit (Sci-tech innovation, Qingdao, China). Activities of starch synthase (SSS), adenosine diphosphate-glucose pyro phosphorylase (AGPase), and cell wall invertase (CWIN), which are three key enzymes involved in the starch synthesis pathway in maize kernels, were determined. In addition, the ribulose diphosphate carboxylase (RuBPCase) activities were measured in the leaves were measured following Zhang et al. (2020a).

Transcriptome Analysis

Transcriptome analysis was conducted on both the leaves and the kernels of the two maize varieties grown under unstressed (CK) and stressed (HT) conditions in a field environment. The total RNA was extracted using TRizol reagent (Invitrogen, Carlsbad, CA) following the manufacturer's directions with three biological repeats tested. The RNA concentration, purity, and integrity were measured using a NanoDrop 2000 (Thermo Fisher Scientific, Wilmington, DE) and the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA), respectively. The input material for RNA sample preparations was 1 μg per sample. The database sequencing libraries were established following the manufacturer's recommendations for the NEBNext® UltraTMRNA Library Prep Kit for Illumina® (NEB, USA). Index codes were added to attribute sequences to each sample. As instructed by the manufacturer, the clustering of the index-coded samples was performed on the cBot Cluster generation system through a TruSeq PE Cluster Kit v4-cBot-HS (Illumina). After cluster generation, the library preparations were sequenced on an Illumina platform and paired-end reads were generated. The sequence analysis was performed using the BMKCloud platform (www.biocloud.net). The data were subjected to strict quality control by deleting low-quality sequence reads. The data considered were reads with a proportion of N higher than 10%, and reads with a quality value of Q ≤ 10 accounted for more than 50% of the total reads. The clean data were mapped to the maize reference genome (B73_RefGen_v2) using HISAT2 (Kim et al., 2015). The gene expression outputs statistical data is given as follows **Supplementary Table S1**. After quality control of sequencing data, 189.04Gb Clean Data were obtained and the minimum of Q30 was 94.73%).

The mapped read numbers and transcript length were normalized. Fragments Per Kilobase of transcript per Million fragments mapped (FPKM; Florea et al., 2013) was used as an index for the gene expression levels in different samples.

The differentially expressed genes (DEGs) were selected based on $\log_2$ (fold change) >1 or $\log_2$ (fold change) <−1 and with statistical significance of $p < 0.05$. The transformed and normalized expression values of the DEGs FPKM by Z-score were used for hierarchical clustering. **Supplementary Table S2** shows the annotations of the enzyme related genes described in this study. **Supplementary Table S3** shows the related annotations of heat shock genes in this study. The annotations of genes involved in photosynthesis were sourced from the database of the National Biotechnology Information Center (NCBI, https://www.ncbi.nlm.nih.gov/).

Quantitative Real Time-PCR

The RNA-seq data were further validated by quantifying the gene expression of a selected number of genes in the XY335 kernels using quantitative real-time PCR analysis (qRT-PCR). The cDNA synthesis from the total RNA was performed using the TRUEscript 1st Strand cDNA SYNTHESIS Kit (Aidlab, Beijing, China), and qRT-PCR are done with 5 × RT Reaction Mix (MedChemExpress, China). The specific primers used in the qRT-PCR are listed in **Supplementary Table S4**. The primers were designed based on gene sequences in the NCBI GenBank database and were synthesized by Biomarker Technologies (Beijing, China). The fluorescence was measured at the end of each cycle for quantification. Using GRMZM2G171060 (Zm00001d000379) as the reference gene, the $2^{-\Delta\Delta Ct}$ method was used to calculate relative gene expression with three technical replicates tested. The qRT-PCR results showed that the transcriptome results were reliable (**Supplementary Figure S4**).

Statistical Analysis and Drawing of Illustrations

A two-tailed Student's *t-test* was used to determine significance levels between the CK and HT in kernel number, ear growth rate,

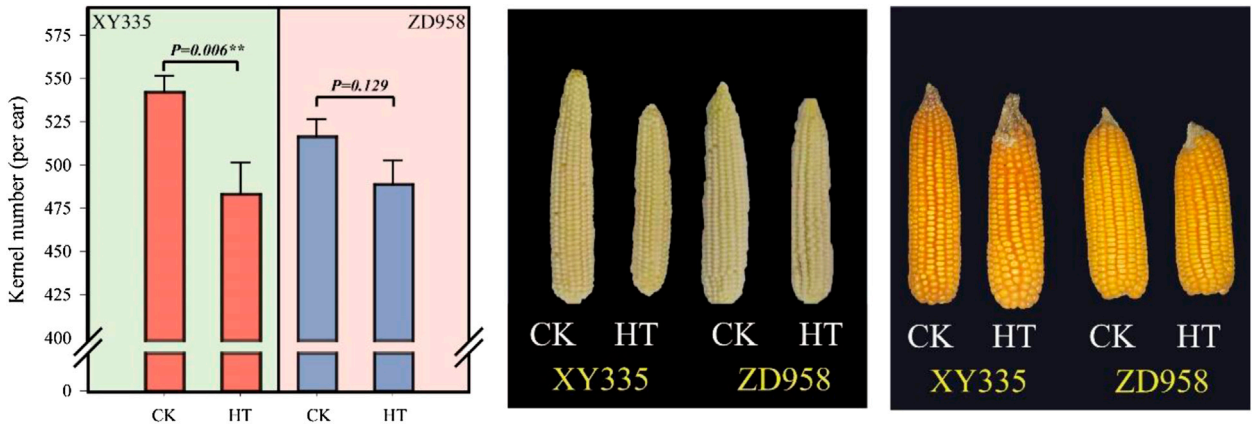


FIGURE 1 | Effects of heat stress on the kernel number of maize varieties (ZD958 and XY335) grown under control (CK) and heat treatment (HT) conditions.

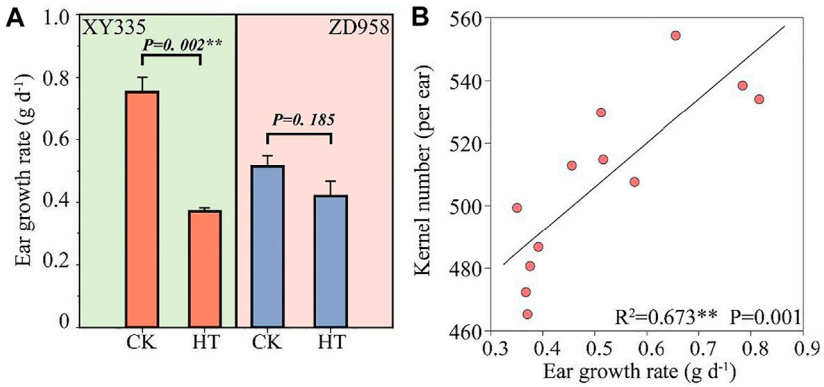


FIGURE 2 | (A) Effects of heat stress on ear growth rate after 5 days of heat treatment. CK and HT indicate the control and heat treatment, respectively. **(B)** Relationship between the ear growth rate after 5 days of heat treatment and the kernel number per plant.

enzymatic activity, sugar content, and photosynthesis rate. Statistical analyses were performed using IBM SPSS Statistics Version 25 and Microsoft Excel 2019.

Figures were drawn using SigmaPlot 12.5 and Adobe Illustrator CC 2020. A heatmap of the DEGs was drawn using the R package Pheatmap. Statistics of pathway enrichment were drawn by the platform BMKCloud (http://www.biocloud.net).

RESULTS

Kernel Number and Ear Growth Under Heat Stress

As shown in **Figure 1**, the kernel number per ear under HT in XY335 was reduced by 10.9% compared to CK, while no significant difference was observed between CK and HT in ZD958. The ear growth rate significantly decreased under high temperatures. The ear growth rate of XY335 and ZD958 under HT was significantly reduced by 50.6 and 18.4% compared to CK, respectively (**Figure 2A**). Moreover, the correlation analysis results showed that high ear growth rate around pollination increased the kernel number (p = 0.001, **Figure 2B**).

Analysis of Gene Expression and Differentially Expressed Genes

Figure 3A shows the results of the transcriptome analyses of DEGs in the HT treatment compared to the CK. The leaf and kernel samples of XY335 had 871 and 12,891 DEGs, respectively. The numbers of DEGs identified in the leaf and kernel samples in ZD958 were 3,208 and 1,720, respectively. Under the same heat stress, the number of DEGs detected in ZD958 was considerably lower than in XY335, with 392 and 511 DGEs overlapping between the leaf and kernel, respectively (**Figure 3B**). The heatmap of heat shock genes showed 20 upregulated genes and five downregulated genes in XY335, while normal expression levels were detected for these genes in ZD958. Remarkably, HSP4 was downregulated in ZD958 but normal in expression XY335 (**Figure 3C**). Therefore, this study showed

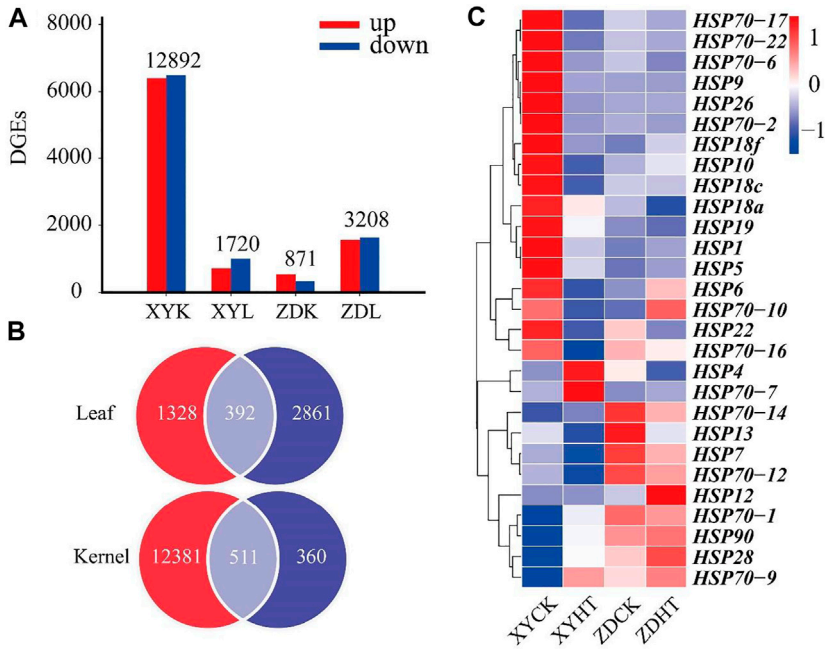


FIGURE 3 | Comparing the differentially expressed genes (DEGs) of maize varieties XY335 and ZD958 grown under both control (CK) and heat (HT) treatments on the 5th day of HT. **(A)** Total numbers of upregulated and downregulated genes **(B)** Venn diagram of the DGEs and **(C)** heatmap of the genes related to heat shock under heat stress.

that XY335 was greatly affected by high temperature, whereas ZD958 was not.

Effect of Heat Stress on Photosynthesis

The leaf net photosynthetic rate (Pn) of XY335 was significantly inhibited by heat stress, whereas the Pn of ZD958 was not obviously reduced (**Figure 4A**). Similarly, RuBPCase activity under HT was significantly decreased by 14.1% compared with CK in XY335. HT lowered the RuBPCase activity by 5.3% less than CK in ZD958 (**Figure 4B**).

A total of 183 genes related to photosynthesis were analyzed, and the two varieties showed different results under the CK and HT treatments (**Supplementary Figure S5**). Among these genes, 131 and 133 genes were downregulated in XY335 and ZD958, respectively. Additionally, 62 genes differed between the two varieties. Unexpectedly, two genes encoding RuBPCase, Zm00001d004894 and Zm00001d052595 were significantly downregulated in both XY335 and ZD958 under HT (**Figure 4C**).

Sugar Metabolism in the Maize Kernels

The soluble sugar in the XY335 kernels under HT decreased by 26.1% compared with CK, while there was no remarkable reduction in soluble sugar content in ZD958 under HT. The kernel starch content in XY335 and ZD958 under HT decreased by 58.5 and 27.2% compared with those of CK, respectively. **Figure 5A** shows that the decrease in kernel starch contents reached a significant level (p= 0.025) in XY335 but not in ZD958 (p= 0.333). Correlation analysis indicated that sufficient soluble sugar (p= *0.056*) and starch (p= 0.021) could increase kernel number (**Figure 5B**). Additionally, soluble sugar content (p = 0.003) and starch content (p < 0.001) were positively correlated with ear growth rate (**Figure 5C**).

Under HT condition, the activities of CWIN and SSS in the kernel did not change significantly, while AGPase activity was found to be sensitive to HT treatment (**Figure 6**). The results showed that the SSS activity under HT was reduced by 8.1% (p= 0.162) and 1.8% (p= 0.300) in XY335 and ZD958, respectively. The AGPase activity in the maize kernels decreased significantly by 10.6% in XY335, while a 5.8% increase was observed in ZD958, though this was not statistically significant (p= 0.093). The RNA-seq results indicated that five SSS genes and three AGPase genes were downregulated under the HT relative to CK. Interestingly, CWIN-related genes exhibited upregulated expression compared to those in the CK group (**Figure 7**).

Additionally, 20 genes related to trehalose synthesis were upregulated, whereas four genes were downregulated. In the metabolic pathways of starch and sucrose, trehalose-synthesized genes were found to be upregulated in XY335 under HT; however, ZD958 showed relatively minor changes (**Figure 8**). From the results, it is clear that starch synthesis was inhibited, while trehalose synthesis was promoted under high temperature stress, which may have resulted in kernel abortion.

DISCUSSION

Previous studies have shown that maize pollens are susceptible to high temperature, usually resulting in kernel abortion (Wang et al., 2019; Wang et al., 2021). Our results indicated that heat stress after pollination still plays a critical role in kernel abortion,

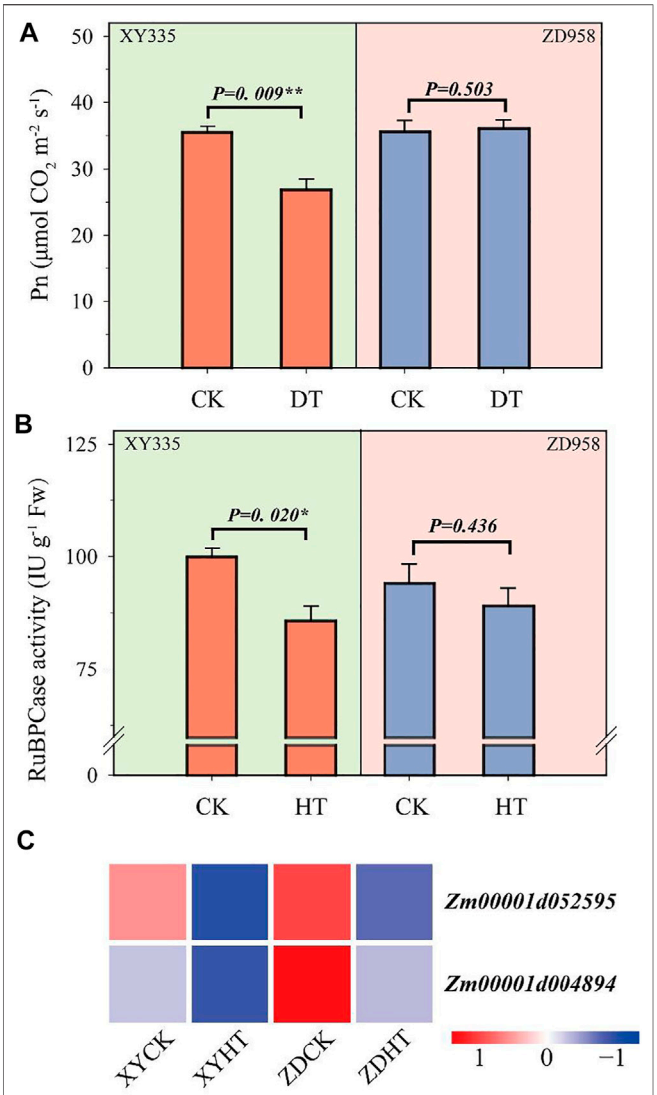


FIGURE 4 | Changes in photosynthesis after 5 days of heat treatment (HT) compared to the control (CK): **(A)** photosynthesis rate in CK and HT; **(B)** ribulose diphosphatecarboxylase (RuBPCase) activity in CK and HT; and **(C)** heatmap of the genes related to RuBPCase under heat stress. Red and blue lines represent upregulated and downregulated genes, respectively.

and it reduces carbohydrate availability and damages carbon metabolism. Additionally, kernel abortion after pollination at a high temperature was found to be variety-specific.

Limited Low Ear Growth Rate in the Early Stage of Maize Grain Filling Reduced the Kernel Number

During the critical growth stage bracketing of silking, stress typically reduces the maize plant/ear growth rate (Rossini et al., 2011, Pagano and Maddonni, 2007; Borras and Vitantonio-Mazzini, 2018). Kernel number per ear is significantly related to plant growth rate around silking and biomass partitioning to the ear during this period (Pagano and Maddonni, 2007). In this experiment, heat stress also decreased maize ear growth rate, especially for XY335. HT after pollination reduced ear leaf photosynthesis and decreased assimilate availability, which reduced ear growth. Moreover, previous research has shown that the ear growth rate at the early stage of kernels growth determined the final grain number (Rossini et al., 2011). In line with previous studies, there was a significant correlation between the kernel number and the ear growth rate during this period in our experiment (**Figure 2B**). As discussed, the kernel number was affected by the restriction of ear growth rate under HT.

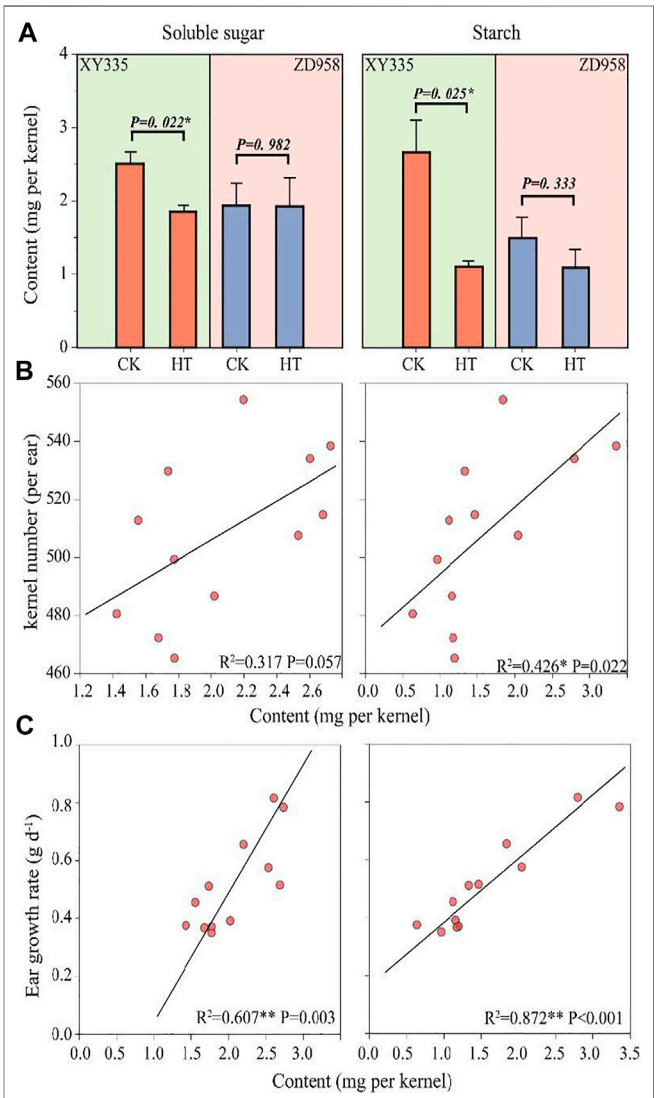


FIGURE 5 | Effects of heat stress on soluble sugar, and starch content in the kernel and relationships between soluble sugar/starch with kernel number and ear growth rate. **(A)** Soluble sugar and starch contents in the maize kernels as affected by heat treatment (HT) after the 5th day of treatment. **(B)** The relationship between the soluble sugar (left)/starch (right) contents with final kernel number. **(C)** The relationships between the soluble sugar (left)/ starch (right) contents with the ear growth rate.

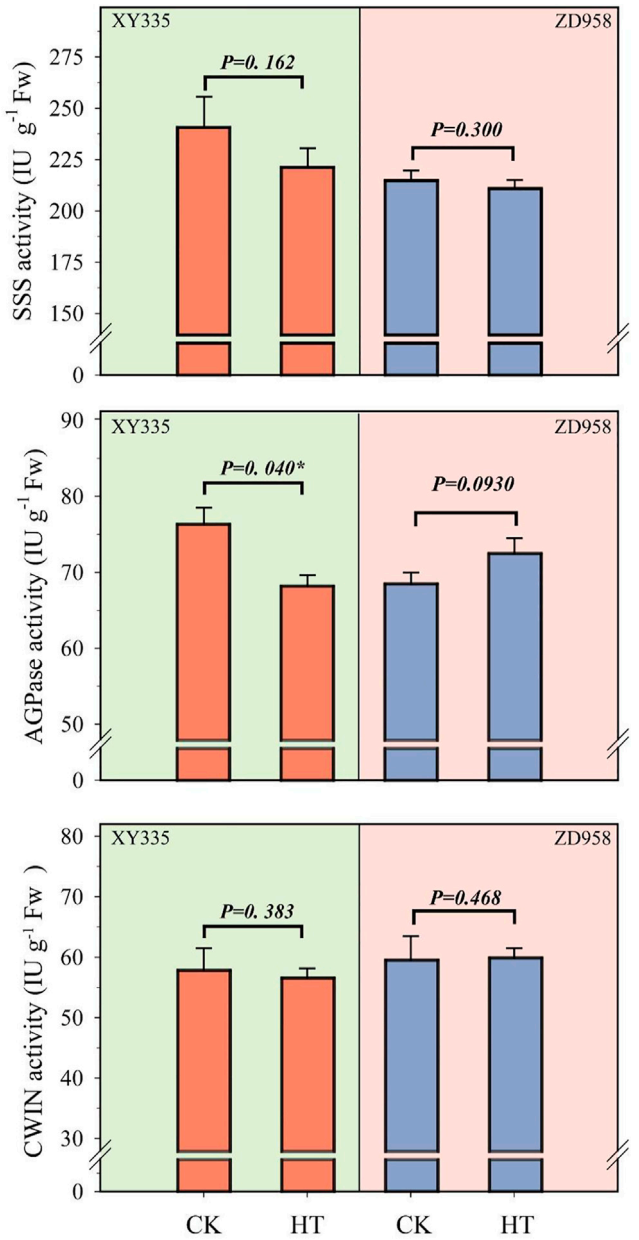


FIGURE 6 | Changes in the activities of starch synthase (SSS), adenosine diphosphate-glucose pyrophosphorylase (AGPase), and cell wall invertase (CWIN) at 5 days after heat treatment.

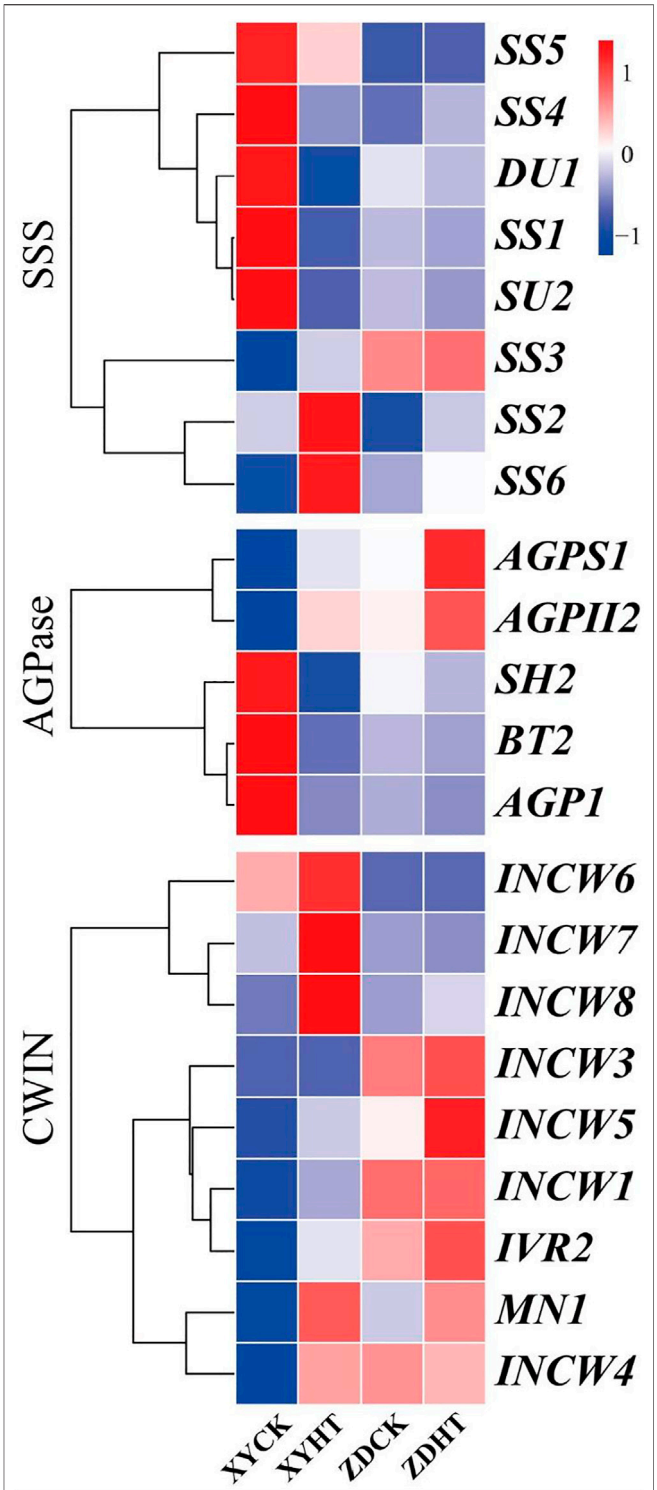


FIGURE 7 | Expression profiles of genes that encode starch synthase (SSS), adenosine diphosphate-glucose pyrophosphorylase (ADGPase), and cell wall invertase (CWIN). **(A)** Gene expression (FPKM fold change) of the candidate genes in maize leaves under CK and HT conditions.

Insufficient Sugar Supply Leads to Restricted Ear Growth Rate After Pollination

Maize is very sensitive to high temperatures during tasseling, flowering, pollination, and kernel filling (Zhang et al., 2020b). The reasons may be that 1) a high temperature causes kernel abortion by destroying pollination processes (Deryng et al., 2014) and 2) a high temperature causes sugar deficiency or insufficient sugar metabolism, leading to kernel abortion (Edreira and Otegui, 2012). A previous study showed that 22.1% of the kernels were aborted during the 15-days high

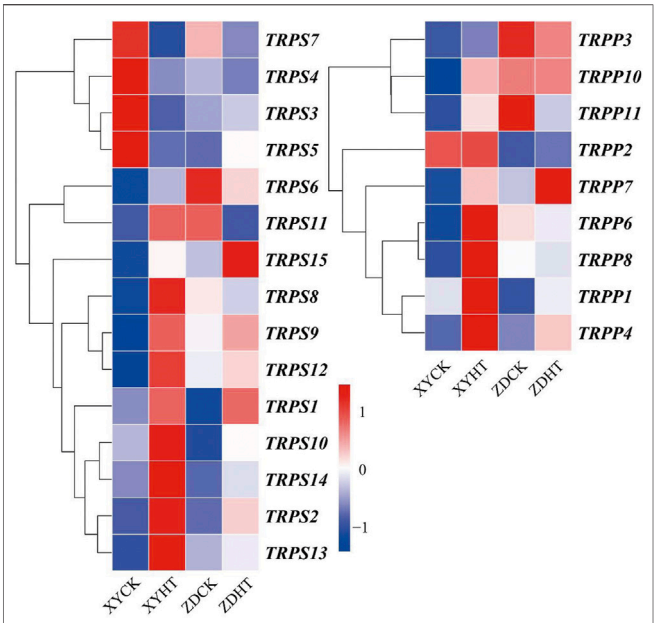


FIGURE 8 | Variation in expression profiles of the genes involved in trehalose biosynthesis.

temperature stress treatment after tasseling (Wang et al., 2021). The results of this study showed that 10.9% of kernels in the XY335 maize variety were aborted after post-pollination heat stress. This also suggested that the abortion of a large percentage of maize kernels can still occur post-pollination under high temperature conditions.

Significant decreases in the leaf photosynthetic rate as well as in soluble sugar and starch contents in the kernels were observed under high-temperature stress (**Figures 4A**, **5**). The poor supply of sugars induces ovary abortion, ultimately affecting kernel formation and yield in maize (Gao et al., 2020; Usmani et al., 2020). Sucrose feeding can reverse the kernel loss induced by drought or shade stress (Hiyane et al., 2010; Zinselmeier et al., 1995) and can partially restore the activity of carbon metabolism-related enzymes, thus restoring some kernel growth (McLaughlin and Boyer, 2004). These findings imply that kernel abortion is associated with an insufficient assimilate supply (Puteh et al., 2014; Edreira and Otegui, 2012). Carbon-related enzymes also play a vital role in kernel abortion (Shen et al., 2020), as demonstrated in this study where kernel abortion was caused by the decreased activity of enzymes related to starch synthesis and reduced levels of starch (**Figure 5B**). Transcriptome analysis also showed that heat stress caused the low expression of starch synthesis-related genes. Overall, the results of this study are consistent with previous studies (Yue et al., 2016; Lambarey et al., 2020; Shen et al., 2020).

Interestingly, no significant change in the activity of CWIN was observed, but its related genes were upregulated. CWINs contribute to sink strength and have been previously reported to exhibit a key role in sucrose import and kernel filling (Millera and Chourey, 1992; Weber et al., 1995; Wang et al., 2008; Morey et al., 2018). We speculated that a low sugar supply promoted the upregulation of CWIN, but we found that heat stress inhibited CWIN activity and reduced sucrose import.

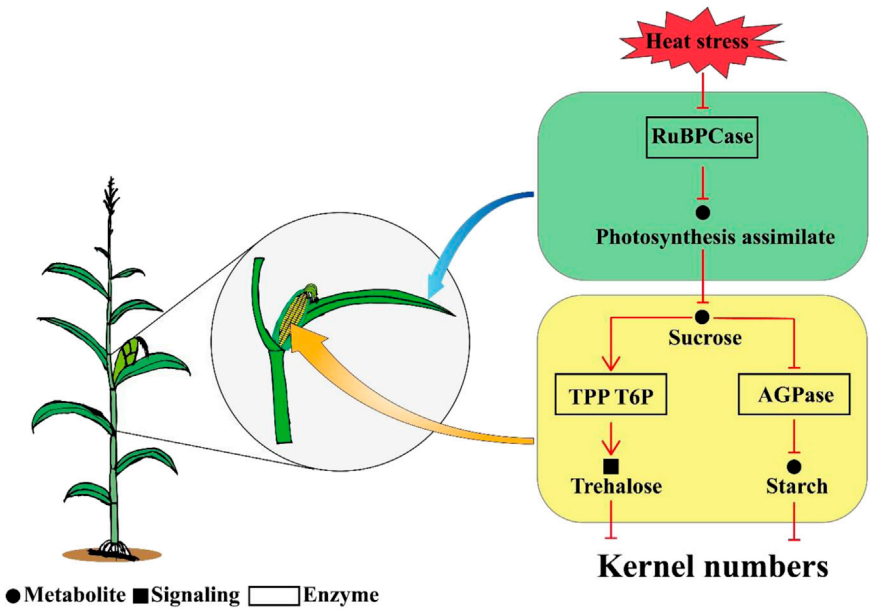


FIGURE 9 | A schematic diagram model for the mechanism of kernel abortion, including transcripts, starch and sucrose metabolism and photosynthesis.

The Trehalose Pathway Affected Kernel Setting Under High Temperature

Trehalose biosynthesis has been confirmed to increase tolerance to multiple abiotic stresses in tobacco, potato, and rice (Schluepmann and Paul, 2009). In plants, trehalose-6-phosphate synthase (TPS) catalyzes UDP-glucose and glucose-6-phosphate to synthesize trehalose-6-phosphate (T6P) (Kumar et al., 2013). T6P is further metabolized to trehalose by trehalose-6-phosphate phosphatase (TPP) (O'Hara et al., 2013). T6P is a critical signaling molecule that integrates sugar status with growth and development in plants (Paul et al., 2017; Gustin et al., 2018; Paul et al., 2020). The results of this study showed that under high temperatures, both the TPS and TPP genes were obviously upregulated in XY335, a heat-sensitive maize variety, while no obvious change was observed in ZD958, a heat-resistant variety. The results suggested that trehalose metabolism played an important role in kernel abortion. The similar changes may occur under shade stress (Liang et al., 2020). Additionally, a reduction in T6P level via the expression of TPP can prevent maize kernel abortion and increase yield under drought stress (Nuccio et al., 2015). The elevated gene expressions of TPS and lower gene expression of TPP in the apical kernels inhibited seed setting (Shen et al., 2020). In this study, both the TPS and TPP genes were upregulated in XY335 under heat stress, but relatively unobvious changes were observed in ZD958. This suggested that the response of trehalose metabolism to heat stress was distinct in the heat-sensitive variety, ultimately leading to kernel abortion under heat stress.

ZD958 has Higher Heat Resistance Than XY335

XY335 is sensitive type to environmental stress, including heat stress whereas ZD958 is resistant to environmental stress (Liu, 2014; Wang et al., 2020a). The yield of the former is typically higher than that of the latter. However, yield performance differs markedly under stress (Berry and Bjorkman, 1980). In a stressful environment, low assimilate availability aggravates kernel abortion (Gao et al., 2020; Shen et al., 2018). However, sugar stored in the stems can serve as a buffer to ensure kernel growth (Milne et al., 2013; Bledsoe et al., 2017). This study showed that the heat-sensitive (XY335) had a higher kernel number than the heat-tolerant variety (ZD958) under control conditions. Under HT conditions, ZD958 produced more kernels compared to XY335. Although the sugar content in the stem was not measured in this study, all the data gathered, including the net photosynthetic rate, DEGs, and final kernel number, demonstrated that ZD958 is a heat-resistant variety.

Overall, the results of this study indicated that kernel abortion was caused by carbohydrate metabolic disorders. Heat stress decreased the RuBPCase activity by downregulating Zm0001d052595 and Zm0001d004894 which restricted photosynthesis and decreased assimilate availability for the kernels. The downregulation of genes related to AGPase and the upregulation of genes related TPP resulted in T6P disrupting the balance between trehalose and starch. Consequently, this study demonstrates that reduced carbohydrate availability leads to kernel abortion under post-pollination heat stress conditions (**Figure 9**).

CONCLUSION

Heat stress after pollination can result in kernel abortion, especially in heat sensitive varieties. Heat stress mainly reduces leaf photosynthesis and RuBPCase activity thus lowering assimilate availability. Ear growth rate was significantly reduced and showed significant relationship with kernel number. Concurrently, the soluble sugar and starch content and key enzyme activity in the kernels were decreased and the related genes also showed obvious downregulation. Additionally, the altered synthetic pathway of trehalose may play a critical role in kernel setting under heat stress. In conclusion, heat stress in maize after pollination results in kernel abortion due to insufficient assimilate availability.

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: NCBI SRA; PRJNA757605.

AUTHOR CONTRIBUTIONS

GZ initiated and designed the research, was a substantial contributor to the preparation of the manuscript. NS performed the experiment, analyzed the data, and wrote the manuscript. CY, DX, and BD provided advice on the experiments. WD, LS, TQ, and ZY contributed to the acquisition of the data and participated in the revision of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Key Research and Development Program of China (2017YFD0300908), the Key Research and Development Program of Hebei Province (20326414D), Startup Fund of Hebei Agricultural University (YJ201827), the Hebei Agriculture Research System (HBCT2018020202), the Funding of State Key Laboratory of North China Crop Improvement and Regulation (NCCIR2020ZZ-17), and the Scientific Research Development Fund of Hebei Agricultural University (JY2019006).

ACKNOWLEDGMENTS

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.728166/full#supplementary-material

Supplementary Figure 1 | The meteorological data during the summer maize growing period and climatic conditions under the experimental treatments: **(A)** daily average temperature and rainfall; **(B)** maximum (dotted line) and average (solid line) temperature inside and outside the greenhouse during 8:00-18:00 every day during the treatment period; and **(C)** the light intensity inside and outside the greenhouse.

Supplementary Figure 2 | Photo of the greenhouse in the field.

Supplementary Figure 3 | **(A)** Location of sampled kernel selection and **(B)** the corresponding location at maturity. The number (white) in the figure is the "kernel number per row," and the red box is the sampling location.

Supplementary Figure 4 | Expression pattern of differentially expressed genes in the kernels under heat treatments (HT).

Supplementary Figure 5 | Expression patterns of DEGs involved in photosynthesis pathways of two maize varieties under the control (CK) and heat (HT) treatments. For each gene, the Fragments Per Kilobase of Transcript per Million fragments mapped (FPKM) fold change in maize leaves under the CK and HT conditions was measured.

Supplementary Figure 6 | Effects of heat stress on the thousand-kernel weight of maize varieties grown under control (CK) and heat (HT) treatments.

Supplementary Table 1 | Sequencing data statistics note.

Supplementary Table 2 | Some of the genes used in this study.

Supplementary Table 3 | Some of the heat shock genes used in this study.

Supplementary Table 4 | List of primer sequences used in this study.

REFERENCES

Abendroth, L. J., Roger, W. E., Matthew, J. B., and Marlay, S. K. (2011). *Corn Growth and Development*. Ames, Iowa: Iowa State University Extension.

Bakhtavar, M. A., Afzal, I., Basra, S. M. A., Ahmad, A.-u. -H., and Noor, M. A. (2015). Physiological Strategies to Improve the Performance of spring maize (*Zea mays* L.) Planted under Early and Optimum Sowing Conditions. *PLoS One* 10, e0124441. doi:10.1371/journal.pone.0124441

Basu, P. S., Pratap, A., Gupta, S., Sharma, K., Tomar, R., and Singh, N. P. (2019). Physiological Traits for Shortening Crop Duration and Improving Productivity of Greengram (*Vigna Radiata* L. Wilczek) under High Temperature. *Front. Plant Sci.* 10, 1508. doi:10.3389/fpls.2019.01508

Ben-Asher, J., Garcia y Garcia, A., and Hoogenboom, G. (2008). Effect of High Temperature on Photosynthesis and Transpiration of Sweet Corn (*Zea mays* L. Var. *Rugosa*). *Photosynt.* 46, 595–603. doi:10.1007/s11099-008-0100-2

Berry, J., and Bjorkman, O. (1980). Photosynthetic Response and Adaptation to Temperature in Higher Plants. *Annu. Rev. Plant Physiol.* 31, 491–543. doi:10.1146/annurev.pp.31.060180.002423

Bita, C. E., and Gerats, T. (2013). Plant Tolerance to High Temperature in a Changing Environment: Scientific Fundamentals and Production of Heat Stress-Tolerant Crops. *Front. Plant Sci.* 4, 273. doi:10.3389/fpls.2013.00273

Bledsoe, S. W., Henry, C., Griffiths, C. A., Paul, M. J., Feil, R., Lunn, J. E., et al. (2017). The Role of Tre6P and SnRK1 in maize Early Kernel Development and Events Leading to Stress-Induced Kernel Abortion. *BMC Plant Biol.* 17, 74. doi:10.1186/s12870-017-1018-2

Borrás, L., and Vitantonio-Mazzini, L. N. (2018). Maize Reproductive Development and Kernel Set under Limited Plant Growth Environments. *J. Exp. Bot.* 69, 3235–3243. doi:10.1093/jxb/erx452

Carberry, P. S., Muchow, R. C., and Mccown, R. L. (1989). Testing the CERES-Maize Simulation Model in a Semi-arid Tropical Environment. *Field Crops Res.* 20, 297–315. doi:10.1016/0378-4290(89)90072-5

Deryng, D., Conway, D., Ramankutty, N., Price, J., and Warren, R. (2014). Global Crop Yield Response to Extreme Heat Stress Under Multiple Climate Change Futures. *Environ. Res. Lett.* 9, 034011. doi:10.1088/1748-9326/9/3/034011

Dong, X., Guan, L., Zhang, P., Liu, X., Li, S., Fu, Z., et al. (2021). Responses of maize with Different Growth Periods to Heat Stress Around Flowering and Early Grain Filling. *Agric. For. Meteorol.* 303, 108378. doi:10.1016/j.agrformet.2021.108378

Rattalino Edreira, J. I., and Otegui, M. E. (2012). Heat Stress in Temperate and Tropical maize Hybrids: Differences in Crop Growth, Biomass Partitioning and Reserves Use. *Field Crops Res.* 130, 87–98. doi:10.1016/j.fcr.2012.02.009

Fahad, S., Bajwa, A. A., Nazir, U., Anjum, S. A., Farooq, A., Zohaib, A., et al. (2017). Crop Production under Drought and Heat Stress: Plant Responses and Management Options. *Front. Plant Sci.* 8, 1147. doi:10.3389/fpls.2017.01147

Florea, L., Song, L., and Salzberg, S. L. (2013). Thousands of Exon Skipping Events Differentiate Among Splicing Patterns in Sixteen Human Tissues. *F1000Res* 2, 188. doi:10.12688/f1000research.2-188.v210.12688/f1000research.2-188.v1

Gao, Z., Sun, L., Ren, J.-H., Liang, X.-G., Shen, S., Lin, S., et al. (2020). Detasseling Increases Kernel Number in maize under Shade Stress. *Agric. For. Meteorol.* 280, 107811. doi:10.1016/j.agrformet.2019.107811

Griffiths, C. A., Sagar, R., Geng, Y., Primavesi, L. F., Patel, M. K., Passarelli, M. K., et al. (2016). Chemical Intervention in Plant Sugar Signalling Increases Yield and Resilience. *Nature* 540, 574–578. doi:10.1038/nature20591

Gustin, J. L., Boehlein, S. K., Shaw, J. R., Junior, W., Settles, A. M., Webster, A., et al. (2018). Ovary Abortion Is Prevalent in Diverse Maize Inbred Lines and Is under Genetic Control. *Sci. Rep.* 8, 13032. doi:10.1038/s41598-018-31216-9

Hanft, J. M., and Jones, R. J. (1986). Kernel Abortion in Maize. *Plant Physiol.* 81, 503–510. doi:10.1104/pp.81.2.503

Hatfield, J. L., and Prueger, J. H. (2015). Temperature Extremes: Effect on Plant Growth and Development. *Weather Clim. Extremes* 10, 4–10. doi:10.1016/j.wace.2015.08.001

Hiyane, R., Hiyane, S., Tang, A. C., and Boyer, J. S. (2010). Sucrose Feeding Reverses Shade-Induced Kernel Losses in maize. *Ann. Bot.* 106, 395–403. doi:10.1093/aob/mcq132

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

Kumar, K., Kumar, M., Kim, S.-R., Ryu, H., and Cho, Y.-G. (2013). Insights into Genomics of Salt Stress Response in rice. *Rice* 6, 27. doi:10.1186/1939-8433-6-27

Lambarey, H., Moola, N., Veenstra, A., Murray, S., and Suhail Rafudeen, M. (2020). Transcriptomic Analysis of a Susceptible African maize Line to fusarium Verticillioides Infection. *Plants* 9, 1112. doi:10.3390/plants9091112

Li, Z., and Howell, S. H. (2021). Heat Stress Responses and Thermotolerance in maize. *Ijms* 22, 948. doi:10.3390/ijms22020948

Liang, X.-G., Gao, Z., Shen, S., Paul, M. J., Zhang, L., Zhao, X., et al. (2020). Differential Ear Growth of Two maize Varieties to Shading in the Field Environment: Effects on Whole Plant Carbon Allocation and Sugar Starvation Response. *J. Plant Physiol.* 251, 153194. doi:10.1016/j.jplph.2020.153194

Liu, Y. (2014). Diurnal Variation and Directivity of Photosynthetic Carbon Metabolism in maize Hybrids under Drought Stress. Doctor Thesis (China). Beijing, China: China Agricultural University.

Lizaso, J. I., Ruiz-Ramos, M., Rodríguez, L., Gabaldon-Leal, C., Oliveira, J. A., Lorite, I. J., et al. (2018). Impact of High Temperatures in Maize: Phenology and Yield Components. *Field Crops Res.* 216, 129–140. doi:10.1016/j.fcr.2017.11.013

McLaughlin, J. E., and Boyer, J. S. (2004). Sugar-responsive Gene Expression, Invertase Activity, and Senescence in Aborting Maize Ovaries at Low Water Potentials. *Ann. Bot.* 94, 675–689. doi:10.1093/aob/mch193

Miller, M. E., and Chourey, P. S. (1992). The Maize Invertase-Deficient Miniature-1 Seed Mutation Is Associated with Aberrant Pedicel and Endosperm Development. *Plant Cell* 4, 297–305. doi:10.1105/tpc.4.3.297

Milne, R. J., Byrt, C. S., Patrick, J. W., and Grof, C. P. L. (2013). Are Sucrose Transporter Expression Profiles Linked with Patterns of Biomass Partitioning in Sorghum Phenotypes?. *Front. Plant Sci.* 4, 223. doi:10.3389/fpls.2013.00223

Morey, S. R., Hirose, T., Hashida, Y., Miyao, A., Hirochika, H., Ohsugi, R., et al. (2018). Genetic Evidence for the Role of a rice Vacuolar Invertase as a Molecular Sink Strength Determinant. *Rice* 11, 6. doi:10.1186/s12284-018-0201-x

Nuccio, M. L., Wu, J., Mowers, R., Zhou, H.-P., Meghji, M., Primavesi, L. F., et al. (2015). Expression of Trehalose-6-Phosphate Phosphatase in maize Ears Improves Yield in Well-Watered and Drought Conditions. *Nat. Biotechnol.* 33, 862–869. doi:10.1038/nbt.3277

O'Hara, L. E., Paul, M. J., and Wingler, A. (2013). How Do Sugars Regulate Plant Growth and Development? New Insight into the Role of Trehalose-6-Phosphate. *Mol. Plant* 6, 261–274. doi:10.1093/mp/sss120

Obaid, A. Y., Sabir, J. S. M., Atef, A., Liu, X., Edris, S., El-Domyati, F. M., et al. (2016). Analysis of Transcriptional Response to Heat Stress in Rhazya Stricta. *BMC Plant Biol.* 16, 252. doi:10.1186/s12870-016-0938-6

Pagano, E., and Maddonni, G. A. (2007). Intra-specific Competition in maize: Early Established Hierarchies Differ in Plant Growth and Biomass Partitioning to the Ear Around Silking. *Field Crops Res.* 101, 306–320. doi:10.1016/j.fcr.2006.12.007

Pan, X., Hasan, M. M., Li, Y., Liao, C., Zheng, H., Liu, R., et al. (2015). Asymmetric Transcriptomic Signatures between the Cob and Florets in the maize Ear under Optimal- and Low-Nitrogen Conditions at Silking, and Functional Characterization of Amino Acid Transporters ZmAAP4 and ZmVAAT3. *Exbotj* 66, 6149–6166. doi:10.1093/jxb/erv315

Paul, M. J., Oszvald, M., Jesus, C., Rajulu, C., and Griffiths, C. A. (2017). Increasing Crop Yield and Resilience with Trehalose 6-phosphate: Targeting a Feast-Famine Mechanism in Cereals for Better Source-Sink Optimization. *J. Exp. Bot.* 68, 4455–4462. doi:10.1093/jxb/erx083

Paul, M. J., Watson, A., and Griffiths, C. A. (2020). Trehalose 6-phosphate Signalling and Impact on Crop Yield. *Biochem. Soc. Trans.* 48, 2127–2137. doi:10.1042/BST20200286

Puteh, A. B., Mondal, M. M. A., Ismail, M. R., and Latif, M. A. (2014). Grain Sterility in Relation to Dry Mass Production and Distribution in Rice (Oryza sativaL.). *Biomed. Res. Int.* 2014, 1–6. doi:10.1155/2014/302179

Rossini, M. A., Maddonni, G. A., and Otegui, M. E. (2011). Inter-Plant Competition for Resources in Maize Crops Grown Under Contrasting Nitrogen Supply and Density: Variability in Plant and Ear Growth. *Field Crops Res.* 121, 373–380. doi:10.1016/j.fcr.2011.01.003

Schluepmann, H., and Paul, M. (2009). Trehalose Metabolites in Arabidopsis-Elusive, Active and central. *Arabidopsis Book* 7, e0122. doi:10.1199/tab.0122

Shen, S., Zhang, L., Liang, X. G., Zhao, X., Lin, S., Qu, L. H., et al. (2018). Delayed Pollination and Low Availability of Assimilates are Major Factors Causing Maize Kernel Abortion. *J. Exp. Bot.* 69, 1599–1613. doi:10.1093/jxb/ery013

Shen, S., Liang, X. G., Zhang, L., Zhao, X., Liu, Y. P., Lin, S., et al. (2020). Intervening in Sibling Competition for Assimilates by Controlled Pollination Prevents Seed Abortion under Postpollination Drought in maize. *Plant Cel Environ.* 43, 903–919. doi:10.1111/pce.13704

Usmani, M. M., Nawaz, F., Majeed, S., Shehzad, M. A., Ahmad, K. S., Akhtar, G., et al. (2020). Sulfate-mediated Drought Tolerance in maize Involves Regulation at Physiological and Biochemical Levels. *Sci. Rep.* 10, 1147. doi:10.1038/s41598-020-58169-2

Wang, E., Wang, J., Zhu, X., Hao, W., Wang, L., Li, Q., et al. (2008). Control of rice Grain-Filling and Yield by a Gene with a Potential Signature of Domestication. *Nat. Genet.* 40, 1370–1374. doi:10.1038/ng.220

Wang, Y., Tao, H., Tian, B., Sheng, D., Xu, C., Zhou, H., et al. (2019). Flowering Dynamics, Pollen, and Pistil Contribution to Grain Yield in Response to High Temperature during maize Flowering. *Environ. Exp. Bot.* 158, 80–88. doi:10.1016/j.envexpbot.2018.11.007

Wang, H.-Q., Liu, P., Zhang, J.-W., Zhao, B., and Ren, B.-Z. (2020a). Endogenous Hormones Inhibit Differentiation of Young Ears in maize (*Zea mays* L.) under Heat Stress. *Front. Plant Sci.* 11, 533046. doi:10.3389/fpls.2020.533046

Wang, J., Zheng, C., Shao, X., Hu, Z., Li, J., Wang, P., et al. (2020b). Transcriptomic and Genetic Approaches Reveal an Essential Role of the NAC Transcription Factor SlNAP1 in the Growth and Defense Response of Tomato. *Hortic. Res.* 7, 209. doi:10.1038/s41438-020-00442-6

Wang, Y., Sheng, D., Zhang, P., Dong, X., Yan, Y., Hou, X., et al. (2021). High Temperature Sensitivity of Kernel Formation in Different Short Periods Around Silking in maize. *Environ. Exp. Bot.* 183, 104343. doi:10.1016/j.envexpbot.2020.104343

Weber, H., Borisjuk, L., Heim, U., Buchner, P., and Wobus, U. (1995). Seed Coat-Associated Invertases of Fava Bean Control Both Unloading and Storage Functions: Cloning of cDNAs and Cell Type-Specific Expression. *Plant Cell* 7, 1835–1846. doi:10.1105/tpc.7.11.1835

Xuan, X., Chen, S., Zhao, S., Yoon, J. Y., Boczkaj, G., and Sun, X. (2020). Carbon Nanomaterials from Metal-Organic Frameworks: A New Material Horizon for CO_2 Reduction. *Front. Chem.* 8, 573797. doi:10.3389/fchem.2020.573797

Yang, D., Li, Y., Shi, Y., Cui, Z., Luo, Y., Zheng, M., et al. (2016). Exogenous Cytokinins Increase Grain Yield of Winter Wheat Cultivars by Improving Stay-Green Characteristics Under Heat Stress. *PLoS One* 11, e0155437. doi:10.1371/journal.pone.0155437

Yue, R., Lu, C., Qi, J., Han, X., Yan, S., Guo, S., et al. (2016). Transcriptome Analysis of Cadmium-Treated Roots in maize (*Zea mays* L.). *Front. Plant Sci.* 7, 1298. doi:10.3389/fpls.2016.01298

Zhang, Y., and Zhao, Y. (2017). Ensemble Yield Simulations: Using Heat-Tolerant and Later-Maturing Varieties to Adapt to Climate Warming. *PLoS One* 12, e0176766. doi:10.1371/journal.pone.0176766

Zhang, Y. Z., and Huang, B. Q. (2018). Identification of Multiple Genes Encoding SnRK1 Subunits in Potato Tuber. *PLoS One* 13 , e0200321. doi:10.1371/journal.pone.0200321

Zhang, M., An, P., Li, H., Wang, X., Zhou, J., Dong, P., et al. (2019). The miRNA-Mediated Post-Transcriptional Regulation of Maize in Response to High Temperature. *Ijms* 20, 1754. doi:10.3390/ijms20071754

Zhang, C. Y., Bai, J., Ding, X. P., Zhang, J. W., Liu, P., Zhao, B., et al. (2020a). Genome-wide Effects of Staggered Planting With Increased Density on the Photosynthetic Characteristics and Yield of Summer Maize. *Sci. Agric. Sin.* 53, 3928–3941. doi:10.3864/j.issn.0578-1752.2020.19.007

Zhang, H., Li, G., Fu, C., Duan, S., Hu, D., and Guo, X. (2020b). Genome-wide Identification, Transcriptome Analysis and Alternative Splicing Events of Hsf Family Genes in Maize. *Sci. Rep.* 10, 8073. doi:10.1038/s41598-020-65068-z

Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., et al. (2017). Temperature Increase Reduces Global Yields of Major Crops in Four Independent Estimates. *Proc. Natl. Acad. Sci. USA* 114, 9326–9331. doi:10.1073/pnas.1701762114

Zinselmeier, C., Lauer, M. J., and Boyer, J. S. (1995). Reversing Drought-Induced Losses in Grain Yield: Sucrose Maintains Embryo Growth in Maize. *Crop Sci.* 35, 1390–1400. doi:10.2135/cropsci1995.0011183X003500050022x

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 in Genetics

ORIGINAL RESEARCH
published: 11 November 2021
doi: 10.3389/fgene.2021.770014

Genome-wide Identification and Evolution of the *PP2C* Gene Family in Eight Rosaceae Species and Expression Analysis Under Stress in *Pyrus bretschneideri*

Guoming Wang[1†], Xun Sun[1†], Zhihua Guo[1], Dirk Joldersma[2], Lei Guo[2], Xin Qiao[1], Kaijie Qi[1], Chao Gu[1] and Shaoling Zhang[1]**

[1]State Key Laboratory of Crop Genetics and Germplasm Enhancement, Centre of Pear Engineering Technology Research, Nanjing Agricultural University, Nanjing, China, [2]Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD, United States

OPEN ACCESS

Edited by:
Suxu Tan,
Michigan State University,
United States

Reviewed by:
Mehanathan Muthamilarasan,
University of Hyderabad, India
Ghulam Qanmber,
Cotton Research Institute (CAAS),
China

***Correspondence:**
Chao Gu
guchao@njau.edu.cn
Shaoling Zhang
slzhang@njau.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 03 September 2021
Accepted: 11 October 2021
Published: 11 November 2021

Citation:
Wang G, Sun X, Guo Z, Joldersma D, Guo L, Qiao X, Qi K, Gu C and Zhang S (2021) Genome-wide Identification and Evolution of the PP2C Gene Family in Eight Rosaceae Species and Expression Analysis Under Stress in Pyrus bretschneideri. Front. Genet. 12:770014. doi: 10.3389/fgene.2021.770014

Type 2C protein phosphatase (PP2C) plays an essential role in abscisic acid (ABA) signaling transduction processes. In the current study, we identify 719 putative *PP2C* genes in eight Rosaceae species, including 118 in Chinese white pear, 110 in European pear, 73 in Japanese apricot, 128 in apple, 74 in peach, 65 in strawberry, 78 in sweet cherry, and 73 in black raspberry. Further, the phylogenetic analysis categorized *PbrPP2C* genes of Chinese white pear into twelve subgroups based on the phylogenic analysis. We observed that whole-genome duplication (WGD) and dispersed gene duplication (DSD) have expanded the Rosaceae *PP2C* family despite simultaneous purifying selection. Expression analysis finds that *PbrPP2C* genes have organ-specific functions. QRT-PCR validation of nine *PbrPP2C* genes of subgroup A indicates a role in ABA-mediated response to abiotic stress. Finally, we find that five *PbrPP2C* genes of subgroup A function in the nucleus. In summary, our research suggests that the *PP2C* family functions to modulate ABA signals and responds to abiotic stress.

Keywords: type 2C protein phosphatase, abscisic acid, pear (*Pyrus bretschneideri*), Rosaceae, abiotic stress

INTRODUCTION

Protein phosphorylation is a fundamental signal that regulates cellular processes, including growth factor responses, hormone responses, metabolic control, and developmental processes, and as with all signals, removal is as important as induction (den Hertog 1999; Kerk et al., 2002; Schweighofer et al., 2007). Protein kinases (PKs) phosphorylate serine (Ser), threonine (Thr), and tyrosine (Tyr)

Abbreviations: PK, protein kinase; PP, protein phosphatase; Ser, serine; Thr, threonine; Tyr, tyrosine; PTP, protein threonine phosphatase; STP, serine/threonine phosphatase; DSPTP, dual-specificity phosphatase; PPP, phosphor-protein phosphatase; PPM, the Mg2+- or Mn2+-dependent protein phosphatase; PP2C, type 2C protein phosphatase; HMM, hidden Markov mode; WGD, Whole-genome duplication; TD, tandem duplication; TRD, transposed duplication; PD, proximal duplication; DSD, dispersed duplication; ML, maximum likelihood; PGDD, llant genome duplication database; GDR, Genome Database for Rosaceae; RPKM, reads per kilobase of exon model per million mapped reads; Ka, non-synonymous substitution rates; Ks, synonymous substitution rates.

residues, whereas protein phosphatases (PPs) can reverse this action by removing the phosphate group (Luan, 1998; Wei et al., 2014). Therefore, the PPs are classified into three groups based on substrate specificity: protein Tyr phosphatases (PTPs), Ser/Thr phosphatases (STPs), and dual-specificity phosphatases (DSPTPs) (Cohen, 1989; Kerk et al., 2008). STPs are further divided into three subgroups: the phosphor-protein phosphatase (PPP), Mg^{2+}- or Mn^{2+}-dependent protein phosphatase (PPM), and aspartate-based protein phosphatases (Cohen 1989; Kerk et al., 2008). The PPP family covers PP1, PP2A, PP4, PP5, PP6, PP7, and PP2B, whereas the PPM family includes type 2C protein phosphatases (PP2Cs) and also pyruvate dehydrogenase phosphatases (Cohen, 1989; Kerk et al., 2008).

PP2Cs modulate and regulate protein kinase signaling cascades in archaea, bacteria, fungi, plants, and animals (Cao et al., 2016). In higher plants, *PP2C* genes were demonstrated to negatively regulate signaling pathways by opposing specific protein kinases (Tähtiharju and Palva, 2001; Yoshida et al., 2006). In *Arabidopsis*, 76 *PP2C* genes have been identified and categorized into ten groups (A–J), with the remaining six being uncategorized. Several *PP2C* genes from subgroup A have been verified as factors in ABA signaling (Hirayama and Umezawa, 2010). AP2C1 of subgroup B interacts with MPK4 or MPK6 to suppress MAPK activates in response to wounding and pathogen stresses (Schweighofer et al., 2007). POL or PLL1 of PP2C of subgroup C interacts with the receptor kinase CLV1 to regulate flower development and maintain stem cell polarity (Song et al., 2008; Gagne and Clark, 2010). AtPP2C6-6 of PP2C of subgroup E interacts with histone acetyl transferase AtGCN5 to modulate stomatal signaling (Servet et al., 2008). WIN2 of PP2C of subgroup F interacts with the bacterial effector HopW1-1 to induce stress response (Lee et al., 2008). Likewise, KAPP of unclustered PP2Cs interacts with RLKs to regulate plant immunity responses and hormonal signaling (Gomez-Gomez et al., 2001).

PP2Cs regulate plant development in both biotic and abiotic stress conditions (Sugimoto et al., 2014; Singh et al., 2016). At the molecular level, the *PP2C* function reflects its role in modulating signals transmitted by abscisic acid (ABA) (Merlot et al., 2001; Yoshida et al., 2006). In the ABA signaling pathway, PP2Cs can inactivate SnRK2 via dephosphorylation, and this inactivation was inhibited by ABA receptors (PYR/PYL/RCRA) (Soon et al., 2012). In *Arabidopsis,* proteins encoded by *PP2CA, ABI1,* and *ABI2* function in tolerance to exposure to salt, drought, and freezing (Strizhov et al., 1997; Merlot et al., 2001), and proteins encoded by *HAB1*, *HAB2,* and *AHG1* negatively regulate SnRK2 kinases required for ABA signaling (Saez et al., 2004; Saez et al., 2008; Umezawa et al., 2009). In *Fagus sylvatica*, *e*ctopic expression of *FsPP2C1* in *Arabidopsis* resulted in ABA insensitivity during seed germination (Gonzalez-Garcia et al., 2003), but ectopic expression of *FsPP2C2* in *Arabidopsis* resulted in enhanced ABA sensitivity and tolerance of abiotic stress in seeds (Reyes et al., 2006). This link between PP2C proteins and ABA signaling is ancient. In moss, *PpABI1A* and *PpABI1B* of subgroup A of *PP2C* function in drought tolerance via downregulation of ABA signaling (Komatsu et al., 2013), and in maize, *ZmPP2C-A10* also regulates drought stress tolerance (Xiang et al., 2017). Taken together, *PP2C* genes in subgroup A have been demonstrated to play key roles in plant development and environmental stresses.

In this study, we identify 719 *PP2C* genes from eight Rosaceae species. At the genomic level, we analyze the expression and phylogeny of the 118 *PP2C* genes found in Chinese white pear and analyze its evolution. At the protein level, we describe protein features and functions, domains of expression, and subcellular localization. Our findings set a foundation to understand the function of *PbrPP2C* genes in mediating responses to various stress conditions in a commercially important family of higher plants.

MATERIALS AND METHODS

Sequence Retrieval Resources and Identification of Type 2C Protein Phosphatase Genes

For identification of *PP2C* genes in pear and other Rosaceae species, the HMM (hidden Markov odel) files were constructed by downloading the seed file PP2C (PF00481) domains from Pfam (http://pfam.xfam.org/) and were searched against the local protein databases using HMMER3 (Finn et al., 2015). The protein sequences of candidate *PP2C* genes were validated by using Interproscan 63.0 (http://www.ebi.ac.uk/InterProScan/) and Pfam (http://pfam.xfam.org/). The methods of screening and identification were identical to a previous report (Qanmber et al., 2019a; Qanmber et al., 2019b). *Arabidopsis PP2C* genes were downloaded from TAIR (http://www.arabidopsis.org/). Chinese white pear (*Pyrus bretschneideri*) and Japanese apricot's (*Prunus mume*) genome sequences were obtained from the Pear Genome Project (http://peargenome.njau.edu.cn/) and *Prunus mume* Genome Project (http://prunusmumegenome.bjfu.edu.cn/index.jsp), respectively. The genome sequences of European pear (*Pyrus communis*), apple (*Malus domestica*), peach (*Prunus persica*), strawberry (*Fragaria vesca*), black raspberry (*Rubus occidentalis*), and sweet cherry (*Prunus avium*) were collected from the Genome database for Rosaceae (GDR) (http://www.rosaceae.org). The obtained PP2C protein sequences were screened for the PP2C catalytic domain by using the Pfam website (https://pfam.xfam.org/).

Phylogenetic, Exon–Intron Structure, and Protein Motif Analysis

The full length protein sequences of *PP2C* were used to perform multiple sequence alignment, and the phylogenetic tree was performed using MEGA7.0 (Kumar et al., 2016) with the maximum likelihood method (ML), a bootstrap of 1,000 replications, and the Jones–Taylor–Thornton (JTT) model. The exon–intron organization of *PbrPP2C* genes was analyzed using CDSs and genomic DNA sequences using the Gene Structure Display Server (GSDS: http://gsds.cbi.pku.edu.cn/). Motif Elicitation (MEME: http://meme.sdsc.edu/meme/itro.

html) was performed to identify conserved motifs of PbrPP2C proteins, with the maximum number of motifs = 20.

Cis-Element Predictions of *Pbr*PP2C

All the *PbrPP2C* promoter sequences (selected as 2000 upstream bp) were downloaded from the Pear Genome Project (http://peargenome.njau.edu.cn/). The *cis*-regulatory elements of *PbrPP2C* were identified by the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/).

Chromosomal Locations, Synteny, and *Ka*/*Ks* Analysis of Type 2C Protein Phosphatase Genes

The chromosomal location information of the *PP2C* family genes was obtained from the genome annotation files. The analysis of synteny among eight Rosaceae genomes was performed by PGDD (http://chibba.agtec.uga.edu/duplication/) (Lee et al., 2013). BLASTP was carried out to identify multiple alignments of protein sequences (e-value < 10–5, top 5 matches) in the eight Rosaceae species. Then, MCScanX was carried out to produce orthologous gene pairs of *PP2C* within each Rosaceae species (Wang et al., 2012). Segmental/whole-genome duplication (WGD), tandem duplication (TD), proximal duplication (PD), transposed duplication (TRD), and dispersed duplication (DSD) in the *PP2C* gene family were identified by using the tools in the MCScanX package (Qiao et al., 2019). Localization and duplicate gene pairs of the *PP2C* genes were visualized using TB tools software (Chen et al., 2020). The values of *Ka* (non-synonymous substitutions) and *Ks* (synonymous substitutions) were calculated using *KaKs*_Calculator 2.0 with default parameters, and the *Ka*/*Ks* ratio was based on a model-averaged method (Wang et al., 2010).

Transcriptome Expression Pattern Analysis in Different Tissues and in Different Ages of Pear Fruit

Previously published and unpublished dynamic RNA-seq data were used to analyze *PP2C* gene expression in different tissues of pear (Qiao et al., 2018; Li et al., 2019), including the pollen, seed, sepal, petal, ovary, bud, stem, leaf, and fruit. The raw RNA-seq reads were cleaned by removing low-quality reads (quality score <15), poly (A/T) tails, and adapter sequences. HISAT2 and feature counts were performed to align clean reads to the reference genome and estimate transcript abundance levels (Liao et al., 2014; Kim et al., 2015). Finally, the values of fragments per kilobase million (FPKM) were used to indicate the expression levels of *PP2C* genes. The heatmap of *PP2C* gene expression was visualized using TB tools software (Chen et al., 2020).

Plant Material and Treatment

"Cuiguan" pear (*Pyrus pyrifolia* Nakai) seeds were collected from the pear germplasm orchard of the Pear Engineering Technology Research Center of Nanjing Agricultural University in Nanjing, China. Seedlings were grown for 5 weeks in a growth chamber, with a photoperiod of 16/8 h and a temperature of 25 ± 1°C. The seedlings were irrigated with 200 mM NaCl and 20% PEG 6000 for salinity and drought abiotic stress, respectively. Seedling leaves were sprayed with 100 μM ABA for ABA treatment. Seedlings were subjected to temperatures of 4°C and 37°C for low and high temperature stress, respectively. The leaves of treated seedling were collected at 0, 6, 12 and 24 h, respectively. The collected samples were quickly frozen in liquid nitrogen and stored at −80°C until further use.

Quantitative Real-Time PCR Analysis

Total RNA was extracted using RNAprep Pure Plant Kit (Tiangen, Beijing, China). The extracted total RNAs were subjected to the first-strand cDNA using TransScript One-Step gDNA Removal and cDNA synthesis Supermix (TransGen, Beijing, China). The primers of 9 *PP2C* genes were designed using Primer Premier 6.0, and the tubulin gene of pear was used as the reference gene. All the primer sequences are listed in **Supplementary Table S5**. QRT-PCR was carried out in LightCycler 480 SYBRGREEN I Master (Roche, United States), and the reaction mixture and cycling program were identical to those of a previous report (Hao et al., 2018). All of the analyses were carried out with three independent biological replicates. The genes expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

Subcellular Localization

For the subcellular localization analysis of PbrPP2Cs, the CDS sequences without the termination codon were amplified and cloned into pCAMBIA1300-35S: CDS-GFP vector. Primers used for cloning are listed in **Supplementary Table S5**. The recombinant plasmids and the control plasmid were transformed into 30-day-old tobacco (*Nicotiana benthamiana*) leaves according to the published protocol (Sparkes et al., 2006). Fluorescence was imaged using a confocal microscope LSM780 (Zeiss LSM 780, Germany).

RESULTS

Identification, Characteristics, and Phylogenetic Relationship of Type 2C Protein Phosphatase Genes

A total of 719 putative *PP2C* genes were identified in eight Rosaceae species: 118 in Chinese white pear, 110 in European pear, 73 in Japanese apricot, 128 in apple, 74 in peach, 65 in strawberry (Haider et al., 2019), 78 in sweet cherry, and 73 in black raspberry (**Supplementary Table S1**). 118 putative *PP2C* genes of pear were arranged as *PbrPP2C1* to *PbrPP2C118* based on phylogenic analysis and the relative position of *Arabidopsis* orthologs. In addition, some members of the *PbrPP2C* gene family have two alternative splice variants, including *Pbr012020*, *Pbr019958*, *Pbr019984*, *Pbr022419,* and *Pbr031084* (**Supplementary Table S1**). Splicing variants played a crucial role in the posttranscriptional regulatory mechanism that modulates transcriptome and proteome diversity, such as alternative splicing

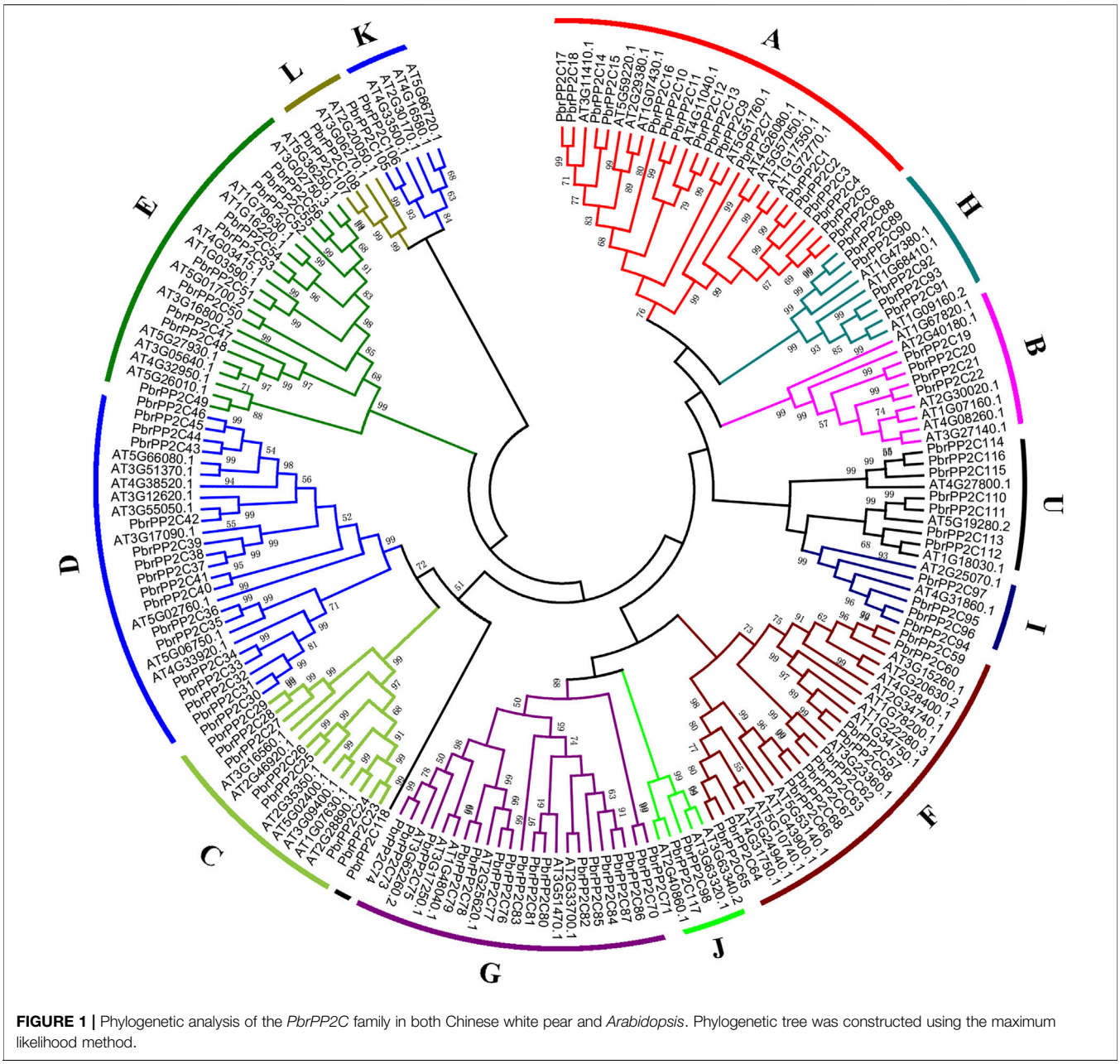


FIGURE 1 | Phylogenetic analysis of the *PbrPP2C* family in both Chinese white pear and *Arabidopsis*. Phylogenetic tree was constructed using the maximum likelihood method.

of *PpDAM1* was important in the pear flower bud dormancy process (Li et al., 2021). The lengths of *PP2C* gene sequences ranged from 203 bp to 21,156 bp. Moreover, the protein molecular weights were 7.5–243 kDa, and the theoretical isoelectric point was from 3.9 to 10.49. The *PP2C* gene ID and the characteristics are shown in **Supplementary Table S1**.

To gain insights into the phylogenetic relationship of the *PP2C* genes in pear, a phylogenetic tree was constructed using MEGA7.0 by adopting the maximum likelihood method (ML) based on multiple sequence alignments of 80 *Arabidopsis PP2C* genes obtained from a previous study (Xue et al., 2008) and 118 pear *PbrPP2C* genes (**Figure 1**). This analysis divided 118 *PbrPP2C* genes into twelve subgroups: subgroups A–L and one unclassified subgroup U (**Figure 1** and **Supplementary Figure S1**). Subgroups A, D, and G contain 18, 17, and 17 genes, respectively. Subgroup J, K, and L contain less.

In addition, the conserved motif analysis of *PbrPP2C* genes supported the phylogenetic analysis and classification (**Supplementary Figure S1**). A total of 20 motifs were identified in all the *PbrPP2C* family members. The PP2C proteins of each subgroup contained similar motifs. Motifs 1 and 5 were detected in almost all of the *PbrPP2C* genes, while Motif 4 was specific to subgroups C and D (**Supplementary Figure S1**). Most members of *PbrPP2C* subgroups contained more than 10 motifs, while a few members had 2–4 motifs, such as *PbrPP2C16*, *PbrPP2C71*, *PbrPP2C85*, *PbrPP2C113*, and

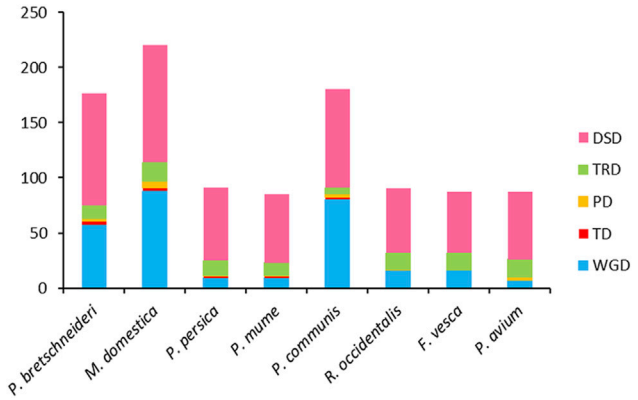


FIGURE 2 | Number of *PP2C* gene pairs derived from different modes of gene duplication in pear and seven other Rosaceae species. *X*-axis represents the species. *Y*-axis represents the number of duplicated gene pairs. WGD: whole-genome duplication, TD: tandem duplication, PD: proximal duplication, TRD: transposed duplication, and DSD: dispersed duplication.

PbrPP2C118. Different subgroups contain their own specific motifs that may lead to the functional divergence of each subgroup. To better understand the structures of *PbrPP2C* genes in pear, exon–intron organizations were compared among different subgroups (**Supplementary Figure S1**). The number of exons in the *PbrPP2C* family members varied from 1 to 21, and 37 genes were annotated in the 3′or 5′UTR region. Interestingly, *PbrPP2C* genes in the same subgroup show more or less similar exon–intron structures (**Supplementary Figure S1**).

Chromosome Location and Collinearity Analysis of the Type 2C Protein Phosphatase Gene Family

To explore the contribution of different gene duplication modes to the expansion and evolution of *PP2C* genes in eight Rosaceae species, a comparative analysis of gene duplication was performed in each genome (**Figure 2**, **Supplementary Table S2**). 1,014 duplicated gene pairs were found in the *PP2C* family members and were assigned to five duplication modes of WGD, PD, TD, TRD, or DSD. WGD is responsible for 8.2–44.4% of *PP2C* gene pairs in the investigated species. Consistent with lineage-specific duplications, Chinese white pear (32.4%), European pear (44.4%), and apple (40.0%) exhibit a relatively high proportion of WGD-derived *PP2C* genes. DSD accounts for the highest number of derived genes (48.2–72.9%), but TD (0–2.4%) and PD (0–2.7%) were observed with low frequency. TRD (3.3–18.4%) of *PP2C* gene pairs shows a high frequency in each of the Rosaceae species, which contributed to the formation of the *PP2C* gene clusters observed. The results are consistent with the inference that different gene duplication modes of *PP2C* gene pairs may have led to the neofunctionalization of ancestral genes.

We located the 719 *PP2C* genes on the chromosomes of each species based on genome annotations. For Chinese white pear, 118 of *PbrPP2C* genes were anchored onto all the 17 chromosomes and scaffolds, with the maximum number of genes detected on Chr 15 (17), followed by 11 genes on Chr 5. However, there was no correlation between the number of genes and the length of chromosomes, and the genes were randomly distributed on each chromosome (**Figure 3A**, **Supplementary Table S2**). *PP2C* genes were also found to be randomly distributed in the other Rosaceae species' genomes. We identified 274 collinearity gene pairs, including 59 pairs in Chinese white pear, 16 pairs in strawberry, 78 pairs in apple, 80 pairs in European pear, 10 pairs in Japanese apricot, 8 pairs in peach, 16 pairs in black raspberry, and 7 pairs in sweet cherry (**Figure 3**, **Supplementary Table S2**). A large number of collinearity gene pairs were identified in Chinese white pear, European pear, and apple species.

Ka and *Ks* Substitutions per Site and *Ka*/*Ks* Analysis for Type 2C Protein Phosphatase Family Genes

The *Ks* value has been widely used to estimate the evolutionary stage of WGD events (Qiao et al., 2015). The mean *Ks* values of WGD-derived gene pairs in apple, European pear, Chinese white pear, Japanese apricot, black raspberry, sweet cherry, strawberry, and peach were 1.70, 1.90, 2.21, 2.40, 2.57, 2.907, 2.93, and 3.48, respectively (**Supplementary Table S3**). The lower *Ks* values of WGD-derived gene pairs in apple and European pear suggested that they were duplicated and retained from more recent WGD events, while sweet cherry, strawberry and peach were derived from more ancient WGD events. The *Ka*/*Ks* ratio is usually used to measure the magnitude and direction of selection pressure, and the *Ka*/*Ks* value refers to selection type: >1 indicate positive selection and <1 indicates purifying selection (Yang 2007). Purifying selection can remove deleterious mutations, and positive selection can induce favorable mutations (Starr et al., 2003). Here, the *Ka*/*Ks* values of *PP2C* orthologous gene pairs were calculated among eight Rosaceae species (**Figure 4**). The *Ka*/*Ks* values of duplicated gene pairs in European pear, Japanese apricot, black raspberry, and strawberry were less than 1, suggesting that *PP2C* genes evolved under strong purifying selection. Several gene pairs with higher *Ka*/*Ks* ratios were identified in apple, Chinese white pear, sweet cherry, and peach, suggesting that these genes may have a complicated evolutionary history. For Chinese white pear, the mean *Ka*/*Ks* ratios of WGD TD, PD, TRD, and DSD, were 0.17, 0.45, 0.98, 0.14, and 0.21, respectively (**Figure 4**). TD and PD had higher *Ka*/*Ks* ratios compared with other molds of duplicated gene pairs, suggesting that they evolved at a higher rate than the other gene pairs.

Analysis of Putative Regulatory *Cis*-elements of the Type 2C Protein Phosphatase Gene Family

The *cis*-elements in promoter regions are closely related to gene transcription, and they play a critical role in plant signal transduction by interacting with their cognate transcription factor. Therefore, to better understand the function of *PbrPP2C* genes, 2.0 kb upstream promoter sequences of

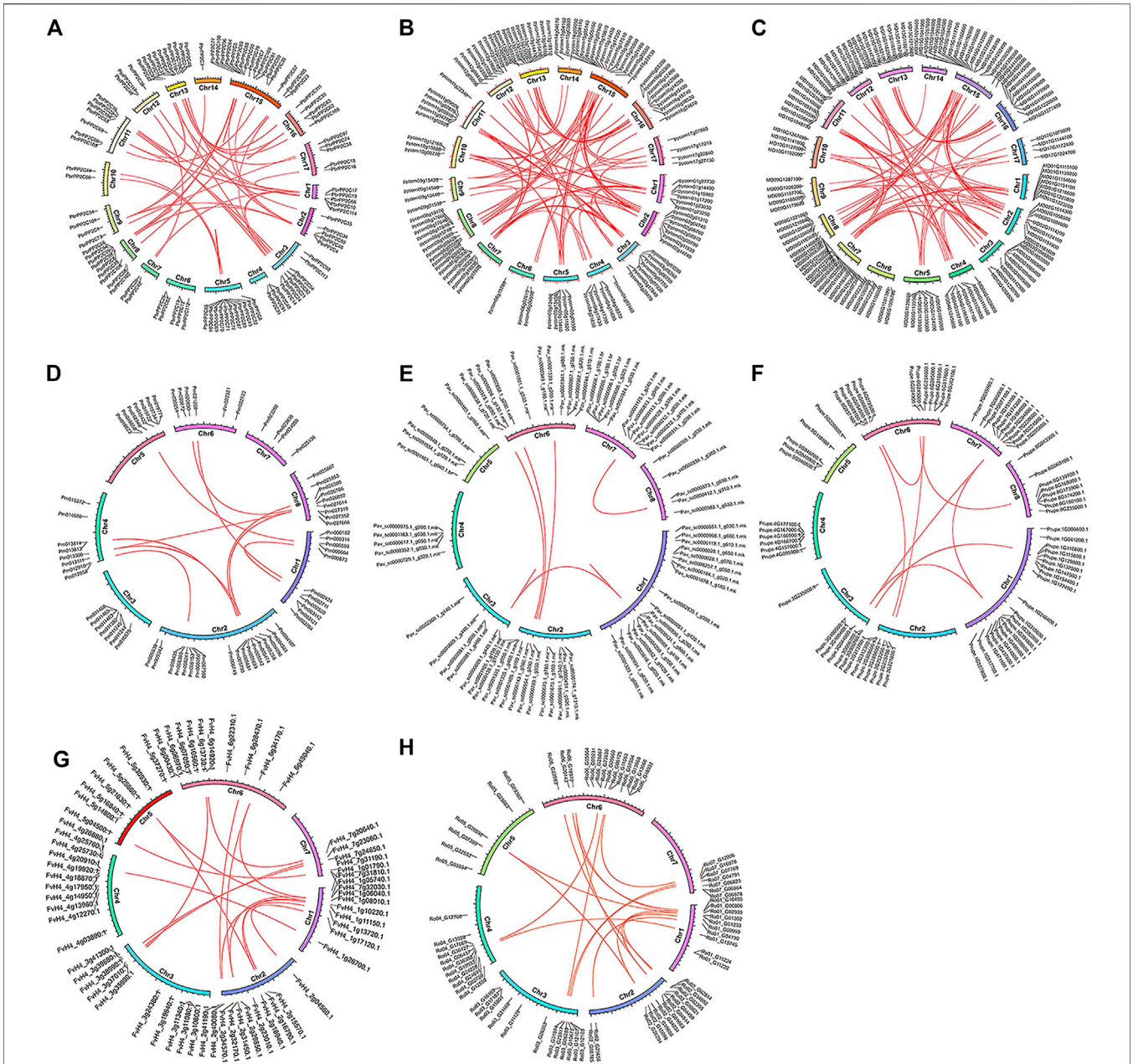


FIGURE 3 | Gene location and collinearity analysis of the *PP2C* gene family. **(A)** Chinese white pear; **(B)** European pear; **(C)** apple; **(D)** Japanese apricot; **(E)** sweet cherry; **(F)** peach; **(G)** strawberry; **(H)** black raspberry. Genes were located on different chromosomes. Red lines represent the syntenic gene pairs.

PbrPP2C genes were downloaded from the Pear Genome Database and analyzed by using PlantCARE. Some common *cis*-regulatory elements are briefly summarized and listed in **Supplementary Table S6**, such as ABRE was involved in ABA responsiveness, ERE was involved in ethylene responsiveness, MBS was the MYB binding site involved in drought-inducibility, and DRE was involved in adverse stress. The result showed that various *cis*-elements were related to plant hormones, light, abiotic stress, essential element, enhancer, circadian factors, and other regulatory stress responses (**Supplementary Table S6**). Consequently, various conserved *cis*-regulatory elements of PbrPP2C genes were crucial in mediating responses to various stress-related hormones or adverse biotic–abiotic stresses.

Expression Profiling of the Type 2C Protein Phosphatase Gene Family in Different Tissues of Pear

To investigate the expression patterns of *PbrPP2C* family genes in different pear tissues, a heatmap was constructed using previously published RNA-seq data including matured pollen, seed, petal, sepal, ovary, stem, bud, leaf, and fruit (**Figure 5** and

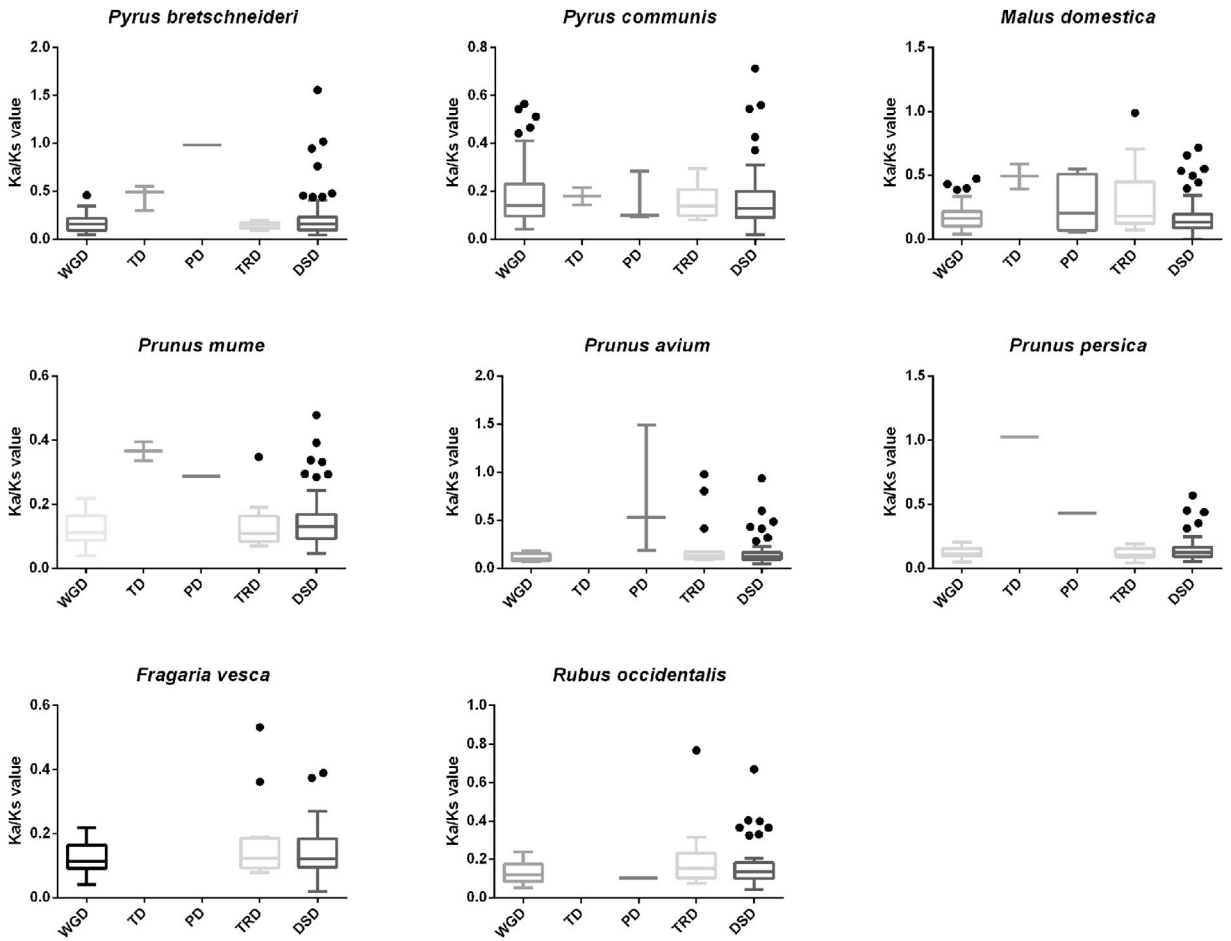


FIGURE 4 | Ka/Ks distribution of eight Rosaceae species. Ka/Ks values were analyzed using coding sequences. *X*-axis represents five different duplication categories. *Y*-axis indicates the Ka/Ks ratio. T-boxplot was also constructed using prism 6.0.

Supplementary Table S4). Most *PbrPP2C* genes displayed a very broad expression range, and several *PbrPP2C* genes showed expressional activation in at least three or more tissues. Five genes (*PbrPP2C71*, *PbrPP2C72*, *PbrPP2C71*, *PbrPP2C102*, *PbrPP2C103,* and *PbrPP2C118*) exhibited very low or no expression. Eighteen *PbrPP2C* genes were found to be highly expressed preferentially in leaves, and five were highly expressed in buds. We identified three genes with ovary-specific expression (*PbrPP2C12*, *PbrPP2C16,* and *PbrPP2C49*), three were pollen-specific (*PbrPP2C9*, *PbrPP2C50,* and *PbrPP2C62*), and one gene exhibited sepal-specific expression (*PbrPP2C69*) (**Figure 5**). This analysis has identified candidate *PbrPP2C* genes that may play specialized roles in different organs' development.

To further verify the role of PP2C-mediated stress response, we analyzed the transcriptome of pear under biotic/abiotic stresses and pathogen treatments based on unpublished and published data of our laboratory (Li et al., 2016; Yang and Huang, 2018). The findings of the heatmaps showed that a large number of *PbrPP2C* genes responded to stress, and *PbrPP2C* genes showed variation in their expression pattern among different treatments (**Supplementary Figure S2**). In addition, most of *PbrPP2Cs* from subgroup A were also differentially upregulated by exposure to cold, drought, salt, and pathogen treatments, such as *PbrPP2C1*, *PbrPP2C4*, *PbrPP2C7*, *PbrPP2C11*, *PbrPP2C17,* and *PbrPP2C18* (**Supplementary Figure S2**). The diversity in the expression profiling of *PbrPP2C* genes may suggest that these *PbrPP2C* genes were stress-responsive.

qRT-PCR Analysis of *Pbr*PP2C

To explore *PbrPP2C* gene expression under different stress conditions and identify *PbrPP2C* genes important for improving tolerance, the seedlings were subjected to heat, cold, drought, NaCl, and ABA treatments. It has been verified that subgroup A *PP2Cs* in *Arabidopsis* and rice were transcriptionally upregulated under exogenous ABA treatment or abiotic stress that stimulates ABA biosynthesis (Xue et al., 2008; Singh et al., 2010). Nine *PbrPP2C* genes from subgroup A were selected to perform qRT-PCR at different time points after various treatments. QRT-PCR after exogenous ABA treatment indicated that four genes (*PbrPP2C10*, *PbrPP2C11*, *PbrPP2C15,* and *PbrPP2C18*) were upregulated at three different time points:

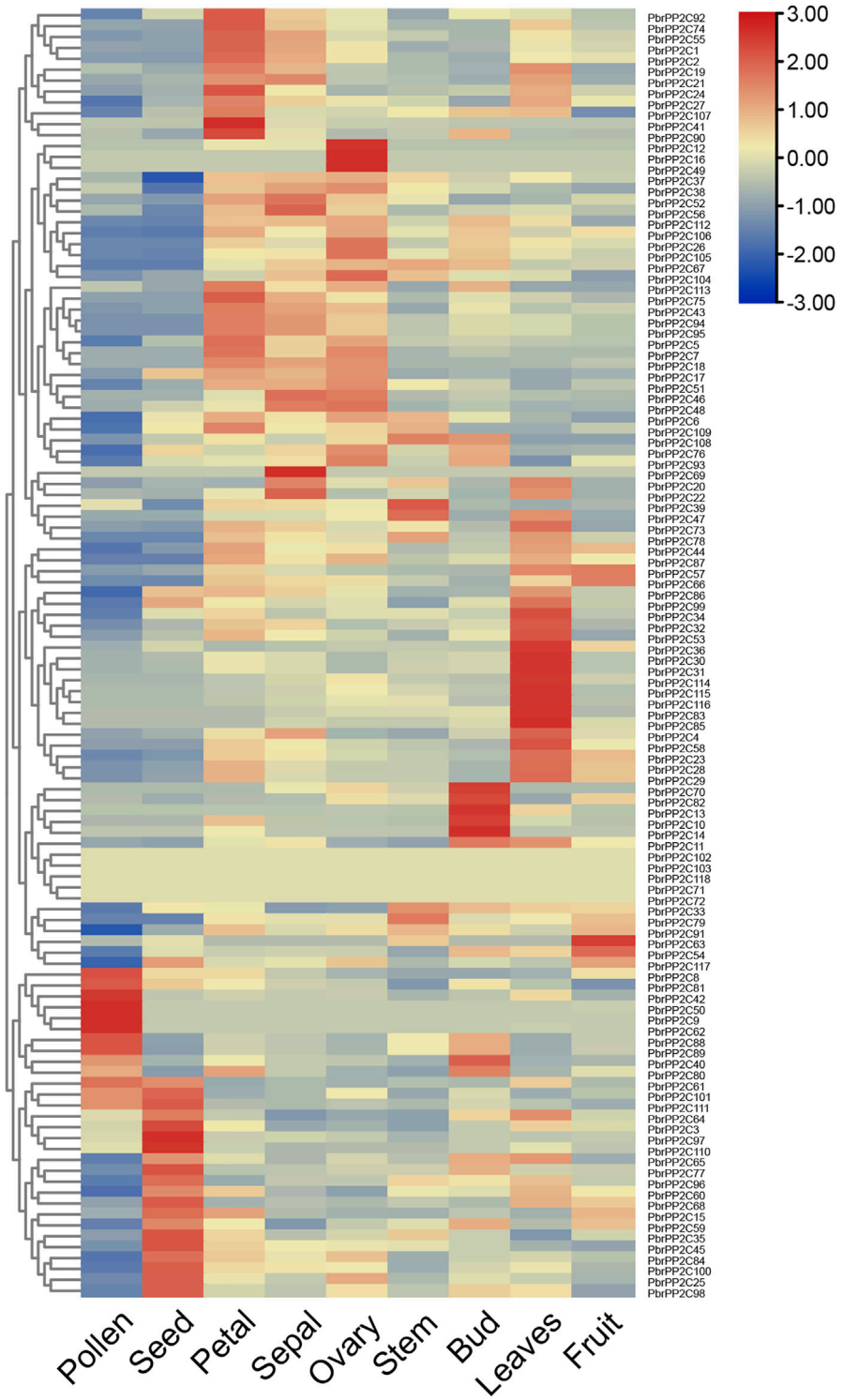


FIGURE 5 | Heatmap of expression profiles (FPKM) for *PP2C* in the nine various tissues of pear (pollen, seed, petal, sepal, ovary, stem, bud, leaf, and fruit). Expression levels are indicated by the color bar.

6, 12, and 24 h. Four more genes (*PbrPP2C1/2*, *PbrPP2C4*, *PbrPP2C6,* and *PbrPP2C17*) were upregulated more than 5-fold in 24 h after exogenous ABA treatment. In contrast, the expression of *PbrPP2C6* and *PbrPP2C7* decreased to less than half that of CK in 6 and 12 h, respectively (**Figure 6**). Nine *PbrPP2C* genes from subgroup A were upregulated in response to more

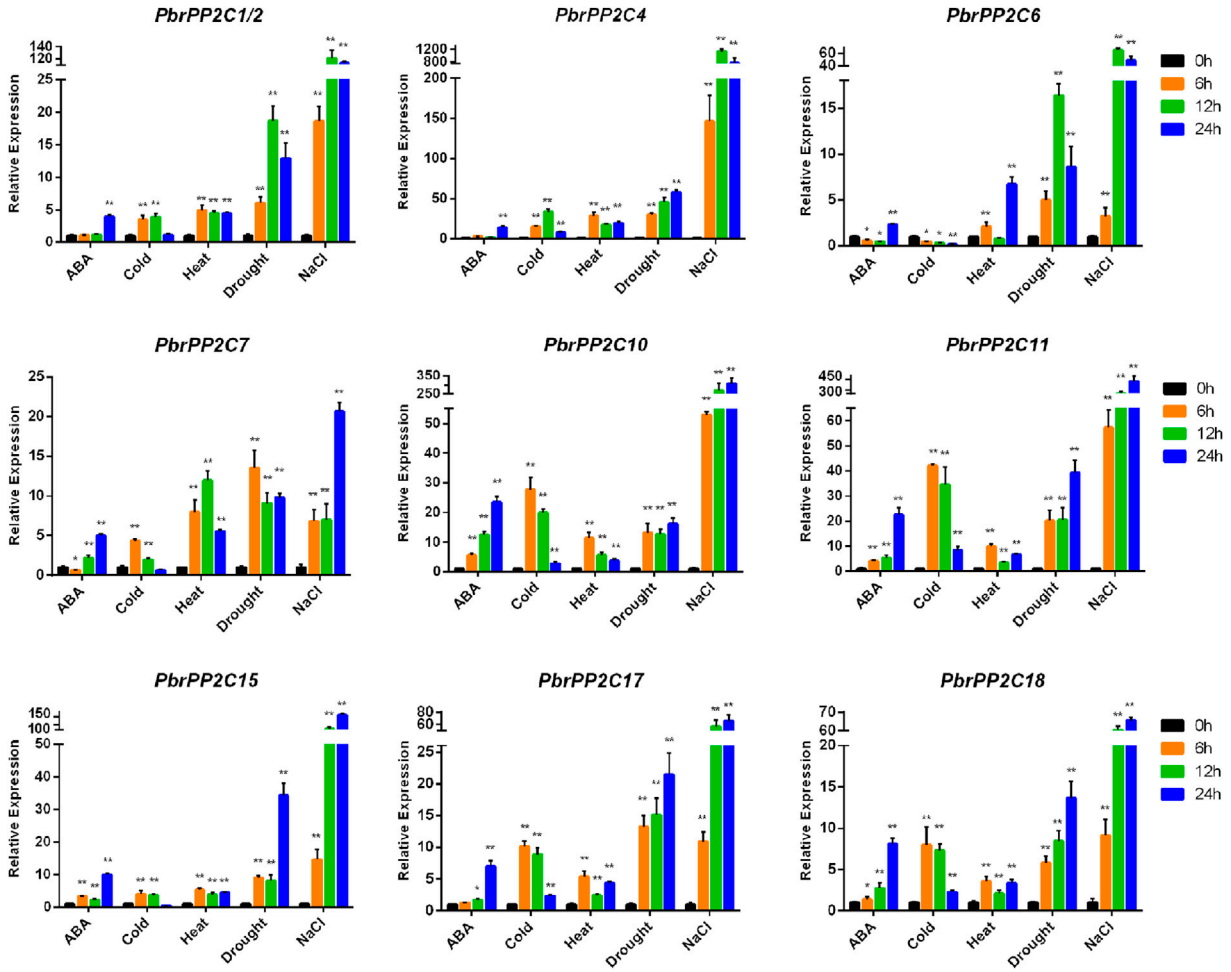


FIGURE 6 | QRT-PCR analysis of subgroup A *PbrPP2C* genes under exposure to heat, cold, drought, NaCl, and ABA treatments in pear seedlings. Standard errors and ANOVA were calculated by applying Student's t-test. Single and double stars stand for the levels of significant difference at p-value < 0.05 and <0.01, respectively.

than one treatment with abiotic stress (**Figure 6**). For example, the *PbrPP2C6* expression level increased more than 3-fold under exposure to heat, drought, and salt treatment, but exposure to cold repressed its expression (**Figure 6**). All nine genes exhibited strongly increased expression levels (from 10-fold to 400-fold greater than CK) in response to NaCl treatment (**Figure 6**). Some results were consistently consistent with the analysis of heatmaps (**Supplementary Figure S2**). Taken together, the expression characteristics of subgroup A *PbrPP2C* genes indicate that these nine genes respond to exogenous ABA and abiotic stress.

Subcellular Localization of PbrPP2C Protein

Previous research has shown that PbrPP2C proteins localize in the nucleus (Haider et al., 2019; Kim et al., 2012). To determine the subcellular localizations of PbrPP2C proteins, five *PbrPP2C* genes (*PbrPP2C1, PbrPP2C4, PbrPP2C7, PbrPP2C10,* and *PbrPP2C15*) were selected from each branch from subfamily A of the *PbrPP2C* family. PbrPP2Cs-GFP recombinant plasmids were introduced into *N. benthamiana* leaves. The fluorescence confirms the localization of five PbrPP2C-GFP fusion proteins in the nucleus (**Figure 7**). ABA receptor proteins, namely PYR/PYL/RCRA, are localized in both the cytoplasm and nucleus, despite the fact that interacting PP2C proteins are localized exclusively in the nucleus (Santiago et al., 2009; Kim et al., 2012). This observation suggests that PbrPP2C may play a transcriptional regulatory role in the nucleus.

DISCUSSION

The landscape of the *PP2C* genes family in the plant kingdom has been characterized in previous research studies, such as in *Arabidopsis* (Kerk et al., 2002), rice (Singh et al., 2010), maize (Wei and Pan 2014), banana (Hu et al., 2017), *Brachypodium distachyon* (Cao et al., 2016), *Brassica rapa* (Khan et al., 2019), and *Gossypium hirsutum* (Shazadee et al., 2019). However, the *PP2C* gene family has not been studied widely in the Rosaceae family. The current study provides a comprehensive analysis of the *PP2C* gene family in eight Rosaceae species, including gene identification, phylogenetic relationships, chromosomal

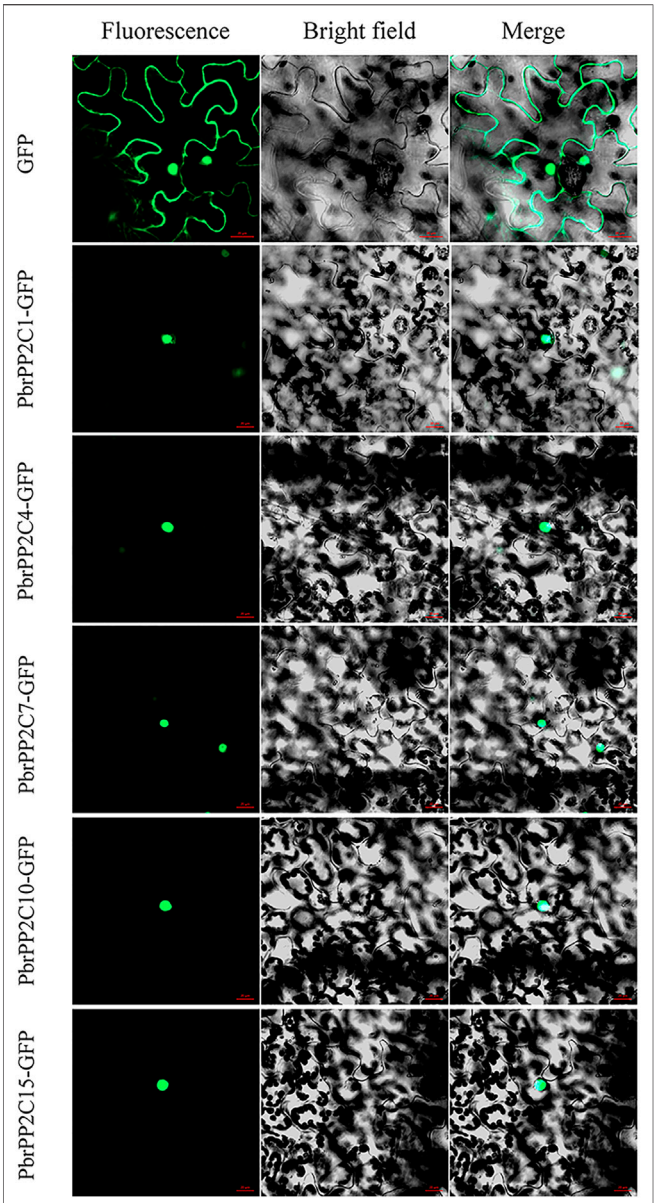


FIGURE 7 | Subcellular localization of the fusion protein PbrPP2Cs-GFP in *N. benthamiana* leaves. Vector 35S-GFP was used as the control. Bar = 20 μm.

localizations, and evolutionary analysis. 719 *PP2Cs* were identified, ranging from 65 genes in strawberry to 128 in apple. The genes exhibited widespread and uneven distribution across Rosaceae chromosomes (**Figure 3**). In Chinese white pear, 118 *PbrPP2C* genes were further categorized into twelve subgroups A–L and one unclassified subgroup according to phylogenetic and evolutionary analysis (**Figure 1**). This classification of *PP2C* genes was consistent with previous studies, such as those conducted on *Arabidopsis* (Kerk et al., 2002) and rice (Singh et al., 2010). *PbrPP2C* genes were found to be expressed in specific organs, providing strong evidence of specialized function. At least eight *AtPP2C* genes from subgroup A were key factors in the ABA signaling network (Hirayama and Umezawa, 2010). *AP2C1* of subgroup B was involved in suppressing MAPK activates in response to wounding or pathogen stresses (Schweighofer et al., 2007). *POL* and *PLL1* of subgroup C were involved in regulating flower development and maintain stem cell polarity (Song et al., 2008; Gagne and Clark, 2010). *PP2C* of subgroup E and *AtPP2C6-6* were involved in modulating stomata signaling (Servet et al., 2008). *WIN2* of subgroup F was involved in inducing the stress response (Lee et al., 2008). Overall, *PP2C* genes of the same subgroup have similar specialized biological functions, although the functions of many *PP2C* subgroups are still unclear.

Gene duplication is the predominant driving force for broad expansion of the gene family, which could obtain new functions and evolutionary processes (Qiao et al., 2019). The different types of gene duplications, including WGD, TD, PD, TRD, and DSD (Doerks et al., 2002; Moore and Purugganan, 2003), contribute differently to the expansion of gene families (Freeling, 2009). In the genome and genetic evolutionary system, WGD is the main driving force of new functions and features of eukaryotic genome evolution (Friedman and Hughes, 2001; Moore and Purugganan, 2003). For instance, the *BES1* and *GhERF* subfamily *B3* group gene families in cotton were expanded primarily though segmental or WGD duplication events (Liu et al., 2018; Lu et al., 2021). *F-box* and heat-shock transcription factor families in pear were expanded primarily through WGD and DSD (Qiao et al., 2015; Wang et al., 2016). In our study, we demonstrated WGD replication events in eight Rosaceae species (**Figure 3**). We also show that WGD and DSD were the driving forces for the expansion of the *PP2C* gene family in Chinese white pear, European pear, and apple (**Figure 2**). Finally, we found that most *Ka*/*Ks* ratios of *PP2C* gene pairs were less than one, suggesting that these genes have experienced strong purifying selection.

Surrounded by various stress factors, such as ABA, drought, salt, heat, cold, and phytohormonal stresses, are the major limiting factors of plant development and agricultural crop production, and the role of ABA signaling in stress adaptation and stress resistance mechanisms has been well documented (Sugimoto et al., 2014; Singh et al., 2016). Group A of *PP2Cs* comprises of negative regulators of ABA signaling by PYL intracellular receptors (Antoni et al., 2012). In *Arabidopsis*, at least six genes of group A *PP2C* (*ABI1*, *ABI2*, *HAB1*, *HAB1*, *HAB2*, *PP2CA,* and *AHG1*) resulted in increasing the ABA sensitivity under various stresses, indicating the diverse outcome in ABA signaling (Merlot et al., 2001; Tähtiharju and Palva, 2001; Umezawa et al., 2009). In *Fagus sylvatica*, two genes of group A *PP2C* (*FsPP2C1* and *FsPP2C2*) resulted in influencing ABA sensitivity and tolerance of abiotic stress in seeds (Gonzalez-Garcia et al., 2003; Reyes et al., 2006). In maize, *ZmPP2C-A10* of subgroup A of *PP2C* was confirmed for its negative regulation in drought stress (Xiang et al., 2017). Although *PP2C* genes in subgroup A have been demonstrated to play key roles in various stress conditions in some species, the role of the key components of ABA signaling against *PP2C* genes is mainly obscure in pear. In the present study, nine

PbrPP2C genes from subgroup A exhibited substantial transcriptional variations when confronted by heat, cold, drought, and NaCl challenges and in response to ABA treatment, indicating their regulatory role in stress tolerance. Gene expression levels exhibited, especially, strong response to salt stress (**Figure 6**). Therefore, the study reveals potential functions of *PP2C* genes in a commercially important angiosperm family. However, validation of the individual gene product's function at the molecular level remains an important step in understanding *PP2C* genes in the Rosacaea family in future.

5. CONCLUSION

In this study, a total of 719 *PP2C* gene family members were first identified in eight Rosaceae species. The *PP2C* gene pairs of Rosaceae species might evolve undergoing strong purifying selection. The 118 *PbrPP2C* genes of Chinese white pear were classified into twelve subgroups according to the phylogenetic relationship gene structure and protein motif pattern. Moreover, qRT-PCR revealed nine candidate genes from subgroup A which might have participated in the plant stress tolerance particularly to ABA, heat, cold, drought, and NaCl stress. Subcellular localization analysis proved the functionality of five *PbrPP2C* genes from each branch of subfamily A in the nucleus. Consequently, our findings provide a foundation for the potential function of *PbrPP2C* genes under various stress conditions.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

GW carried out the experiments and wrote the manuscript. CG and SZ designed the experiment and revised the manuscript. XQ contributed to collinearity analysis and the Perl script. XS, XG and ZG performed the experiments. DJ, LG and KQ directed and revised the manuscript. All authors have read and approved the final manuscript.

FUNDING

This work was financially supported by Jiangsu Agriculture Science and Technology Innovation Fund (CX (19)2028), the National Natural Science Foundation of China (31830081), the National Key Research and Development Program of China (2018YFD1000107), the Fundamental Research Funds for the Central Universities (JCQY201901), the Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions, the Earmarked Fund for China Agriculture Research System (CARS-28).

ACKNOWLEDGMENTS

The authors thank the laboratory members for assistance. This work was supported by the high-performance computing platform of Bioinformatics Center, Nanjing Agricultural University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.770014/full#supplementary-material

REFERENCES

Antoni, R., Gonzalez-Guzman, M., Rodriguez, L., Rodrigues, A., Pizzio, G. A., and Rodriguez, P. L. (2012). Selective Inhibition of Clade A Phosphatases Type 2C by PYR/PYL/RCAR Abscisic Acid Receptors. *Plant Physiol.* 158, 970–980. doi:10.1104/pp.111.188623

Cao, J., Jiang, M., Li, P., and Chu, Z. (2016). Genome-Wide Identification and Evolutionary Analyses of the PP2C Gene Family with Their Expression Profiling in Response to Multiple Stresses in Brachypodium Distachyon. *BMC Genomics* 17, 175. doi:10.1186/s12864-016-2526-4

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* 13, 1194–1202. doi:10.1016/j.molp.2020.06.009

Cohen, P. (1989). The Structure and Regulation of Protein Phosphatases. *Annu. Rev. Biochem.* 58, 453–508. doi:10.1146/annurev.bi.58.070189.002321

den Hertog, J. (1999). Protein-Tyrosine Phosphatases in Development. *Mech. Develop.* 85, 3–14. doi:10.1016/s0925-4773(99)00089-1

Doerks, T., Copley, R. R., Schultz, J., Ponting, C. P., and Bork, P. (2002). Systematic Identification of Novel Protein Domain Families Associated with Nuclear Functions. *Genome Res.* 12, 47–56. doi:10.1101/gr.203201

Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., et al. (2015). HMMER Web Server: 2015 Update. *Nucleic Acids Res.* 43, W30–W38. doi:10.1093/nar/gkv397

Freeling, M. (2009). Bias in Plant Gene Content Following Different Sorts of Duplication: Tandem, Whole-Genome, Segmental, or by Transposition. *Annu. Rev. Plant Biol.* 60, 433–453. doi:10.1146/annurev.arplant.043008.092122

Friedman, R., and Hughes, A. L. (2001). Pattern and Timing of Gene Duplication in Animal Genomes. *Genome Res.* 11, 1842–1847. doi:10.1101/gr.200601

Gagne, J. M., and Clark, S. E. (2010). The Arabidopsis Stem Cell Factor Poltergeist Is Membrane Localized and Phospholipid Stimulated. *Plant Cell.* 22, 729–743. doi:10.1105/tpc.109.068734

Gomez-Gomez, L., Bauer, Z., and Boller, T. (2001). Both the Extracellular Leucine-Rich Repeat Domain and the Kinase Activity of FLS2 Are Required for Flagellin Binding and Signaling in Arabidopsis. *The Plant Cell* 13, 1155–1163. doi:10.2307/3871370

González-García, M. P., Rodríguez, D., Nicolás, C., Rodríguez, P. L., Nicolás, G., and Lorenzo, O. (2003). Negative Regulation of Abscisic Acid Signaling by the Fagus Sylvatica FsPP2C1 Plays a Role in Seed Dormancy Regulation and Promotion of Seed Germination. *Plant Physiol.* 133, 135–144. doi:10.1104/pp.103.025569

Haider, M. S., Khan, N., Pervaiz, T., Zhongjie, L., Nasim, M., Jogaiah, S., et al. (2019). Genome-Wide Identification, Evolution, and Molecular

Characterization of the PP2C Gene Family in woodland Strawberry. *Gene* 702, 27–35. doi:10.1016/j.gene.2019.03.025

Hao, P.-P., Wang, G.-M., Cheng, H.-Y., Ke, Y.-Q., Qi, K.-J., Gu, C., et al. (2018). Transcriptome Analysis Unravels an Ethylene Response Factor Involved in Regulating Fruit Ripening in Pear. *Physiol. Plantarum* 163, 124–135. doi:10.1111/ppl.12671

Hirayama, T., and Umezawa, T. (2010). The PP2C-SnRK2 Complex. *Plant Signaling Behav.* 5, 160–163. doi:10.4161/psb.5.2.10460

Hu, W., Yan, Y., Shi, H., Liu, J., Miao, H., Tie, W., et al. (2017). The Core Regulatory Network of the Abscisic Acid Pathway in Banana: Genome-Wide Identification and Expression Analyses during Development, Ripening, and Abiotic Stress. *BMC Plant Biol.* 17, 145. doi:10.1186/s12870-017-1093-4

Kerk, D., Templeton, G., and Moorhead, G. B. (2008). Evolutionary Radiation Pattern of Novel Protein Phosphatases Revealed by Analysis of Protein Data from the Completely Sequenced Genomes of Humans, Green Algae, and Higher Plants. *Plant Physiol.* 146, 351–367. doi:10.1104/pp.107.111393

Kerk, D., Bulgrien, J., Smith, D. W., Barsam, B., Veretnik, S., and Gribskov, M. (2002). The Complement of Protein Phosphatase Catalytic Subunits Encoded in the Genome of Arabidopsis. *Plant Physiol.* 129, 908–925. doi:10.1104/pp.004002

Khan, N., Ke, H., Hu, C. M., Naseri, E., Haider, M. S., Ayaz, A., et al. (2019). Genome-Wide Identification, Evolution, and Transcriptional Profiling of PP2C Gene Family in Brassica Rapa. *Biomed. Res. Int.* 2019, 15. doi:10.1155/2019/2965035

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, H., Hwang, H., Hong, J.-W., Lee, Y.-N., Ahn, I. P., Yoon, I. S., et al. (2012). A rice Orthologue of the ABA Receptor, OsPYL/RCAR5, Is a Positive Regulator of the ABA Signal Transduction Pathway in Seed Germination and Early Seedling Growth. *J. Exp. Bot.* 63, 1013–1024. doi:10.1093/jxb/err338

Komatsu, K., Suzuki, N., Kuwamura, M., Nishikawa, Y., Nakatani, M., Ohtawa, H., et al. (2013). Group A PP2Cs Evolved in Land Plants as Key Regulators of Intrinsic Desiccation Tolerance. *Nat. Commun.* 4, 2219. doi:10.1038/ncomms3219

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi:10.1093/molbev/msw054

Lee, M. W., Jelenska, J., and Greenberg, J. T. (2008). Arabidopsis Proteins Important for Modulating Defense Responses to *Pseudomonas syringae* that Secrete HopW1-1. *Plant J.* 54, 452–465. doi:10.1111/j.1365-313x.2008.03439.x

Lee, T. H., Tang, H., Wang, X., and Paterson, A. H. (2013). PGDD: A Database of Gene and Genome Duplication in Plants. *Nucleic Acids Res.* 41, D1152–D1158. doi:10.1093/nar/gks1104

Li, J., Yan, X., Ahmad, M., Yu, W., Song, Z., Ni, J., et al. (2021). Alternative Splicing of the Dormancy-Associated MADS-Box Transcription Factor Gene PpDAM1 Is Associated with Flower Bud Dormancy in 'Dangshansu' Pear (Pyrus Pyrifolia white Pear Group). *Plant Physiol. Biochem.* 166, 1096–1108. doi:10.1016/j.plaphy.2021.07.017

Li, K. Q., Xu, X. Y., and Huang, X. S. (2016). Identification of Differentially Expressed Genes Related to Dehydration Resistance in a Highly Drought-Tolerant Pear, Pyrus Betulaefolia, as through RNA-Seq. *PloS one* 11, e0149352. doi:10.1371/journal.pone.0149352

Li, Q., Qiao, X., Yin, H., Zhou, Y., Dong, H., Qi, K., et al. (2019). Unbiased Subgenome Evolution Following a Recent Whole-Genome Duplication in Pear (Pyrus Bretschneideri Rehd.). *Hortic. Res.* 6, 34. doi:10.1038/s41438-018-0110-6

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, Z., Qanmber, G., Lu, L., Qin, W., Liu, J., Li, J., et al. (2018). Genome-Wide Analysis of BES1 Genes in Gossypium Revealed Their Evolutionary Conserved Roles in Brassinosteroid Signaling. *Sci. China Life Sci.* 61, 1566–1582. doi:10.1007/s11427-018-9412-x

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods* 25, 402–408. doi:10.1006/meth.2001.1262

Lu, L., Qanmber, G., Li, J., Pu, M., Chen, G., Li, S., et al. (2021). Identification and Characterization of the ERF Subfamily B3 Group Revealed GhERF13.12 Improves Salt Tolerance in Upland Cotton. *Front. Plant Sci.* 12, 705883. doi:10.3389/fpls.2021.705883

Luan, S. (1998). Protein Phosphatases and Signaling Cascades in Higher Plants. *Trends Plant Sci.* 3, 271–275. doi:10.1016/s1360-1385(98)01258-8

Merlot, S., Gosti, F., Guerrier, D., Vavasseur, A., and Giraudat, J. (2001). The ABI1 and ABI2 Protein Phosphatases 2C Act in a Negative Feedback Regulatory Loop of the Abscisic Acid Signalling Pathway. *Plant J.* 25, 295–303. doi:10.1046/j.1365-313x.2001.00965.x

Moore, R. C., and Purugganan, M. D. (2003). The Early Stages of Duplicate Gene Evolution. *Proc. Natl. Acad. Sci.* 100, 15682–15687. doi:10.1073/pnas.2535513100

Qanmber, G., Liu, J., Yu, D., Liu, Z., Lu, L., Mo, H., et al. (2019a). Genome-Wide Identification and Characterization of the PERK Gene Family in Gossypium Hirsutum Reveals Gene Duplication and Functional Divergence. *Int. J. Mol. Sci.* 20, 1750. doi:10.3390/ijms20071750

Qanmber, G., Lu, L., Liu, Z., Yu, D., Zhou, K., Huo, P., et al. (2019b). Genome-Wide Identification of GhAAI Genes Reveals that GhAAI66 Triggers a Phase Transition to Induce Early Flowering. *J. Exp. Bot.* 70, 4721–4736. doi:10.1093/jxb/erz239

Qiao, X., Li, M., Li, L., Yin, H., Wu, J., and Zhang, S. (2015). Genome-Wide Identification and Comparative Analysis of the Heat Shock Transcription Factor Family in Chinese white Pear (Pyrus Bretschneideri) and Five Other Rosaceae Species. *BMC Plant Biol.* 15, 12. doi:10.1186/s12870-014-0401-5

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., et al. (2019). Gene Duplication and Evolution in Recurring Polyploidization-Diploidization Cycles in Plants. *Genome Biol.* 20, 38–23. doi:10.1186/s13059-019-1650-2

Qiao, X., Yin, H., Li, L., Wang, R., Wu, J., Wu, J., et al. (2018). Different Modes of Gene Duplication Show Divergent Evolutionary Patterns and Contribute Differently to the Expansion of Gene Families Involved in Important Fruit Traits in Pear (Pyrus Bretschneideri). *Front. Plant Sci.* 9, 161. doi:10.3389/fpls.2018.00161

Reyes, D., Rodri´guez, D., Gonza´lez-Garci´a, M. P., Lorenzo, O., Nicola´s, G., Garci´a-Marti´nez, J. L., et al. (2006). Overexpression of a Protein Phosphatase 2C from Beech Seeds in Arabidopsis Shows Phenotypes Related to Abscisic Acid Responses and Gibberellin Biosynthesis. *Plant Physiol.* 141, 1414–1424. doi:10.1104/pp.106.084681

Saez, A., Apostolova, N., Gonzalez-Guzman, M., Gonzalez-Garcia, M. P., Nicolas, C., Lorenzo, O., et al. (2004). Gain-of-Function and Loss-of-Function Phenotypes of the Protein Phosphatase 2C HAB1 Reveal its Role as a Negative Regulator of Abscisic Acid Signalling. *Plant J.* 37, 354–369. doi:10.1046/j.1365-313x.2003.01966.x

Saez, A., Rodrigues, A., Santiago, J., Rubio, S., and Rodriguez, P. L. (2008). HAB1-SWI3B Interaction Reveals a Link between Abscisic Acid Signaling and Putative SWI/SNF Chromatin-Remodeling Complexes in Arabidopsis. *Plant Cel.* 20, 2972–2988. doi:10.1105/tpc.107.056705

Santiago, J., Rodrigues, A., Saez, A., Rubio, S., Antoni, R., Dupeux, F., et al. (2009). Modulation of Drought Resistance by the Abscisic Acid Receptor PYL5 through Inhibition of Clade A PP2Cs. *Plant J.* 60, 575–588. doi:10.1111/j.1365-313x.2009.03981.x

Schweighofer, A., Kazanaviciute, V., Scheikl, E., Teige, M., Doczi, R., Hirt, H., et al. (2007). The PP2C-Type Phosphatase AP2C1, Which Negatively Regulates MPK4 and MPK6, Modulates Innate Immunity, Jasmonic Acid, and Ethylene Levels in Arabidopsis. *Plant Cel.* 19, 2213–2224. doi:10.1105/tpc.106.049585

Servet, C., Benhamed, M., Latrasse, D., Kim, W., Delarue, M., and Zhou, D.-X. (2008). Characterization of a Phosphatase 2C Protein as an Interacting Partner of the Histone Acetyltransferase GCN5 in Arabidopsis. *Biochim. Biophys. Acta (Bba) - Gene Regul. Mech.* 1779, 376–382. doi:10.1016/j.bbagrm.2008.04.007

Shazadee, H., Khan, N., Wang, J., Wang, C., Zeng, J., Huang, Z., et al. (2019). Identification and Expression Profiling of Protein Phosphatases (PP2C) Gene Family in Gossypium Hirsutum L. *Int. J. Mol. Sci.* 20, 1395. doi:10.3390/ijms20061395

Singh, A., Giri, J., Kapoor, S., Tyagi, A. K., and Pandey, G. K. (2010). Protein Phosphatase Complement in rice: Genome-Wide Identification and

Transcriptional Analysis under Abiotic Stress Conditions and Reproductive Development. *BMC Genomics* 11, 435. doi:10.1186/1471-2164-11-435

Singh, A., Pandey, A., Srivastava, A. K., Tran, L.-S. P., and Pandey, G. K. (2016). Plant Protein Phosphatases 2C: From Genomic Diversity to Functional Multiplicity and Importance in Stress Management. *Crit. Rev. Biotechnol.* 36, 1023–1035. doi:10.3109/07388551.2015.1083941

Song, S.-K., Hofhuis, H., Lee, M. M., and Clark, S. E. (2008). Key Divisions in the Early Arabidopsis Embryo Require POL and PLL1 Phosphatases to Establish the Root Stem Cell Organizer and Vascular axis. *Develop. Cel* 15, 98–109. doi:10.1016/j.devcel.2008.05.008

Soon, F. F., Ng, L. M., Zhou, X. E., West, G. M., Kovach, A., Tan, M. H., et al. (2012). Molecular Mimicry Regulates ABA Signaling by SnRK2 Kinases and PP2C Phosphatases. *Science* 335, 85–88. doi:10.1126/science.1215106

Sparkes, I. A., Runions, J., Kearns, A., and Hawes, C. (2006). Rapid, Transient Expression of Fluorescent Fusion Proteins in Tobacco Plants and Generation of Stably Transformed Plants. *Nat. Protoc.* 1, 2019–2025. doi:10.1038/nprot.2006.286

Starr, T. K., Jameson, S. C., and Hogquist, K. A. (2003). Positive Andnegativeselection Oft Cells. *Annu. Rev. Immunol.* 21, 139–176. doi:10.1146/annurev.immunol.21.120601.141107

Strizhov, N., Ábrahám, E., Ökrész, L., Blickling, S., Zilberstein, A., Schell, J., et al. (1997). Differential Expression of Two P5CS Genes Controlling Proline Accumulation during Salt-Stress Requires ABA and Is Regulated by ABA1, ABI1 and AXR2 in Arabidopsis. *Plant J.* 12, 557–569. doi:10.1111/j.0960-7412.1997.00557.x

Sugimoto, H., Kondo, S., Tanaka, T., Imamura, C., Muramoto, N., Hattori, E., et al. (2014). Overexpression of a Novel Arabidopsis PP2C Isoform, AtPP2CF1, Enhances Plant Biomass Production by Increasing Inflorescence Stem Growth. *J. Exp. Bot.* 65, 5385–5400. doi:10.1093/jxb/eru297

Tähtiharju, S., and Palva, T. (2001). Antisense Inhibition of Protein Phosphatase 2C Accelerates Cold Acclimation in *Arabidopsis thaliana*. *Plant J.* 26, 461–470. doi:10.1046/j.1365-313x.2001.2641048.x

Umezawa, T., Sugiyama, N., Mizoguchi, M., Hayashi, S., Myouga, F., Yamaguchi-Shinozaki, K., et al. (2009). Type 2C Protein Phosphatases Directly Regulate Abscisic Acid-Activated Protein Kinases in Arabidopsis. *Proc. Natl. Acad. Sci.* 106, 17588–17593. doi:10.1073/pnas.0907095106

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies. *Genomics, Proteomics & Bioinformatics* 8, 77–80. doi:10.1016/s1672-0229(10)60008-3

Wang, G.-M., Yin, H., Qiao, X., Tan, X., Gu, C., Wang, B.-H., et al. (2016). F-box Genes: Genome-Wide Expansion, Evolution and Their Contribution to Pollen Growth in Pear (Pyrus Bretschneideri). *Plant Sci.* 253, 164–175. doi:10.1016/j.plantsci.2016.09.009

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: A Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity. *Nucleic Acids Res.* 40, e49. doi:10.1093/nar/gkr1293

Wei, K., and Pan, S. (2014). Maize Protein Phosphatase Gene Family: Identification and Molecular Characterization. *BMC Genomics* 15, 773. doi:10.1186/1471-2164-15-773

Wei, K., Wang, Y., and Xie, D. (2014). Identification and Expression Profile Analysis of the Protein Kinase Gene Superfamily in maize Development. *Mol. Breed.* 33, 155–172. doi:10.1007/s11032-013-9941-x

Xiang, Y., Sun, X., Gao, S., Qin, F., and Dai, M. (2017). Deletion of an Endoplasmic Reticulum Stress Response Element in a ZmPP2C-A Gene Facilitates Drought Tolerance of Maize Seedlings. *Mol. Plant* 10, 456–469. doi:10.1016/j.molp.2016.10.003

Xue, T., Wang, D., Zhang, S., Ehlting, J., Ni, F., Jakab, S., et al. (2008). Genome-Wide and Expression Analysis of Protein Phosphatase 2C in Rice and Arabidopsis. *BMC Genomics* 9, 550. doi:10.1186/1471-2164-9-550

Yang, T., and Huang, X.-S. (2018). Deep Sequencing-Based Characterization of Transcriptome of Pyrus Ussuriensis in Response to Cold Stress. *Gene* 661, 109–118. doi:10.1016/j.gene.2018.03.067

Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi:10.1093/molbev/msm088

Yoshida, T., Nishimura, N., Kitahata, N., Kuromori, T., Ito, T., Asami, T., et al. (2006). ABA-Hypersensitive Germination3 Encodes a Protein Phosphatase 2C (AtPP2CA) that Strongly Regulates Abscisic Acid Signaling during Germination Among Arabidopsis Protein Phosphatase 2Cs. *Plant Physiol.* 140, 115–126. doi:10.1104/pp.105.070128

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 11 January 2022
doi: 10.3389/fgene.2021.778788

Comparative Transcriptome Analyses of Gayal *(Bos frontalis)*, Yak *(Bos grunniens)*, and Cattle *(Bos taurus)* Reveal the High-Altitude Adaptation

Jun Ma, Tianliu Zhang, Wenxiang Wang, Yan Chen, Wentao Cai, Bo Zhu, Lingyang Xu, Huijiang Gao, Lupei Zhang, Junya Li* and Xue Gao*

Laboratory of Molecular Biology and Bovine Breeding, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China

OPEN ACCESS

Edited by:
Suxu Tan,
Michigan State University,
United States

Reviewed by:
Ikhide G. Imumorin,
Georgia Institute of Technology,
United States
Loan To Nguyen,
University of Queensland, Australia

****Correspondence:***
Junya Li
lijunya@caas.cn
Xue Gao
gaoxue76@126.com

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *17 September 2021*
Accepted: *06 December 2021*
Published: *11 January 2022*

Citation:
Ma J, Zhang T, Wang W, Chen Y, Cai W, Zhu B, Xu L, Gao H, Zhang L, Li J and Gao X (2022) Comparative Transcriptome Analyses of Gayal (Bos frontalis)*, Yak* (Bos grunniens)*, and Cattle* (Bos taurus) *Reveal the High-Altitude Adaptation. Front. Genet. 12:778788. doi: 10.3389/fgene.2021.778788*

Gayal and yak are well adapted to their local high-altitude environments, yet the transcriptional regulation difference of the plateau environment among them remains obscure. Herein, cross-tissue and cross-species comparative transcriptome analyses were performed for the six hypoxia-sensitive tissues from gayal, yak, and cattle. Gene expression profiles for all single-copy orthologous genes showed tissue-specific expression patterns. By differential expression analysis, we identified 3,020 and 1,995 differentially expressed genes (DEGs) in at least one tissue of gayal *vs.* cattle and yak *vs.* cattle, respectively. Notably, we found that the adaptability of the gayal to the alpine canyon environment is highly similar to the yak living in the Qinghai-Tibet Plateau, such as promoting red blood cell development, angiogenesis, reducing blood coagulation, immune system activation, and energy metabolism shifts from fatty acid β-oxidation to glycolysis. By further analyzing the common and unique DEGs in the six tissues, we also found that numerous expressed regulatory genes related to these functions are unique in the gayal and yak, which may play important roles in adapting to the corresponding high-altitude environment. Combined with WGCNA analysis, we found that *UQCRC1* and *COX5A* are the shared differentially expressed hub genes related to the energy supply of myocardial contraction in the heart-related modules of gayal and yak, and *CAPS* is a shared differential hub gene among the hub genes of the lung-related module, which is related to pulmonary artery smooth muscle contraction. Additionally, *EDN3* is the unique differentially expressed hub gene related to the tracheal epithelium and pulmonary vasoconstriction in the lung of gayal. *CHRM2* is a unique differentially expressed hub gene that was identified in the heart of yak, which has an important role in the autonomous regulation of the heart. These results provide a basis for further understanding the complex transcriptome expression pattern and the regulatory mechanism of high-altitude domestication of gayal and yak.

Keywords: gayal, yak, differentially expressed genes, co-expression, high-altitude adaptation, hypoxia

INTRODUCTION

Species living at high altitudes are exposed to strict selection pressures and physiological challenges owing to harsh environmental conditions, such as thin air, cold temperatures, ultraviolet exposure, and low pressure (Miao et al., 2015). Despite the harsh conditions, numerous species including Tibetan pigs (Jia et al., 2016; Zhang B. et al., 2017), Tibetan sheep (Zhang et al., 2013), Tibetan chickens (Gou et al., 2014), and yak (Qiu et al., 2012; Qi et al., 2019) have evolved unique physiological characteristics, such as superior blood oxygen transport system and high metabolic efficiency, to adapt to the harsh living pressure (Lan et al., 2018), as have the native Tibetan people (Ge et al., 2012; Tashi et al., 2014). At present, published studies have identified *EPAS1*, *EGLN1,* and *PPARA* genes that play important roles in high-altitude adaptation (Haasl and Payseur, 2016; Heinrich et al., 2019). Exploring the molecular mechanisms underlying hypoxia adaptation has long garnered attention.

Gayal, also known as Drung cattle (*Bos frontalis*) (Winter et al., 1986; Uzzaman et al., 2014), is a unique semi-wild cattle breed, mainly found in the typical alpine valleys and subtropical rain forests in the Drung and Nujiang river basins of Yunnan Province, southwestern China. Studies have shown that the level of hemoglobin in gayal, as well as the number of red blood cells and white blood cells, are equivalent to those of yak, which is the physiological basis for the ability of gayal to resist invasion by bacteria, viruses, and parasites in the alpine environment (Tian et al., 1998). Yak (*Bos grunniens*), a native mammal on the Qinghai-Tibet Plateau and its adjacent regions, provides meat and other necessities for Tibetans. Compared with lowland cattle, yaks have a larger alveolar area per unit area, thinner alveolar spacing, thinner gas-blood barrier (Wei and Yu, 2008), and stronger expression of some genes related to oxygen supply as well as defense system under hypoxia pressure (Wang et al., 2016). Therefore, a thorough understanding of the genetic basis of the physiological characteristics of gayal and yak will provide insight into their adaptation to the high-altitude environment.

With recent rapid progress in next-generation genome sequencing (NGS) technologies, high-throughput RNA-Sequencing (RNA-seq) technology represents a powerful and cost-efficient approach to explore species domestication, breeding, and genetic molecular mechanisms in organisms (Salleh et al., 2017; Keren et al., 2018). RNA-seq can detect and quantify gene expression with digital measurements, which are especially sensitive for low-expressed genes (Mortazavi et al., 2008). In addition, it is also used to improve gene annotation, discover novel genes or transcripts, and survey alternative splicing (AS) events at single-nucleotide resolution (Sultan et al., 2008; Ozsolak and Milos, 2011). Thus far, transcriptome studies on gayal and yak have been widely conducted. The muscle transcriptome analysis of Indian Mithun showed that hub genes including *MTMR3, CUX1, LONRF3, PLXNB2, KMT2C, ZRSR2Y, PRR14, USP9Y, SLMAP,* and *KANSL2* might contribute to muscle growth (Mukherjee et al., 2020). Transcriptomic analysis of yaks living at different altitudes indicates PI3K-Akt, HIF-1, focal adhesion, and ECM–receptor interaction pathway were significantly enriched, and the *EPAS1* expression increases with altitude (Qi et al., 2019). Comparative transcriptome analysis of four organs (heart, kidney, liver, and lung) between yak and cattle revealed that DEGs associated with the blood supply system, modulation of cardiac contractility, vascular smooth muscle proliferation, and the glutamate receptor system probably play an important role in yak adaptation to hypoxia environments (Wang et al., 2016). However, gayal is a species that is well adapted to their local alpine valley environment, and few transcriptomic studies focusing on regulation analysis of gayal in adaptation to the local high-altitude environment have so far not been reported. In addition, studies have found specific adaptive genes and mechanisms are distinct between species and populations (Heinrich et al., 2019; Pamenter et al., 2020). Therefore, we seek to understand how changes in altitude affect the transcriptional regulation of gayal, and whether the relevant transcriptional regulation mechanism is the same as that of yak, which shares a bovine subfamily.

In this study, we characterized gene expression profiles using RNA-seq data from six major tissues including heart, lung, liver, kidney, spleen, and muscle among gayal, yak, and cattle. We performed a comparative transcriptome analysis of six tissues in gayal *vs.* cattle and yak *vs.* cattle, respectively, and identified candidate DEGs related to adaptability to the plateau environment, as well as analyzed their functions and expression patterns. In addition, weighted gene co-expression network analysis (WGCNA) was further used to uncover the hub genes that regulate tissue function in high-altitude adaptation. The results could lay a foundation for further explorations of the genetic changes underlying hilly adaptations in subtropical mammals.

MATERIALS AND METHODS

Sample Collection

Six tissues (heart, kidney, liver, lung, skeletal muscle, and spleen) were collected from three adult female gayal living in the semi-wild preserved field at an altitude of 2,300–3,500 m in the Drung and Nujiang river basins. All tissue samples were taken and immediately snap-frozen in liquid nitrogen. In addition, the corresponding tissue transcriptome data of yak and cattle were retrieved from a previous publication (Tang et al., 2017) from the GEO databases under the accession numbers GSE93855 and GSE77020 (**Supplementary Table S1**). In addition, one gayal multi-tissue transcriptome data that we previously published was also included in this study to expand the sample size, which was stored in National Genomics Data Center under the accession code PRJCA002143. An overview of the samples used and the associated statistics are provided in **Supplementary Table S1**.

RNA Extraction, Library Construction, and Transcriptome Sequencing

Total RNA was isolated and purified with the TRIzol reagent (Life Technologies, Carlsbad, United States) according to the

manufacturer's protocols. RNA purity was qualified using NanoDrop ND 2000 spectrophotometer at 260 and 280 nm (Thermo Fisher Scientific, Wilmington, United States), and RNA integrity was assessed using Agilent 2,100 Bioanalyzer (Agilent Technologies, Palo Alto, United States). The OD260/280 ratios of all samples were greater than 1.8, and the RNA integrity number (RIN) value of >7.0. Then, mRNA was broken down into small fragments using a magnesium RNA cleavage module (NEB, Ipswich, United States). Random primers and related reverse transcriptase were used to synthesize the first-strand complementary DNA (cDNA). Subsequently, the second-strand cDNA was synthesized. The average insert size for the final cDNA library was 300 ± 50 bp. Finally, we conducted the 2 × 150 bp paired-end sequencing (PE150) on an Illumina Novaseq™ 6,000 (LC-Bio Technology CO., Ltd., Hangzhou, China) following the supplier's recommended protocol.

Identification of Orthologous Genes

To identify orthologous genes among gayal, yak and cattle, we first downloaded protein sequence data of yak and cattle from the Ensemble genome browser 104 (http://asia.ensembl.org/index.html) and combined with the protein sequence data of the gayal reference genome assembled by ourselves (the accession code of National Genomics Data Center is PRJCA004132) for downstream analysis. We used ORTHOFINDER v2.2.6 (Emms and Kelly, 2015) software to delimit orthologous groups, in which all predicted protein sequences were compared using a BLAST all-against-all search (Camacho et al., 2009). The single-copy genes, duplicated genes, and species-specific genes were extracted from the ORTHOFINDER output, respectively. Among them, the 1:1:1 homologous genes of the three species were defined as the conserved single-copy homologous genes of each species, which were retained for the following analysis.

Data Quality Control, Processing, and Normalization of Gene Expression Levels

To ensure the accuracy of subsequent biological information analysis, the quality of the raw data was first checked using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). According to the evaluation results of raw reads, the software FASTP (https://github.com/OpenGene/fastp) with default parameters was used to remove adapters, reads with a high proportion of unknown nucleotides, and low-quality sequences (≤Q20) to obtain high-quality clean reads. To avoid biases due to sequence divergence across bovine subfamily genomes, we downloaded yak (BosGru_v2.0, http://ftp.ensembl.org/pub/release-104/fasta/bos_mutus/) and cattle (ARS_UCD1_2, http://ftp.ensem -bl.org/pub/release-104/fasta/bos_taurus/) reference genomes from Ensemble database, and retrieved gayal (Drung_v1.2) reference genomes from National Genomics Data Center (NGDC) (https://bigd.big.ac.cn/). Then, the clean reads for each sample were mapped to the corresponding reference genome using HISAT2 software (Pertea et al., 2016). FeatureCount from the Subread package (Version 2.0.0) (Liao et al., 2014) was applied to calculate the numbers of reads mapped to each gene. Read pairs were counted only if they were uniquely mapped and properly paired. The gene expression level estimates may be biased among species due to factors such as the size of mRNA transcript and the qualities of different genome annotations (Blake et al., 2020). To avoid these issues, we used a custom script to only retain reads that mapped to the 1:1 orthologous which can be used for each of the three genomes (Blake et al., 2020). Taking into account the difference in sequencing depth of samples from different sequencing platforms, we first filtered out genes with low expression, and only keep genes with expression level CPM> 1 in at least half of the samples (Blake et al., 2020). We normalized the raw read counts using the weighted trimmed mean of M-values algorithm (TMM) in the edgeR package (Robinson et al., 2010). Then scaling factors were used to scale the expression levels of all orthologous. The normalized expression data for the trimmed orthologous were used for subsequent gene expression analysis.

Gene Expression Profile Analysis

After performing normalization, we obtained the TMM-normalized $\log_2$-transformed CPM values for each gene. The principal component analysis (PCA) was performed with the prcomp function in R using normalized $\log_2$-transformed CPM values. The hierarchical clustering of Pearson's correlation coefficients for each pair of samples was performed using the complete-linkage agglomerative method on the correlation distance matrix and generated symmetrical heatmap plot using the R gplots package.

Differential Expression Pattern Analysis

To characterize the high-altitude adaptive gene expression pattern of gayal and yak, we performed differential expressed analyses among heart, kidney, liver, lung, skeletal muscle, and spleen tissues in gayal *vs.* cattle group and yak *vs.* cattle group by using R package edgeR (Robinson et al., 2010). Genes with Benjamini Hochberg false discovery rate (FDR) values less than 0.05 and fold change more than 2 were identified as DEGs. If the $\log_2$ fold changes more than 1 were defined as up-regulated DEGs, or less than -1 were defined as down-regulated. We performed Gene Ontology (GO) function annotation and KEGG pathway enrichment analysis for DEGs by the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 using the *Bos taurus* genome as the background set (Huang Da et al., 2009). The significantly enriched GO and KEGG terms using a 0.05 cutoff for the *p*-value with Benjamini-Hochberg test adjustment. The jvenn and UpSetR R package (version 1.4.0) were used to identify and visualize the DEGs shared by two comparison groups across the multiple tissues and between two comparison groups in the same tissue, respectively (Conway et al., 2017). In addition, to better understand the hypoxia adaptation of species, we also collected a set of known or potential high-altitude adaptation candidate genes from previously published literature (Zhang et al., 2014; Wang et al., 2016; Peng et al., 2017).

Construction of Gene Co-Expression Network

The adaptation of species to the hypoxia environment is a complex biological process, which involves the coordinated regulation of

multiple genes. WGCNA is a feasible method to clarify the regulatory relationship between genes and identify important hub genes based on gene expression profiles to explore the mechanism of hypoxia adaptation (Wang et al., 2021). Therefore, to further identify the relevant regulatory gene modules related to the functions of each tissue and explore the core driver genes, we used the CPM values of filtered genes to perform WGCNA analysis on three species, respectively (Langfelder and Horvath, 2008). The hierarchical cluster analysis on all tissues was carried out using the hclust function. The soft-power threshold beta was determined by the function "sft$powerEstimate" (Chen et al., 2019). The minimal gene module size was set to 30 to obtain appropriate modules, and the threshold to merge similar modules was set to 0.25. Gene modules were detected based on the TOM matrix. The correlations between modules eigengene and tissues were investigated using the Pearson correlation, and only modules with correlation coefficients >0.7 and $p < 0.001$ were considered tissue-relevant modules (Touma et al., 2016). Here, we designate the highest top 5% of genes as the hub genes based on the intramodular connectivity in each functional module, which is calculated by summing the connection strengths with other intramodular genes (Yang et al., 2014). Functional enrichment analyses for tissue-related module genes were performed using DAVID v6.8 using the *Bos taurus* genome as the background set, and the *p*-values were corrected for multiple testing using Benjamini-Hochberg correction (Huang Da et al., 2009).

RESULTS

Summary Statistics of RNA-Seq Data

The transcriptome data of six tissues (heart, kidney, liver, lung, skeletal muscle, and spleen) from four gayals, three yaks, and three cattle were used in this study (**Supplementary Table S1**). One gayal heart sample (gayal3_heart) was excluded due to low sequencing quality. In total, ~327.55 G clean data has been generated after removing adapter-embedded or low-quality reads. We calculated Q20 (99% base call accuracy), Q30 (99.9% base call accuracy), and GC-content of the clean data. The average values of Q20 and Q30 in the samples were 97.04% (ranging from 94.60 to 98.47%) and 92.38% (ranging from 89.39 to 95.43%), respectively (**Supplementary Table S2**). The average base GC content was 50.80%, ranging from 47.69 to 55.03%, indicating that the base composition and quality of the sequencing data were qualified. Then the high-quality reads of each sample were mapped to the corresponding reference genome (i.e., Drung_v1.2, BosGru_v2.0, ARS_UCD1_2) using the HISAT2 software (Pertea et al., 2016). The average reads mapping rates in gayal, yak, and cattle were 86.35, 93.60, and 95.58%, respectively. Detailed statistics and mapping information for each sample were summarized in **Supplementary Table S2**.

Ortholog Identification and Gene Expression Levels Normalization

To reduce the effect of chromosome number difference (*Bos frontalis*, 2n = 58; and *Bos grunniens*, 2n = 60; *Bos taurus*, 2n = 60) and the qualities of different genome annotations among three bovine species, we first used the blast function of ORTHOFINDER2 (Emms and Kelly, 2015) to identify orthologous genes among the gayal, yak and cattle. A total of 19,174 protein orthogroups were identified. Among them, only 6,371 orthogroups were high-confidence single-copy orthogroups of three species (**Supplementary Table S3**). Then the corresponding gene names were extracted by a custom script and retained reads that mapped to the single-copy homologous genes of three species genome, and the ensemble gene names of cattle were uniformly used for subsequent analysis. In the 6,371 single-copy orthologous gene expression datasets, genes with low expression (CPM <1) in more than half of the samples were further excluded, and 4,712 genes were obtained (Blake et al., 2020). Then we normalized 4,712 single-copy orthologous genes using the weighted TMM of edgeR (Hao et al., 2019).

Gene Expression Profiles Across Tissues and Bovine Species

To investigate broad gene expression patterns in multiple tissues across different bovine species, we performed hierarchical cluster analysis based on the gene expression of six tissues among gayal, yak, and cattle. As shown in **Figure 1A**, the tissue-specific expression pattern of lung, liver, spleen, and kidney shows that the same tissues of different bovine species gather earlier than different tissues of the same bovine species. The result indicated that tissue differentiation may precede species differentiation, which was consistent with previous studies (Necsulea and Kaessmann, 2014; Hao et al., 2019). However, the samples of heart and muscle from each species were more similar than the same tissue between different species (species-specific expression pattern), which reflects striated muscular tissues may have similar expression profiles (Merkin et al., 2012; Tang et al., 2017; Feng et al., 2020) (**Figure 1A**). In addition, our expression profiles suggested that the differences in global gene expression among tissues were more significant than that among altitudes or the same tissue of different species, which was consistent with the results of previous studies (Necsulea and Kaessmann, 2014; Tang et al., 2017; Hao et al., 2019). Furthermore, principal component analysis (PCA) was performed on the gene expression of 59 samples. The score plot of PCA analysis for six main organs shows that 45.4% of the variance can be explained by the first two principal components, accounting for 26.4 and 19% of the variance, respectively (**Figure 1B**). In essence, 59 samples were clustered based on similar tissue types, which confirmed that tissue differentiation may precede species differentiation.

Gene Expression Changes Accompanying High-Altitude Environment

To detect expression changes of gayal and yak in response to the high-altitude stress, we used edgeR (Robinson et al., 2010) to identify DEGs within each tissue. According to the tissue and species we considered, 1,218 to 1,585 interspecies DEGs were identified in the comparison of six type tissues in gayal and cattle

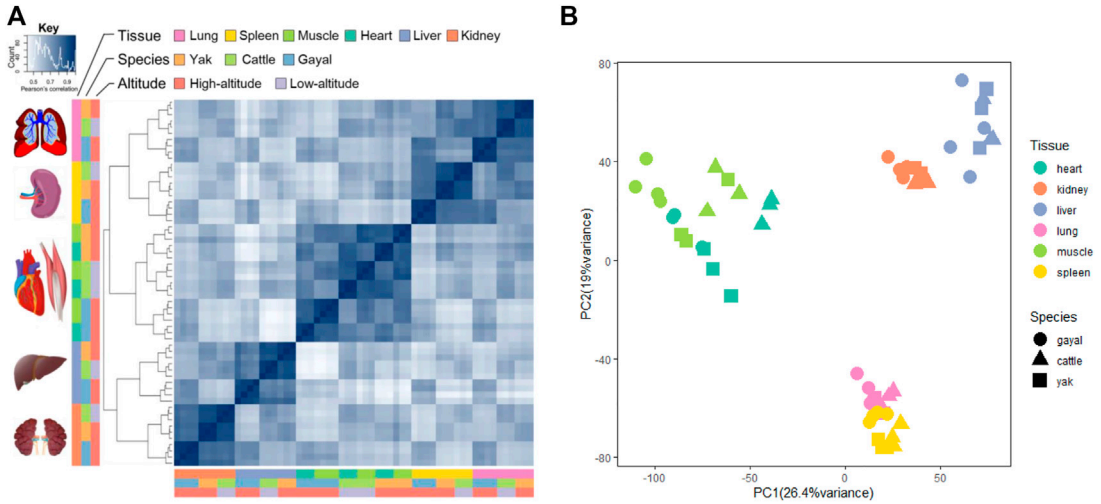


FIGURE 1 | Gene expression patterns across the three bovine species: yak, gayal, and cattle. **(A)** symmetrical heatmap of Pearson's correlation coefficients between all pairs of samples across all genes. The key is hierarchically clustered (complete method, correlation distance). **(B)** PCA of the log-transformed normalized expression levels of all orthologs across all species and tissues. Species are represented by point shape; tissues are represented by point color.

TABLE 1 | The number of DEGs identified in six tissues for gayal and yak compared with low-altitude cattle.

Group	Up- or down-regulated	Heart	Liver	Spleen	Lung	Kidney	Muscle
Gayal VS cattle	Up	618	764	599	512	586	560
	Down	967	812	839	706	766	929
	Total	1,585	1,576	1,438	1,218	1,352	1,489
Yak VS cattle	Up	278	202	313	387	279	367
	Down	537	341	508	532	443	446
	Total	815	543	821	919	722	813

(**Table 1**). Among the six tissues, the heart has the largest number of DEGs (618 up-regulated and 967 down-regulated genes). By upset analysis of all DEGs in each tissue, 446 DEGs shared in six tissues were identified, and each gene showed consistent up-regulation or down-regulation in all tissues (**Figure 2A**). In the yak *vs.* cattle group, we identified 1,995 DEGs among six types of tissues. The number of DEGs among tissues ranges from 543 in liver tissue (202 up-regulated and 341 down-regulated genes) to 919 in lung tissue (387 up-regulated and 532 down-regulated genes) (**Table 1**). There were 154 DEGs with consistent regulation direction shared among the six tissues, of which 64 genes were also present in all tissues of the gayal *vs.* cattle group (**Figure 2B**).

To further investigate the potential biological function of DEGs for environmental adaptation in gayal and yak, we performed functional annotation analysis based on the DAVID database. In the gayal *vs.* cattle group, DEGs were annotated in energy metabolism, cardiovascular system, blood coagulation system, and immune system among tissues (**Supplementary Table S4**). As such, the DEGs of the heart were enriched in the oxidation-reduction process (GO: 0055114), positive regulation of angiogenesis (GO:0045766), positive regulation of blood coagulation (GO:0030194). Muscle DEGs were enriched in blood coagulation (GO:0007596), patterning of blood vessels (GO:0001569), skeletal muscle cell differentiation (GO:0035914) (**Figure 3**). We also observed DEGs in the gayal *vs.* cattle group were significantly enriched in hypoxia response, metabolism, and immune system pathway. For instance, hematopoietic cell lineage (bta04640) was enriched in four tissues (heart, spleen, lung, and kidney) and galactose metabolism (bta00052) was enriched in four tissues (heart, liver, spleen, and lung) (**Supplementary Table S4**). In the yak *vs.* cattle group, DEGs identified across six tissues pairwise comparisons were also enriched in the cardiovascular system, energy metabolism, and blood coagulation system, which included response to oxidative stress (GO:0006979) and oxidation-reduction process (GO:0055114) in the heart, and patterning of blood vessels (GO:0001569) and positive regulation of canonical wnt signaling pathway (GO:0090263) in the muscle (**Figure 4**). Similar results were found in the KEGG analysis, pathways related to energy metabolism and the immune system were significantly enriched (**Supplementary Table S4**).

To explore the similarities and differences of expression regulation genes related to environmental adaptability among gayal and yak, we further analyzed the shared and unique DEGs

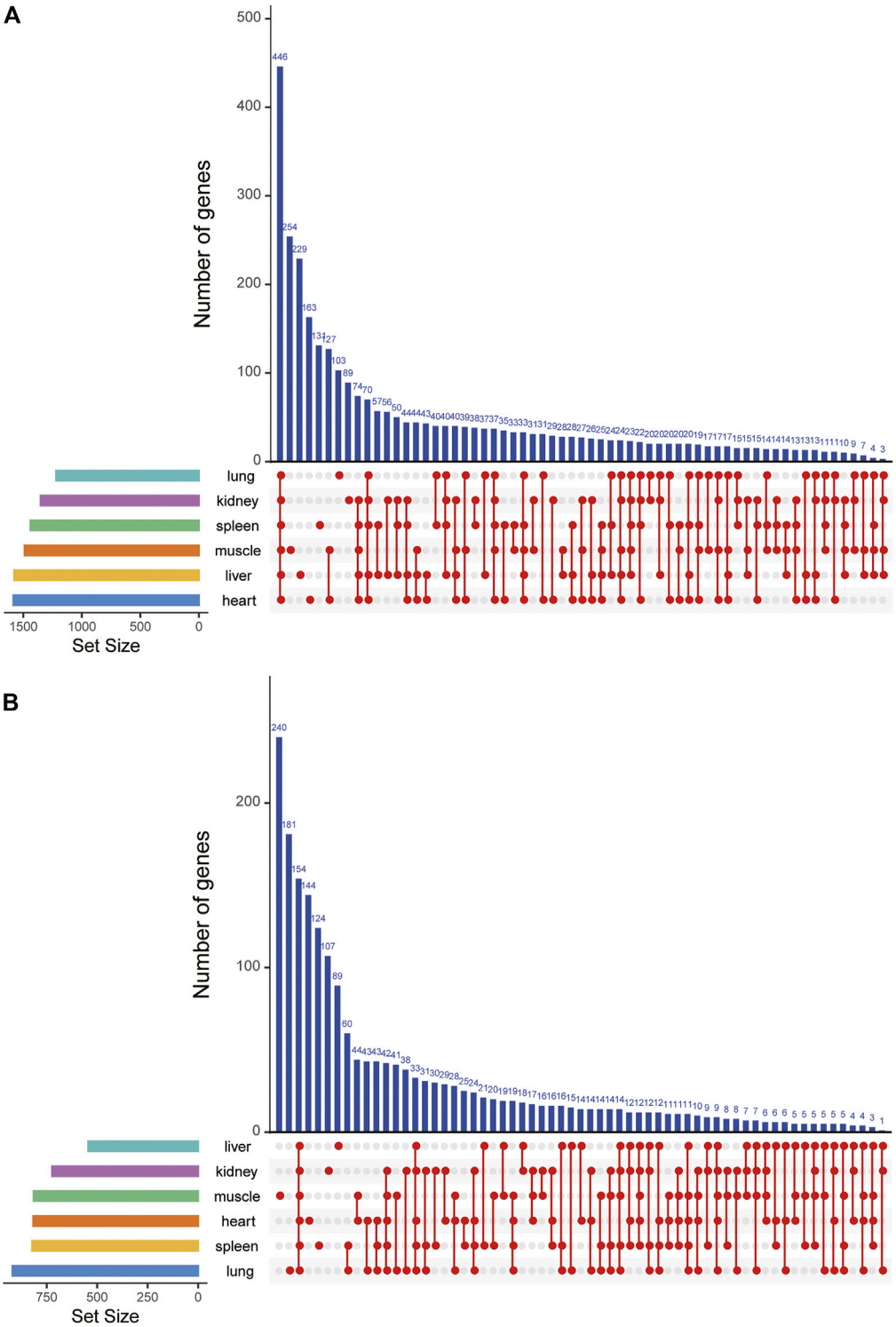


FIGURE 2 | DEGs upset map of all DEGs in six tissues of **(A)** gayal *vs.* cattle group and **(B)** yak *vs.* cattle group.

between the gayal *vs.* cattle group and the yak *vs.* cattle group. Among the six tissues of gayal and yak, the heart and lung shared the largest number of DEGs, with 397 (101 up-regulated and 296 down-regulated genes) and 396 genes (146 up-regulated and 250 down-regulated genes), respectively (**Figure 5**). In addition, we have identified 822–1,331 and 298–500 unique DEGs in the

FIGURE 3 | GO functional and KEGG pathway enrichment analysis of DEGs in the **(A)** heart, **(B)** liver, **(C)** spleen, **(D)** lung, **(E)** kidney, and **(F)** muscle for the gayal *vs.* cattle group.

six tissues of gayal and yak compared with lowland cattle, respectively (**Figure 5**). By functional annotation analysis, we observed the shared DEGs of the heart were enriched in the oxidation-reduction process (GO:0055114), and positive regulation of blood coagulation (GO:0030194). Lung-shared DEGs were involved in platelet activation (GO:0030168), and

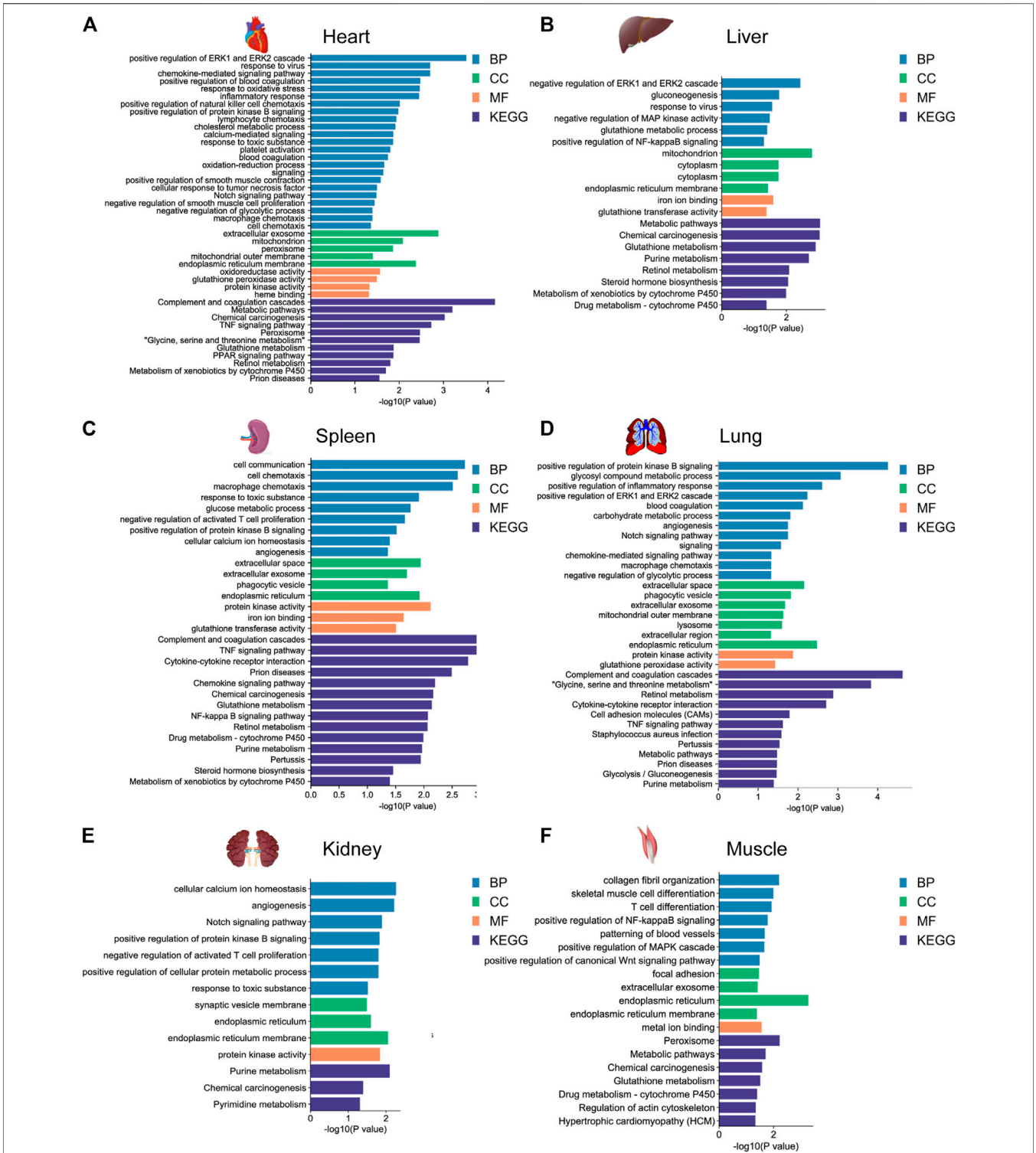


FIGURE 4 | GO functional and KEGG pathway enrichment analysis of DEGs in the **(A)** heart, **(B)** liver, **(C)** spleen, **(D)** lung, **(E)** kidney, and **(F)** muscle for the yak *vs.* cattle group.

the intrinsic apoptotic signaling pathway in response to DNA damage (GO:0008630). Liver-shared DEGs were enriched in innate immune response (GO:0045087). Spleen-shared DEGs were enriched in chemotaxis (GO:0006935). And the shared DEGs in the muscle were enriched in skeletal muscle cell differentiation (GO:0035914) (**Supplementary Table S4**).

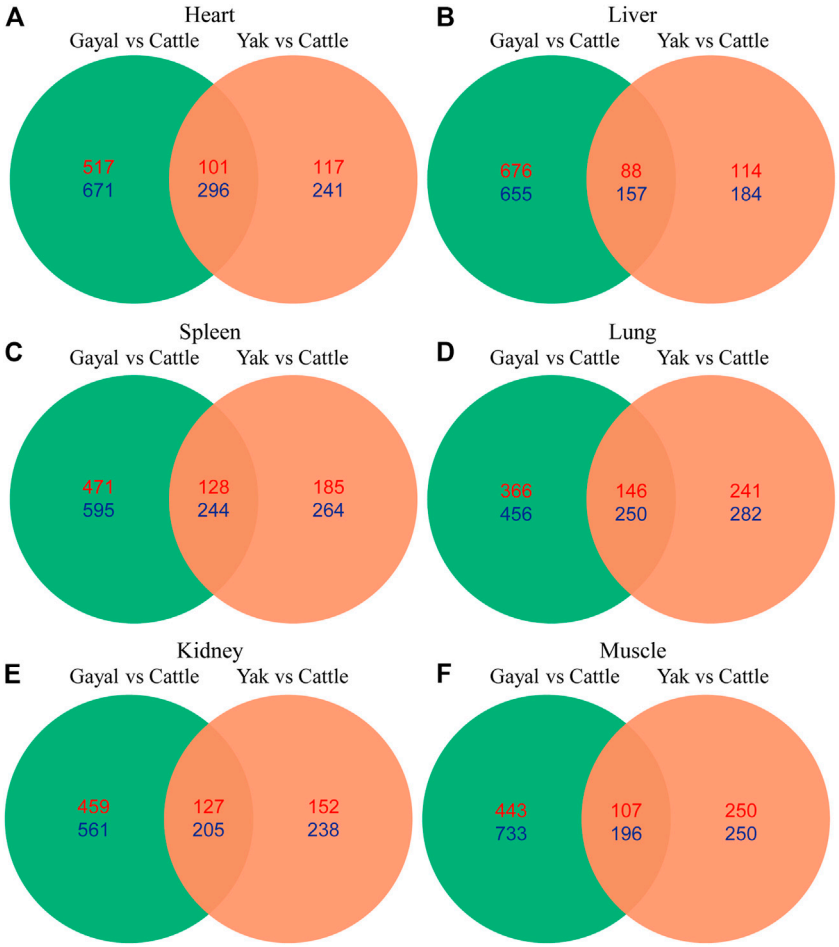


FIGURE 5 | Venn diagram indicating differentially expressed genes that were shared among the two high- and low-altitude pairs in **(A)** heart, **(B)** liver, **(C)** spleen, **(D)** lung, **(E)** kidney, and **(F)** muscle. Numbers in red and blue indicate genes up- and down-regulated in the high-altitude Bovina species relative to low-altitude cattle.

As for the unique DEGs of the gayal *vs.* cattle group, we found that five unique DEGs in the heart related to apoptosis (bta04210), include *CASP9, ENDOG, AIFM1, CASP6, TNFSF10, RIPK1, FASLG, and TNF*. Seven unique genes (*BLOC1S6, SERPINE2, TFPI2, HPS4, SERPING1, GP1BA, and TFPI*) are associated with blood coagulation (GO:0007596) in the lung. Spleen-unique DEGs were enriched in lymphocyte chemotaxis (GO:0048247) and inflammatory response (GO: 0006954). Furthermore, we found that positive regulation of angiogenesis (GO:0045766) was significantly enriched in specific genes in the lung, kidney, and spleen; hematopoietic cell lineage (bta04640) was also significantly enriched in the unique genes of the heart, spleen, lung, and kidney. In addition, we found that the unique DEGs in the heart, liver, and spleen were also related to galactose metabolism (bta00052). These unique DEGs may play an important role in the adaptation of gayal to the alpine and valley environment (**Supplementary Table S4**). Among the unique DEGs of the yak *vs.* cattle group, we found that four genes (*CD40LG, SRF, F2R, and F5*) related to platelet activation (GO:0030168) in the heart. Blood coagulation (GO:0007596) related genes (*PROCR, THBD, F10, PROS1, F3, and KNG1*) were also significantly enriched in the lung. The unique genes in the spleen were enriched in immune-related items, such as the NOD-like receptor signaling pathway (bta04621), and TNF signaling pathway (bta04668). In addition, cell adhesion molecules (CAMs) (bta04514) were uniquely enriched in the heart and lung of yak, which was reported as the intensity of hypoxia increases lead to marked reduced cell adhesion (Kaiser et al., 2012) (**Supplementary Table S4**).

Expression Regulation of High-Altitude Related Genes in the DEGs

To further establish the hypoxia adaptation of gayal and yak, we collected a candidate gene set with known or potential functions in hypoxia responses from previously published literature (Zhang et al., 2014; Peng et al., 2017). Of those, 295 and 187 genes were DEGs in at least one tissue of gayal and yak, respectively (**Supplementary Table S5**). Among which the heart has the largest number of high-altitude related genes in both comparison groups DEGs (149 in gayal *vs.* cattle group and 85

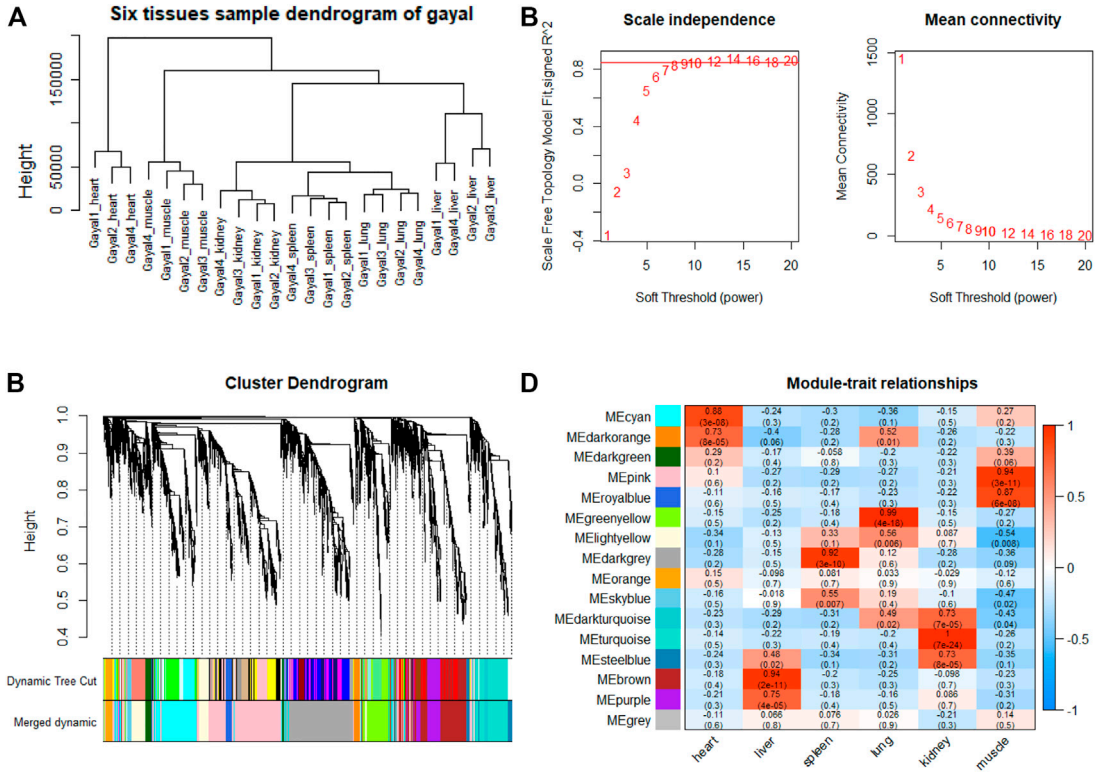


FIGURE 6 | WGCNA analysis of gayal tissue samples. **(A)** Hcluster diagram of gayal tissue samples. **(B)** Plot of scale-free topology and mean connectivity in regard to soft-thresholding power for samples. **(C)** Hierarchical clustering tree (dendrogram) of genes based on co-expression network analysis. **(D)** Heatmap of the correlation between the module eigengenes and tissues of gayal.

genes in yak *vs.* cattle group). Notably, in the altitude-related gene set, we found that several positively selected genes related to hypoxia response in Tibetans were differentially expressed in two comparison groups. For example, *EGLN1*, also known as *PHD2*, as a known target gene of the hypoxia-sensing pathway (Haasl and Payseur, 2016; Heinrich et al., 2019; Storz, 2021), showed a consistent down-regulation across all tissues of yak and gayal compared with lowland cattle, but the expression level and pattern vary in different tissues. *HFE* (Homeostatic Iron Regulator), a gene involved in indirectly regulating iron balance in animals via reducing transferrin-mediated iron uptake (Salter-Cid et al., 1999), was uniquely down-regulated in all tissues of gayal except for kidney compared with the lowland cattle. Additionally, *HOXB6*, a gene associated with the regulation of the erythropoietic system, was particularly up-regulated in the spleen and kidney of the yak (Zimmermann and Rich, 1997).

Tissue-specific Module Detection and Hub Genes Identification

To explore genes and co-expression modules with similar expression profiles related to high-altitude adaptation in various tissues, WGCNA analysis was independently performed on the three species based on the CPM values of 4,712 filtered genes. The sample clustering dendrogram of the gayal, yak, and cattle are shown in **Figures 6A**, **7A**, **8A**, respectively. In both species, all samples were included in the clusters. The soft-power threshold of 10, 18, and 16 (scale-free R^2 of 0.85) were selected for further analysis of the gayal, yak, and cattle, respectively (**Figures 6B**, **7B**, **8B**). Next, the gene modules were detected based on the topological overlap matrix (TOM), and sixteen modules were detected under the gayal, while ten modules were detected under the yak (**Figures 6C**, **7C**, **8C**). To identify tissue-specific modules, we calculated the correlation coefficients between the module and each tissue (**Figures 6D**, **7D**, **8D**). Further, GO and KEGG analyses were conducted to investigate the biological function of the genes in each tissue-related module.

Indeed, the relevant modules enriched in six tissues are highly related to the tissue-specific functions. For example, in the tissue-related modules of gayal, the MEcyan module was significantly correlated with heart (r = 0.88, $p = 3 \times 10^{-8}$), which was enriched in functional annotations such as negative regulation of cardiac muscle cell apoptotic process (GO:0010667), and pathway in MEcyan module were involved in cardiac muscle contraction (bta04260). Lung-related module (MEgreenyellow, r = 0.99, $p = 4 \times 10^{-18}$) was enriched in HIF-1 signaling pathway (bta04066) (**Supplementary Table S6**). Similar results were found in the yak, MEgreenyellow was the most significantly correlated with the heart of yak (r = 0.98, $p = 3 \times 10^{-12}$), which was enriched in functional annotations such as positive regulation of cardiac muscle cell proliferation (GO:0060045), positive regulation of

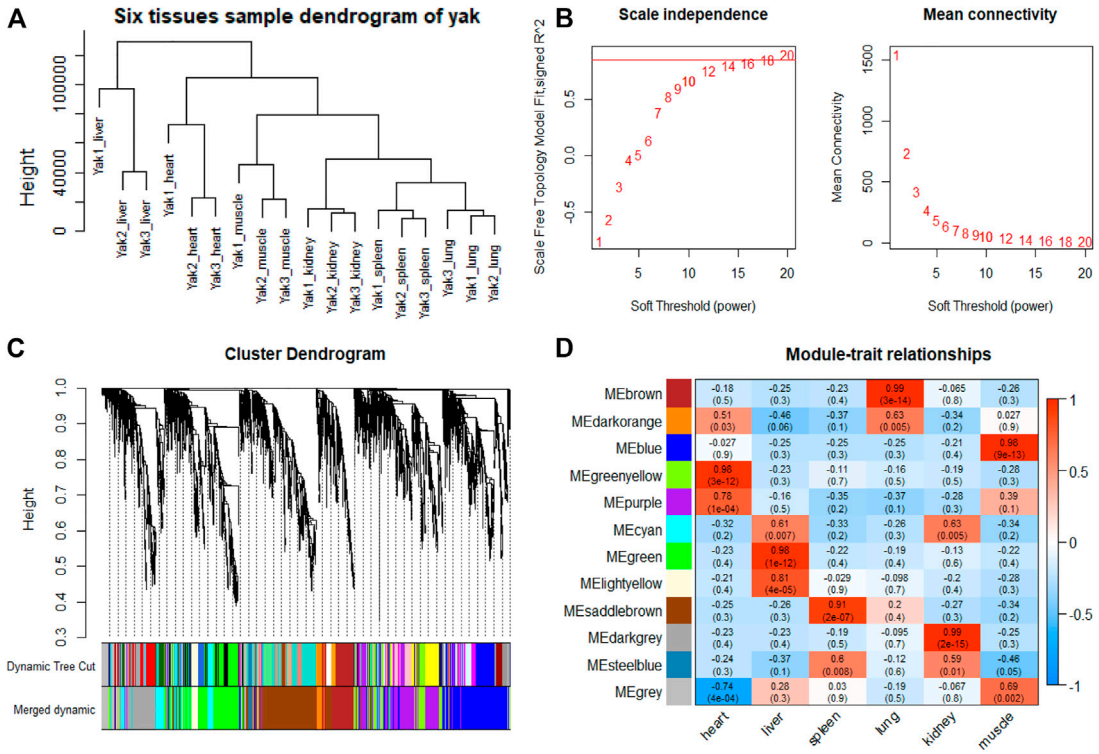


FIGURE 7 | WGCNA analysis of yak tissue samples. **(A)** Hcluster diagram of yak tissue samples. **(B)** Plot of scale-free topology and mean connectivity in regard to soft-thresholding power for samples. **(C)** Hierarchical clustering tree (dendrogram) of genes based on co-expression network analysis. **(D)** Heatmap of the correlation between the module eigengenes and tissues of yak.

osteoblast differentiation (GO:0045669), and KEGG including Adrenergic signaling in cardiomyocytes (bta04261). The module of MEbrown was significantly correlated with the lung of yak (r = 0.99, $p = 3 \times 10^{-14}$), which was involved in response to hypoxia (GO:0001666) (**Supplementary Table S6**). For low-altitude cattle, the heart-related module (MElightyellow, r = 0.99, $p = 1 \times 10^{-15}$) was enriched in cardiac right ventricle morphogenesis (GO:0003215), negative regulation of cardiac muscle cell apoptotic process (GO:0010667), and heart development (GO:0007507). Lung-related module (MEdarkorange, r = 0.99, $p = 7 \times 10^{-16}$) was enriched in lung development (GO:0030324). Additionally, in the tissue-related module of each species, the terms related to bile secretion were found in the liver-specific module, functions related to ion transport and absorption are enriched in the kidney-specific module, and immune and skeletal muscle cell differentiation-related functions were enriched in the spleen-specific and muscle-specific module, respectively (**Supplementary Table S6**).

Furthermore, genes with the highest intramodular connectivity were selected as hub genes in each tissue-related module in both gayal and yak. We identified 4–51, 6–44, and 2–46 hub genes in six tissue-related modules of gayal, yak, and cattle, respectively (**Supplementary Table S7**). Among the heart-related modules of the three species, there are fourteen common hub genes between gayal and yak, of which five genes are involved in cardiac contraction (bta04260), including *UQCRC1, UQCRQ, UQCRFS1, UQCR11*, and *COX5A*. Among them, *UQCRC1*, *UQCRQ*, *UQCRFS1*, and *UQCR11* are important subunit of mitochondrial complex III (Fernández-Vizarra and Zeviani, 2015). Mitochondria play a key role in cardioprotection and heart contraction energy needed for pumping blood to oxygenate the body organs (Lemieux and Hoppel, 2009). *UQCRC1* directly affects mitochondria function related to cardioprotection (Yi et al., 2020). *COX5A* is the component of the cytochrome c oxidase, which is central to oxidative phosphorylation and ATP generation (Huang et al., 2019). However, no common hub gene was found between gayal and low-altitude cattle, and only four hub genes are shared by yak and cattle, including *KCNJ8*, *MTUS2*, *P4HTM*, and *BTG3*. *KCNJ8* encodes a subunit of an ATP-sensitive potassium channel and regulates potassium inward rectifier, which is related to heart development (Erginel-Unaltuna et al., 1998). There are one (*COX6A2*) and two (*CYC1* and *UQCR10*) non-common hub genes in the heart-related modules of gayal and yak, respectively, which are involved in cardiac contraction. In addition, four hub genes (*HAND2*, *CHRM2*, *EDNRA*, and *ZFPM2*) associated with heart development and regulation were uniquely identified in the greenyellow module of the yak heart. Among these, *CHRM2* was an up-regulated DEG in the yak heart ($\log_2$ fold changes of 1.14), which plays a fundamental role in autonomic regulation of the heart, such as heart rate recovery after maximal exercise (Hautala et al., 2006). *HAND2* is essential for cardiac

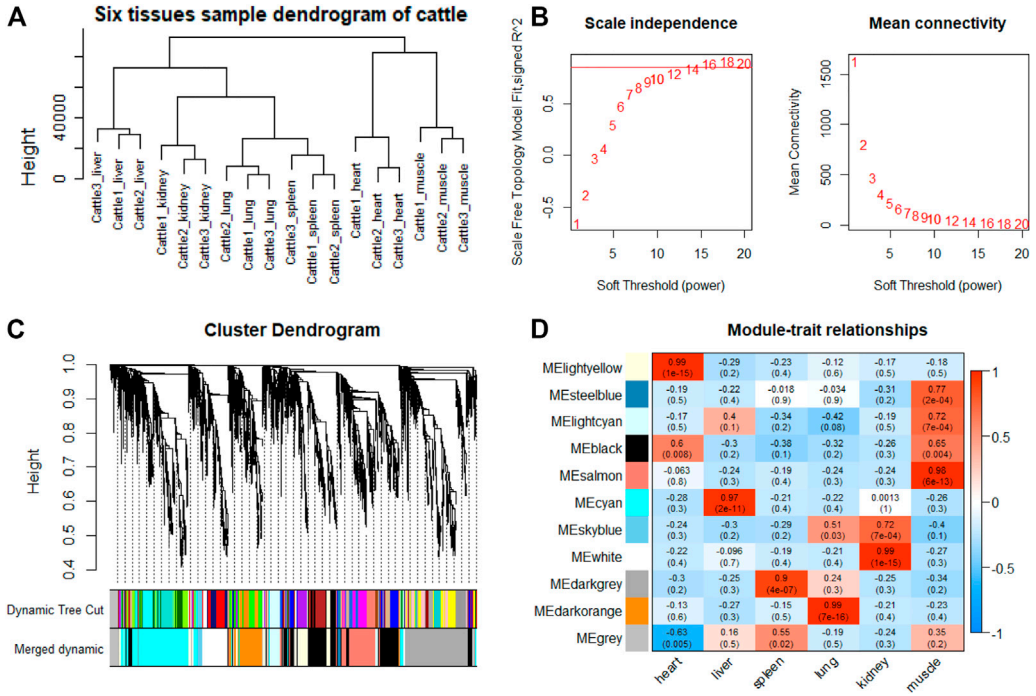


FIGURE 8 | WGCNA analysis of cattle tissue samples. **(A)** Hcluster diagram of cattle tissue samples. **(B)** Plot of scale-free topology and mean connectivity in regard to soft-thresholding power for samples. **(C)** Hierarchical clustering tree (dendrogram) of genes based on co-expression network analysis. **(D)** Heatmap of the correlation between the module eigengenes and tissues of cattle.

morphogenesis, vascular development, and the regulation of angiogenesis (Anderson et al., 2016). *EDNRA* encodes the endothelin 1 (*EDN1*) receptor, a gene involved in the vasoconstriction mechanism (Sharma et al., 2014), as well as closely related to pulmonary hypertension and HIF activity, which is positively selected in Tibetan (Simonson et al., 2010). *ZFPM2*, regulates the activity of GATA family proteins, plays a vital function in cardiac morphogenesis and coronary vascular development (Luo et al., 2020).

Among the hub genes of the lung-related module, six common hub genes were identified among gayal and yak, including *CAPS*, *TSNAXIP1*, *GPRIN2*, *SPEF1*, *BMP3*, *and EFHC1*. Gayal and yak have two (*BMP3* and *SEC14L3*) and three (*BMP3*, *DNAAF1*, and *N4BP3*) hub genes in common with low-altitude cattle respectively. Among them, *BMP3*, a hub gene shared by all three bovine subfamily species, has been reported to play a regulatory role in the morphogenesis and/or function of the human lung (Vukicevic et al., 1994). *CAPS* encodes a calcium-binding protein and plays a role in the regulation of Ca^{2+} transport (Pan et al., 2019), which may be correlated with pulmonary artery smooth muscle contraction (Waypa and Schumacker, 2002). *SEC14L3* is an up-regulated DEG in the gayal lung and plays an important role in maintaining the homeostasis of airway epithelial cells (Shan et al., 2009). *DNAAF1* gene is associated with lung development (Duquesnoy et al., 2009). As for the non-common hub genes in the lung-related module of gayal, we discovered several genes (*EDN3*, *GJA5*, and *CCDC39*) related to the tracheal epithelium and pulmonary vasoconstriction. Such as *EDN3*, an endothelium-derived vasoconstrictor peptide, which was up-regulated in the lungs of gayal, and may play an important role in enhancing pulmonary vasoconstrictor response (Masaki, 2000). *GJA5* is the predominant gap junction protein present in vascular endothelium, plays an important role in coupling between cells in the vascular wall (Ebong et al., 2006). *CCDC39* was related to lung development (Blanchon et al., 2012). Therefore, these genes related to promoting cardiac blood circulation and pulmonary vasoconstriction may play crucial roles in the environmental adaptability of gayal and yak.

DISCUSSION

In this study, we explored the high-altitude adaptability of Bovidae using the transcriptome data of six tissues from gayal and published data on yak and lowland cattle (Tang et al., 2017). Compared with numerous previous studies that only used the same reference genomes (yak or cattle genome) for different bovine subfamily species (Tang et al., 2017; Lan et al., 2018; Xin et al., 2019; Wang et al., 2020), here we optimized the experimental design, using the corresponding reference genomes for three species to reduce the gene sequence differences of different species due to phylogenetic evolution, and only retained reads that mapped to the 1:1 orthologous gene which had high-confidence alignments across the three genomes for subsequent analysis. To reduce the difference in the

sequencing depth of transcripts from different sequencing platforms, we used the TMM standardized method of edgeR for correction with reference to the method of Blake et al. (Blake et al., 2020). Consistent with previous studies (Necsulea and Kaessmann, 2014; Tang et al., 2017; Hao et al., 2019), we found that the differences in global gene expression among tissues were more significant than that among altitudes or the same tissue of different species. Among them, four tissues (lung, liver, spleen, and kidney) showed tissue-specific expression patterns, while bovine species-specific expression patterns were shown in the tissue of heart and muscle (Necsulea and Kaessmann, 2014; Tang et al., 2017; Hao et al., 2019). In addition, we identified numerous DEGs in gayal and yak compared with low-altitude cattle, which might have an important role in high altitude adaptation. WGCNA was further used to explore the core genes of each tissue that regulate tissue function in the high-altitude adaptation.

Expression Regulation of Genes Involved in the Regulation of Blood Cell Development

Hypoxia is one of the most serious challenges faced by high-altitude animals, and oxygen supply is directly related to the development of red blood cells (Zhang D. et al., 2017). Studies in mice (Li et al., 2011) and yak (Xin et al., 2019) have shown an increase in the number of red blood cells and platelets under hypoxia, but the number of granulocyte-macrophage progenitor cells periodically declined. In this study, similar results were found in the gayal. The hematopoietic cell lineage pathway was uniquely enriched in multiple tissues (heart, kidney, lung, and spleen) of gayal *vs.* the cattle group. Among these, *IL11RA* is a member of the hematopoietic cytokine receptor family (Ng et al., 2021), which was uniquely up-regulated by 1.33 times in the heart of gayal. *GP1BA* is the alpha subunit of platelet surface membrane glycoprotein Ib (Luo et al., 2007), were 9.58, 3.90, 7.77, 11.41, 12.38, 12.39-fold higher in the tissues (corresponding to heart, spleen, lung, kidney) of gayal than in the cattle, suggesting that the number of red blood cells and platelets might be increased. As for the regulatory mechanisms underlying granulocyte-macrophage progenitor cells, we observed three genes (*IL1B, CSF1, and IL6R*) were down-regulated in multi tissues of gayal, which might be a reason for the decreased proportion of granulocytes. Among them, *IL1B* is produced by activated macrophages as a proprotein (Lappas, 2013), which was down-regulated 1.94-fold in the lung of gayal. *CSF1* controls the production, differentiation, and function of macrophages (Hamilton et al., 2016), displaying $\log_2$ fold changes of −1.06, −1.10, and −1.76 in the heart, spleen, and muscle, respectively. In addition, Interleukin 6 is involved in the maturation of B cells. *IL6R* encodes a subunit of the interleukin 6 (*IL6*) receptor complex (Liu et al., 2017) and was found down-regulated in all tissues of gayal, which suggests lymphocytes may decrease, which is consistent with the previous results on the hypoxia adaptation of yak (Xin et al., 2019).

In addition, in the high-altitude related gene set that we searched (**Supplementary Table S5**), several crucial positive selection genes related to erythrocyte development were found differential expression in the gayal and yak. *EGLN1,* a key positive selection gene in the HIF pathway (Bigham et al., 2010), was found consistently down-regulated in all tissues of gayal, which contribute to enhancing HIF1-α activity and has a role in increased Hb levels via erythropoiesis under hypoxia conditions (Epstein et al., 2001; Xiang et al., 2013). *HFE* is a positively selected gene in Tibetans (Yi et al., 2010), which indirectly regulates iron balance by reducing the entry of transferrin into cells (Salter-Cid et al., 1999). Compared with lowland cattle, the appropriate down-regulation of *HFE* gene expression in gayal can accelerate iron storage to facilitate the synthesis of hemoglobin to cope with the reduced oxygen partial pressure (Muckenthaler et al., 2020). In the high-altitude related DEG set of yak, we also found *EGLN1* was consistently down-regulated in all tissues of yak. In addition, *HOXB6* is considered to be a marker for erythropoiesis (Magli et al., 1997), and its expression was uniquely up-regulated 1.85 and 1.11-fold in the spleen and kidney of the yak, respectively, which may contribute to yak adaptation to the hypoxia environment (Xin et al., 2019).

Expression Regulation of Genes Involved in Angiogenesis

The regulation of angiogenesis by hypoxia links blood oxygen supply to metabolism (Pugh and Ratcliffe, 2003). In this study, the shared and unique DEGs involved in blood vessel and vascular development were identified. For example, glutamyl aminopeptidase (*ENPEP*), a member of the M1 family of endopeptidases involved in blood pressure regulation and blood vessel formation (Holmes et al., 2017), was up-regulated in the lungs of gayal and yak with the $\log_2$ fold changes of 2.78 and 2.98, respectively. Additionally, we found that *SRPX2* and *ESM1* were up-regulated in the spleen of gayal and yak relative to lowland cattle, and *SAT1* and *SERPINE1* genes were up-regulated in the kidney and muscle of gayal and yak relative to lowland cattle, respectively, which are angiogenesis promoting genes (Aitkenhead et al., 2002; Mohr et al., 2017; Xiao et al., 2020). *HEY1* is the downstream effector of Notch signaling required for cardiovascular development (Aitkenhead et al., 2002) with $\log_2$ fold changes of 1.08 and 1.71 in the heart and lung of gayal, respectively, while displaying $\log_2$ fold changes of 1.25 and 1.03 in the yak, respectively. In addition, *MEOX2* has been reported to have a role in inhibiting angiogenesis (Dhahri et al., 2020), and was down-regulated in all 6 tissues of gayal and the kidney of yak. Research has found that VHL-deficient mice show increased vasculogenesis in embryonic cells (Wang et al., 2017), and the *VHL* gene was down-regulated in expression in all six tissues of yak and down-regulated in the lung of gayal. In addition, we have identified two unique DEGs (*RAMP2* and *TGFBI*) in gayal that promote angiogenesis. *RAMP2* mediates the pro-angiogenic effect (Albertin et al., 2010), which was up-regulated in all six tissues of gayal. *TGFBI* is a matricellular protein-coding gene that plays an important role in tumor angiogenesis (Fico and Santamaria-Martínez, 2020), which was up-regulated in the liver and kidney of gayal with $\log_2$ fold changes of 1.34 and 1.45, respectively. Additionally, DEGs (*THY1* and *PLCD3*) that promote angiogenesis were also uniquely found in yak. *THY1* was expressed on mouse tumor-associated lymphatic vessels and

blood vessels (Jurisic et al., 2010), displaying $\log_2$ fold changes of 1.63 and 1.03 in the spleen and lung of yak. *PLCD3* induce angiogenesis in human endothelium (Kim et al., 2002), which was up-regulated in all six tissues, displaying $\log_2$ fold changes of 1.77, 1.83, 2.07, 3.20, 2.12, and 1.15 in the heart, liver, spleen, lung, kidney, and muscle of yak, respectively. This suggests that angiogenesis may be conducive to the adaptation of gayal and yak to a high-altitude environment.

Expression Regulation of Genes Involved in the Immune System

Increasing evidence shows that the harsh environment of hypoxia may regulate the immune system (Mishra and Ganju, 2010; Gaur et al., 2020). In the present study, we found that the immune system of the spleen in gayal and yak was highly regulated. DEGs of the spleen in gayal were significantly enriched in lymphocytic chemotaxis, inflammatory response, and cytokine receptor interaction pathways (**Supplementary Table S4**), and the DEGs of the spleen in yak were also significantly enriched in TNF signaling pathway, cytokine receptor interaction, and chemokine signaling pathway (**Supplementary Table S4**). In the spleen of gayal *vs.* cattle group, chemokines and their receptors like *CCL4, CCL5, CCL14, CCL16,* and *CXCR4* displayed $\log_2$ fold changes of 2.33, 2.78, 1.24, 1.83, and 1.18, respectively, which mediates chemokinetic inflammatory response (Loetscher et al., 1994; Ariel et al., 2000); Interleukins like *IL12RB2* promotes the proliferation of T-cells as well as NK cells (Presky et al., 1996), *IL23A* contribute to the production of proinflammatory cytokines (Hoeve et al., 2006), and *IL12RB2* and *IL23A* with $\log_2$ fold changes of 1.77 and 1.09, respectively; tumor necrosis factors like *TNFRSF8, TNFRSF4,* and *LTBR* displayed $\log_2$ fold changes of 4.06, 1.41, and 1.32, respectively, which activated T and B cell expression (Nishikori et al., 2005). *CD40LG* regulates B cell function by engaging *CD40* on the B cell surface (Henn et al., 1998), which was up-regulated 2.63-fold in the spleen of gayal. In the spleen of the yak *vs.* cattle group, *CXCR4* and *CD40LG* were also up-regulated, with $\log_2$ fold changes of 1.52 and 2.83, respectively. *CXCL14* which fulfills a unique role in antimicrobial immunity was up-regulated 2.15-fold in the spleen of yak (Maerki et al., 2009). These changes might help gayal and yak resist inflammation and disease in high-altitude harsh environments.

High altitude pulmonary edema (HAPE) is a serious life-threatening disease in humans and animals characterized by uneven vasoconstriction of pulmonary blood vessels due to hypoxia (Stream and Grissom, 2008). It has been confirmed that HAPE is related to coagulation activation and fibrin formation enhancement (Mannucci et al., 2002). In addition, hypoxia can lead to increased platelet activity, significantly increasing the risk of thrombosis (Shilei et al., 2016). In this study, we found that a large number of DEGs are related to platelet activation and blood coagulation in the two comparison groups. Coagulation factors play an important role in the endogenous and exogenous blood coagulation process in animals (Thiel et al., 2017). In this study, compared with low-altitude cattle, the expression levels of coagulation factor II (*F2*) and coagulation factor V (*F5*) in the lung of gayal were separately reduced 1.26 and 1.64-fold, and displayed $\log_2$ fold changes of −2.33 and −1.23 in the lung of yak, respectively. In addition, coagulation factor X (*F10*) was found uniquely down-regulated 5.62 times in the lungs of yak. Meanwhile, serine protease inhibitors are inhibitors of blood coagulation factors (Seixas et al., 2007), and we have found the expression level of *SERPINA5, SERPINF2* and *SERPING1* were uniquely increased in the lung of gayal, and the corresponding $\log_2$ fold changes of 1.80, 3.06, and 1.48, which may play an important role in the mechanism of pulmonary edema resistance, thus reducing the risk of pulmonary edema. In addition, impaired decomposition of bradykinin and its metabolites may be an important cause of pulmonary edema (De Maat et al., 2020). We found that the expression of the kininogen 1 (*KNG1*) gene was down-regulated by 2.46 times in the lung of yak, while down-regulated by 12.39 and 7.38-fold in the heart of gayal and yak than that of lowland cattle, respectively. These results suggest yak and gayal have developed a mechanism to prevent the occurrence of severe pulmonary edema.

Expression Regulation of Genes Involved in Energy Metabolism

High energy metabolism or reduced energy turnover is essential for survival under hypoxia (Qiu et al., 2012). In this study, mitochondrion-related categories were widely enriched in the tissues of gayal and yak (**Supplementary Table S4**). Mitochondria is the main place for aerobic respiration of cells, which provides ATP for organisms through the oxidation of sugar, fats, and amino acids (Solaini et al., 2010). Similarly, several pathways related to carbohydrate and fat energy metabolism were enriched, such as glycolysis/gluconeogenesis in yak lung, oxidative phosphorylation, and galactose metabolism in multiple tissues of gayal, and PPAR signaling pathway in the heart of gayal and yak. Among the genes related to these functions, *TPI1* is essential for efficient energy production and plays an important role in glycolysis (Shimoda et al., 2012). *G6PC* (glucose-6-phosphatase) is a key enzyme of gluconeogenesis (Jia et al., 2012). We found that the gene expression of *TPI1* was up-regulated in all tissues of both gayal and yak, and *G6PC* was down-regulated in three tissues (heart, spleen, and lung) of both gayal and yak, which is conducive to the production of ATP through strengthened glycolysis and reduced gluconeogenesis. Oxidative phosphorylation is an important process for ATP syntesis in aerobic organisms. We found that most genes related to oxidative phosphorylation were up-regulated in gayal, such as *COX6A2, NDUFA9, COX8B, NDUFB7, NDUFA11, NDUFB5, UQCR10, NDUFS8, UQCRQ, PPA1,* and *NDUFS3*. In addition, the significant down-regulation of *COX5B* found in all tissues of gayal and yak may be beneficial to hypoxia adaptation (Trueblood et al., 1988). Additionally, galactose metabolism (bta00052) was uniquely enriched in the tissues (lung, liver, spleen, and muscle) of gayal. Galactose is a kind of monosaccharide, most of which are converted into glucose in the liver, and then incorporated into glycogen or used for energy metabolism by glycolysis. Lactase (*LCT*), a key enzyme involved

in the conversion of galactose to glucose was up-regulated in all six tissues of gayal. The PPAR signaling pathway, a classic pathway in lipid catabolism was significantly enriched in the heart of gayal and yak. *EHHADH*, a dehydrogenase in the β-oxidase system (Yeh et al., 2006), was down-regulated 4.02 and 2.69 times in the heart of gayal and yaks. Additionally, the utilization of fatty acids first needs to be activated to produce fatty acyl CoA under the catalysis of the ACSL and SLC27A family (Bowman et al., 2016). In this study, we found that *SLC27A2* and *SLC27A6* of the SLC27A gene family were down-regulated 6.24 and 3.24 times in the heart of gayal, and *SLC27A2* and *SLC27A4* were down-regulated 6.81 and 2.08 times in yak heart, respectively. And the *ACSL3* and *ACSL4* genes expression of ACSL gene family members were found uniquely decreased in the heart of gayal. Therefore, our results suggest gayal and yak may adapt to high altitude environments by strengthening carbohydrate metabolism and reducing fatty acid degradation although the related regulation genes may be different between gayal and yak.

Despite the fact that we found numerous genes that may be related to the high-altitude adaptation of gayal and yak, we must admit that the limitation of age differences between gayal and other species, the gayal being a younger species, may affect the experimental results obtained to some extent. Therefore, it is necessary to conduct comparative transcriptome analysis on more bovine species of the same age that are combined with physiologic experiments to verify our findings, and better clarify the high-altitude adaptation mechanism of bovine subfamily species.

CONCLUSION

This study shows a comprehensive multi tissues transcriptome expression landscape of high- and low-altitude bovine populations and reveals the gene expression regulation associated with high-altitude acclimatization. The gene expression profiles of gayal, yak, and cattle showed tissue-specific expression patterns. The comparative analysis of six hypoxia-related tissues of high- and low-altitude bovine subfamily species highlights numerous DEGs and terms underlying associated with hypoxia. Notably, categories related to angiogenesis, blood coagulation, energy metabolism, and the immune system were commonly enriched in the tissue DEGs of gayal *vs.* cattle group and yak *vs.* cattle group, and we found that many expression regulatory genes related to these functions in gayal and yak are different, which may serve as an important regulatory mechanism for gayal and yak to adapt to the corresponding local environment. In addition, we also found numerous hub genes related to myocardial contraction, energy metabolism, and pulmonary vasoconstriction by WGCNA analysis. Overall, our study lays a foundation for further study on the environmental adaptability of mammals in subtropical plateau and provides a valuable basis for resource utilization in high-altitude areas.

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: NCBI PRJNA783860.

ETHICS STATEMENT

The animal study was reviewed and approved by All procedures are approved by the Scientific Research Department of the Institute of Animal Science, Chinese Academy of Agricultural Sciences, and strictly follow the guidelines formulated by the Chinese Animal Protection Commission.

AUTHOR CONTRIBUTIONS

JM, TZ, and XG conceived and designed the experiments. JM analyzed the data and wrote the paper. YC, WW, and BZ collected the data. WC contributed analysis tools, LZ, LX, HG, JL, and XG discussed and improved the article. All authors read and approved the final manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (31572376), the Agricultural Science and Technology Innovation Program in Chinese Academy of Agricultural Sciences (ASTIP-IAS03), the Science and Technology Project of Inner Mongolia Autonomous Region (2020GG0210), and the China Agriculture Research System of MOF and MARA (CARS-37). The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.778788/full#supplementary-material

REFERENCES

Aitkenhead, M., Wang, S.-J., Nakatsu, M. N., Mestas, J., Heard, C., and Hughes, C. C. W. (2002). Identification of Endothelial Cell Genes Expressed in an *In Vitro* Model of Angiogenesis: Induction of ESM-1, βig-h3, and NrCAM. *Microvasc. Res.* 63, 159–171. doi:10.1006/mvre.2001.2380

Anderson, K. M., Anderson, D. M., Mcanally, J. R., Shelton, J. M., Bassel-Duby, R., and Olson, E. N. (2016). Transcription of the Non-coding RNA Upperhand Controls Hand2 Expression and Heart Development. *Nature* 539, 433–436. doi:10.1038/nature20128

Ariel, A., Lider, O., Brill, A., Cahalon, L., Savion, N., Varon, D., et al. (2000). Induction of Interactions between CD44 and Hyaluronic Acid by a Short Exposure of Human T Cells to Diverse Pro-inflammatory Mediators. *Immunology* 100, 345–351. doi:10.1046/j.1365-2567.2000.00059.x

Bigham, A., Bauchet, M., Pinto, D., Mao, X., Akey, J. M., Mei, R., et al. (2010). Identifying Signatures of Natural Selection in Tibetan and Andean Populations Using Dense Genome Scan Data. *Plos Genet.* 6, e1001116. doi:10.1371/journal.pgen.1001116

Blake, L. E., Roux, J., Hernando-Herraez, I., Banovich, N. E., Perez, R. G., Hsiao, C. J., et al. (2020). A Comparison of Gene Expression and DNA Methylation Patterns across Tissues and Species. *Genome Res.* 30, 250–262. doi:10.1101/gr.254904.119

Blanchon, S., Legendre, M., Copin, B., Duquesnoy, P., Montantin, G., Kott, E., et al. (2012). Delineation ofCCDC39/CCDC40mutation Spectrum and Associated Phenotypes in Primary Ciliary Dyskinesia. *J. Med. Genet.* 49, 410–416. doi:10.1136/jmedgenet-2012-100867

Bowman, T. A., O'keeffe, K. R., D'aquila, T., Yan, Q. W., Griffin, J. D., Killion, E. A., et al. (2016). Acyl CoA Synthetase 5 (ACSL5) Ablation in Mice Increases Energy Expenditure and Insulin Sensitivity and Delays Fat Absorption. *Mol. Metab.* 5, 210–220. doi:10.1016/j.molmet.2016.01.001

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: Architecture and Applications. *BMC Bioinformatics* 10, 421. doi:10.1186/1471-2105-10-421

Chen, S., Yang, D., Lei, C., Li, Y., Sun, X., Chen, M., et al. (2019). Identification of Crucial Genes in Abdominal Aortic Aneurysm by WGCNA. *PeerJ* 7, e7873. doi:10.7717/peerj.7873

Conway, J. R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R Package for the Visualization of Intersecting Sets and Their Properties. *Bioinformatics* 33, 2938–2940. doi:10.1093/bioinformatics/btx364

De Maat, S., De Mast, Q., Danser, A. H. J., Van De Veerdonk, F. L., and Maas, C. (2020). Impaired Breakdown of Bradykinin and its Metabolites as a Possible Cause for Pulmonary Edema in COVID-19 Infection. *Semin. Thromb. Hemost.* 46, 835–837. doi:10.1055/s-0040-1712960

Dhahri, W., Dussault, S., Légaré, É., Rivard, F., Desjarlais, M., Mathieu, R., et al. (2020). Reduced Expression of microRNA-130a Promotes Endothelial Cell Senescence and Age-dependent Impairment of Neovascularization. *Aging* 12, 10180–10193. doi:10.18632/aging.103340

Duquesnoy, P., Escudier, E., Vincensini, L., Freshour, J., Bridoux, A.-M., Coste, A., et al. (2009). Loss-of-function Mutations in the Human Ortholog of Chlamydomonas Reinhardtii ODA7 Disrupt Dynein Arm Assembly and Cause Primary Ciliary Dyskinesia. *Am. J. Hum. Genet.* 85, 890–896. doi:10.1016/j.ajhg.2009.11.008

Ebong, E. E., Kim, S., and Depaola, N. (2006). Flow Regulates Intercellular Communication in HAEC by Assembling Functional Cx40 and Cx37 gap Junctional Channels. *Am. J. Physiology-Heart Circulatory Physiol.* 290, H2015–H2023. doi:10.1152/ajpheart.00204.2005

Emms, D. M., and Kelly, S. (2015). OrthoFinder: Solving Fundamental Biases in Whole Genome Comparisons Dramatically Improves Orthogroup Inference Accuracy. *Genome Biol.* 16, 157. doi:10.1186/s13059-015-0721-2

Epstein, A. C. R., Gleadle, J. M., Mcneill, L. A., Hewitson, K. S., O'rourke, J., Mole, D. R., et al. (2001). *C. elegans* EGL-9 and Mammalian Homologs Define a Family of Dioxygenases that Regulate HIF by Prolyl Hydroxylation. *Cell* 107, 43–54. doi:10.1016/s0092-8674(01)00507-4

Erginel-Unaltuna, N., Yang, W.-P., and Blanar, M. A. (1998). Genomic Organization and Expression of KCNJ8/Kir6.1, a Gene Encoding a Subunit of an ATP-Sensitive Potassium Channel. *Gene* 211, 71–78. doi:10.1016/s0378-1119(98)00086-9

Feng, S., Ma, J., Long, K., Zhang, J., Qiu, W., Li, Y., et al. (2020). Comparative microRNA Transcriptomes in Domestic Goats Reveal Acclimatization to High Altitude. *Front. Genet.* 11, 809. doi:10.3389/fgene.2020.00809

Fernández-Vizarra, E., and Zeviani, M. (2015). Nuclear Gene Mutations as the Cause of Mitochondrial Complex III Deficiency. *Front. Genet.* 6, 1. doi:10.3389/fgene.2015.00134

Fico, F., and Santamaria-Martínez, A. (2020). TGFBI Modulates Tumour Hypoxia and Promotes Breast Cancer Metastasis. *Mol. Oncol.* 14, 3198–3210. doi:10.1002/1878-0261.12828

Gaur, P., Saini, S., Ray, K., Asanbekovna, K. N., Akunov, A., Maripov, A., et al. (2020). Temporal Transcriptome Analysis Suggest Modulation of Multiple Pathways and Gene Network Involved in Cell-Cell Interaction during Early Phase of High Altitude Exposure. *PLoS One* 15, e0238117. doi:10.1371/journal.pone.0238117

Ge, R.-L., Simonson, T. S., Cooksey, R. C., Tanna, U., Qin, G., Huff, C. D., et al. (2012). Metabolic Insight into Mechanisms of High-Altitude Adaptation in Tibetans. *Mol. Genet. Metab.* 106, 244–247. doi:10.1016/j.ymgme.2012.03.003

Gou, W., Peng, J., Wu, Q., Zhang, Q., Zhang, H., and Wu, C. (2014). Expression Pattern of Heme Oxygenase 1 Gene and Hypoxic Adaptation in Chicken Embryos. *Comp. Biochem. Physiol. B: Biochem. Mol. Biol.* 174, 23–28. doi:10.1016/j.cbpb.2014.05.005

Guidolin, G., Sorato, E., Oselladore, B., Mascarin, A., Tortorella, C., and Guidolin, D. (2010). Involvement of Vascular Endothelial Growth Factor Signaling in CLR/RAMP1 and CLR/RAMP2-mediated Pro-angiogenic Effect of Intermedin on Human Vascular Endothelial Cells. *Int. J. Mol. Med.* 26, 289–294. doi:10.3892/ijmm_00000464

Haasl, R. J., and Payseur, B. A. (2016). Fifteen Years of Genomewide Scans for Selection: Trends, Lessons and Unaddressed Genetic Sources of Complication. *Mol. Ecol.* 25, 5–23. doi:10.1111/mec.13339

Hamilton, J. A., Cook, A. D., and Tak, P. P. (2016). Anti-colony-stimulating Factor Therapies for Inflammatory and Autoimmune Diseases. *Nat. Rev. Drug Discov.* 16, 53–70. doi:10.1038/nrd.2016.231

Hao, Y., Xiong, Y., Cheng, Y., Song, G., Jia, C., Qu, Y., et al. (2019). Comparative Transcriptomics of 3 High-Altitude Passerine Birds and Their Low-Altitude Relatives. *Proc. Natl. Acad. Sci. USA* 116, 11851–11856. doi:10.1073/pnas.1819657116

Hautala, A. J., Rankinen, T., Kiviniemi, A. M., Mäkikallio, T. H., Huikuri, H. V., Bouchard, C., et al. (2006). Heart Rate Recovery after Maximal Exercise Is Associated with Acetylcholine Receptor M2 (CHRM2) Gene Polymorphism. *Am. J. Physiology-Heart Circulatory Physiol.* 291, H459–H466. doi:10.1152/ajpheart.01193.2005

Heinrich, E. C., Wu, L., Lawrence, E. S., Cole, A. M., Anza-Ramirez, C., Villafuerte, F. C., et al. (2019). Genetic Variants at the EGLN1 Locus Associated with High-altitude Adaptation in Tibetans Are Absent or Found at Low Frequency in highland Andeans. *Ann. Hum. Genet.* 83, 171–176. doi:10.1111/ahg.12299

Henn, V., Slupsky, J. R., Gräfe, M., Anagnostopoulos, I., Förster, R., Müller-Berghaus, G., et al. (1998). CD40 Ligand on Activated Platelets Triggers an Inflammatory Reaction of Endothelial Cells. *Nature* 391, 591–594. doi:10.1038/35393

Hoeve, M. A., Savage, N. D. L., De Boer, T., Langenberg, D. M. L., de Waal Malefyt, R., Ottenhoff, T. H. M., et al. (2006). Divergent Effects of IL-12 and IL-23 on the Production of IL-17 by Human T Cells. *Eur. J. Immunol.* 36, 661–670. doi:10.1002/eji.200535239

Holmes, R. S., Spradling Reeves, K. D., and Cox, L. A. (2017). Mammalian Glutamyl Aminopeptidase Genes (ENPEP) and Proteins: Comparative Studies of a Major Contributor to Arterial Hypertension. *J. Data Mining Genomics Proteomics* 08. doi:10.4172/2153-0602.1000211

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources. *Nat. Protoc.* 4, 44–57. doi:10.1038/nprot.2008.211

Huang, Y., Li, S.-n., Zhou, X.-y., Zhang, L.-x., Chen, G.-x., Wang, T.-h., et al. (2019). The Dual Role of AQP4 in Cytotoxic and Vasogenic Edema Following Spinal Cord Contusion and its Possible Association with Energy Metabolism via COX5A. *Front. Neurosci.* 13, 584. doi:10.3389/fnins.2019.00584

Jia, C., Kong, X., Koltes, J. E., Gou, X., Yang, S., Yan, D., et al. (2016). Gene Co-expression Network Analysis Unraveling Transcriptional Regulation of High-Altitude Adaptation of Tibetan Pig. *PLoS One* 11, e0168161. doi:10.1371/journal.pone.0168161

Jia, Y., Cong, R., Li, R., Yang, X., Sun, Q., Parvizi, N., et al. (2012). Maternal Low-Protein Diet Induces Gender-dependent Changes in Epigenetic Regulation of the Glucose-6-Phosphatase Gene in Newborn Piglet Liver. *J. Nutr.* 142, 1659–1665. doi:10.3945/jn.112.160341

Jurisic, G., Iolyeva, M., Proulx, S. T., Halin, C., and Detmar, M. (2010). Thymus Cell Antigen 1 (Thy1, CD90) Is Expressed by Lymphatic Vessels and Mediates Cell Adhesion to Lymphatic Endothelium. *Exp. Cel Res.* 316, 2982–2992. doi:10.1016/j.yexcr.2010.06.013

Kaiser, R., Friedrich, D., Chavakis, E., Böhm, M., and Friedrich, E. B. (2012). Effect of Hypoxia on Integrin-Mediated Adhesion of Endothelial Progenitor Cells. *J. Cel. Mol. Med.* 16, 2387–2393. doi:10.1111/j.1582-4934.2012.01553.x

Keren, L., Ke, M., Tiandong, C., Jinwei, Z., Wanling, Q., Yujie, W., et al. (2018). Transcriptome Differences in Frontal Cortex between Wild Boar and Domesticated Pig. *Anim. Sci. J. = Nihon chikusan Gakkaiho* 89.

Kim, Y.-M., Kim, Y.-M., Lee, Y. M., Kim, H.-S., Kim, J. D., Choi, Y., et al. (2002). TNF-related Activation-Induced Cytokine (TRANCE) Induces Angiogenesis through the Activation of Src and Phospholipase C (PLC) in Human Endothelial Cells. *J. Biol. Chem.* 277, 6799–6805. doi:10.1074/jbc.M109434200

Lan, D., Xiong, X., Ji, W., Li, J., Mipam, T.-D., Ai, Y., et al. (2018). Transcriptome Profile and Unique Genetic Evolution of Positively Selected Genes in Yak Lungs. *Genetica* 146, 151–160. doi:10.1007/s10709-017-0005-8

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

Lappas, M. (2013). NOD1 and NOD2 Regulate Proinflammatory and Prolabor Mediators in Human Fetal Membranes and Myometrium via Nuclear Factor-Kappa B1. *Biol. Reprod.* 89, 14. doi:10.1095/biolreprod.113.110056

Lemieux, H., and Hoppel, C. L. (2009). Mitochondria in the Human Heart. *J. Bioenerg. Biomembr* 41, 99–106. doi:10.1007/s10863-009-9211-0

Li, P., Huang, J., Tian, H.-j., Huang, Q.-y., Jiang, C.-h., and Gao, Y.-q. (2011). Regulation of Bone Marrow Hematopoietic Stem Cell Is Involved in High-Altitude Erythrocytosis. *Exp. Hematol.* 39, 37–46. doi:10.1016/j.exphem.2010.10.006

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, A., Wang, Y., Sahana, G., Zhang, Q., Liu, L., Lund, M. S., et al. (2017). Genome-wide Association Studies for Female Fertility Traits in Chinese and Nordic Holsteins. *Sci. Rep.* 7, 8487. doi:10.1038/s41598-017-09170-9

Loetscher, M., Geiser, T., O'reilly, T., Zwahlen, R., Baggiolini, M., and Moser, B. (1994). Cloning of a Human Seven-Transmembrane Domain Receptor, LESTR, that Is Highly Expressed in Leukocytes. *J. Biol. Chem.* 269, 232–237. doi:10.1016/s0021-9258(17)42339-8

Luo, S.-Z., Mo, X., Afshar-Kharghan, V., Srinivasan, S., Lo´pez, J. A., and Li, R. (2007). Glycoprotein Ibα Forms Disulfide Bonds with 2 Glycoprotein Ibβ Subunits in the Resting Platelet. *Blood* 109, 603–609. doi:10.1182/blood-2006-05-024091

Luo, Y., Wang, X., Ma, L., Ma, Z., Li, S., Fang, X., et al. (2020). Bioinformatics Analyses and Biological Function of lncRNA ZFPM2-AS1 and ZFPM2 G-ene in H-epatocellular C-arcinoma. *Oncol. Lett.* 19, 3677–3686. doi:10.3892/ol.2020.11485

Maerki, C., Meuter, S., Liebi, M., Mühlemann, K., Frederick, M. J., Yawalkar, N., et al. (2009). Potent and Broad-Spectrum Antimicrobial Activity of CXCL14 Suggests an Immediate Role in Skin Infections. *J. Immunol.* 182, 507–514. doi:10.4049/jimmunol.182.1.507

Magli, M. C., Largman, C., and Lawrence, H. J. (1997). Effects ofHOX Homeobox Genes in Blood Cell Differentiation. *J. Cel. Physiol.* 173, 168–177. doi:10.1002/(sici)1097-4652(199711)173:2<168::aid-jcp16>3.0.co;2-c

Mannucci, P. M., Gringeri, A., Peyvandi, F., Di Paolantonio, T., and Mariani, G. (2002). Short-term Exposure to High Altitude Causes Coagulation Activation and Inhibits Fibrinolysis. *Thromb. Haemost.* 87, 342–343.

Masaki, T. (2000). The Endothelin Family: an Overview. *J. Cardiovasc. Pharmacol.* 35, S3–S5. doi:10.1097/00005344-200000002-00002

Merkin, J., Russell, C., Chen, P., and Burge, C. B. (2012). Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. *Science* 338, 1593–1599. doi:10.1126/science.1228186

Miao, F., Guo, Z., Xue, R., Wang, X., and Shen, Y. (2015). Effects of Grazing and Precipitation on Herbage Biomass, Herbage Nutritive Value, and Yak Performance in an Alpine Meadow on the Qinghai-Tibetan Plateau. *Plos One* 10, e0127275. doi:10.1371/journal.pone.0127275

Mishra, K. P., and Ganju, L. (2010). Influence of High Altitude Exposure on the Immune System: a Review. *Immunological Invest.* 39, 219–234. doi:10.3109/08820131003681144

Mohr, T., Haudek-Prinz, V., Slany, A., Grillari, J., Micksche, M., and Gerner, C. (2017). Proteome Profiling in IL-1β and VEGF-Activated Human Umbilical Vein Endothelial Cells Delineates the Interlink between Inflammation and Angiogenesis. *PLoS One* 12, e0179065. doi:10.1371/journal.pone.0179065

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. doi:10.1038/nmeth.1226

Muckenthaler, M. U., Mairbäurl, H., and Gassmann, M. (20201985). Iron Metabolism in High-Altitude Residents. *J. Appl. Physiol.* 129, 920–925. doi:10.1152/japplphysiol.00019.2020

Mukherjee, S., Mukherjee, A., Jasrotia, R. S., Jaiswal, S., Iquebal, M. A., Longkumer, I., et al. (2020). Muscle Transcriptome Signature and Gene Regulatory Network Analysis in Two Divergent Lines of a Hilly Bovine Species Mithun (*Bos frontalis*). *Genomics* 112, 252–262. doi:10.1016/j.ygeno.2019.02.004

Necsulea, A., and Kaessmann, H. (2014). Evolutionary Dynamics of Coding and Non-coding Transcriptomes. *Nat. Rev. Genet.* 15, 734–748. doi:10.1038/nrg3802

Ng, B., Widjaja, A. A., Viswanathan, S., Dong, J., Chothani, S. P., Lim, S., et al. (2021). Similarities and Differences between IL11 and IL11RA1 Knockout Mice for Lung Fibro-Inflammation, Fertility and Craniosynostosis. *Sci. Rep.* 11, 14088. doi:10.1038/s41598-021-93623-9

Nishikori, M., Ohno, H., Haga, H., and Uchiyama, T. (2005). Stimulation of CD30 in Anaplastic Large Cell Lymphoma Leads to Production of Nuclear Factor-kappaB P52, Which Is Associated with Hyperphosphorylated Bcl-3. *Cancer Sci.* 96, 487–497. doi:10.1111/j.1349-7006.2005.00078.x

Ozsolak, F., and Milos, P. M. (2011). RNA Sequencing: Advances, Challenges and Opportunities. *Nat. Rev. Genet.* 12, 87–98. doi:10.1038/nrg2934

Pamenter, M. E., Hall, J. E., Tanabe, Y., and Simonson, T. S. (2020). Cross-Species Insights into Genomic Adaptations to Hypoxia. *Front. Genet.* 11, 1. doi:10.3389/fgene.2020.00743

Pan, H., Xiang, H., Wang, J., Wei, Z., Zhou, Y., Liu, B., et al. (2019). CAPS Mutations Are Potentially Associated with Unexplained Recurrent Pregnancy Loss. *Am. J. Pathol.* 189, 124–131. doi:10.1016/j.ajpath.2018.09.010

Peng, Y., Cui, C., He, Y., OuzhuluobuZhang, H., Zhang, H., Yang, D., et al. (2017). Down-Regulation ofEPAS1Transcription and Genetic Adaptation of Tibetans to High-Altitude Hypoxia. *Mol. Biol. Evol.* 34, msw280–830. doi:10.1093/molbev/msw280

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level Expression Analysis of RNA-Seq Experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667. doi:10.1038/nprot.2016.095

Presky, D. H., Yang, H., Minetti, L. J., Chua, A. O., Nabavi, N., Wu, C.-Y., et al. (1996). A Functional Interleukin 12 Receptor Complex Is Composed of Two -type Cytokine Receptor Subunits. *Proc. Natl. Acad. Sci.* 93, 14002–14007. doi:10.1073/pnas.93.24.14002

Pugh, C. W., and Ratcliffe, P. J. (2003). Regulation of Angiogenesis by Hypoxia: Role of the HIF System. *Nat. Med.* 9, 677–684. doi:10.1038/nm0603-677

Qi, X., Zhang, Q., He, Y., Yang, L., Zhang, X., Shi, P., et al. (2019). The Transcriptomic Landscape of Yaks Reveals Molecular Pathways for High Altitude Adaptation. *Genome Biol. Evol.* 11, 72–85. doi:10.1093/gbe/evy264

Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z., et al. (2012). The Yak Genome and Adaptation to Life at High Altitude. *Nat. Genet.* 44, 946–949. doi:10.1038/ng.2343

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

Salleh, M. S., Mazzoni, G., Höglund, J. K., Olijhoek, D. W., Lund, P., Løvendahl, P., et al. (2017). RNA-seq Transcriptomics and Pathway Analyses Reveal Potential Regulatory Genes and Molecular Mechanisms in High- and Low-Residual Feed Intake in Nordic Dairy Cattle. *BMC Genomics* 18, 258. doi:10.1186/s12864-017-3622-9

Salter-Cid, L., Brunmark, A., Li, Y., Leturcq, D., Peterson, P. A., Jackson, M. R., et al. (1999). Transferrin Receptor Is Negatively Modulated by the Hemochromatosis Protein HFE: Implications for Cellular Iron Homeostasis. *Proc. Natl. Acad. Sci.* 96, 5434–5439. doi:10.1073/pnas.96.10.5434

Seixas, S., Suriano, G., Carvalho, F., Seruca, R., Rocha, J., and Di Rienzo, A. (2007). Sequence Diversity at the Proximal 14q32.1 SERPIN Subcluster: Evidence for Natural Selection Favoring the Pseudogenization of SERPINA2. *Mol. Biol. Evol.* 24, 587–598. doi:10.1093/molbev/msl187

Shan, L., Kawakami, T., Asano, S., Noritake, S., Yoshimoto, D., Yamashita, K., et al. (2009). Inverse Relationship between Sec14l3 mRNA/protein Expression and Allergic Airway Inflammation. *Eur. J. Pharmacol.* 616, 293–300. doi:10.1016/j.ejphar.2009.06.055

Sharma, M., Singh, S. B., and Sarkar, S. (2014). Genome Wide Expression Analysis Suggests Perturbation of Vascular Homeostasis during High Altitude Pulmonary Edema. *PLoS One* 9, e85902. doi:10.1371/journal.pone.0085902

Shilei, C., Changhong, D., Mingqiang, S., Gaomei, Z., Yang, X., Ke, Y., et al. (2016). Sympathetic Stimulation Facilitates Thrombopoiesis by Promoting Megakaryocyte Adhesion, Migration, and Proplatelet Formation. *Blood* 127.

Shimoda, Y., Han, J., Kawada, K., Smaoui, A., and Isoda, H. (20122012). Metabolomics Analysis ofCistus monspeliensisLeaf Extract on Energy Metabolism Activation in Human Intestinal Cells. *J. Biomed. Biotechnol.* 2012, 1–7. doi:10.1155/2012/428514

Simonson, T. S., Yang, Y., Huff, C. D., Yun, H., Qin, G., Witherspoon, D. J., et al. (2010). Genetic Evidence for High-Altitude Adaptation in Tibet. *Science* 329, 72–75. doi:10.1126/science.1189406

Solaini, G., Baracca, A., Lenaz, G., and Sgarbi, G. (2010). Hypoxia and Mitochondrial Oxidative Metabolism. *Biochim. Biophys. Acta (Bba) - Bioenerg.* 1797, 1171–1177. doi:10.1016/j.bbabio.2010.02.011

Storz, J. F. (2021). High-Altitude Adaptation: Mechanistic Insights from Integrated Genomics and Physiology. *Mol. Biol. Evol.* 38, 2677–2691. doi:10.1093/molbev/msab064

Stream, J. O., and Grissom, C. K. (2008). Update on High-Altitude Pulmonary Edema: Pathogenesis, Prevention, and Treatment. *Wilderness Environ. Med.* 19, 293–303. doi:10.1580/07-weme-rev-173.1

Sultan, M., Schulz, M. H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., et al. (2008). A Global View of Gene Activity and Alternative Splicing by Deep Sequencing of the Human Transcriptome. *Science* 321, 956–960. doi:10.1126/science.1160342

Tang, Q., Gu, Y., Zhou, X., Jin, L., Guan, J., Liu, R., et al. (2017). Comparative Transcriptomics of 5 High-Altitude Vertebrates and Their Low-Altitude Relatives. *Gigascience* 6, 1–9. doi:10.1093/gigascience/gix105

Tashi, T., Feng, T., Koul, P., Amaru, R., Hussey, D., Lorenzo, F. R., et al. (2014). High Altitude Genetic Adaptation in Tibetans: No Role of Increased Hemoglobin-Oxygen Affinity. *Blood Cell Mol. Dis.* 53, 27–29. doi:10.1016/j.bcmd.2014.02.003

Thiel, A., Mogel, H., Bruggisser, J., Baumann, A., Wyder, M., Stoffel, M., et al. (2017). Effect of *Clostridium perfringens* β-Toxin on Platelets. *Toxins* 9, 336. doi:10.3390/toxins9100336

Tian, Y. B., He, S. Y., and Ge, C. R. (1998). Gayal. *J. Yellow Cattle Sci.* 6-13, 39.

Touma, M., Kang, X., Zhao, Y., Cass, A. A., Gao, F., Biniwale, R., et al. (2016). Decoding the Long Noncoding RNA during Cardiac Maturation. *Circ. Cardiovasc. Genet.* 9, 395–407. doi:10.1161/circgenetics.115.001363

Trueblood, C. E., Wright, R. M., and Poyton, R. O. (1988). Differential Regulation of the Two Genes Encoding *Saccharomyces cerevisiae* Cytochrome C Oxidase Subunit V by Heme and the HAP2 and REO1 Genes. *Mol. Cel Biol* 8, 4537–4540. doi:10.1128/mcb.8.10.4537-4540.1988

Uzzaman, M. R., Bhuiyan, M. S. A., Edea, Z., and Kim, K.-S. (2014). Semi-domesticated and Irreplaceable Genetic Resource Gayal (*Bos frontalis*) Needs Effective Genetic Conservation in Bangladesh: A Review. *Asian Australas. J. Anim. Sci.* 27, 1368–1372. doi:10.5713/ajas.2014.14159

Vukicevic, S., Helder, M. N., and Luyten, F. P. (1994). Developing Human Lung and Kidney Are Major Sites for Synthesis of Bone Morphogenetic Protein-3 (Osteogenin). *J. Histochem. Cytochem.* 42, 869–875. doi:10.1177/42.7.8014470

Wang, J., Chai, Z., Deng, L., Wang, J., Wang, H., Tang, Y., et al. (2020). Detection and Integrated Analysis of lncRNA and mRNA Relevant to Plateau Adaptation of Yak. *Reprod. Dom Anim.* 55, 1461–1469. doi:10.1111/rda.13767

Wang, K., Yang, Y., Wang, L., Ma, T., Shang, H., Ding, L., et al. (2016). Different Gene Expressions between Cattle and Yak Provide Insights into High-Altitude Adaptation. *Anim. Genet.* 47, 28–35. doi:10.1111/age.12377

Wang, T., Guo, Y., Liu, S., Zhang, C., Cui, T., Ding, K., et al. (2021). KLF4, a Key Regulator of a Transitive Triplet, Acts on the TGF-β Signaling Pathway and Contributes to High-Altitude Adaptation of Tibetan Pigs. *Front. Genet.* 12, 1. doi:10.3389/fgene.2021.628192

Wang, Y., Chen, D.-Q., Chen, M.-Y., Ji, K.-Y., Ma, D.-X., and Zhou, L.-F. (2017). Endothelial Cells by Inactivation of VHL Gene Direct Angiogenesis, Not Vasculogenesis via Twist1 Accumulation Associated with Hemangioblastoma Neovascularization. *Sci. Rep.* 7, 5463. doi:10.1038/s41598-017-05833-9

Waypa, G. B., and Schumacker, P. T. (2002). O2 Sensing in Hypoxic Pulmonary Vasoconstriction: the Mitochondrial Door Re-opens. *Respir. Physiol. Neurobiol.* 132, 81–91. doi:10.1016/s1569-9048(02)00051-4

Wei, Q., and Yu, H. X. (2008). Comparison of Histological Structure of Pulmonary Alveoli between 180 Days Old Yak and plain Cattle. *J. Qinghai Univ. (Natural Sci. Edition)* 1, 36–39.

Winter, H., Mayr, B., Schleger, W., Dworak, E., Krutzler, J., and Kalat, M. (1986). Genetic Characterisation of the Mithun (*Bos frontalis*) and Studies of Spermatogenesis, Blood Groups and Haemoglobins of its Hybrids with *Bos indicus*. *Res. Vet. Sci.* 40, 8–17. doi:10.1016/s0034-5288(18)30479-x

Xiang, K., OuzhuluobuPeng, Y., Peng, Y., Yang, Z., Zhang, X., Cui, C., et al. (2013). Identification of a Tibetan-specific Mutation in the Hypoxic Gene EGLN1 and its Contribution to High-Altitude Adaptation. *Mol. Biol. Evol.* 30, 1889–1898. doi:10.1093/molbev/mst090

Xiao, Y., Li, C., Wang, H., and Liu, Y. (2020). LINC00265 Targets miR-382-5p to Regulate SAT1, VAV3 and Angiogenesis in Osteosarcoma. *aging* 12, 20212–20225. doi:10.18632/aging.103762

Xin, J.-W., Chai, Z.-X., Zhang, C.-F., Zhang, Q., Zhu, Y., Cao, H.-W., et al. (2019). Transcriptome Profiles Revealed the Mechanisms Underlying the Adaptation of Yak to High-Altitude Environments. *Sci. Rep.* 9, 7558. doi:10.1038/s41598-019-43773-8

Yang, Y., Han, L., Yuan, Y., Li, J., Hei, N., and Liang, H. (2014). Gene Co-expression Network Analysis Reveals Common System-Level Properties of Prognostic Genes across Cancer Types. *Nat. Commun.* 5, 3231. doi:10.1038/ncomms4231

Yeh, C.-S., Wang, J.-Y., Cheng, T.-L., Juan, C.-H., Wu, C.-H., and Lin, S.-R. (2006). Fatty Acid Metabolism Pathway Play an Important Role in Carcinogenesis of Human Colorectal Cancers by Microarray-Bioinformatics Analysis. *Cancer Lett.* 233, 297–308. doi:10.1016/j.canlet.2005.03.050

Yi, T., Wu, X., and Li, H. (2020). Ubiquinol-cytochrome C Reductase Core Protein 1 Overexpression Protects H9c2 Cardiac Cells against Mimic Ischemia/reperfusion Injury through PI3K/Akt/GSK-3β Pathway. *Biochem. Biophysical Res. Commun.* 529, 904–909. doi:10.1016/j.bbrc.2020.06.089

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., et al. (2010). Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. *Science* 329, 75–78. doi:10.1126/science.1190371

Zhang, B., Chamba, Y., Shang, P., Wang, Z., Ma, J., Wang, L., et al. (2017a). Comparative Transcriptomic and Proteomic Analyses Provide Insights into the Key Genes Involved in High-Altitude Adaptation in the Tibetan Pig. *Sci. Rep.* 7, 3654. doi:10.1038/s41598-017-03976-3

Zhang, C., Wang, G., Wang, J., Ji, Z., Liu, Z., Pi, X., et al. (2013). Characterization and Comparative Analyses of Muscle Transcriptomes in Dorper and Small-Tailed Han Sheep Using RNA-Seq Technique. *PLoS One* 8, e72686. doi:10.1371/journal.pone.0072686

Zhang, D., Yu, M., Hu, P., Peng, S., Liu, Y., Li, W., et al. (2017b). Genetic Adaptation of Schizothoracine Fish to the Phased Uplifting of the Qinghai-Tibetan Plateau. *G3 (Bethesda)* 7, 1267–1276. doi:10.1534/g3.116.038406

Zhang, W., Fan, Z., Han, E., Hou, R., Zhang, L., Galaverni, M., et al. (2014). Hypoxia Adaptations in the Grey Wolf (*Canis lupus* Chanco) from Qinghai-Tibet Plateau. *Plos Genet.* 10, e1004466. doi:10.1371/journal.pgen.1004466

Zimmermann, F., and Rich, I. N. (1997). Mammalian Homeobox B6 Expression Can Be Correlated with Erythropoietin Production Sites and Erythropoiesis during Development, but Not with Hematopoietic or Nonhematopoietic Stem Cell Populations. *Blood* 89, 2723–2735. doi:10.1182/blood.v89.8.2723

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 in Genetics

ERRATUM
published: 07 March 2022
doi: 10.3389/fgene.2022.868475

Erratum: Comparative Transcriptome Analysis of Gayal (*Bos frontalis*), Yak (*Bos grunniens*), and Cattle (*Bos taurus*) Reveal the High-Altitude Adaptation

Frontiers Production Office *

Frontiers Media SA, Lausanne, Switzerland

Keywords: gayal, yak, differentially expressed genes, co-expression, high-altitude adaptation, hypoxia

OPEN ACCESS

Approved by:
Frontiers Editorial Office, Frontiers Media SA, Switzerland

***Correspondence:**
Frontiers Production Office production.office@frontiersin.org

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *02 February 2022*
Accepted: *02 February 2022*
Published: *07 March 2022*

Citation:
Frontiers Production Office (2022) Erratum: Comparative Transcriptome Analysis of Gayal (Bos frontalis), Yak (Bos grunniens), and Cattle (Bos taurus) Reveal the High-Altitude Adaptation. Front. Genet. 13:868475. doi: 10.3389/fgene.2022.868475

An Erratum on

Comparative Transcriptome Analysis of Gayal (*Bos frontalis*), Yak (*Bos grunniens*), and Cattle (*Bos taurus*) Reveal the High-Altitude Adaptation
by Ma, J., Zhang, T., Wang, W., Chen, Y., Cai, W., Zhu, B., Xu, L., Gao, H., Zhang, L., Li, J., and Gao, X. (2022). Front. Genet. 12:778788. doi:10.3389/fgene.2021.778788

Due to a production error, a version of the article with grammatical mistakes was published.

The publisher apologizes for this mistake. The original version of this article has been updated.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 24 January 2022
doi: 10.3389/fvets.2021.801278

Comprehensive Analysis of Long Non-coding RNA and mRNA Transcriptomes Related to Hypoxia Adaptation in Tibetan Sheep

Zengkui Lu[1,2], *Chao Yuan*[1,2], *Jianye Li*[1,2], *Tingting Guo*[1,2], *Yaojing Yue*[1,2], *Chune Niu*[1,2], *Jianbin Liu*[1,2]* *and Bohui Yang*[1,2]*

[1] Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Sciences, Lanzhou, China, [2] Sheep Breeding Engineering Technology Research Center of Chinese Academy of Agricultural Sciences, Lanzhou, China

Tibetan sheep have lived on the Qinghai-Tibet Plateau for a long time, and after long-term natural selection, they have shown stable genetic adaptability to high-altitude environments. However, little is known about the molecular mechanisms of the long non-coding (lnc)RNAs involved in the adaptation of Tibetan sheep to hypoxia. Here, we collected lung tissues from high-altitude Tibetan sheep and low-altitude Hu sheep for RNA sequencing to study the regulatory mechanisms of the lncRNAs and mRNAs in the adaptation of Tibetan sheep to hypoxia. We identified 254 differentially expressed lncRNAs and 1,502 differentially expressed mRNAs. We found 20 pairs of cis-regulatory relationships between 15 differentially expressed lncRNAs and 14 protein-coding genes and two pairs of trans-regulatory relationships between two differentially expressed lncRNAs and two protein-coding genes. These differentially expressed mRNAs and lncRNA target genes were mainly enriched in pathways related to lipid metabolism and immune function. Interaction network analysis showed that 17 differentially expressed lncRNAs and 15 differentially expressed mRNAs had an interactive relationship. Additionally, we used six differentially expressed lncRNAs and mRNAs to verify the accuracy of the sequencing data via qRT-PCR. Our results provide a comprehensive overview of the expression patterns of the lncRNAs and mRNAs involved in the adaptation of Tibetan sheep to hypoxia, laying a foundation for further analysis of the adaptations of plateau animals.

Keywords: Tibetan sheep, lung, hypoxic adaptation, long non-coding RNAs, transcriptome

OPEN ACCESS

Edited by:
Turgay Unver,
FicusBio, Turkey

Reviewed by:
Xin Qi,
Ocean University of China, China
Qianjun Zhao,
Institute of Animal Sciences, Chinese Academy of Agricultural Sciences (CAAS), China

***Correspondence:**
Jianbin Liu
liujianbin@caas.cn
Bohui Yang
yangbohui@caas.cn

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Veterinary Science

Received: *25 October 2021*
Accepted: *20 December 2021*
Published: *24 January 2022*

Citation:
Lu Z, Yuan C, Li J, Guo T, Yue Y, Niu C, Liu J and Yang B (2022) Comprehensive Analysis of Long Non-coding RNA and mRNA Transcriptomes Related to Hypoxia Adaptation in Tibetan Sheep. Front. Vet. Sci. 8:801278. doi: 10.3389/fvets.2021.801278

INTRODUCTION

The Qinghai-Tibet Plateau is the highest plateau worldwide. It is known as the "Third Pole" of the earth and is known for its low oxygen, low temperatures, and strong ultraviolet radiation. Along with human migration and settlement, many domestic animals have reproduced for generations in these harsh living conditions and have thus adapted unique and distinctive characteristics at the morphological, physiological and genetic levels (1–3). Scientists have used large-scale omics data to reveal the adaptive genetic mechanisms of domestic animals on the Qinghai-Tibet Plateau

and identified a number of plateau-adaptive candidate genes, including *EPAS1* (endothelial PAS domain-containing protein 1), *EGLN1* (Egl-9 homolog 1) and *VEGF* (vascular endothelial growth factor), in the HIF (hypoxia inducible factor) hypoxia-inducing pathway (4–7).

Long non-coding RNA (lncRNA) is non-coding RNA that contains >200 nucleotides. LncRNA has important roles in many life activities such as the dosage compensation effect, epigenetic regulation, cell cycle regulation and cell differentiation regulation (8, 9). Studies have shown that lncRNA can bind to HIF-1α and activate its expression, thereby playing an important role in hypoxia-induced tumor cells (10–14). LncRNA may also play an important role in the adaptability of domestic animals to plateaus. Recent studies have shown that lncRNA is involved in the adaptability of Tibetan chickens and yaks to plateaus; however, the regulatory mechanisms of this lncRNA remain largely unknown (15, 16).

To reveal the potential role of lncRNA in the adaptability of domestic animals to plateaus, we selected sheep from different altitudes and collected lung tissue for transcriptomic sequencing. We used an integrated analysis of mRNA and lncRNA data to explore the molecular mechanisms of Tibetan sheep's adaptability to plateaus, find the genes and regulatory pathways related to this adaptability, and provide a theoretical basis for Tibetan sheep production and breeding.

MATERIALS AND METHODS

Sample Collection

We collected lung tissues from six healthy 18-month-old rams, of which, three were from the Qinghai-Tibet Plateau (Zashijia sheep, Qumalai County, Yushu Tibetan Autonomous Prefecture, Qinghai, China, altitude ∼4,800 m), and three were from the plains of China (Hu sheep, Minqin County, Wuwei, Gansu, China, altitude ∼1,400 m). All six sheep grazed and received supplemented feed and were euthanized by an intravenous injection of phenobarbital solution (Fatal-Plus, 10 mg/kg body weight, Vortech Pharmaceuticals, MI, USA). Tissues were collected from the middle lobe of the right lung and stored in liquid nitrogen, then used for RNA extraction.

RNA Extraction, Library Construction, and Sequencing

We used TRIzol reagent (Invitrogen, Carlsbad, CA, USA) to extract the total RNA and remove rRNA per the manufacturer's instructions. RNA samples that qualified through quality inspection were used for library construction in accordance with the instructions provided by Illumina (Illumina, CA, USA). The RNA was randomly interrupted into short 200–500-nt fragments, and the first cDNA strand was synthesized using random primers. Buffer, dNTPs (dUTP instead of dTTP), RNase H and DNA polymerase I were added to synthesize the second cDNA strand. The cDNA was purified using the QiaQuick PCR kit, then EB buffer was added to elute the proteins, the ends were repaired, and base A and sequencing adapters were added. Uracil-N-glycosylase was used to degrade the second chain. Agarose gel electrophoresis was used to select the appropriate fragments for PCR amplification. Qualified libraries were sequenced using the HiSeq^{TM} 4000 platform.

Raw Data Filtering, Comparison, and Splicing

After sequencing, the raw data (raw reads) were obtained, and fastp software was used for quality control to obtain clean reads. Quality control was performed by removing reads containing adapters, reads with an N content > 10%, reads with only A bases, and low-quality reads (the number of bases with a quality value of $Q \leq 20$ accounted for the total number of reads above 50). We used Bowtie2 software to compare clean reads to the rRNA database and remove the ribosomal reads. We used HISAT2 software to compare the clean reads to the sheep reference genome (Oar_v4.0), set the parameter to -rna-strandness RF, and set the remaining parameters to the default values. We used Stringtie software to assemble the clean reads compared to the sheep reference genome and used Cuffmerge software to merge each transcript. We used Cuffcompare software to compare the assembled transcripts with known transcript types (NCBI Refseq, Ensembl transcripts, UCSC), identify known mRNAs, and screen new non-coding transcripts (novel ncRNA).

lncRNA Identification and Target Gene Prediction

To screen for new non-coding transcripts (novel ncRNA), we filtered transcripts with ≥200 bp and ≥2 exons. We used CPC2, CNCI and Pfam software to predict the coding ability of the new transcripts and took the intersection of the transcripts without coding potential as candidate lncRNAs. The lncRNA target genes included cis-target and trans-target genes. mRNA 50 kb upstream and downstream of the lncRNA were used as the cis-target, and mRNA whose expression correlation coefficient was >0.9 was used as the trans-target.

Identification of Differentially Expressed mRNAs and lncRNAs

For the FPKM value, we used StringTie software to calculate the mRNA and lncRNA expression levels, with the equation FPKM = total fragments/mapped reads (millions) × exon length (kB). We used DESeq2 software to analyze the differences between the mRNA and lncRNA, standardize the reads counts, calculate the hypothesis test probability (*p*-value), and perform multiple hypothesis test corrections to obtain the false discovery rate (FDR). From the difference analysis, $P < 0.05$ and a fold change > 1.5 were used to screen significantly differentially expressed mRNA and lncRNA.

Functional Enrichment Analysis of Differentially Expressed Genes

We mapped the differential genes to each term in the gene ontology (GO) database (http://www.geneontology.org/), calculated the number of differential genes in each term, and obtained a list of differential genes with specific GO functions. The hypergeometric distribution test was used to calculate the *P*-value for the significantly enriched GO functions, then the

P-value was corrected using the Benjamini-Hochberg multiple test (FDR). GO items with FDR $\leq$ 0.05 were considered significantly enriched. We used the Kyoto Encyclopedia of Genes and Genomes (KEGG) database to annotate and classify the differentially expressed genes with pathway functions. The KEGG pathway functional enrichment method is similar to the GO functional enrichment analysis.

Construction of the lncRNA-mRNA Network

We used the interaction relationship in the STRING protein interaction database (http://string-db.org) to analyze the differential gene interaction network. The differential gene set was extracted from the database, and the interaction network diagram was constructed and visualized using Cytoscape software.

Validation of lncRNA and mRNA Expression by qRT-PCR

To test the accuracy of the sequencing results, we selected six differentially expressed mRNA and lncRNA for qRT-PCR verification. We used a cDNA synthesis kit to reverse transcribe the extracted total RNA into cDNA. We used Oligo7 software to design primers and perform specific detections in NCBI. We used the TransStart Green qPCR SuperMix and LightCycler 480 II instruments to perform qRT-PCR and ran each sample in triplicate to ensure the accuracy of the quantitative results. The $2^{-\Delta\Delta Ct}$ method was used to calculate the relative expression of the target genes, and *ACTB* (beta-actin) was used as the internal reference gene. **Supplementary Table 1** lists the primers used in this study.

Statistical Analysis

All data are presented as the mean $\pm$ standard deviation and were analyzed using Student's t-test in SPSS software. $p < 0.05$ was considered statistically significant.

RESULTS

Overview of the Sequencing Data

Using the Illumina HiSeq™ 4000 platform, the six constructed libraries were paired-end sequenced. **Table 1** shows the data quality control and genome comparison results. The six libraries produced 499,954,984 raw reads; 498,979,644 clean reads remained after quality control. The average rate of the clean reads was 99.81%. The GC content of the six samples was 47.57–49.25%, which was consistent with the base composition law of Q20 $\geq$ 97.63% and Q30 $\geq$ 93.17%. We used HISAT2 software to compare the clean reads to the sheep reference genome. Approximately 92.47% of the reads could be accurately compared, and the match rate was high. Only data aligned to the sheep reference genome were used for further bioinformatics analysis.

lncRNA and mRNA Feature Analysis

An average of 17,904 expressed genes and 5,615 lncRNAs were identified from the six libraries. Using CNCI, CPC2, and Pfam to predict the coding ability of the new transcripts yielded 1,740 new lncRNAs (**Figure 1A**), including 357 sense, 89 antisense, 31 intronic, 37 bidirectional, 941 intergenic, and 285 other lncRNAs (**Figure 1B**). Most lncRNAs had two or three exons, which was significantly less than the number of exons in the mRNA (**Figure 1C**). The length distribution range of the lncRNA and mRNA was basically the same, but the lncRNA was longer than the mRNA (**Figure 1D**). The lncRNA expression level was lower than that of the mRNA (**Figure 1E**).

Identification and Analysis of Differentially Expressed lncRNA and mRNA

We identified 254 differentially expressed lncRNAs between the Tibetan and Hu sheep, of which, 123 were upregulated, and 131 were downregulated (**Figures 2A,C**; **Supplementary Table 2**). The five lncRNAs with the most significant differential expression were MSTRG.19949.1, MSTRG.6422.3, MSTRG.6796.1, MSTRG.16435.3, and MSTRG.19804.1. We identified 1502 differentially expressed mRNAs between the Tibetan and Hu sheep, of which 469 were upregulated and 1,033 were downregulated (**Figures 2B,D**; **Supplementary Table 3**). The five most significantly differentially expressed mRNAs were *NKIRAS2* (NK-κB inhibitor-interacting Ras-like 2), *SEPW1* (selenoprotein W), *PDZK1* (PDZ domain-containing 1), *PEAK1* (pseudopodium enriched atypical kinase 1), and *KIAA1549*. Among these, *NKIRAS2* is involved in immune response, *SEPW1* is involved in oxidative stress, and *PDZK1* is involved in fat metabolism. Additionally, the differentially expressed lncRNA and mRNA were combined into one group, and the differentially upregulated and downregulated lncRNA and mRNA were combined into one group (**Figures 2E,F**). The differentially expressed lncRNA and mRNA had good reproducibility in the groups.

Prediction of lncRNA Target Genes and Analysis of lncRNA-mRNA Interaction

The 254 differentially expressed lncRNAs were predicted for cis- and trans-target genes (**Table 2**). We found 20 pairs of cis-regulatory relationships between 15 differentially expressed lncRNAs and 14 protein-coding genes, of which, three lncRNAs were upstream of the target gene, and 17 lncRNAs were downstream. There were also two pairs of trans-regulatory relationships between the two differentially expressed lncRNAs and the two protein-coding genes. Among these target genes, *SH2D7* (SH2 domain containing) and *IL12A* (interleukin-12 A) are involved in immune response, *LPAR1* (lysophosphatidic acid receptor 1) and *FKBP10* (FK506 binding protein 10) are involved in signal transduction, *STEAP4* (six-transmembrane epithelial antigen of prostate 4) is involved in oxidative stress, *KCNG3* (voltage-gated channel subfamily G member 3) is involved in ion transport, and *PHACTR1* (phosphatase and actin regulator 1) is involved in pulse pressure regulation.

We constructed an interaction network based on these differentially expressed lncRNAs and mRNAs, and 17 differentially expressed lncRNAs and 15 differentially expressed

TABLE 1 | Summary of sequenced RNA-seq data.

Items	HS1	HS2	HS3	TS1	TS2	TS3
Raw datas	81,317,688	81,303,748	81,256,760	79,261,536	88,127,470	88,687,782
Clean datas	81,181,820 (99.83%)	81,140,622 (99.80%)	81,085,206 (99.79%)	79,104,972 (99.80%)	87,948,784 (99.80%)	88,518,240 (99.81%)
GC content	49.25%	47.57%	47.58%	47.96%	48.22%	47.74%
Q20 (%)	97.87%	97.68%	97.63%	98.56%	98.55%	98.57%
Q30 (%)	93.63%	93.29%	93.17%	95.38%	95.37%	95.40%
Multiple mapped	2,698,512 (3.32%)	2,282,439 (2.81%)	2,239,924 (2.76%)	2,357,259 (2.98%)	2,885,038 (3.28%)	2,691,472 (3.04%)
Unique mapped	72,640,831 (89.50%)	72,720,281 (89.65%)	72,944,114 (90.01%)	71,800,901 (90.80%)	79,348,585 (90.26%)	80,450,762 (90.92%)
Unmapped	5,820,851 (7.17%)	6,108,764 (7.53%)	5,856,384 (7.23%)	4,917,942 (6.22%)	5,678,065 (6.46%)	5,344,200 (6.04%)
Total mapped	75,339,343 (92.83%)	75,002,720 (92.47%)	75,184,038 (92.77%)	74,158,160 (93.78%)	82,233,623 (93.54%)	83,142,234 (93.96%)

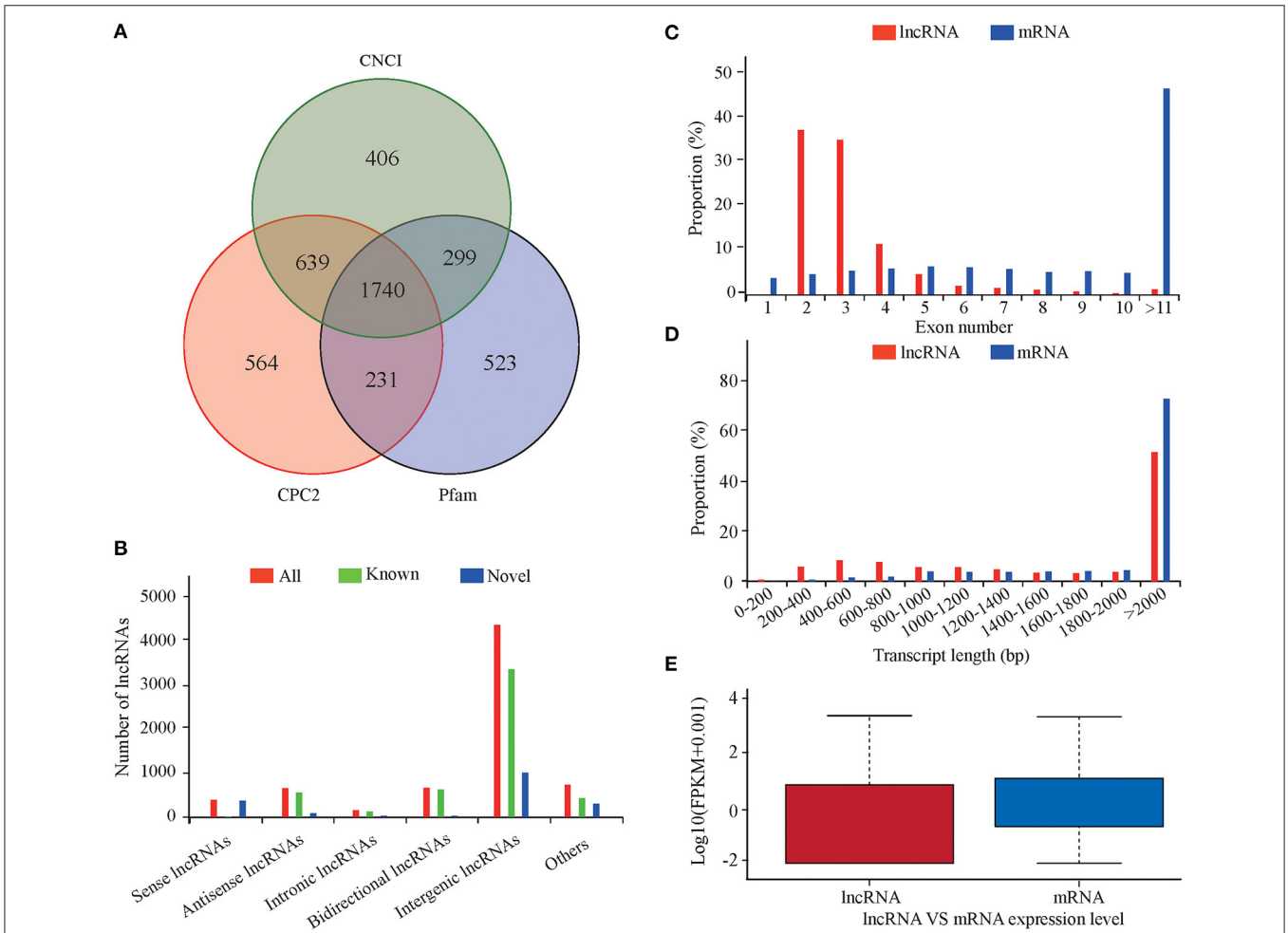


FIGURE 1 | Characteristics of lncRNAs and mRNAs in the lung tissues of sheep at different altitudes. **(A)** Screening of candidate lncRNAs by CNCI and CPC2. **(B)** Data on lncRNA transcript types. **(C)** Distribution of exon numbers in lncRNAs and mRNAs. **(D)** Length distribution of lncRNAs and mRNAs. **(E)** Expression levels of lncRNAs and mRNAs.

mRNAs had an interaction relationship (**Figure 3**). Among the differentially expressed lncRNAs, MSTRG.10206.4, MSTRG.10206.3, XR_001435094.1, MSTRG.10206.1 and XR_001022162.2 interacted with two mRNAs, and the rest of the lncRNAs interacted with only one mRNA. Among the differentially expressed mRNAs, MSTRG.10204 interacted with four lncRNAs, MSTRG.10208 interacted with four lncRNAs, and the remaining mRNAs interacted with one lncRNA. Among these, there were more cis-interaction relationships than trans-interaction relations.

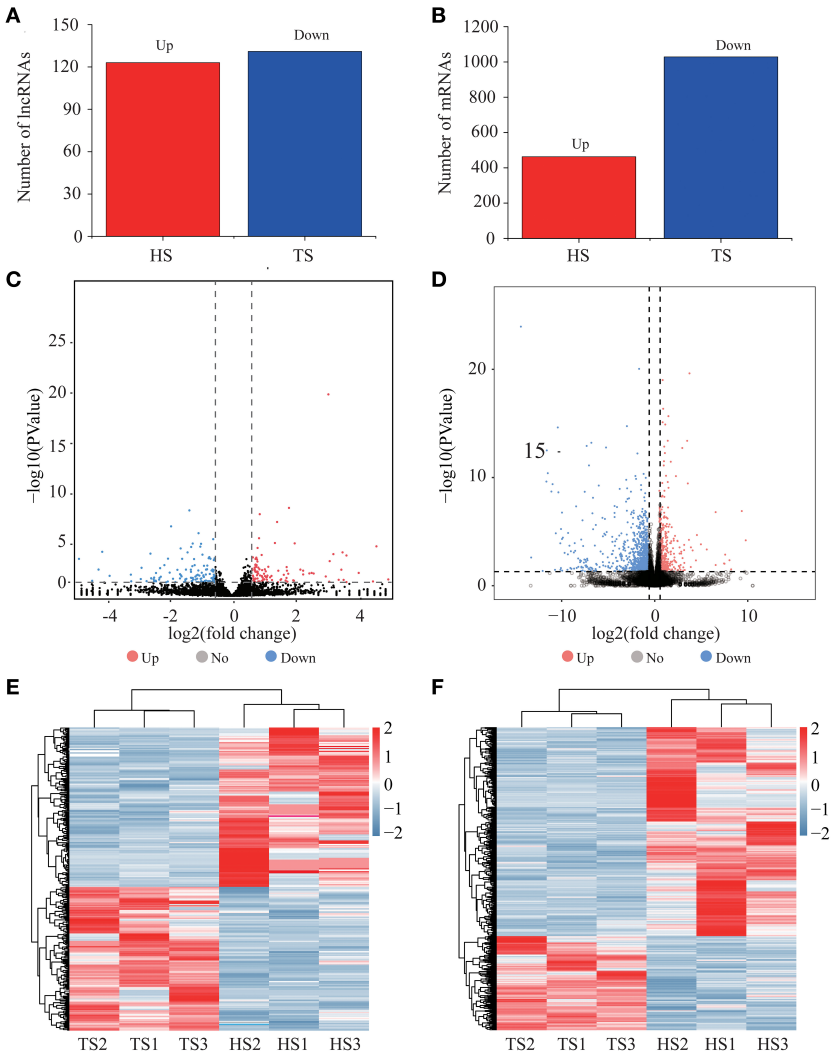


FIGURE 2 | Expression analysis of lncRNAs and mRNAs. **(A)** Numbers of upregulated and downregulated differentially expressed lncRNAs. **(B)** Numbers of upregulated and downregulated differentially expressed mRNAs. **(C)** Volcano plots displaying differentially expressed lncRNAs. **(D)** Volcano plots displaying differentially expressed mRNAs. **(E)** Heatmap of differentially expressed lncRNAs. **(F)** Heatmap of differentially expressed mRNAs.

GO and KEGG Enrichment Analyses

The target genes of the differentially expressed lncRNAs were significantly associated with 38 GO terms, most of which were related to material metabolic processes, including ferric iron transport, trivalent inorganic cation transport, and sulfate transport (**Figure 4A**; **Supplementary Table 4**). In KEGG analysis, there were 18 significantly enriched pathways, including the RIG-I-like receptor signaling pathway, taste transduction, and the Toll-like receptor signaling pathway (**Figure 4B**; **Supplementary Table 5**). Differentially expressed mRNA was significantly enriched in 215 GO terms, most of which were related to immune function and stress response (**Figure 4C**; **Supplementary Table 6**). KEGG analysis was significantly enriched in 37 pathways, including the PPAR signaling pathway involved in fat metabolism and the NF-κβ, NOD-like receptor and Toll-like receptor signaling pathways involved in immune recognition and response (**Figure 4D**; **Supplementary Table 7**).

qRT-PCR Validation of Differential lncRNA and mRNA Expression

Six differentially expressed lncRNAs and mRNAs were selected for qRT-PCR verification (**Figure 5**). The qRT-PCR and RNA-seq results showed similar expression trends, confirming the reliability of the transcriptomic sequencing data.

DISCUSSION

Tibetan sheep have undergone natural selection for tens of thousands of years in the low-oxygen environment of the Qinghai-Tibet Plateau. Specifically, they have developed cardiopulmonary functions and strong low-oxygen adaptability.

TABLE 2 | Differentially expressed lncRNAs and their targeted mRNAs.

Type	lncRNA ID	chr	Gene ID	Up/down stream	Symbol
cis	XR_001026700.1	NC_019459.2	ncbi_443346	Upstream	LPAR1
	MSTRG.2514.1	NC_019459.2	MSTRG.2515	Downstream	–
	MSTRG.4734.2	NC_019460.2	ncbi_101115443	Downstream	KCNG3
	MSTRG.5378.1	NC_019460.2	ncbi_101119634	Downstream	SLC26A10
	MSTRG.6522.1	NC_019461.2	ncbi_101111187	Downstream	STEAP4
	MSTRG.10206.1	NC_019466.2	MSTRG.10204	Downstream	–
	MSTRG.10206.1	NC_019466.2	MSTRG.10208	Downstream	–
	XR_001435094.1	NC_019466.2	MSTRG.10204	Downstream	–
	XR_001435094.1	NC_019466.2	MSTRG.10208	Downstream	–
	MSTRG.10206.3	NC_019466.2	MSTRG.10204	Downstream	–
	MSTRG.10206.3	NC_019466.2	MSTRG.10208	Downstream	–
	MSTRG.10206.4	NC_019466.2	MSTRG.10204	Downstream	–
	MSTRG.10206.4	NC_019466.2	MSTRG.10208	Downstream	–
	MSTRG.10898.1	NC_019467.2	ncbi_101107613	Upstream	LOC101107613
	XR_001435802.1	NC_019472.2	ncbi_101118897	Downstream	LOC101118897
	MSTRG.16852.1	NC_019475.2	ncbi_101114647	Downstream	SH2D7
	XR_001022162.2	NC_019475.2	ncbi_105603299	Upstream	LOC105603299
	XR_001022162.2	NC_019475.2	ncbi_105601883	Downstream	LOC105601883
	MSTRG.17316.1	NC_019476.2	ncbi_101111131	Downstream	EXOG
	MSTRG.18414.1	NC_019477.2	ncbi_101115323	Downstream	PHACTR1
antisence	MSTRG.11948.1	NC_019468.2	ncbi 101108933	DOWNSTREAM	FKBP10
	MSTRG.1821.1	NC_019458.2	ncbi 443064	DOWNSTREAM	IL12A

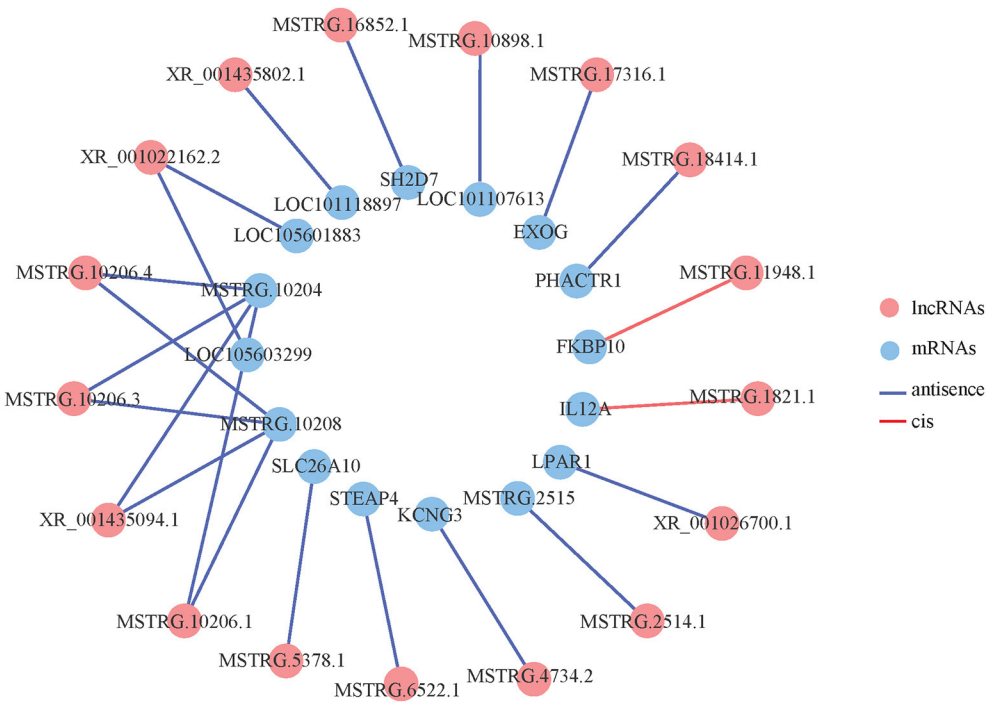


FIGURE 3 | Co-expression network of differentially expressed lncRNAs and differentially expressed mRNAs.

Under hypoxic stress, different areas of the body show different response mechanisms, forming a unique hypoxia adaptation strategy. In-depth study of the genetic mechanisms of Tibetan sheep in response to high-altitude hypoxic environmental stress can provide a theoretical basis for solving problems with animals entering Tibet, adapting to the high-altitude hypoxic

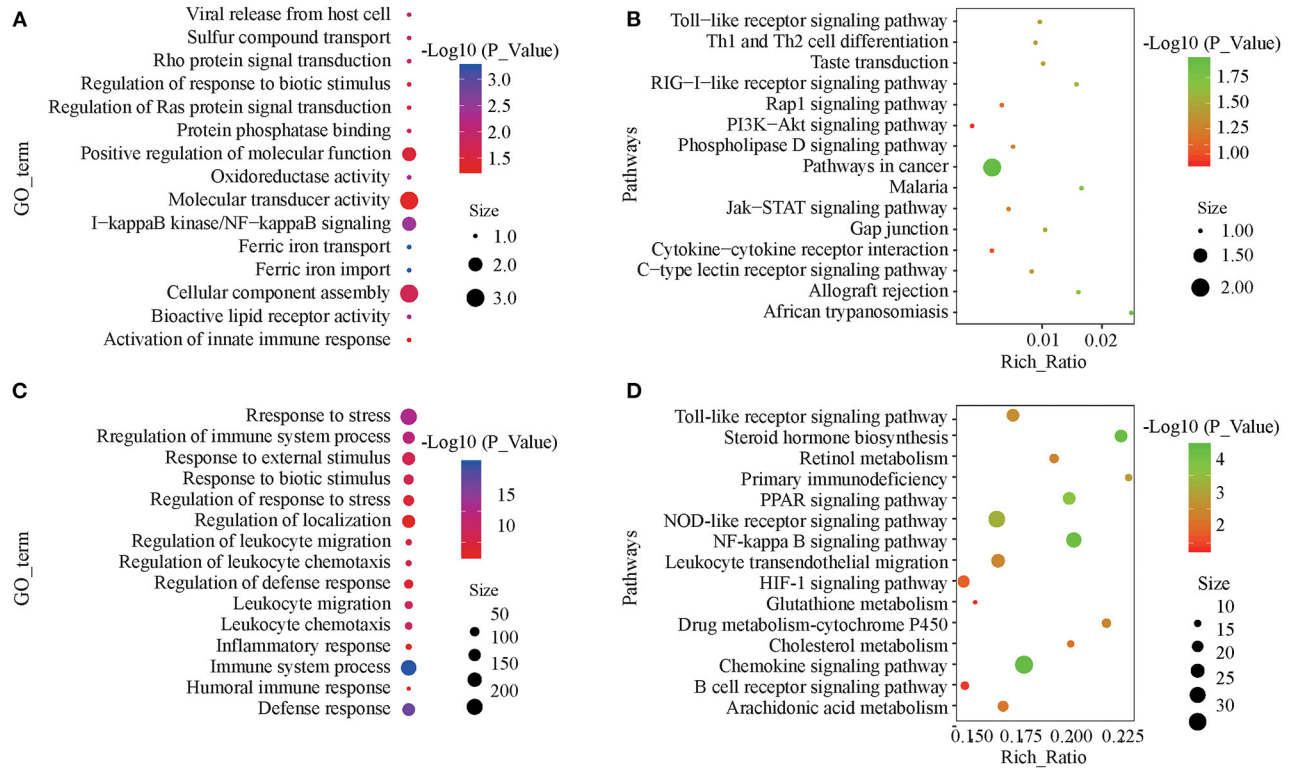


FIGURE 4 | GO and KEGG enrichment analyses of differentially expressed lncRNAs and mRNAs. **(A)** Annotated GO terms of target genes of differentially expressed lncRNAs. **(B)** Enriched KEGG pathways of target genes of differentially expressed lncRNAs. **(C)** Annotated GO terms of differentially expressed mRNAs. **(D)** Enriched KEGG pathways of differentially expressed mRNAs.

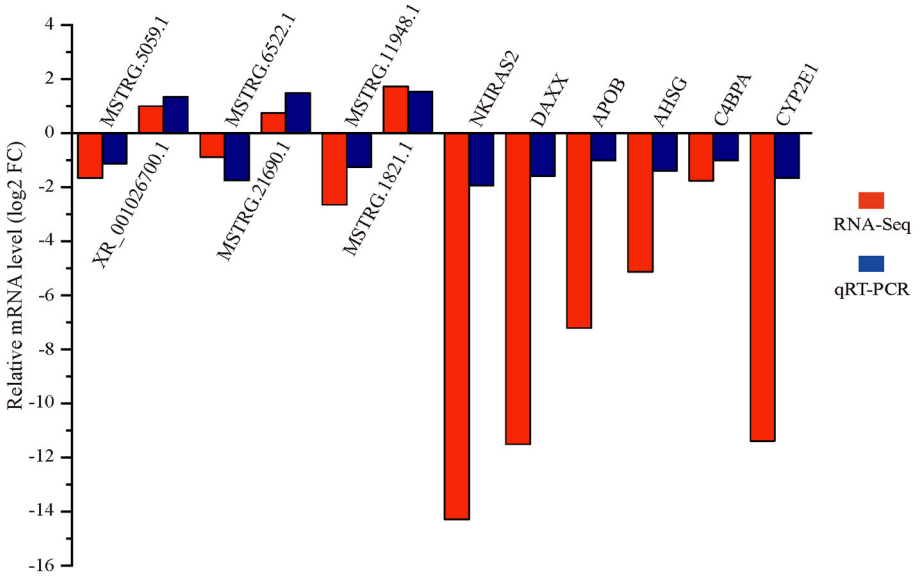


FIGURE 5 | Verification of differentially expressed lncRNAs and mRNAs by qRT-PCR.

environment and maintaining normal production performance in plains areas. Studying these animals can also help reveal their excellent genetic resources. Thus, we analyzed the complex life processes of Tibetan sheep and the phenotype that allows them to adapt to high-altitude hypoxic conditions at the transcriptomic level. Screening and regulating the genes related

to hypoxia adaptation in Tibetan sheep will enable constructing a transcriptional expression regulatory network to analyze the genetic mechanisms of hypoxia adaptation in Tibetan sheep. GO and KEGG analyses of differentially expressed genes revealed the significant enrichment of pathways such as energy metabolism, immune response, oxidative stress, digestion and metabolism, and body temperature regulation. Strong selection on these pathways has also been performed in the study of adaptability of Tibetan pigs and yaks to high-altitude hypoxia (17–19).

Studies have shown that high-altitude hypoxic environments have an important impact on energy metabolism in animals (15, 20, 21). Similarly, we found many differentially expressed genes and signaling pathways related to glucose and lipid metabolism. Among the most significant top 20 differentially expressed genes, five were related to lipid metabolism (*PDZK1*, *APOB*, *AHSG*, *DAXX*, and *C4BPA*) (22–26). Additionally, in the KEGG enrichment analysis, the PPAR signaling pathway related to fat metabolism was significantly enriched. Studies on the high-altitude adaptability of Tibetan pigs have also found the significant enrichment of energy metabolism pathways. Fat is the main form of energy storage in animals' bodies. Under hypoxic conditions, the body's fat metabolism changes significantly, and the body intentionally increases its lipid oxidation and phosphorylated ATP synthesis levels with increased heat production to adapt to alpine environments (27–29). Similarly, some genes involved in fat metabolism, such as LPAR1, is also found in lncRNA target genes (30). The PPAR gene is a candidate gene for adaptation to high-altitude hypoxia. It participates in and is regulated by the HIF pathway and is closely related to the production of ATP in animals (31, 32). The activated PPAR gene can enhance fatty acid oxidation and upregulate the mitochondrial β-oxidation process. Therefore, under hypoxic conditions, Tibetan sheep can adapt to the plateau environment by increasing energy metabolism. In addition, there are certain differences in the structure of alveoli among animals inhabiting different altitudes. Type II alveoli are essential for normal lung function and regeneration after hypoxic injury. Respiration depends on surfactants produced and secreted by type II alveoli, mainly composed of phospholipids (33). Some FGF family members (*FGF11*, *FGF14*, and *FGF4*) were included among the differentially expressed genes. Studies have shown that FGF can stimulate the proliferation of type II alveolar cells, which in turn affects lipid homeostasis on the alveolar surface (34). This may also be caused by the enrichment of differential genes in fat metabolism.

The immune system is the body's defense barrier, and exposure to high-altitude and hypoxic environments can change the body's immune functions (35). At an altitude of 3,000 m, the number of circulating dendritic cells in human plasma is significantly reduced (36). At an altitude of 5,000 m, plasmacytoid dendritic cells in human plasma are reduced, and the TNF-α and IL-6 contents are significantly increased (37). We found that several genes, including NKIRAS2, IKBKE, and TRIM7, were involved in immunoregulation and were lncRNA target genes (38–40). Additionally, some immune-related signaling pathways, such as the NF-κβ, Toll-like receptor, and B-cell-receptor signaling pathways, were also enriched. NF-κβ is an important immune response regulator involved in innate immunity and adaptive immune responses. The differentially expressed genes identified in this study, including *NKIRAS2*, *IKBKE* (inhibitor of nuclear factor kappa-B kinase subunit epsilon) and *TRIM7* (tripartite motif containing 7), are key factors in the NF-κβ signaling pathway. After recognizing pathogenic microorganisms, Toll-like receptors activate NF-κβ through the MyD88 or TRIF-dependent signaling pathways, then initiate innate and adaptive immunity (41, 42). After activation, NF-κB regulates the expression of cytokines such as IL-1, IL-6, and IL-8 and releases them outside the cell to exert an early immune response effect (43, 44). This study found that the expression levels of genes such as *NF-κB2*, *TLR2* (toll-like receptor 2), *IL1A*, *IL1R2*, and *IL18* were significantly lower in Tibetan sheep than in Hu sheep. The downregulation of *TLR2* expression in the lungs of Tibetan sheep may reduce the binding of *IL1A*, *IL1R2*, and *IL18*, thereby reducing downstream TNF-α activity (45). Decreased TNF-α activity may weaken NF-κB-induced kinase activity (NIK). Correspondingly, the decrease in NIK activity may lead to a decrease in the kinase activity inhibited by NF-κB, and the inhibition of NF-κB kinase activity leads to a decrease in the activity of the I-κB kinase (IKK) complex (46). This cascade reaction attenuates the phosphorylation of NF-κB complex inhibitor I-κB, thereby slowing its dissociation from NF-κB and reducing the level of NF-κB, which may also cause the downregulation of TNF-α expression in the lungs of Tibetan sheep (47). Our results indicate that the NF-κβ signaling pathway and its key factors play important roles in the immunoregulation of Tibetan sheep.

Animals living in high-altitude areas for a long time, under low pressure, low oxygen, low temperature and strong ultraviolet stimulation, will experience varying degrees of oxidative stress and produce large amounts of reactive oxygen and nitrogen species (48, 49). The selenoprotein, *SEPW1*, has glutathione-dependent antioxidant activity (50). *GPX1* (glutathione peroxidase-1) is a selenoprotein that can oxidize glutathione, remove excess hydroxyl free radicals and peroxides, and protect cells from oxidative stress (51). The lncRNA target gene *STEAP4* is an iron-copper oxidoreductase that uses the reducing power provided by NADPH to reduce Fe^{3+} and Cu^{2+} to Fe^{2+} and Cu^{+}, which is essential for maintaining many cellular processes (52). Tibetan sheep have been living in a hypoxic environment for many years, necessitating physiological and genetic adaptation to oxidative stress. Therefore, compared with Tibetan sheep, Hu sheep are more sensitive to oxidative stress and have higher antioxidant activity. Compared with the level in Hu sheep, *GPX1* is expressed at a significantly lower level in Tibetan sheep, which is consistent with the lower GPX activity of Tibetans living on plateaus (53). In contrast, we found that the expression of *CAT*, a gene related to antioxidant enzyme activity, was significantly higher in Tibetan sheep than in Hu sheep.

Compared with Hu sheep, Tibetan sheep have always been under traditional grazing management and have a strong ability to adapt to the low-oxygen environment of the plateau. The Qinghai-Tibet Plateau is extremely deficient in

forage in winter and spring, and Tibetan sheep can maintain normal reproduction even if their nutritional intake is severely insufficient, indicating that Tibetan sheep have developed unique digestive and metabolic functions. We found two related genes involved in bile acid metabolism, *SLC26A10* and *IL12A*, among the lncRNA target genes. *SLC26A10*, one of the main carriers involved in the enterohepatic circulation of bile acids, ingests plasma-bound bile salts into liver cells in a sodium-dependent manner (54, 55). *IL12A* has an immunomodulatory effect, is significantly involved in the susceptibility to primary biliary cholangitis, and can be used as a molecular target for clinical diagnoses (56, 57). Bile acid is an important component of bile. A compound with an amphiphilic molecular structure produced by cholesterol metabolism in the liver plays an important role in lipid digestion, absorption and metabolism. This suggests that Tibetan sheep can digest and absorb lipids rationally through bile acids as a defense against the harm caused by the high-altitude hypoxic environment.

Tibetan sheep have strong cold tolerance and can grow and develop normally on the Qinghai-Tibet Plateau more than 3,000 m above sea level. A series of genes involved in the regulation of body temperature were discovered in this study, such as *UCP3* (uncoupling protein 3) and *HTR4* (5-hydroxytryptamine receptor 4). Studies have shown that *UCP3* can significantly increase the oxidative respiration rate of Tibetan pig fat cells, indicating that it is an important gene in enabling Tibetan pigs to resist cold and heat (58). Another study also showed that the cold-resistance mechanism of pigs does not simply involve muscle tremor and fever, but activation of the *UCP3* protein (59). By upregulating the expression of the *UCP3* gene, Tibetan pigs promote the browning of subcutaneous white fat, increase fat production, and maintain body temperature balance. Moreover, the expression level of *UCP3* in Tibetan sheep was shown to be significantly higher than that in Hu sheep, indicating that Tibetan sheep can also mediate subcutaneous fat browning and increase the body's heat production through *UCP3*. Studies have shown that 5-HT participates in the regulation of animal body temperature (60). When an animal is in a high-temperature environment, the 5-HT neuroendocrine system is activated. 5-HT binds to 5-HTR and activates the downstream cAMP and cGMP signaling pathways, resulting in increased activity of heat-sensitive nerves and then the inhibition of heat production (61). In addition to participating in body temperature regulation, *HTR4* also plays an important role in animal gastrointestinal sensitivity and food intake (62). In contrast to the case of Hu sheep, the Qinghai-Tibet Plateau forages that Tibetan sheep live on are characterized by high fiber content, meaning that these animals have a long history of nutritional stress. These factors affect the food intake and ruminant activities of Tibetan sheep, and then affect the calories produced by metabolism to maintain body temperature stability. During this period, *HTR4* may have played an important regulatory role.

CONCLUSION

Here, we systematically identified the expression profiles of lncRNA and mRNA involved in the adaptation process of Tibetan sheep to hypoxia. Functional enrichment and interaction network analysis results showed that lncRNA and mRNA may participate in the adaptation of Tibetan sheep to hypoxia via lipid metabolism. These results provide valuable resources for studying lncRNA and mRNA involved in the adaptation of animals to plateau hypoxia and can help clarify the molecular mechanisms of the adaptation of animals to plateaus.

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**.

ETHICS STATEMENT

The animal study was reviewed and approved by Institutional Animal Care and Use Committee of Lanzhou Institute of Husbandry and Pharmaceutical Science of Chinese Academy of Agricultural Sciences (Approval No. NKMYD201805; Approval Date: 18 October 2018).

AUTHOR CONTRIBUTIONS

ZL, JLiu, and BY conceived and designed the study. CY, JLi, TG, CN, and YY collected the samples. ZL performed the experiments, analyzed the data, and wrote the paper. ZL, JLi, and BY contributed to revisions of the manuscript. All authors read and approved the manuscript.

FUNDING

This research work was supported by the Chinese Academy of Agricultural Sciences of Technology Innovation Project (25-LZIHPS-07 and CAAS-ZDRW202106), the National Natural Science Foundation of China (32002170), and the Special Fund of the Chinese Central Government for Basic Scientific Research Operations in Commonweal Research Institutes (1610322020004).

ACKNOWLEDGMENTS

We thank the Guangzhou Genedenovo Biotechnology Co., Ltd. for assisting in RNA sequencing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2021.801278/full#supplementary-material

REFERENCES

1. York JM, Scadeng M, McCracken KG, Milsom WK. Respiratory mechanics and morphology of Tibetan and Andean high-altitude geese with divergent life histories. *J Exp Biol.* (2018) 221:jeb170738. doi: 10.1242/jeb.170738
2. Storz JF, Runck AM, Sabatino SJ, Kelly JK, Ferrand N, Moriyama H, et al. Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. *Proc Natl Acad Sci USA*. (2009) 106:14450–5. doi: 10.1073/pnas.0905224106
3. Yang J, Jin Z-B, Chen J, Huang X-F, Li X-M, Liang Y-B, et al. Genetic signatures of high-altitude adaptation in Tibetans. *Proc Natl Acad Sci USA*. (2017) 114:4189–94. doi: 10.1073/pnas.1617042114
4. Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, et al. Genetic evidence for high-altitude adaptation in Tibet. *Science.* (2010) 329:72–5. doi: 10.1126/science.1189406
5. Liu X, Zhang Y, Li Y, Pan J, Wang D, Chen W, et al. EPAS1 gain-of-function mutation contributes to high-altitude adaptation in Tibetan horses. *Mol Biol Evol.* (2019) 36:2591–603. doi: 10.1093/molbev/msz158
6. Lorenzo FR, Huff C, Myllymäki M, Olenchock B, Swierczek S, Tashi T, et al. A genetic mechanism for Tibetan high-altitude adaptation. *Nat Genet.* (2014) 46:951–6. doi: 10.1038/ng.3067
7. Cai W, Liu S, Liu Z, Hou S, Lv Q, Cui H, et al. Downregulation of lung miR-203a-3p expression by high-altitude hypoxia enhances VEGF/Notch signaling. *Aging.* (2020) 12:4247–67. doi: 10.18632/aging.102878
8. Zhu J, Fu H, Wu Y, Zheng X. Function of lncRNAs and approaches to lncRNA-protein interactions. *Sci China Life Sci.* (2013) 56:876–85. doi: 10.1007/s11427-013-4553-6
9. Mercer TR, Mattick JS. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Biol.* (2013) 20:300–7. doi: 10.1038/nsmb.2480
10. Liu M, Zhong J, Zeng Z, Huang K, Ye Z, Deng S, et al. Hypoxia-induced feedback of HIF-1α and lncRNA-CF129 contributes to pancreatic cancer progression through stabilization of p53 protein. *Theranostics.* (2019) 9:4795–810. doi: 10.7150/thno.30988
11. Ma C-N, Wo L-L, Wang D-F, Zhou C-X, Li J-C, Zhang X, et al. Hypoxia activated long non-coding RNA HABON regulates the growth and proliferation of hepatocarcinoma cells by binding to and antagonizing HIF-1 alpha. *RNA Biol.* (2021) 18:1791–806. doi: 10.1080/15476286.2020.1871215
12. Hua Q, Mi B, Xu F, Wen J, Zhao L, Liu J, et al. Hypoxia-induced lncRNA-AC020978 promotes proliferation and glycolytic metabolism of non-small cell lung cancer by regulating PKM2/HIF-1α axis. *Theranostics.* (2020) 10:4762–78. doi: 10.7150/thno.43839
13. Zhang Z, Fang E, Rong Y, Han H, Gong Q, Xiao Y, et al. Hypoxia-induced lncRNA CASC9 enhances glycolysis and the epithelial-mesenchymal transition of pancreatic cancer by a positive feedback loop with AKT/HIF-1α signaling. *Am J Cancer Res.* (2021) 11:123–37.
14. Wang X, Wang Y, Li L, Xue X, Xie H, Shi H, et al. A lncRNA coordinates with Ezh2 to inhibit HIF-1α transcription and suppress cancer cell adaption to hypoxia. *Oncogene.* (2020) 39:1860–74. doi: 10.1038/s41388-019-1123-9
15. Zhang Y, Su W, Zhang B, Ling Y, Kim WK, Zhang H. Comprehensive analysis of coding and non-coding RNA transcriptomes related to hypoxic adaptation in Tibetan chickens. *J Anim Sci Biotechnol.* (2021) 12:60. doi: 10.1186/s40104-021-00582-2
16. Wang J, Chai Z, Deng L, Wang J, Wang H, Tang Y, et al. Detection and integrated analysis of lncRNA and mRNA relevant to plateau adaptation of Yak. *Reprod Domest Anim.* (2020) 55:1461–9. doi: 10.1111/rda.13767
17. Zhang B, Chamba Y, Shang P, Wang Z, Ma J, Wang L, et al. Comparative transcriptomic and proteomic analyses provide insights into the key genes involved in high-altitude adaptation in the Tibetan pig. *Sci Rep.* (2017) 7:3654. doi: 10.1038/s41598-017-03976-3
18. Wang T, Guo Y, Liu S, Zhang C, Cui T, Ding K, et al. KLF4, a key regulator of a transitive triplet, acts on the TGF-β signaling pathway and contributes to high-altitude adaptation of Tibetan pigs. *Front Genet.* (2021) 12:628192. doi: 10.3389/fgene.2021.628192
19. Ge Q, Guo Y, Zheng W, Zhao S, Cai Y, Qi X. Molecular mechanisms detected in yak lung tissue via transcriptome-wide analysis provide insights into adaptation to high altitudes. *Sci Rep.* (2021) 11:7786. doi: 10.1038/s41598-021-87420-7
20. Feng S, Ma J, Long K, Zhang J, Qiu W, Li Y, et al. Comparative microRNA transcriptomes in domestic goats reveal acclimatization to high altitude. *Front Genet.* (2020) 11:809. doi: 10.3389/fgene.2020.00809
21. Xin JW, Chai ZX, Zhang CF, Yang YM, Zhang Q, Zhu Y, et al. Transcriptome analysis identified long non-coding RNAs involved in the adaption of yak to high-altitude environments. *R Soc Open Sci.* (2020) 7:200625. doi: 10.1098/rsos.200625
22. Trigatti BL. SR-B1 and PDZK1: partners in HDL regulation. *Curr Opin Lipidol.* (2017) 28:201–8. doi: 10.1097/MOL.0000000000000396
23. Sirwi A, Hussain MM. Lipid transfer proteins in the assembly of apoB-containing lipoproteins. *J Lipid Res.* (2018) 59:1094–102. doi: 10.1194/jlr.R083451
24. Stefan N, Hennige AM, Staiger H, Machann J, Schick F, Kröber SM, et al. Alpha2-heremans-schmid glycoprotein/fetuin-A is associated with insulin resistance and fat accumulation in the liver in humans. *Diabetes Care.* (2006) 29:853–7. doi: 10.2337/diacare.29.04.06.dc05-1938
25. Li TP, Sun SW, Xiong GZ, Qiu F, Yang DM, Sun SY, et al. Direct interaction of daxx and androgen receptor is required for their regulatory activity in cholesterol biosynthesis. *Pharmacology.* (2021) 106:29–36. doi: 10.1159/000506488
26. Jiang P, Wang XN, Wang Y, Li AN, Xiao H, Guo PC, et al. Difference analysis of C4BPA gene expression in mammary tissue of dairy cows. *Chin J Vet Sci Jum.* (2016) 36:1032–5+43. doi: 10.16303/j.cnki.1005-4545.2016.06.28
27. Cheviron ZA, Bachman GC, Connaty AD, McClelland GB, Storz JF. Regulatory changes contribute to the adaptive enhancement of thermogenic capacity in high-altitude deer mice. *Proc Natl Acad Sci USA*. (2012) 109:8635–40. doi: 10.1073/pnas.1120523109
28. Gangwar A, Paul S, Ahmad Y, Bhargava K. Intermittent hypoxia modulates redox homeostasis, lipid metabolism associated inflammatory processes and redox post-translational modifications: benefits at high altitude. *Sci Rep.* (2020) 10:7899. doi: 10.1038/s41598-020-64848-x
29. Meir JU, York JM, Chua BA, Jardine W, Hawkes LA, Milsom WK. Reduced metabolism supports hypoxic flight in the high-flying bar-headed goose. *Elife.* (2019) 8:e44986. doi: 10.7554/eLife.44986
30. Zhang X, Li M, Yin N, Zhang J. The expression regulation and biological function of autotaxin. *Cells.* (2021) 10:939. doi: 10.3390/cells10040939
31. Finck BN, Bernal-Mizrachi C, Han DH, Coleman T, Sambandam N, LaRiviere LL, et al. A potential link between muscle peroxisome proliferator- activated receptor-alpha signaling and obesity-related diabetes. *Cell Metab.* (2005) 1:133–44. doi: 10.1016/j.cmet.2005.01.006
32. Lefebvre P, Chinetti G, Fruchart JC, Staels B. Sorting out the roles of PPAR alpha in energy metabolism and vascular homeostasis. *J Clin Investig.* (2006) 116:571–80. doi: 10.1172/JCI27989
33. Sever N, Miličić G, Bodnar NO, Wu X, Rapoport TA. Mechanism of lamellar body formation by lung surfactant protein B. *Molecular Cell.* (2021) 81:49–66. doi: 10.1016/j.molcel.2020.10.042
34. Mason RJ. Biology of alveolar type II cells. *Respirology.* (2006) 11:12–5. doi: 10.1111/j.1440-1843.2006.00800.x
35. Feuerecker M, Crucian BE, Quintens R, Buchheim J-I, Salam AP, Rybka A, et al. Immune sensitization during 1 year in the Antarctic high-altitude Concordia environment. *Allergy.* (2019) 74:64–77. doi: 10.1111/all.13545
36. Rohm I, Aderhold N, Ratka J, Goebel B, Franz M, Pistulli R, et al. Hypobaric hypoxia in 3000 m altitude leads to a significant decrease in circulating plasmacytoid dendritic cells in humans. *Clin Hemorheol Microcirc.* (2016) 63:257–65. doi: 10.3233/CH-152035
37. Yilmaz A, Ratka J, Rohm I, Pistulli R, Goebel B, Asadi Y, et al. Decrease in circulating plasmacytoid dendritic cells during short-term systemic normobaric hypoxia. *Eur J Clin Invest.* (2016) 46:115–22. doi: 10.1111/eci.12416
38. Sarais F, Rebl H, Verleih M, Ostermann S, Krasnov A, Köllner B, et al. Characterisation of the teleostean κB-Ras family: the two members NKIRAS1 and NKIRAS2 from rainbow trout influence the activity of NF-κB in opposite ways. *Fish Shellfish Immunol.* (2020) 106:1004–13. doi: 10.1016/j.fsi.2020.08.052

39. Li Q, Verma IM. NF-kappaB regulation in the immune system. *Nat Rev Immunol.* (2002) 2:725–34. doi: 10.1038/nri910
40. Lu M, Zhu X, Yang Z, Zhang W, Sun Z, Ji Q, et al. E3 ubiquitin ligase tripartite motif 7 positively regulates the TLR4-mediated immune response via its E3 ligase domain in macrophages. *Mol Immunol.* (2019) 109:126–33. doi: 10.1016/j.molimm.2019.01.015
41. Kawai T, Akira S. Signaling to NF-kappaB by Toll-like receptors. *Trends Mol Med.* (2007) 13:460–9. doi: 10.1016/j.molmed.2007.09.002
42. Anthoney N, Foldi I, Hidalgo A. Toll and Toll-like receptor signalling in development. *Development.* (2018) 145:156018. doi: 10.1242/dev.156018
43. Fields JK, Günther S, Sundberg EJ. Structural basis of IL-1 family cytokine signaling. *Front Immunol.* (2019) 10:1412. doi: 10.3389/fimmu.2019.01412
44. Qin Y, Li H, Qiao J. TLR2/MyD88/NF-κB signalling pathway regulates IL-8 production in porcine alveolar macrophages infected with porcine circovirus 2. *J Gen Virol.* (2016) 97:445–52. doi: 10.1099/jgv.0.000345
45. Lu YC, Yeh WC, Ohashi PS. LPS/TLR4 signal transduction pathway. *Cytokine.* (2008) 42:145–51. doi: 10.1016/j.cyto.2008.01.006
46. Kawai T, Akira S. TLR signaling. *Semin Immunol.* (2007) 19:24–32. doi: 10.1016/j.smim.2006.12.004
47. Takeda K, Akira S. TLR signaling pathways. *Semin Immunol.* (2004) 16:3–9. doi: 10.1016/j.smim.2003.10.003
48. Gaur P, Prasad S, Kumar B, Sharma SK, Vats P. High-altitude hypoxia induced reactive oxygen species generation, signaling, and mitigation approaches. *Int J Biometeorol.* (2021) 65:601–15. doi: 10.1007/s00484-020-02037-1
49. Dosek A, Ohno H, Acs Z, Taylor AW, Radak Z. High altitude and oxidative stress. *Respir Physiol Neurobiol.* (2007) 158:128–31. doi: 10.1016/j.resp.2007.03.013
50. Jeong Dw, Kim TS, Chung YW, Lee BJ, Kim IY. Selenoprotein W is a glutathione-dependent antioxidant in vivo. *FEBS Lett.* (2002) 517:225–8. doi: 10.1016/S0014-5793(02)02628-5
51. Lei XG, Cheng W-H, McClung JP. Metabolic regulation and function of glutathione peroxidase-1. *Annu Rev Nutr.* (2007) 27:41–61. doi: 10.1146/annurev.nutr.27.061406.093716
52. Scarl RT, Lawrence CM, Gordon HM, Nunemaker CS. STEAP4: its emerging role in metabolism and homeostasis of cellular iron and copper. *J Endocrinol.* (2017) 234:123–34. doi: 10.1530/JOE-16-0594
53. Imai H, Kashiwazaki H, Suzuki T, Kabuto M, Himeno S, Watanabe C, et al. Selenium levels and glutathione peroxidase activities in blood in an andean high-altitude population. *J Nutr Sci Vitaminol.* (1995) 41:349–61. doi: 10.3177/jnsv.41.349
54. Jani M, Beéry E, Heslop T, Tóth B, Jagota B, Kis E, et al. Kinetic characterization of bile salt transport by human NTCP (SLC10A1). *Toxicol In Vitro.* (2018) 46:189–93. doi: 10.1016/j.tiv.2017.10.012
55. Claro da Silva T, Polli JE, Swaan PW. The solute carrier family 10 (SLC10): beyond bile acid transport. *Mol Aspects Med.* (2013) 34:252–69. doi: 10.1016/j.mam.2012.07.004
56. Hirschfield GM, Liu X, Xu C, Lu Y, Xie G, Lu Y, et al. Primary biliary cirrhosis associated with HLA, IL12A, and IL12RB2 variants. *N Engl J Med.* (2009) 360:2544–55. doi: 10.1056/NEJMoa0810440
57. Carbone M, Mells GF, Alexander GJ, Westbrook RH, Heneghan MA, Sandford RN, et al. Calcineurin inhibitors and the IL12A locus influence risk of recurrent primary biliary cirrhosis after liver transplantation. *Am J Transplant.* (2013) 13:1110–1. doi: 10.1111/ajt.12132
58. Lin J, Cao C, Tao C, Ye R, Dong M, Zheng Q, et al. Cold adaptation in pigs depends on UCP3 in beige adipocytes. *J Mol Cell Biol.* (2017) 9:364–75. doi: 10.1093/jmcb/mjx018
59. Li M, Tian S, Jin L, Zhou G, Li Y, Zhang Y, et al. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat Genet.* (2013) 45:1431–8. doi: 10.1038/ng.2811
60. Cox B, Lee TF. Further evidence for a physiological role for hypothalamic dopamine in thermoregulation in the rat. *J Physiol.* (1980) 300:7–17. doi: 10.1113/jphysiol.1980.sp013147
61. Boulant JA. Role of the preoptic-anterior hypothalamus in thermoregulation and fever. *Clin Infect Dis.* (2000) 31 (Suppl. 5):157–61. doi: 10.1086/317521
62. Lucas G. Serotonin receptors, type 4: a new hope? *Curr Drug Targets.* (2009) 10:1085–95. doi: 10.2174/138945009789735200

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 in Genetics

ORIGINAL RESEARCH
published: 28 January 2022
doi: 10.3389/fgene.2021.804481

Comparison of the Agronomic, Cytological, Grain Protein Characteristics, as Well as Transcriptomic Profile of Two Wheat Lines Derived From Wild Emmer

Fangyi Gong[1], Tiangang Qi[1], Tian Zhang[1], Yusen Lu[1], Jia Liu[1], Xiaoying Zhong[1], Jingshu He[1], Yunfang Li[2], Youliang Zheng[1,3], Dengcai Liu[1,3], Lin Huang[1] and Bihua Wu[1,3]**

[1]State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Triticeae Research Institute, Sichuan Agricultural University, Chengdu, China, [2]Chengdu Institute of Biology, Chinese Academy of Science, Chengdu, China, [3]Key Laboratory of Crop Genetic Resources and Improvement, Ministry of Education, Sichuan Agricultural University, Chengdu, China

OPEN ACCESS

Edited by:
Suxu Tan,
Michigan State University,
United States

Reviewed by:
Xiaojun Nie,
Northwest A and F University, China
Satinder Kaur,
Punjab Agricultural University, India

***Correspondence:**
Lin Huang
lhuang@sicau.edu.cn
Bihua Wu
wubihua2017@126.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 29 October 2021
Accepted: 13 December 2021
Published: 28 January 2022

Citation:
Gong F, Qi T, Zhang T, Lu Y, Liu J, Zhong X, He J, Li Y, Zheng Y, Liu D, Huang L and Wu B (2022) Comparison of the Agronomic, Cytological, Grain Protein Characteristics, as Well as Transcriptomic Profile of Two Wheat Lines Derived From Wild Emmer. Front. Genet. 12:804481. doi: 10.3389/fgene.2021.804481

Two advanced wheat lines BAd7-209 and BAd23-1 without the functional gene *GPC-B1* were obtained from a cross between common wheat cultivar Chuannong 16 (CN16) and wild emmer wheat accession D97 (D97). BAd7-209 showed superior quality parameters than those of BAd23-1 and CN16. We found that the components of glutenins and gliadins in BAd7-209 and BAd23-1 were similar, whereas BAd7-209 had higher amount of glutenins and gliadins than those of BAd23-1. RNA sequencing analysis on developing grains of BAd7-209 and BAd23-1 as well as their parents revealed 382 differentially expressed genes (DEGs) between the high–grain protein content (GPC) (D97 + BAd7-209) and the low-GPC (CN16 + BAd23-1) groups. DEGs were mainly associated with transcriptional regulation of the storage protein genes, protein processing in endoplasmic reticulum, and protein export pathways. The upregulated gluten genes and transcription factors (e.g., NAC, MYB, and bZIP) may contribute to the high GPC in BAd7-209. Our results provide insights into the potential regulation pathways underlying wheat grain protein accumulation and contribute to make use of wild emmer for wheat quality improvement.

Keywords: wild emmer wheat, GPC, transcriptome, processing quality, gluten

INTRODUCTION

Wheat (*Triticum aestivum* L.) is one of the most important staple crops, which can be processed into a wide range of products such as bread, noodle, and biscuit. Grain protein content (GPC) is an important quality trait in wheat and determines the nutritional value and processing quality (Shewry et al., 1995, 2009). However, GPC and grain yield-related traits are usually negative correlated (Groos et al., 2003), which hampered their simultaneous improvement in conventional wheat breeding program. The characterization and transferring of gene(s)/quantitative trait loci (QTLs) from wheat wild relatives is an effective strategy in the development of elite varieties with high GPC and/or yield (Krugman et al., 2018; Liu et al., 2019; Xiang et al., 2019; Gong et al., 2021).

Wild emmer (*T. turgidum* ssp. *dicoccoides*, 2n = 4x = 28, AABB), the tetraploid ancestor of common wheat, provides a valuable reservoir of genetic variation for GPC (Uauy et al., 2006; Liu

TABLE 1 | Agronomic characteristics of BAd7-209, BAd23-1, CN16, and D97

Materials	Plant height (cm)	Number of spikelets per spike (no.)	Number of spikes (no.)	1,000-Kernel weight (g)
BAd7-209	128.25 ± 4.44b	17.60 ± 1.06b	8.80 ± 1.97a	47.8 ± 0.8a
BAd23-1	106.32 ± 5.37c	17.13 ± 1.41b	8.27 ± 1.98a	41.81 ± 1.6b
D97	143 ± 4.3a	11.50 ± 1.66c	7.82 ± 0.41b	19.15 ± 0.68c
CN16	83.0 ± 3.2d	18.60 ± 1.55a	8.5 ± 0.7a	43.4 ± 1.3ba

Note: The letters a, b, c, and d indicate the significant difference at 0.05 level with Turkey's two-way test.

et al., 2019). A number of QTLs affecting GPC were reported in wild emmer wheat (Joppa et al., 1997; Gonzalez-Hernandez et al., 2004; Uauy et al., 2006; Fatiukha et al., 2019; Liu et al., 2019). To date, however, only *Gpc-B1* on chromosome 6BS has been cloned. The introgression of *Gpc-B1* in wheat breeding programs can significantly improve GPC while reduce grain-yield related traits in some wheat lines and environments (Uauy et al., 2006; Chen et al., 2017). Therefore, it is desirable to identify GPC-QTLs that are not negative correlated or less correlated with grain yield-related traits.

GPC was regulated by a plethora of genes and easily affected by environment (Liu et al., 2019). In forward genetics, identification of candidate genes related to GPC is a time-consuming and laborious process. RNA sequencing (RNA-seq) provides high-resolution methods for deciphering quality traits and quantifying expression levels of candidate genes on a genome-wide scale (Furtado et al., 2015; Rangan et al., 2017). Currently, RNA-seq has been utilized to study the differentially expressed genes (DEGs) and regulation networks that associated with wheat grain protein accumulation (Cantu et al., 2011; Gong et al., 2021).

Grain protein synthesis in cereal crops were determined by several pathways, mainly including transcriptional regulation of the storage protein genes (glutenin and gliadin) and protein processing in endoplasmic reticulum (ER) and Golgi apparatus. Transcription factors (TFs) belong to bZIP, Dof, MYB, and NAC families have been widely reported in transcriptional regulation of grain protein genes in rice, maize, and barley (Suzuki et al., 1998; Onodera et al., 2001; Diaz et al., 2002; Shewry and Halford, 2002; Zhang et al., 2019). Previous studies demonstrated that the protein processing in ER and Golgi apparatus had crucial role in grain protein synthesis in rice (Takemoto et al., 2002; Wang et al., 2016; Ren et al., 2020), whereas their function for GPC was less reported in wheat.

In our previous studies, the agronomically stable advanced wheat lines were developed from a cross between common wheat cultivar Chuannong 16 (CN16) and wild emmer accession D97 (D97) followed by successive selfing (Jiang et al., 2017; Wang et al., 2018; Liu et al., 2019). Some advanced lines with simultaneous improvement of GPC and thousand-kernel weight (TKW) were obtained. In the present study, two sister lines, BAd7-209 and BAd23-1, with desirable agronomic traits were obtained from CN16×D97. These two lines showed contrasting GPC, while both of them did not contain the functional *Gpc-B1*. A comparison of the transcriptomes of developing grains from BAd7-209, BAd23-1, CN16, and D97 revealed candidate genes and regulation pathways that may be contributed to wheat grain protein accumulation.

MATERIALS AND METHODS

Plant Materials

Two sister lines BAd7-209 and BAd23-1 as well as their parents D97 and CN16 were kept at the Triticeae Research Institute, Sichuan Agricultural University, Chengdu, China. CN16 and BAd23-1 were characterized as low-gluten wheat lines, whereas BAd7-209 and D97 were high-GPC lines (**Table 1**). Wheat plants were grown in the field with three biological replicates (10 rows each replicate) at the experimental field of the Sichuan Agricultural University over two wheat growing seasons (2016 and 2017) at Wenjiang (2016WJ and 2017WJ) and Chongzhou (2016CZ and 2017CZ). Individuals were planted 10-cm apart in a 2-m row with 30 cm between rows. A compound fertilizer [N: P: K (15: 15: 15)] was applied before sowing at a rate of 450 kg per hectare. Developing grains were sampled at 10, 14, 18, 22, 26, 30, 34, and 38 days after pollination (DAP) and snap-frozen in liquid nitrogen and then kept at −80°C for RNA-seq.

Characterization of Agronomic Traits and Karyotype

Agronomic traits of BAd7-209 and BAd23-1 and their parents were measured as listed in **Table 1**. The traits of plant height (PH), number of spikelets per spike (NSp), and spike number per plant (SN) were averaged by 10 plants. The weight of 300 randomly selected seeds (GB/T 5519-2008, 2008) was recorded to represent the 1,000-kernel weight (TKW) as described by Wang et al. (2018).

At least 30 root-tip cells of BAd7-209 and BAd23-1 and their parents D97 and CN16 were observed for detection the chromosome number using the methods as described by Zhang et al. (2007).

Single-Nucleotide Polymorphism (SNP) Genotyping

Genomic DNA from leaves was isolated using a plant genomic DNA kit (Tiangen Biotech, Beijing Co. Ltd. Beijing, China). Chip-based genotyping was conducted using the wheat 55 K SNP array (www.capit albio.com). The flanking sequence of each SNP was used to map onto the bread wheat reference

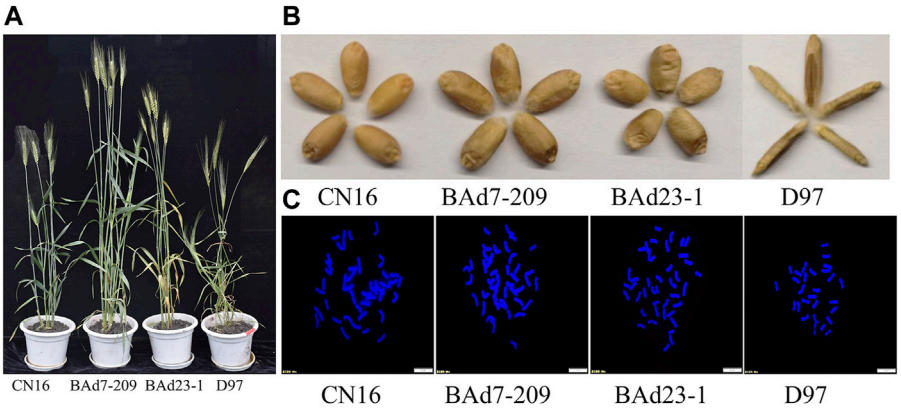


FIGURE 1 | Agronomic traits and chromosome patterns of CN16, BAd7-209, BAd23-1, and D97. **(A)** Plants and **(B)** seeds of CN16, BAd7-209, BAd23-1, and D97. **(C)** The number of root-tip chromosomes; scale bar: 10 um.

TABLE 2 | The major quality parameters in BAd7-209 and BAd23-1 and their parents.

Parameter	BAd7-209	BAd23-1	CN16	D97
Grain protein content (%)	15.8 ± 1.0b	13.8 ± 0.3c	12.6 ± 0.4d	23.70 ± 0.46a
Albumin (%)	3.31 ± 0.03b	3.045 ± 0.57b	3.455 ± 0.22b	6.32 ± 0a
Globumin (%)	0.27 ± 0.16b	0.245 ± 0.05b	0.245 ± 0.09b	2.19 ± 0a
gliadin (%)	3.06 ± 0.41b	2.3 ± 0.23c	2.8 ± 0b	6.77 ± 0a
glutenin (%)	6.34 ± 0.18b	4.78 ± 0.16c	5.025 ± 0.40b	6.48 ± a
Wet glutein content (%)	39.49 ± 0.55a	24.89 ± 0.41b	28.01 ± 0.51b	ND
Sedimentation volume (ml)	41.3 ± 0.42a	22.5 ± 0.71c	26.6 ± 0.57b	ND
Hydroscopic rate (%)	64.6 ± 2.55a	55.8 ± 1.41a	53.6 ± 3.39a	ND
Stabilization time (min)	7.93 ± 0.11a	3.9 ± 0.14b	1.61 ± 0.08c	ND

Note: ND, not determined; the letters a, b, c, and d indicate the significant difference at 0.05 level with Turkey's two-way test.

sequence (https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Assemblies/v1.0/), using BLASTN with E-value threshold of 10^{-10} and a maximum mismatch of one base. SNP markers showed homozygous genotype among the parents, and advanced wheat lines were used to estimate D97 introgressions. The ratios of same SNP to the total SNPs scored between D97 and two advanced lines were computed using a sliding window of 10 Mb and step length of 1 Mb as described by Hao et al. (2019). Graphical representations were constructed using the R package ggplot2 (Wickham 2016).

Quality Parameters Determination

Mature seeds were harvested for measurement of wheat quality parameters (**Supplementary Table S1**). Seeds of D97, CN16, BAd7-209, and BAd23-1 were conditioned to 14% moisture and milled using the Chopin CD1 AUTO (Renault, Boulogne-Billancourt, France) (Wang et al., 2018). Wet gluten content (WGC) was determined following the American Association of Cereal Chemists (AACC) method 38-12A. The sedimentation volume (SV) was measured following the AACC 44-15A (FOSS, Denmark). GPC (0.5 g) was recorded according to the AACC 39-10. Hydroscopic rate (HR) was measured following the AACC-54-21. All of the experiments were repeated three times.

Grain Protein Component Measurement

Albumins, globulins, glutenins, and gliadins were extracted from mature grains using the methods as described by Duan and Zhao, 2004 and Wan et al. (2000) with slight modifications. Protein content was tested using Kjeldahl method (Kjeltec™ 8400). Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate the glutenins following the method of Hu et al. (2010) using the protein extraction buffer consisting of 62.5 mM Tris-HCl, pH 6.8, 10% (v/v) glycerol, 2% (w/v) SDS, 0.002% (w/v) bromophenol blue, and 1.5% (w/v) dithiothreitol (DTT).

RNA-Seq Analysis

Total RNA was extracted using a plant RNA extraction kit v1.5 (Biofit Biotechnologies, Chengdu). The RNA concentration and integrity were assessed using an Qubit®RNA assay kit on a Qubit® 2.0 Flurometer (Life Technologies, CA, United States) and an RNA Nano 6000 assay kit, respectively. RNAs from different grain developing stages were pooled in equi-molar concentrations (1 μg of RNA per sample) for cDNA library construction (three biological replicates per library). The sequencing libraries were completed using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, United States) and then sequenced using the

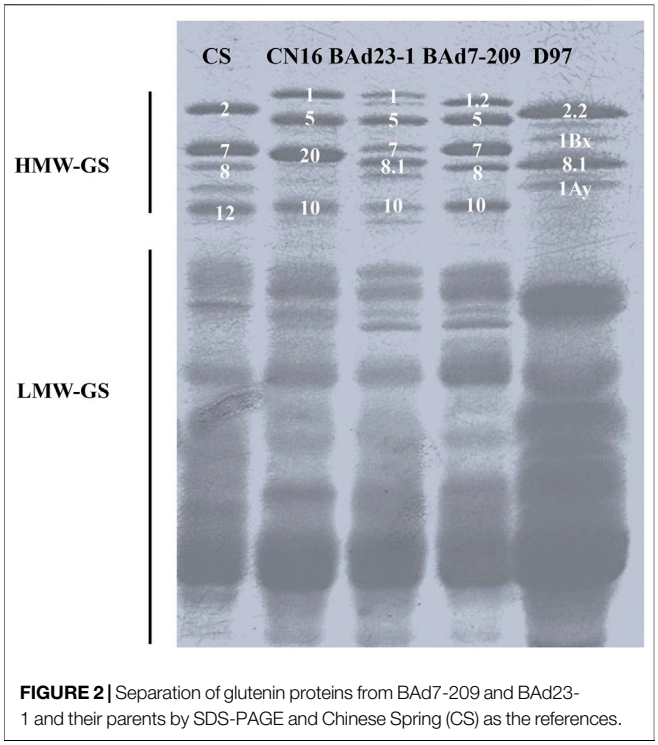


FIGURE 2 | Separation of glutenin proteins from BAd7-209 and BAd23-1 and their parents by SDS-PAGE and Chinese Spring (CS) as the references.

Illumina Hiseq platform (Novogene Bioinformatics Technology Co. Ltd. Beijing, China).

The adaptors, reads with more than 10% N, and reads with phred quality scored Q < 20 from RNA-seq raw data were removed using Trimmomatic (Bolger et al., 2014). The generated clean reads from each library were assessed using the Q20, Q30, and GC contents and aligned to Chinese Spring reference genome (https://urgi.versailles.inra.fr/blast_iwgsc/) using Hisat2 (Kim et al., 2019). FPKM (Elowitz et al., 2002; Jiang and Wong, 2009) was estimated to represent the expression level. DESeq was used for differential expression analysis as described by Wang et al. (2010). Genes with adjusted FDR <0.001 and $|\log_2^{FC}| \geq 2$ found by EBSeq (Leng et al., 2013) were considered as DEGs. Short Time-series Expression Miner (STEM) software (v1.3.11) was used to explore expression patterns of DEGs as described by Ernst and Bar-Joseph (2006) with $\log_2$ standardization, p-value ≤ 0.05, and $|\log_2^{FC}| \geq 2$. Functional annotation of the DEGs was performed using Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) databases (Kanehisa et al., 2004).

Prediction of Transcription Factors and Cis-elements

The putative plant TFs and cis-elements were predicted using iTAK software (Zheng et al., 2016) and PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Rombauts et al., 1999).

Quantitative Real-Time PCR

Quantitative real-time PCR (qRT-PCR) was performed to validate the RNA-seq data. Gene-specific primers were designed using Primer3Plus software (http://www.primer3plus.com/) (Hung and Weng, 2016). Reactions were performed using the Bio-Rad CFX96 Real-Time PCR System (Bio-Rad, United States) in 10 μl of reaction volume containing 2 ng of cDNA, 5 μl of 1 × SYBR Premix ExTaq (TaKaRa), 0.5 μl (300 nM) of each primer, and ddH_2O up to 10 μl (Wang et al., 2020). The $2^{-\Delta\Delta CT}$ method (Schmittgen 2002) was used to quantify the gene expression with endogenous *GAPDH* as the internal reference.

Statistical Analysis

Statistical analyses were conducted using SPSS version 22 software (SPSS Inc. Chicago, IL, United States). The morphological traits and quality parameters were compared by analysis of variance (ANOVA) complemented by Tukey's test.

RESULTS

Agronomic and Karyotype Characteristics

The NSp of BAd7-209 and BAd23-1 were less than that of CN16 ($p < 0.05$), while higher than D97 ($p < 0.05$). The BAd7-209 resembled the BAd23-1 with respect to spike and SN. We have observed significantly differences in PH and TKW among BAd7-209, BAd23-1, and CN16 (**Figure 1**; **Table 1**). The D97 had significantly less NSp and TKW, while significantly higher PH than that of advanced wheat lines (**Table 2**).

The BAd7-209 had higher SV (41.3%), WGC (39.49%), HR (64.6%), and stabilization time (7.93%) than those of BAd23-1 and CN16. The GPC in D97 was highest (23.70%), followed by BAd7-209 (15.8%), BAd23-1 (13.8%), and CN16 (12.6%) (**Table 2**). The WGC, SV, HR, and GPC levels of BAd7-209 were higher than the strong-gluten wheat and those of BAd23-1 were between weak-gluten and medium-gluten wheat.

Cytological observations indicated that both BAd23-1 and BAd7-209 have 42 chromosomes (**Figure 1C**), which reached the genetic background of common wheat. SNP genotyping analyses revealed 150 and 108 potential D97 segments on chromosomes of BAd23-1 and BAd7-209. The number of CN16 segments on BAd7-209 and BAd23-1 chromosomes was 212 and 232 (**Supplementary Figure S1**).

Grain Protein Component and Glutenin Subunits

The total amount of glutenins, gliadins, albumins, and globulins in D97 was higher than those of CN16. There were no differences in albumin and globulin contents between BAd23-1 and BAd7-209. The glutenin and gliadin contents of BAd7-209 were significantly higher than those of BAd23-1 (**Table 2**).

SDS-PAGE analysis showed different high-molecular-glutenin subunits (HMW-GSs) in CN16 (1Ax1, 1Dx5, 1Dy10, and 1Bx+1By20) and D97 (1Ax2.2, 1Bx, 1By8.1, and 1Ay). The BAd7-209 possessed five HWM-GSs (1Ax1.2, 1Bx7 + 1By8, and 1Dx5 + 1Dy10), which were consistent with previous reports (Jiang et al., 2017; Xiang et al., 2019). The BAd23-1 had five HMW-GSs, including 1Bx7 and 1Dx5+1Dy10 that are

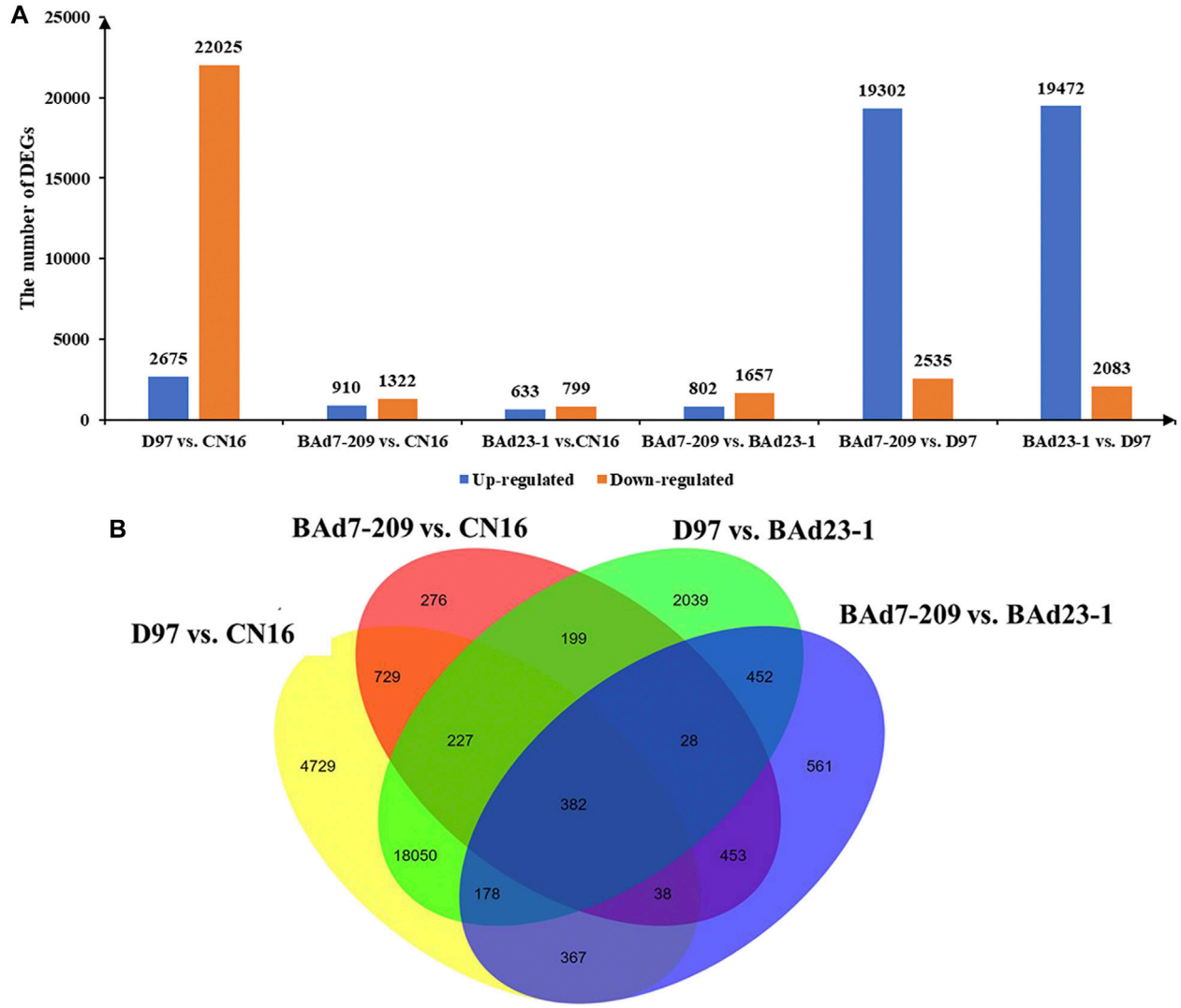


FIGURE 3 | The histograms and Venn diagrams of DEGs in different comparisons. **(A)** Up- and downregulated DEGs in BAd7-209 and BAd23-1 and their parents. **(B)** The Venn diagram of DEGs in the high-GPC (D97 + BAd7-209) vs. low-GPC (CN16 + BAd23-1) groups.

identical to BAd7-209, and one different 1Ax subunit (1Ax1) and one 1By subunit (1By8.1) (**Figure 2**).

The expression levels of 1Ax and 1By in high-GPC group (D97 + BAd7-209) were higher than those of low-GPC group (CN16 + BAd23-1). The composition of low-molecular-glutenin subunits (LMW-GSs) in CN16, BAd7-209, and BAd23-1 was similar. The wild emmer D97 showed less LWM-GS subunits probably due to the absence of the D genome. The expression levels of LMW-GSs in high-GPC group (D97 + BAd7-209) were higher than those of low-GPC group (CN16 + BAd23-1) (**Figure 2**).

Analysis of RNA-Seq Data

RNA-seq generated 189.93 million raw reads from developing grains of BAd7-209, BAd23-1, CN16, and D97 pools. A total of 94.96 million clean reads were retained and further mapped to the *T. aestivum* cDNA database (IWGSC1.0). The mapping ratios ranged from 79.42% to 84.33%, 86.92% to 88.12%, 85.76% to 89.33%, and 79.15% to 80.89% in BAd7-209, BAd23-1, CN16, and D97, respectively. The GC contents among replicates were almost identical, and the Q30 was over 90% in each library (**Supplementary Table S2**). These results demonstrate that the RNA-seq data were qualified for subsequent analysis.

Differential Expression Analysis

In total, 62,404 genes mapped to IWGSC1.0 genome were expressed (FPKM≥1) in D97, CN16, BAd7-209, and BAd23-1 libraries. A total of 24,700, 2,332, 21,555, and 2,459 DEGs ($|\log_2^{FC}|\geq 2$) were found in D97 vs. CN16, BAd7-209 vs. CN16, D97 vs. BAd23-1, and BAd7-209 vs. BAd23-1, respectively (**Figure 3A**). The comparison between the high-GPC group (D97 + BAd7-209) and low-GPC group (CN16 + BAd23-1) revealed 382

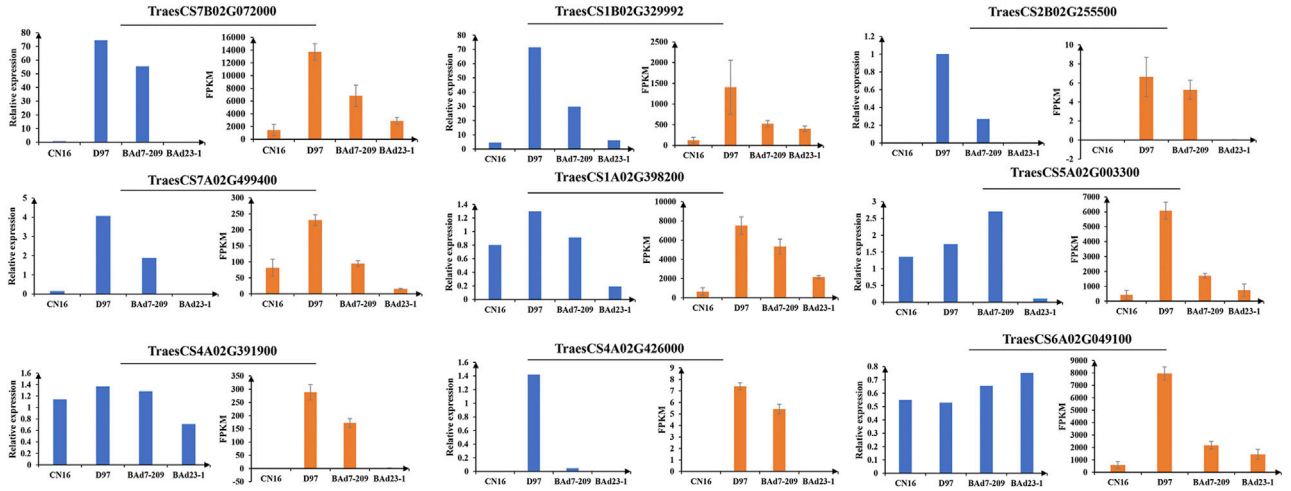


FIGURE 4 | Validation of DEGs by qRT-PCR.

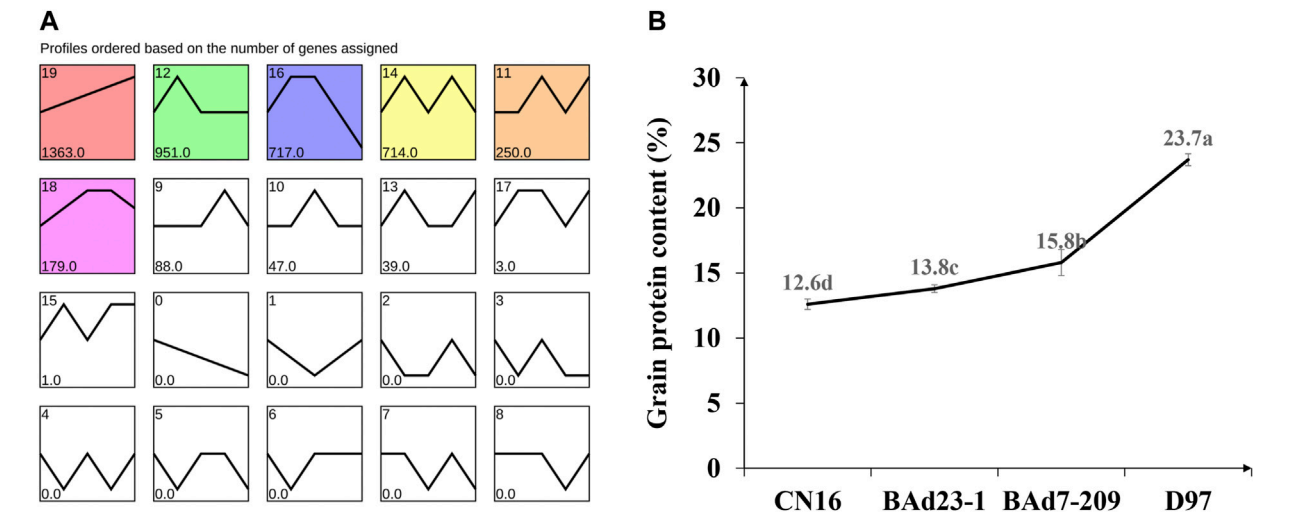


FIGURE 5 | Temporal expression patterns of DEGs **(A)** and grain protein content in CN16, BAd23-1, BAd7-209, and D97 **(B)**. Small letters a, b, c, and d indicate the significant difference at 0.05 level with Tukey's two-way test.

DEGs (**Figure 3B**), of which 148 genes were common expressed, 34 genes only expressed in D97 and BAd7-209, whereas 56 genes only expressed in CN16 and BAd23-1. Nine DEGs encoding HWM-GS, alpha-gliadin, nodulin protein, trypsin inhibitor, purothionin, and pre-mRNA-splicing factor proteins were selected to perform qRT-PCR (**Supplementary Table S3**). The expression changes of seven DEGs were quite consistent with those obtained from the RNA-seq (**Figure 4**, **Supplementary Table S3**).

All genes were further analyzed with the STEM software (v1.3.11) (Ernst and Bar-Joseph, 2006) to obtain the temporal expression patterns. Nineteen expression profiles were clustered, and seven profiles (9, 10, 0, 19, 4, 7, and 6) were regarded as significantly changed ($p \leq 0.05$) (**Figure 5A**). The profile 19 had similar tendency to the GPC in advanced lines and the parents (**Figure 5B**), indicating the positive role in regulation of grain protein accumulation.

The Functional Annotation of DEGs

A total of 459 DEGs in profile 19 were subjected to GO and KEGG analyses. On the basis of GO terms for those genes, three categories can be classified: biological process, cellular component, and molecular function. The GO term metabolic process (GO: 0008152, 172 DEGs), cellular process (GO: 0009987, 131 DEGs), and single-organism process (GO: 0044699, 93 DEGs) were highly enriched in biological process; those of cell (GO: 0005623, 137 DEGs), cell part (GO: 0044464, 136 DEGs), and organelle (GO: 0043226, 105 DEGs) were dominant in cellular component; and those of binding (GO:

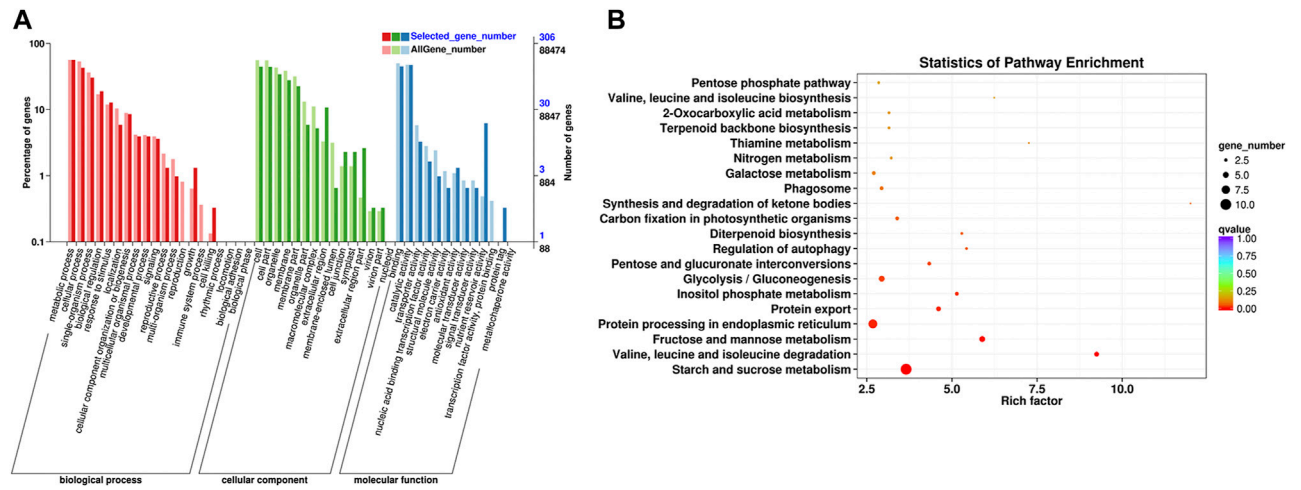


FIGURE 6 | Functional annotation of the DEGs based on GO and KEGG pathway enrichment analyses. **(A)** GO classification of the DEGs. **(B)** KEGG classification of the DEGs.

TABLE 3 | Cis-element in the promoter region of grain protein gene.

Geneid	FPKM				Cis-element				
	CN16	D97	BAd7-209	BAd23-1	GCN4	P-box	AACA	ACGCAA/G	Description
TraesCS6A02G049100	590	7,951.33	2,176.12	1830.75	+	+	–	+	α/β-gliadin
TraesCS6A02G049400	173.93	2,891.6	627.41	880.7	+	+	–	–	α/β-gliadin
TraesCS6A02G049500	18.94	797.11	151.39	179.42	+	–	–	–	α/β-gliadin
TraesCS6B02G065600	63.88	905.52	420.37	370.42	+	–	–	–	α/β-gliadin
TraesCS6B02G065749	112.64	2,186.8	573.67	784.51	+	–	–	–	α/β-gliadin
TraesCS6B02G065800	229.42	1,581.61	806.87	914.95	+	–	–	–	α/β-gliadin
TraesCS1A02G007344	614.17	5,175.81	2,263	1,685.48	+	+	+	+	γ-gliadin
TraesCS1A02G007400	485.43	10,810.48	1783.58	1,352.3	+	+	–	+	γ-gliadin
TraesCS1A02G007405	98.74	2,160.94	443.16	346.52	+	+	–	+	γ-gliadin
TraesCS1B02G329992	211.76	1,405.30	523.24	452.47	–	+	+	+	HMW-Glutenin
TraesCS1A02G007934	34.05	452.88	234.47	110.35	+	+	–	–	LMW-Glutenin
TraesCS1A02G008000	717.6	4,225.24	3,249.7	2,760.87	+	+	–	–	LMW-Glutenin
TraesCS1A02G066100	95.21	1786.88	835.5	357.43	+	–	+	–	12S globulin
TraesCS1A02G317500	226.07	3,879.86	1,045.96	769.24	+	–	–	–	19 kDa globulin
TraesCS1B02G330000	201.04	2,660.14	1,076.59	816.59	+	–	–	–	19 kDa globulin
TraesCS1A02G007700	881.24	12,190.49	3,893.48	2,723.56	+	+	+	–	Avenin-like
TraesCS7A02G035200	11.92	151	45.15	36.84	+	+	–	–	Avenin-like

0005488, 138 DEGs), catalytic activity (GO: 0003824, 145 DEGs), and nutrient reservoir activity (GO: 0045735, 19 DEGs) were primary in molecular function (**Figure 6A**, **Supplementary Table S4**).

KEGG pathway enrichment analysis showed that starch and sucrose metabolism (ko00500, 10 DEGs), protein processing in ER (ko04141, eight DEGs), fructose and mannose metabolism (ko00051, five DEGs), glycolysis/gluconeogenesis (ko00010, five DEGs), protein export (ko03060, four DEGs), inositol phosphate metabolism (ko00562, three DEGs), pentose and glucuronate interconversions (ko00040, three DEGs), valine, leucine, and isoleucine degradation (ko002800, one DEG) were significantly enriched pathways (**Figure 6B**, **Supplementary Table S5**). Eight DEGs in pathway of protein processing in ER are related to protein transport protein Sec61 (Sec61, 3 genes), heat shock protein (HSP, 3 genes), SKP1-like protein (SKP1, 1 gene), and E3 ubiquitin-protein ligase RMA3 (RMA3, 1 gene). Four genes in pathway of protein export are related to Sec61 (3 genes) and signal recognition particle 54 kDa protein 2-like (SRP, 1 gene).

Characterization of DEGs Encoded Gluten and TFs

Twelve out of 459 genes in profile 19 were related to α/β-gliadins (six genes), γ-gliadins (three genes), LMW-GSs (two genes), and HWM-GSs (one gene) (**Table 3**). All these genes were upregulated in the high-GPC group compared to the low-GPC group. Twenty-eight genes belonging to 16 TF families, such as

TABLE 4 | The list of TFs identified from the 459 DEGs in profile 19.

GeneName	GeneID	TF family	FPKM			
			CN16	BAd23-1	BAd7-209	D97
TaAUX0996	*Triticum_aestivum*_newGene_30,996	AUX/IAA	0	0	0.03	1.13
TaB30200	TraesCS3B02G530200	B3	0.01	0.54	0.14	3.37
SPA	TraesCS1A02G329900	bZIP	5.32	9.01	12.48	23.15
TabZIP8700	TraesCS3A02G378700	bZIP	1.08	1.55	1.93	4.86
TaC2H20600	TraesCS5B02G490600	C2H2	0.31	1.79	2.43	5.27
TaC3H8000	TraesCS2B02G138000	C3H	0.8	1.99	4.5	9.65
TaCSD9500	TraesCS6A02G069500	CSD	23.33	57.2	84.36	115.58
TaDBP8500	TraesCS6B02G398500	DBP	8.3	17.93	27.02	37.26
TaFAR079	*Triticum_aestivum*_newGene_56,079	FAR1	0.26	0.52	0.84	1.62
TaGA8200	TraesCS6B02G138200	GARP-G2-like	0.03	0.04	0.13	0.97
TaGA1000	TraesCS7A02G411000	GARP-G2-like	0.04	0.19	0.94	1.16
TaGR5900	TraesCS4A02G185900	GRAS	0.44	0.88	1.28	2
Taju6200	TraesCS5A02G166200	Jumonji	0	0.84	1.44	1.59
TaMA1500	TraesCS4B02G351500	MADS-M-type	1.41	1.36	2.8	5.67
TaMYB0300	TraesCS2A02G370300	MYB	7.12	12.29	17.97	28.99
TaMYB2000	TraesCS2B02G252000	MYB	1.45	1.63	2.49	4.86
TaMYB7800	TraesCS2B02G387800	MYB	0.75	1.96	4.3	5.94
TaGAMyb	TraesCS3A02G336500	MYB	3.54	8.04	12.93	16.15
TaNAC7700	TraesCS3A02G377700	NAC	0.19	0.26	0.5	1.31
TaNAC2900	TraesCS3B02G092900	NAC	0.43	1.77	4.14	6.29
TaNAC2400	TraesCS7A02G152400	NAC	0.52	2.63	4.18	6.16
TaNAC4700	TraesCS7A02G194700	NAC	6.33	32.32	37.62	56.61
TaNAC9500	TraesCS7A02G349500	NAC	0.08	0.25	0.18	0.55
TaNAC9100	TraesCS7A02G569100	NAC	5.76	23.67	34.2	56.1
TaNAC9300	TraesCS7A02G569300	NAC	0.04	0.54	1.25	3.03
TaNAC6300	TraesCS7B02G056300	NAC	3.18	11.97	30.07	64.68
TaNF6700	TraesCS7A02G336700	NF-YC	52.73	113.72	199.94	192.12
Ta3500	TraesCSU02G193500	Others	0	0	0.03	1.82

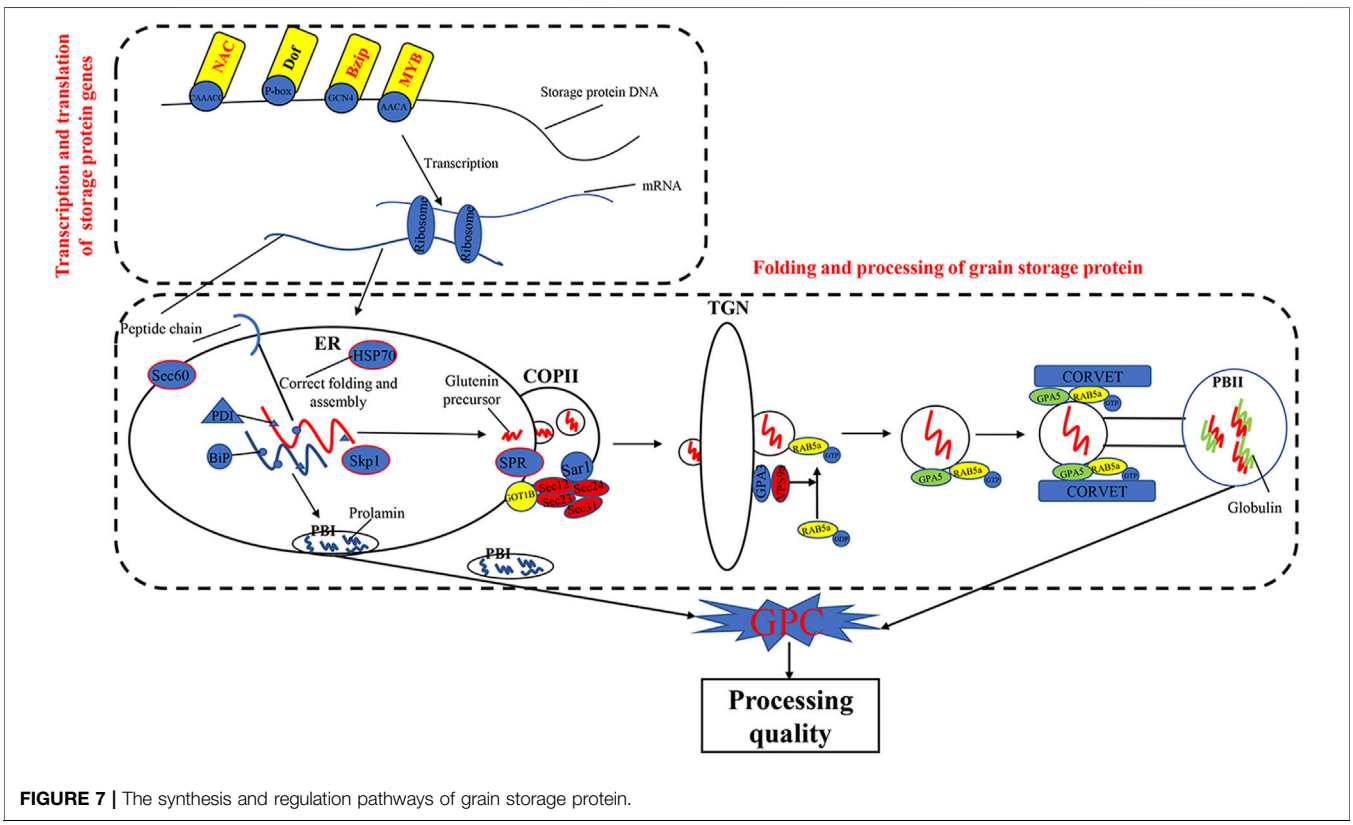


FIGURE 7 | The synthesis and regulation pathways of grain storage protein.

NAC, MYB, GARP-G2-like, bZIP, and C2H2, were identified (**Table 4**). To understand the expression patterns, the expression levels of TFs in different tissues sampled at different time points were retrieved from public available RNA-seq data of IWGSC (Thomas et al., 2014). We found that seven genes (e.g., *TabZIP8700*, *TaC2H20600*, and *TaMYB0300*) were expressed in different tissues at different time points. Ten genes (e.g., *SPA*, *TaC3H8000*, *TaGAMyb*, and *TaNAC019*) were only expressed in grains. Two genes (*TaMYB7800* and *TaNAC9500*) were root specific. The others were not detected in all tissues (**Supplementary Table S6**).

DISCUSSION

GPC is an important quality trait in common wheat. The wild emmer gene *GPC-B1* positively impacts protein, Zn, and Fe in wheat grain (Uauy et al., 2006). In the current study, two sister wheat lines BAd7-209 and BAd23-1 derived from wild emmer showed contrasting GPC, while both of them did not contain the *Gpc-B1*. Processing quality parameters of BAd7-209 were significantly higher than those of BAd23-1. These two lines had genetic background of common wheat and introgression segments from wild emmer D97. Our results indicate the presence of other wild emmer gene(s) that contributed to high GPC in BAd7-209.

Previous reports revealed that the components and expression levels of glutenins and gliadins could affect GPC (Xi and Zheng, 2011; Jiang et al., 2017; Wang et al., 2018; Xiang et al., 2019; Shen et al., 2020). In this study, similar glutenins subunits were identified in BAd7-209 and BAd23-1, whereas BAd7-209 had significant higher amount of glutenins and gliadins than BAd23-1. Nineteen DEGs including gliadin (nine genes), glutenin (three genes), globulin (three genes), avenin-like (three genes), and serpin (two genes) were enriched in the GO term nutrient reservoir activity (GO: 0045735). These genes were upregulated in the high GPC (D97 + BAd7-209) vs. the low-GPC (CN16 + BAd23-1) groups. Previous reports revealed that overexpression gliadin and glutenin genes could increase GPC in wheat (Guo et al., 2015; Li et al., 2019) and rice (Cho et al., 2019). Serpin genes were positively associated with wheat grain development (Benbow et al., 2019) and seed germination (Dong et al., 2015). The results in this study together with previous reports (Jiang et al., 2017; Wang et al., 2018; Xiang et al., 2019; Gong et al., 2021) show that the GPC is closely associated with the expression levels of glutenin and gliadin genes in wheat.

The KEGG pathways of processing in ER (ko04141) and protein export (ko03060) were previously reported to play important roles in the folding and maturation processing of grain storage protein in cereals (Crofts et al., 1998; Foresti et al., 2003; Strasser, 2018; Yu et al., 2018; Yu et al., 2020; Wang et al., 2016; Zhu et al., 2014) (**Figure 7**). In the present study, we found that DEGs, such as HSP, Sec61, SKP1, RMA3, and SRP, were significantly enriched in the two pathways (ko04141 and ko03060). Previous studies have shown that Hsp70 (Dupont et al., 1998) was a key chaperone for the processing and polymerization processes of grain storage proteins in ER. The Sec proteins (e.g., *Sec13*, *Sec23*, *Sec24*, and *Sec31*) were involved in transferring proteins from the ER to the Golgi apparatus (D'Arcangelo et al., 2013). SKP1 protein is a key member of the SCF (SKP–cullin–F-box protein) E3 ligase complex and mediates the regulation of plant ABA sensitivity (Hu et al., 2013; Rao et al., 2018). In wheat plants, ABA treatment could accelerate nitrogen remobilization from vegetative organs to grains (Saeed, 2012; Xie et al., 2004). In addition, the SRP proteins are essential for protein translocation across the ER (Walter and Blobel, 1982). These results demonstrated that the pathways of protein processing in ER and protein export and related DEGs are important in wheat grain protein accumulation.

TFs were involved in transcriptional regulation of gluten genes through binding their cis-elements in promoters (Albani et al., 1997; Xi and Zheng, 2011). For example, the bZIP TFs could bind GCN4 motifs of gliadin and glutenin genes in maize (*O2*) (Schmidt et al., 1992; Vicente-Carbajosa et al., 1998). The Dof TFs (*PBF*) could activate α-gliadin and *Glu-1* genes in wheat (Zhu et al., 2018) by binding P-box elements. MYB TFs were specifically bound to the AACA motifs of gluten genes in rice (*OsMYB5*) (Suzuki et al., 1998), barley (*HvGAMYB*, *HvMYBS2*, and *HvMCB1*) (Gubler et al., 1995; Diaz et al., 2002), and wheat (*TaGAMyb*) (Guo et al., 2015). The *TaNAC019* activated the expression of HMW-GS gene by binding to the [AT]NNNNNNNNNN [ATC][CG] A [CA]GN [ACT]A motif in the promoter region (Gao et al., 2021). In the present study, we have identified 28 TFs, such as NAC, MYB, GARP-G2-like, bZIP, and C_2H_2, from the high-GPC and low-GPC groups. Some TFs (e.g., *TaC3H8000*, *TaNAC2400*, *TaNAC4700*, and *TaNF6700*) were specifically expressed in grains. In addition, we have identified either NAC (ACGCAA/G), MYB (AACA), bZIP (GCN4), or Dof (P-box) motifs in the promoters of gluten genes that were differentially expressed in the high-GPC vs. low-GPC groups. Taken together, our results indicate that some TFs are involved in regulation of gluten and gliadin genes that contributed to the GPC accumulation in BAd7-209 (**Figure 7**).

CONCLUSION

In the present study, we have characterized the agronomic, cytological, grain protein characteristics, and transcriptomic profile of two advanced wheat lines (BAd7-209 and BAd23) from a cross between CN16 and D97. We found that BAd7-209 and BAd23 had genetic background of common wheat and introgression segments from wild emmer D97. The two advanced lines had contrasting GPC, while both absence of the functional *GPC-B1*. BAd7-209 had superior processing quality parameters and higher amount of glutenins and gliadins than those of BAd23-1, while their glutenin and gliadin subunits were similar. RNA-seq revealed that the contrasting GPC in BAd7-209 and BAd23-1 may be closely associated with the expression levels of glutenin and gliadin

genes which regulated by TFs. The protein processing in ER and protein export pathways and related DEGs are important in wheat grain protein accumulation.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available in NCBI under accession number PRJNA777016.

AUTHOR CONTRIBUTIONS

FG: formal analysis and writing—original draft preparation. TQ: investigation. YLu: investigation. JL: investigation. XZ: investigation. JH: investigation. YLi: investigation. YZ: supervision. DL: supervision. LH: writing and editing. BW: conceptualization, project administration, funding acquisition.

FUNDING

This work was financially supported by the Key Research and Development Program of Sichuan Province, China (2021YFYZ0002) and the International Cooperation Program of Science & Technology Department of Sichuan Province (2021YFH0110).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.804481/full#supplementary-material

REFERENCES

Albani, D., Hammond-Kosack, M. C., Smith, C., Conlan, S., Colot, V., Holdsworth, M., et al. (1997). The Wheat Transcriptional Activator *SPA*: a Seed-specific bZIP Protein that Recognizes the GCN4-like Motif in the Bifactorial Endosperm Box of Prolamin Genes. *Plant Cell* 9, 171–184. doi:10.1105/tpc.9.2.171

Benbow, H. R., Jermiin, L. S., and Doohan, F. M. (2019). Serpins: Genome-Wide Characterisation and Expression Analysis of the Serine Protease Inhibitor Family In *Triticum aestivum*. *G3.* 9, 2709–2722. doi:10.1534/g3.119.400444

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170

Cantu, D., Pearce, S. P., Distelfeld, A., Christiansen, M. W., Uauy, C., Akhunov, E., et al. (2011). Effect of the Down-Regulation of the High Grain Protein Content (GPC) Genes on the Wheat Transcriptome during Monocarpic Senescence. *BMC Genomics* 12, 492. doi:10.1186/1471-2164-12-492

Chen, X.-y., Song, G.-q., Zhang, S.-j., Li, Y.-l., Gao, J., Shahidul, I., et al. (2017). The Allelic Distribution and Variation Analysis of the *NAM-B1* Gene in Chinese Wheat Cultivars. *J. Integr. Agric.* 16, 1294–1303. doi:10.1016/S2095-3119(16)61459-4

Cho, K., Jo, Y. M., Lim, S. H., Kim, J. Y., Han, O., and Lee, J. Y. (2019). Overexpressing Wheat Low-Molecular-Weight Glutenin Subunits in Rice (*Oryza Sativa* L. *Japonica Cv*. Koami) Seeds. *Biotech*. 9, 49. 3. doi:10.1007/s13205-019-1579-x

Crofts, A. J., Leborgne-Castel, N., Pesca, M., Vitale, A., and Denecke, J. (1998). *BiP* and Calreticulin Form an Abundant Complex that Is Independent of Endoplasmic Reticulum Stress. *The Plant Cell* 10, 813–824. doi:10.1105/tpc.10.5.813

D'Arcangelo, J. G., Stahmer, K. R., and Miller, E. A. (2013). Vesicle-mediated export from the ER: COPII Coat Function and Regulation. *Biochim. Biophys. Acta (Bba) - Mol. Cel Res.* 1833, 2464–2472. doi:10.1016/j.bbamcr.2013.02.003

Diaz, I., Vicente-Carbajosa, J., Abraham, Z., Martinez, M., Isabel-La Moneda, I., Carbonero, P., et al. (2002). The GAMYB Protein from Barley Interacts with the DOF Transcription Factor BPBF and Activates Endosperm-specific Genes during Seed Development. *Plant J.* 29, 453–464. doi:10.1046/j.0960-7412.2001.01230.x

Dong, K., Zhen, S., Cheng, Z., Hui, C., Ge, P., and Yan, Y. (2015). Proteomic Analysis Reveals Key Proteins and Phosphoproteins Upon Seed Germination of Wheat (*Triticum aestivum* L.). *Front. Plant Sci.*, 6, 1017. doi:10.3389/fpls.2015.01017

Duan, S. E., and Zhao, W. M. (2004). Rapid Separation and SDS-PAGE Analysis of Wheat Glutenin Subunits. *J. Shaanxi Normal Univ.* 32, 77–79.

Dupont, F. M., Hurkman, W. J., Tanaka, C. K., and Chan, R. (1998). *BiP, HSP70, NDK* and *PDI* in Wheat Endosperm. I. Accumulation of mRNA and Protein during Grain Development. *Physiologia Plantarum* 103, 70–79. doi:10.1034/j.1399-3054.1998.1030109.x

Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002). Stochastic Gene Expression in a Single Cell. *Science* 297, 1183–1186. doi:10.1126/science.1070919

Ernst, J., and Bar-Joseph, Z. (2006). STEM: A Tool for the Analysis of Short Time Series Gene Expression Data. *BMC Bioinform.* 7, 191. doi:10.1186/1471-2105-7-191

Fatiukha, A., Lupo, I., Lidzbarsky, G., Klymiuk, V., Korol, A. B., Pozniak, C., et al. (2019). Grain Protein Content QTLs Identified in a Durum × Wild Emmer Wheat Mapping Population Tested in Five Environments. *Theor. Appl. Genet.* 133, 119–131. doi:10.1101/601773

Foresti, O., Frigerio, L., Holkeri, H., de Virgilio, M., Vavassori, S., and Vitale, A. (2003). A Phaseolin Domain Involved Directly in Trimer Assembly Is a Determinant for Binding by the Chaperone BiP. *The Plant Cell Online* 15, 2464–2475. doi:10.2307/387197710.1105/tpc.013052

Furtado, A., Bundock, P. C., Banks, P. M., Fox, G., Yin, X., and Henry, R. J. (2015). A Novel Highly Differentially Expressed Gene in Wheat Endosperm Associated with Bread Quality. *Sci. Rep.* 5, 10446. doi:10.1038/srep10446

Gao, Y., An, K., Guo, W., Chen, Y., Zhang, R., Zhang, X., et al. (2021). The Endosperm-specific Transcription Factor *TaNAC019* Regulates Glutenin and Starch Accumulation and its Elite Allele Improves Wheat Grain Quality. *The Plant Cell* 33, 603–622. doi:10.1093/plcell/koaa040

Gong, F., Huang, L., Qi, T., Tang, G., Liu, J., Xiang, L., et al. (2021). Comparative Analysis of Developing Grain Transcriptome Reveals Candidate Genes and Pathways Improving GPC in Wheat Lines Derived from Wild Emmer. *J. Appl. Genet.* 62, 17–25. doi:10.1007/s13353-020-00588-y

Gonzalez-Hernandez, J. L., Elias, E. M., and Kianian, S. F. (2004). Mapping Genes for Grain Protein Concentration and Grain Yield on Chromosome 5B of Triticum Turgidum (L.) Var. Dicoccoides. *Euphytica* 139, 217–225. doi:10.1007/s10681-004-3157-4

Groos, C., Robert, N., Bervas, E., and Charmet, G. (2003). Genetic Analysis of Grain Protein-Content, Grain Yield and Thousand-Kernel Weight in Bread Wheat. *Theor. Appl. Genet.* 106, 1032–1040. doi:10.1007/s00122-002-1111-1

Gubler, F., Kalla, R., Roberts, J. K., and Jacobsen, J. V. (1995). Gibberellin-regulated Expression of a Myb Gene in Barley Aleurone Cells: Evidence for Myb Transactivation of a High-pI Alpha-Amylase Gene Promoter. *Plant Cell* 7, 1879–1891. doi:10.2307/387019510.1105/tpc.7.11.1879

Guo, W., Yang, H., Liu, Y., Gao, Y., Ni, Z., Peng, H., et al. (2015). The Wheat Transcription Factor TaGAMyb Recruits Histone Acetyltransferase and Activates the Expression of a High-Molecular-Weight Glutenin Subunit Gene. *Plant J.* 84, 347–359. doi:10.1111/tpj.13003

Hao, M., Zhang, L., Zhao, L., Dai, S., Li, A., Yang, W., et al. (2019). A Breeding Strategy Targeting the Secondary Gene Pool of Bread Wheat: Introgression from a Synthetic Hexaploid Wheat. *Theor. Appl. Genet.* 132, 2285–2294. doi:10.1007/s00122-019-03354-9

Hu, X.-g., Wu, B.-h., Yan, Z.-h., Liu, D.-c., Wei, Y.-m., and Zheng, Y.-l. (2010). Characterization of a Novel *1Ay* Gene and its Expression Protein in *Triticum Urartu*. *Agric. Sci. China* 9, 1543–1552. doi:10.1016/S1671-2927(09)60250-9

Hu, D. L., Chen, Q. Z., Zhang, C. J., Wang, Y., Zhang, B. J., and Tang, C. M. (2013). Identification of Cotton Skp1-Like Gene Ghskp1 and its Function in Seed Germination and Taproot Growth in Tobacco. *Canadian J. Plant Sci.* 93, 817–825. doi:10.1139/CJPS2012-312

Hung, J. H., and Weng, Z. (2016). Designing Polymerase Chain Reaction Primers Using Primer3Plus. *Cold Spring Harbor Protoc.* pdb, prot093096. doi:10.1101/pdb.prot093096

Jiang, H., and Wong, W. H. (2009). Statistical Inferences for Isoform Expression in RNA-Seq. *Bioinformatics* 25, 1026–1032. doi:10.1093/bioinformatics/btp113

Jiang, Z.-L., Wu, B.-H., Wang, Z.-Z., Hu, J.-L., Yuan, J., Chen, H.-L., et al. (2017). Enriching Novel *Glu-Ax* Alleles and Significantly Strengthening Gluten Properties of Common Wheat through Wide Hybridization with Wild Emmer. *J. Cereal Sci.* 76, 271–279. doi:10.1016/j.jcs.2017.04.018

Joppa, L. R., Du, C., Hart, G. E., and Hareland, G. A. (1997). Mapping Gene(s) for Grain Protein in Tetraploid Wheat (Triticum Turgidum L.) Using a Population of Recombinant Inbred Chromosome Lines. *Crop Sci.* 37, 1586–1589. doi:10.2135/cropsci1997.0011183X003700050030x

Kanehisa, M., Goto, S., Kawashima, S., and Okuno, Y. (2004). The KEGG Resource for Deciphering the Genome. *Nucleic Acids Res.* 32, 277D–280D. doi:10.1093/nar/gkh063

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

Krugman, T., Nevo, E., Beharav, A., Sela, H., and Fahima, T. (2018). The Institute of Evolution Wild Cereal Gene Bank at the University of Haifa. *Isr. J. Plant Sci.* 65, 129–146. doi:10.1163/22238980-00001065

Leng, N., Dawson, J. A., Thomson, J. A., Ruotti, V., Rissman, A. I., Smits, B. M. G., et al. (2013). EBSeq: an Empirical Bayes Hierarchical Model for Inference in RNA-Seq Experiments. *Bioinformatics* 29, 1035–1043. doi:10.1093/bioinformatics/btt087

Liu, J., Huang, L., Wang, C., Liu, Y., Yan, Z., Wang, Z., et al. (2019). Genome-wide Association Study Reveals Novel Genomic Regions Associated with High Grain Protein Content in Wheat Lines Derived from Wild Emmer Wheat. *Front. Plant Sci.* 10, 464. doi:10.3389/fpls.2019.00464

Onodera, Y., Suzuki, A., Wu, C.-Y., Washida, H., and Takaiwa, F. (2001). A Rice Functional Transcriptional Activator, *RISBZ1*, Responsible for Endosperm-Specific Expression of Storage Protein Genes Through GCN4 Motif. *J. Biol. Chem.* 276, 14139–14152. doi:10.1074/jbc.M007405200

Rangan, P., Furtado, A., and Henry, R. J. (2017). The Transcriptome of the Developing Grain: a Resource for Understanding Seed Development and the Molecular Control of the Functional and Nutritional Properties of Wheat. *BMC Genomics* 18, 766. doi:10.1186/s12864-017-4154-z

Rao, V., Petla, B. P., Verma, P., Salvi, P., and Majee, M. (2018). Arabidopsis Skp1-Like Protein13 (ASK13) Positively Regulates Seed Germination and Seedling Growth Under Abiotic Stress. *J. Exp. Botany*. 69, 3899–3915. doi:10.1093/jxb/ery191Saeedipour

Rombauts, S., Dehais, P., VanMontagu, M., and Rouze, P. (1999). Plantcare, a Plant Cis-Acting Regulatory Element Database. *Nucleic Acids Res.* 27, 295–296. doi:10.1093/nar/27.1.295

Saeed, S. (2012). Relationship of Endogenous Aba and Iaa to Accumulation of Grain Protein and Starch in two Winter Wheat Cultivars Under Post-Anthesis Water Deficit. *J. Agricult. Sci.* 4, 147–156. doi:10.5539/jas.v4n2p147

Schmidt, R. J., Ketudat, M., Aukerman, M. J., and Hoschek, G. (1992). Opaque-2 Is a Transcriptional Activator that Recognizes a Specific Target Site in 22-kD Zein Genes. *Plant Cell* 4, 689–700. doi:10.1105/tpc.4.6.689

Schmittgen, T. D. (2002). Real-time Quantitative PCR. *Methods* 25, 383–385. doi:10.1101/gr.6.10.986

Shen, L., Luo, G., Song, Y., Xu, J., Ji, J., Zhang, C., et al. (2021). A Novel NAC Family Transcription Factor *SPR* Suppresses Seed Storage Protein Synthesis in Wheat. *Plant Biotechnol. J.* 19, 992–1007. doi:10.1111/pbi.13524

Shewry, P. R., and Halford, N. G. (2002). Cereal Seed Storage Proteins: Structures, Properties and Role in Grain Utilization. *J. Exp. Bot.* 53, 947–958. doi:10.1093/jexbot/53.370.947

Shewry, P. R., Tatham, A. S., Barro, F., Barcelo, P., and Lazzeri, P. (1995). Biotechnology of Breadmaking: Unraveling and Manipulating the Multi-Protein Gluten Complex. *Nat. Biotechnol.* 13, 1185–1190. doi:10.1038/nbt1195-1185

Shewry, P. R., Underwood, C., Wan, Y., Lovegrove, A., Bhandari, D., Toole, G., et al. (2009). Storage Product Synthesis and Accumulation in Developing Grains of Wheat. *J. Cereal Sci.* 50, 106–112. doi:10.1016/j.jcs.2009.03.009

Strasser, R. (2018). Protein Quality Control in the Endoplasmic Reticulum of Plants. *Annu. Rev. Plant Biol.* 69, 147–172. doi:10.1146/annurev-arplant-042817-040331

Suzuki, A., Wu, C.-Y., Washida, H., and Takaiwa, F. (1998). Rice MYB Protein *OSMYB5* Specifically Binds to the AACA Motif Conserved Among Promoters of Genes for Storage Protein Glutelin. *Plant Cel Physiol.* 39, 555–559. doi:10.1093/oxfordjournals.pcp.a029404

Takemoto, Y., Coughlan, S. J., Okita, T. W., Satoh, H., Ogawa, M., and Kumamaru, T. (2002). The rice Mutant *Esp2* Greatly Accumulates the Glutelin Precursor and Deletes the Protein Disulfide Isomerase. *Plant Physiol.* 128, 1212–1222. doi:10.1104/pp.010624

Thomas, M., Lise, H., Manuel, S., and Matthias, P. (2014). A Chromosome-Based Draft Sequence of the Hexaploid Bread Wheat (*Triticum aestivum*) Genome. *Science* 345, 1251788. doi:10.1126/science.1251788

Uauy, C., Brevis, J. C., and Dubcovsky, J. (2006). The High Grain Protein Content Gene *Gpc-B1* Accelerates Senescence and Has Pleiotropic Effects on Protein Content in Wheat. *J. Exp. Bot.* 57, 2785–2794. doi:10.1093/jxb/erl047

Vicente-Carbajosa, J., Onate, L., Lara, P., Diaz, I., and Carbonero, P. (1998). Barley BLZ1: a bZIP Transcriptional Activator that Interacts with Endosperm-specific Gene Promoters. *Plant J.* 13, 629–640. doi:10.1111/j.1365-313X.1998.00068.x

Wan, Y., Liu, K., Wang, D., and Shewry, P. R. (2000). High-molecular-weight Glutenin Subunits in the *Cylindropyrum* and Vertebrata Section of the Aegilops Genus and Identification of Subunits Related to Those Encoded by the Dx Alleles of Common Wheat. *Theor. Appl. Genet.* 101, 879–884. doi:10.1007/s001220051556

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R Package for Identifying Differentially Expressed Genes from RNA-Seq Data. *Bioinformatics* 26, 136–138. doi:10.1093/bioinformatics/btp612

Wang, Y., Liu, F., Ren, Y., Wang, Y., Liu, X., Long, W., et al. (2016). GOLGI TRANSPORT 1B Regulates Protein export from the Endoplasmic Reticulum in rice Endosperm Cells. *Plant Cell* 28, 2850–2865. doi:10.1105/tpc.16.00717

Wang, Y. F., Huang, L., Luo, W., Jin, Y. R., Gong, F. Y., He, J. S., et al. (2020). Transcriptome Analysis Provides Insights into the Mechanisms Underlying Wheat Cultivar Shumai126 Responding to Stripe Rust. *Gene*. 768, 145290. doi:10.1016/j.gene.2020.145290

Wang, Z., Huang, L., Wu, B., Hu, J., Jiang, Z., Qi, P., et al. (2018). Characterization of an Integrated Active Glu-1ay Allele in Common Wheat from Wild Emmer and its Potential Role in Flour Improvement. *Int. J. Mol. Sci.* 19, 923. doi:10.3390/ijms19040923

Walterand Blobel. (1982). Signal Recognition Particle Contains A 7S RNA Essential for Protein Translocation Across the Endoplasmic Reticulum. *Nature*. 299, 691–698. doi:10.1038/299691a0

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Xi, D.-M., and Zheng, C.-C. (2011). Transcriptional Regulation of Seed Storage Protein Genes inArabidopsisand Cereals. *Seed Sci. Res.* 21, 247–254. doi:10.1017/S0960258511000237

Xiang, L., Huang, L., Gong, F., Liu, J., Wang, Y., Jin, Y., et al. (2019). Enriching LMW-GS Alleles and Strengthening Gluten Properties of Common Wheat through Wide Hybridization with Wild Emmer. *3 Biotech.* 9, 355. doi:10.1007/s13205-019-1887-1

Xie, Z., Dong, J., Dai, T., Qi, J., and Cao, W. (2004). Effects of Exogenous Aba and Cytokinin on Leaf Photosynthesis and Grain Protein Accumulation in Wheat Ears Cultured *In Vitro*. *Plant Growth Regul.* 44, 25–32. doi:10.1007/s10725-004-1880-4

Yu, Z., Islam, S., She, M., Diepeveen, D., Zhang, Y., Tang, G., et al. (2018). Wheat Grain Protein Accumulation and Polymerization Mechanisms Driven by

Nitrogen Fertilization. *Plant J.* 96, 1160–1177. doi:10.111/tpj.1409610.1111/tpj.14096

Zhang, L.-Q., Yen, Y., Zheng, Y.-L., and Liu, D.-C. (2007). Meiotic Restriction in Emmer Wheat Is Controlled by One or More Nuclear Genes that Continue to Function in Derived Lines. *Sex. Plant Reprod.* 20, 159–166. doi:10.1007/s00497-007-0052-x

Zhang, Z., Dong, J., Ji, C., Wu, Y., and Messing, J. (2019). NAC-type Transcription Factors Regulate Accumulation of Starch and Protein in maize Seeds. *Proc. Natl. Acad. Sci. USA* 116, 11223–11228. doi:10.1073/pnas.1904995116

Zheng, Y., Jiao, C., 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

Zhu, J., Hao, P., Chen, G., Han, C., Li, X., Zeller, F. J., et al. (2014). Molecular Cloning, Phylogenetic Analysis, and Expression Profiling of Endoplasmic Reticulum Molecular Chaperone BiP Genes from Bread Wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 14, 260. doi:10.1186/s12870-014-0260-0

Zhu, J., Fang, L., Yu, J., Ying, Z., and Xia, G. (2018). 5-Azacytidine Treatment and Tapbf-D Over-Expression Increases Glutenin Accumulation Within the Wheat Grain by Hypomethylating the Glu-1 Promoters. *Theoretical and Applied Genetics* 131, 735–746. doi:10.1007/s00122-017-3032-z

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 in Genetics

ORIGINAL RESEARCH
published: 15 March 2022
doi: 10.3389/fgene.2022.828292

Genome-Wide DNA Methylation and Its Effect on Gene Expression During Subclinical Mastitis in Water Buffalo

*Varij Nayan[1†], Kalpana Singh[2†], Mir Asif Iquebal[2†], Sarika Jaiswal[2], Anuradha Bhardwaj[3], Chhama Singh[1], Tanvi Bhatia[1], Sunil Kumar[1], Rakshita Singh[1], M. N. Swaroop[1], Rajesh Kumar[1], S. K. Phulia[1], Anurag Bharadwaj[1], T. K. Datta[1], Anil Rai[2] and Dinesh Kumar[2]**

[1]ICAR-Central Institute for Research on Buffaloes, Hisar, India, [2]Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistical Research Institute, New Delhi, India, [3]ICAR-National Research Centre on Equines, Hisar, India

Subclinical mastitis (SCM) in buffalo is one of the most challenging paradoxes for the dairy sector with very significant milk production losses and poses an imminent danger to milch animal's milk-producing ability. We present here the genome-wide methylation specific to SCM in water buffalo and its consequential effect on the gene expression landscape for the first time. Whole-genome DNA methylation profiles from peripheral blood lymphocytes and gene expression profiles from milk somatic cells of healthy and SCM cases were catalogued from the MeDIP-Seq and RNA-Seq data. The average methylation in healthy buffaloes was found to be higher than that in the SCM-infected buffaloes. DNA methylation was abundant in the intergenic region followed by the intronic region in both healthy control and SCM groups. A total of 3,950 differentially methylated regions (DMRs) were identified and annotated to 370 differentially methylated genes (DMGs), most of which were enriched in the promoter region. Several important pathways were activated due to hypomethylation and belonged to the *Staphylococcus aureus* infection, Th17 cell differentiation, and antigen processing and presentation pathways along with others of defense responses. DNA methylome was compared with transcriptome to understand the regulatory role of DNA methylation on gene expression specific to SCM in buffaloes. A total of 4,778 significant differentially expressed genes (DEGs) were extracted in response to SCM, out of which 67 DMGs were also found to be differentially expressed, suggesting that during SCM, DNA methylation could be one of the epigenetic regulatory mechanisms of gene expression. Genes like CSF2RB, LOC102408349, C3 and PZP like, and CPAMD8 were found to be downregulated in our study, which are known to be involved in the immune response to SCM. Association of DNA methylation with transposable elements, miRNAs, and lncRNAs was also studied. The present study reports a buffalo SCM web resource (BSCM2TDb) available at http://webtom.cabgrid.res.in/BSCM2TDb that catalogues all the mastitis-related information of the analyses results of this study in a single place. This will be of immense use to buffalo researchers to understand the host–pathogen interaction involving SCM, which is required in endeavors of mastitis control and management.

Keywords: water buffalo, subclinical mastitis, DNA methylation, gene expression, web resource

OPEN ACCESS

Edited by:
Rani Alex,
Indian Council of Agricultural Research (ICAR), India

Reviewed by:
Jatin Shrinet,
Florida State University, United States
Ikhide G. Imumorin,
Georgia Institute of Technology, United States

****Correspondence:***
Dinesh Kumar
dinesh.kumar@icar.gov.in

[†]These authors contributed equally to this work.

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *03 December 2021*
Accepted: *31 January 2022*
Published: *15 March 2022*

Citation:
Nayan V, Singh K, Iquebal MA, Jaiswal S, Bhardwaj A, Singh C, Bhatia T, Kumar S, Singh R, Swaroop MN, Kumar R, Phulia SK, Bharadwaj A, Datta TK, Rai A and Kumar D (2022) Genome-Wide DNA Methylation and Its Effect on Gene Expression During Subclinical Mastitis in Water Buffalo. Front. Genet. 13:828292. doi: 10.3389/fgene.2022.828292

INTRODUCTION

Water buffalo (*Bubalus bubalis*) has proven to be the "bank on hooves" by reshaping the landscape of agrarian livelihood in South and Southeast Asian Countries. In India, buffaloes contributed 49% (35% indigenous buffaloes and 14% nondescriptive buffaloes) towards the total milk pool amounting to a massive 91.817 million tonnes of milk. A whopping 44.76 million buffaloes were in milk with an average yield of 5.62 kg/day pan-India in the year 2018–2019 (Basic Animal Husbandry Statistics, 2019). However, mastitis remained a major constraint with huge production and economic losses. Mastitis is considered to be one of the expensive diseases affecting dairy cattle worldwide in terms of production losses (Sinha et al., 2014; Ruegg and Erskine, 2015; Beniæ et al., 2018). Mastitis is caused by varied pathogens leading to the development of subclinical/chronic (25%–65%) or clinical (~5%) infections (Dufour et al., 2012; AHDB, 2016; Wang et al., 2020) worldwide. The economic loss caused by subclinical mastitis (SCM) is often greater than that caused by clinical mastitis (Kirkpatrick and Olson, 2015). SCM is the inflammation of the mammary gland that does not create visible changes in the milk or the udder. It affects the dairy industry by reducing milk production, decreasing milk quality, and suppressing reproductive performance (Khan and Khan, 2006; Ahmad et al., 2011). An estimated loss of more than $1 billion per year was reported by the United States dairy industry in 1999 (Ott, 1999), increasing to $2 billion per year by 2006. The loss was estimated to be 48€/1,000 L in Ireland (Geary et al., 2013). In India, the annual loss due to mastitis has been estimated to the tune of Rs. 71,651.5 million per year (Sudhan and Sharma, 2010; Rao et al., 2017). Apart from economic losses, SCM has distinct importance in public health due to the risk of antibiotic resistance by consumption of milk with antibiotic residues accumulated due to the indiscriminate use of antibiotics (De Vliegher et al., 2012) for SCM treatment.

Staphylococcus aureus is the major cause of SCM in dairy cattle, which causes asymptomatic, persistent, antibiotic-resistant, and reoccurring infections (Song et al., 2016). SCM is caused by a variety of pathogens that can establish chronic infections and include *Escherichia coli*, *Pseudomonas aeruginosa*, *P. mendocina*, *S. chromogenes*, *S. epidermidis*, *Bacillus cereus*, *Klebsiella pneumoniae*, and *Shigella flexneri* (Hoque et al., 2015; Hoque et al., 2020). Similar causal organisms were also reported in buffalo mastitis (Fagiolo and Lai, 2007).

Previous studies showed the role of epigenetics in influencing traits related to health, growth, production, and development in cattle (Ibeagha-Awemu and Zhao, 2015; Reynolds et al., 2017; Sun et al., 2019). The potential contribution of epigenetic regulation in mechanisms of mastitis infection development, especially the role of DNA methylation in the regulation of mammary gland health, has been reported in the case of cattle (Song et al., 2016; Ju et al., 2020; Wang et al., 2020; Zhou et al., 2020). No such study has been reported in water buffalo. There is no web genomic resource for water buffalo mastitis disease with a list of differentially expressed genes (DEGs) and differentially methylated genes (DMGs) to help in understanding the molecular events.

DNA methylation is an important epigenetic regulator of gene expression and chromatin structure (Niazi et al., 2016) that provides stability to the genome by methylation of transposable elements (TEs). DNA methylation is catalyzed by a family of DNA methyltransferases (Dnmts) that forms 5-methylcytosine (5-mC) (Ruzov et al., 2011). Current genome-scale approaches for the determination of DNA methylation are largely based on the detection of 5-mC. MeDIP-Seq is a popular enrichment technique for the methylation status of cytosines captured by noncovalent bonding of 5-mC and antibodies (Taiwo et al., 2012; Li et al., 2015a). It has been used in numerous studies including the first mammalian methylome (Down et al., 2008) and the first cancer methylome (Feber et al., 2011).

Owing to the importance of water buffalo and losses caused by SCM, the present study focuses on the extraction of DMGs from whole-genome methylome (MeDIP-Seq) analysis and DEGs from transcriptome (RNA-Seq) analysis to understand the molecular mechanism of epigenetic regulation of gene expression involving DNA methylation specific to SCM in water buffalo. The present study also explores the association of DNA methylation with TEs, miRNAs, and lncRNAs. All the findings of the study are provided in the form of a user-friendly web resource, Buffalo Subclinical Mastitis Methylome–Transcriptome database (BSCM2TDb), available at http://webtom.cabgrid.res.in/BSCM2TDb. The present study is the first whole-genome methylome study specific to SCM in water buffalo of Murrah breed. The findings of this study can be used to understand the molecular regulation of mastitis disease as well as to identify candidate epigenetic markers related to the disease. It will help buffalo breeders in breed improvement and disease management programs.

MATERIALS AND METHODS

Ethics Statement

This study was approved by the Institute Animal Ethics Committee of the ICAR-Central Institute for Research on Buffaloes (ICAR-CIRB), Hisar.

Determination of SCM in Buffaloes

A total of 138 lactating Indian Murrah breed water buffaloes from the ICAR-CIRB animal farm (coordinates: 29o10′49.40″N, 75o42′24.87″E) were screened for the incidence of SCM. The milk samples were collected aseptically, noninvasively, and during the normal milking process. Milk was collected under clean and sterile conditions. After the initial milk from the teats was ignored, milk was aseptically collected in sterile containers without touching the container. The sample containers were clearly labeled as fore right (FR), fore left (FL), rear left (RL), and rear right (RR), bearing the animal numbers.

For ascertaining the cases of SCM, the criteria provided by the International Dairy Federation were adopted, and milk samples were subjected to somatic cell count (SCC), bacteriological examination, and antimicrobial sensitivity testing. The milk

samples were tested by the cow-side test of the California Mastitis Test (CMT). A total of eight CMT-positive cases for SCM were found. The milk samples found positive with CMT were subjected to SCCs, bacteriological culture, and sensitivity tests for ascertaining the confirmed SCM cases. SCCs were done for the CMT-positive samples and the healthy controls as well, using Newman's stain by adopting the method given by Schalm et al. (1971). SCC of positive cases found was done by making slides and counting manually. Cases with SCCs of $>5 \times 10^5$ cells/ml were designated as subclinical. For all positive cases with CMT, 10 ml milk samples were immediately sent for microbiological culture at the Department of Veterinary Microbiology, LUVAS, Hisar, India, for confirmation of mastitis, identification of microorganisms, and susceptibility to different antibiotics. For bacteriological studies, the milk samples were streaked on both 5% sheep blood agar and MacConkey's lactose agar plates and kept for incubation at 37°C for 24–48 h. The bacterial colonies were further subcultured on blood agar plates and identified by Gram's staining based on bacterial morphology and colony characteristics. Furthermore, the disc diffusion method (Bauer et al., 1966) was used for antimicrobial sensitivity test based on the zone-size interpretation chart and categorized as sensitive, intermediate, and resistant. The results were recorded for the antimicrobials, namely, enrofloxacin, penicillin G, streptomycin, amoxicillin, oxytetracycline, chloramphenicol, moxifloxacin, levofloxacin, ampicillin, gentamicin, neomycin, amikacin, cloxacillin, and cefoperazone.

Sample Collection, DNA Isolation, and Preparation of MeDIP-Seq Libraries

The blood samples were collected by jugular venipuncture from the same five SCM-infected lactating Murrah buffaloes (SCM 1–5) and six healthy Murrah buffaloes (C1–C6) as control in EDTA containing blood collection tubes. Furthermore, DNA isolation and purification were performed using the QIAamp blood DNA mini kit (Qiagen) in all samples from DNA Xperts Private Limited, Noida, India, utilizing the protocol adapted by Li et al. (2010) with few modifications and followed for preparing MeDIP-Seq libraries. Library preparation was done by using the NEBNext® Enzymatic Methyl-Seq Kit, followed by immunoprecipitation enriched for methylated DNA fragments using MeDIP buffer and 5-mC-specific monoclonal antibody. It was followed by a quality check by loading 1 µl of sample on an Agilent Technologies 2100 Bioanalyzer using a DNA-specific chip. Finally, 11 MeDIP-Seq libraries were obtained using the Illumina HiSeq 2500 instrument (Illumina Inc., United States) with 2× 50 bp paired-end (PE) sequencing.

Sample Collection, RNA Isolation, and Preparation of RNA-Seq Libraries

For RNA-Seq libraries, the milk samples were collected from six SCM-infected (SCM 1–6) udder quarters of five SCM-infected Murrah buffaloes and six healthy (C1–C6) udder quarters from six healthy Murrah buffaloes. These samples were centrifuged in 50 ml tubes at $1,500 \times g$ for 20 min at 4°C for total somatic cell isolation and later preserved. RNA was isolated by DNA Xperts Private Limited, Noida, India, using the QIAamp RNA blood mini kit (Qiagen) from suspended somatic cells. Furthermore, quality was checked using the Agilent 2100 Bioanalyzer, and electrophoresis was performed on formaldehyde with 1% agarose gel. The polyA-containing mRNA molecules were purified using oligo-dT attached magnetic beads. Finally, cDNA libraries were prepared as per the Illumina TruSeq RNA library preparation protocol (Illumina Inc., United States) and sequenced using the Illumina HiSeq 2500 instrument to obtain 12 RNA-Seq libraries (2× 100 bp PE).

MeDIP-Seq Data Analysis

Raw PE reads of MeDIP-Seq libraries were passed through a quality check using a FASTQC at Phred score ≥30 and GC distribution >40% and adapter trimming. The processed PE reads were aligned with the buffalo reference genome (GCF_003121395.1 available at https://www.ncbi.nlm.nih.gov/assembly/GCF_003121395.1/) using Bowtie 2 (Langmead and Salzberg, 2012). These aligned reads were used for the correlation analysis using deepTools (Ramirez et al., 2014) and identification of methylated peaks using MACS2 (Zhang et al., 2008) with a *p*-value of 0.01 in all the libraries. The extracted methylated peaks were used for genomic annotation of methylated peaks into genomic regions such as intron, exon, promoter (−1 kb to +100 bp), transcription termination sites (TTS) (−100 bp to +1 kb), and intergenic regions using HOMER (Heinz et al., 2010) utilizing a buffalo RefSeq annotation file. The DMRs in the SCM group were compared to those of the control group using diffReps (Shen et al., 2013) at a *p*-value of 0.001 and $|\text{Log2FC}| \geq 2$. Later, DMGs were extracted from DMRs utilizing the buffalo RefSeq annotation file by the Perl script. Functional annotation of DMGs was performed by gene ontology (GO) analysis using BLAST2GO (Conesa and Götz, 2008), and KEGG pathway analysis was performed using the clusterProfiler package (Yu et al., 2012) of Bioconductor for DMGs at a *p*-value <0.1.

RNA-Seq Data Analysis

For RNA-Seq data analysis, all 12 RNA-Seq data libraries were passed through quality check by FASTQC v0.11.5 (Andrews, 2010) using the parameters of a Phred score ≥20 and GC distribution >40%. The bases with a <20 Phred score were trimmed using FASTX-Toolkit v0.0.14 (https://github.com/agordon/fastx_toolkit). This was followed by transcriptome assembly using Trinity v2.2.0 (Grabherr et al., 2011). The differentially expressed regions (DERs) were extracted using DESeq2 of the R package v1.26.0 (Love et al., 2014). For the extraction of DEGs, open reading frames (ORFs) were predicted using TransDecoder v5.5.0 (Haas and Papanicolaou, 2019), and annotation was performed using Trinotate v3.2.0 at $|\log_2\text{FC}| \geq 03$ and FDR <0.05 (Bryant et al., 2017). Alignment of DMGs with the buffalo reference genome was performed to extract the coordinate information using BLAST with a parameter e-value of 1e−30 using the Perl script. Finally, the in-house Perl scripts were used to get buffalo gene IDs to compare with DMGs with the help of the coordinate information within the buffalo RefSeq annotation file.

To understand the effect of DNA methylation on gene regulation during SCM in buffalo, the results of DNA methylation analysis (MeDIP-Seq) were compared with the results of transcriptome analysis (RNA-Seq). The DMGs and

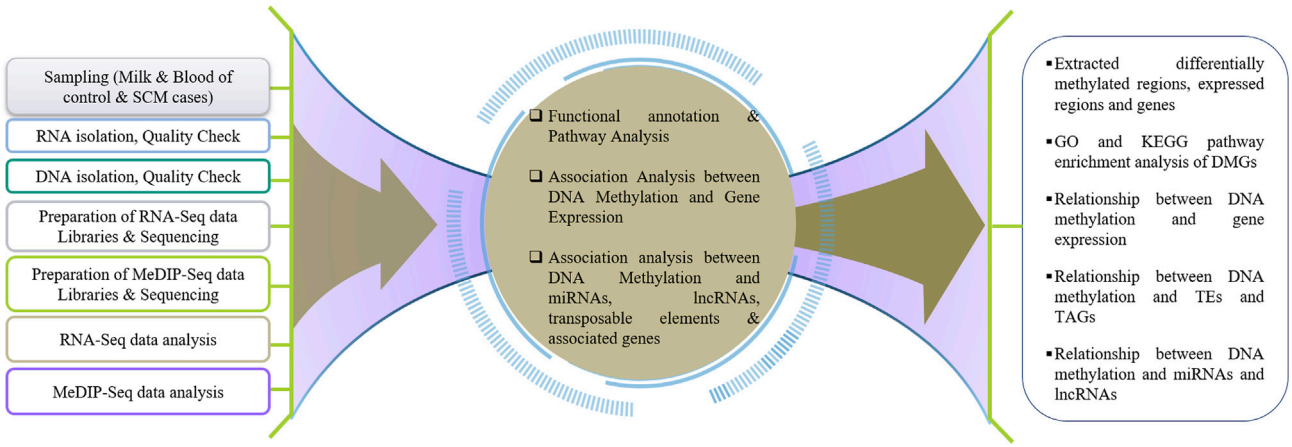


FIGURE 1 | Schematic workflow towards understanding the genome-wide DNA methylation and its effect on gene expression during SCM in water buffalo.

DEGs were compared to find the genes repressed by hypermethylation and expressed by hypomethylation, i.e., a negative relationship as reported by Lou et al. (2014) in the promoter and TTS regions, while a positive relationship in case of exon and intron as reported by Suzuki and Bird (2008), Ball et al. (2009), and Lev Maor et al. (2015). **Figure 1** represents the pipeline for understanding genome-wide DNA methylation and its effect on gene expression during SCM in buffalo.

Association of Methylation with TEs, lncRNAs, and miRNAs

In order to study the role of DNA methylation in genome stability, TEs were searched, which were found to overlap with DMRs using Censor (Huda and Jordan, 2009). The probable TE-associated genes (TAGs) were also extracted by comparing TEs with the buffalo RefSeq annotation file using the Perl script. Furthermore, pre-miRNA sequences of known buffalo miRNAs were mapped with DMRs using BLAST (Altschul et al., 1990) to see the methylation in putative miRNA genes, which were termed as methylation-regulated miRNA genes. Methylation-regulated miRNAs were used to study their target mRNAs from DEGs specific to SCM by using psRNATarget (Dai et al., 2018) to see the consequent indirect effect of DNA methylation through the regulation of miRNA expression. The psRNATarget also predicts the mode of action of miRNAs on their target such as cleavage or binding, which disables their targets for further action. Moreover, functional annotation showed that a few DMGs were encoding lncRNAs, which were considered methylated lncRNA genes, and their respective lncRNAs were termed as methylation-regulated lncRNAs. Targets of methylation-regulated lncRNAs were identified using LncTar (Li et al., 2015b) to see the consequent indirect effect to DNA methylation through regulation of lncRNA expression.

Development of the Web Resource, BSCM2TDb

BSCM2TDb is a three-tier architecture-based relational database, freely accessible at http://webtom.cabgrid.res.in/BSCM2TDb/. All the analyses results like DMRs, DMGs, DM-lncRNAs (methylation-regulated lncRNAs), DM-miRNAs (methylation-regulated miRNAs), DM-TEs, and DM-TAGs from MeDIP-Seq data analysis along with DEGs and DM-DEGs from RNA-Seq data analysis were catalogued and stored in the backend in a MySQL database. The web interface was developed in PHP and launched by the Apache 2 server. BSCM2TDb contains tabs, *namely*, Home, Statistics, Data, Tutorial, and Team.

RESULTS

MeDIP-Seq Data Analysis

A total of ~745 million reads were found in the MeDIP-Seq libraries with an average of ~67.7 million reads in each library. A total of 33.8 million PE reads passed through quality check with an average of 47% GC content (**Supplementary Table S1**). An average of ~71% alignment was attained for all libraries with the buffalo RefSeq genome. Pearson's correlation coefficients among all the libraries were in the range of 0.71–0.99 (**Supplementary Figure S1**).

Peak Calling and their Genomic Annotation

A total of 189,474 peaks were extracted from the control group with an average of 31,579 peaks, while 154,803 peaks were extracted from the SCM group with an average of 30,960 peaks (**Supplementary Table S2**). The result of genomic annotation of these peaks showed that the maximum number of peaks belonged to the intergenic region (75%–76%), followed by the intron (~14%), promoter (~5%), TTS (~4%), and exon (<1%) (**Supplementary Figure S2**).

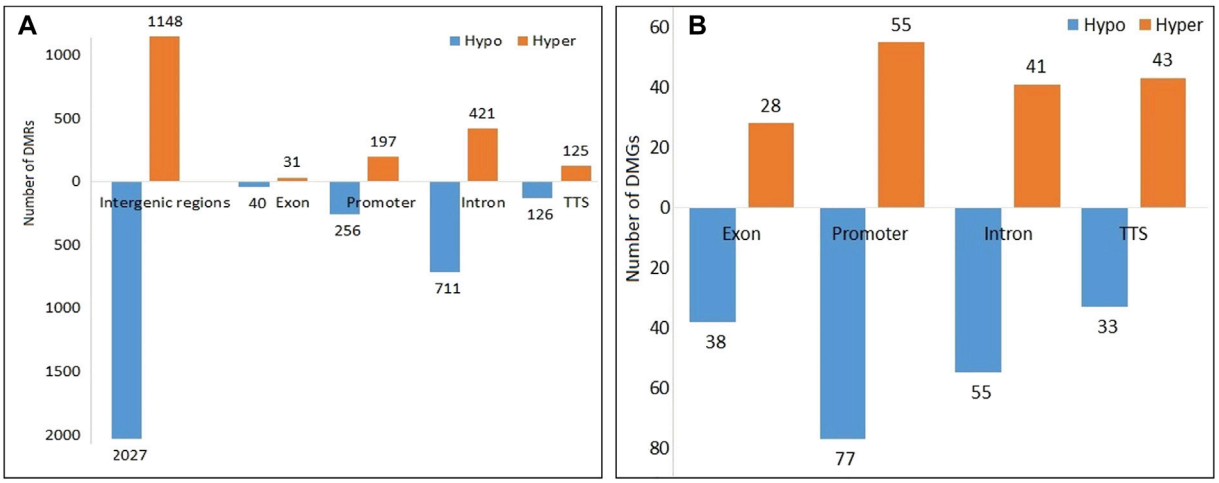


FIGURE 2 | Hypomethylated and hypermethylated **(A)** DMRs and **(B)** DMGs in the SCM group in comparison to the control group.

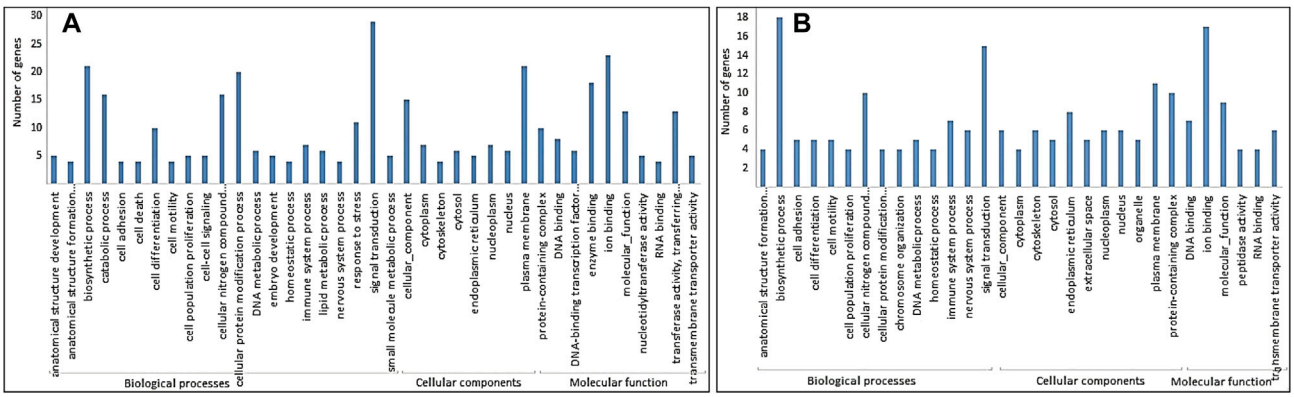


FIGURE 3 | GO terms in three classes: biological processes, cellular components, and molecular function for **(A)** hypomethylated DMGs and **(B)** hypermethylated DMGs.

Identification of DMRs and DMGs

A total of 3,950 DMRs were extracted from the peaks of the SCM group with respect to peaks of the control group, out of which 2,451 DMRs were hypomethylated while the remaining 1,449 were hypermethylated. The distribution of methylation counts in the form of histograms is shown in a Circos diagram (**Supplementary Figure S3**) found within DMRs for the control and SCM groups in all 25 chromosomes. Out of 3,950 DMRs, 71, 1,132, 2,043, 251, and 453 DMRs belonged to exon, intron, intergenic regions, TTS, and promoter regions, respectively. **Figure 2A** shows hypomethylated and hypermethylated DMRs within various genomic regions.

A total of 370 DMGs were extracted from 3,950 DMRs, out of which 169 DMGs were hypermethylated and 201 DMGs were hypomethylated. Genomic distribution of 370 DMGs showed that 66, 132, 96, and 76 DMGs belonged to exons, promoters, introns, and TTS regions, respectively (**Figure 2B**). Functionally, it was found that out of these 370 DMGs, 140, 217, and 1 DMGs were encoding for lncRNAs, proteins, and snRNA, respectively, along with the remaining 12 pseudogenes.

GO and KEGG Pathway Enrichment Analyses of DMGs

The GO analysis of DMGs showed that the biological process was the largest class in both hypomethylated and hypermethylated DMGs. Molecular function was the least abundant in hypermethylated DMGs, while cellular component was the least abundant in hypomethylated DMGs (**Figures 3A,B**). More biological processes were enriched with hypomethylated DMGs than with hypermethylated DMGs. The most enriched GO terms in hypermethylated DMGs in decreasing order were biosynthetic process, ion binding, signal transduction, cellular nitrogen compound, plasma membrane, and protein-containing complex. The most enriched GO terms in hypomethylated DMGs

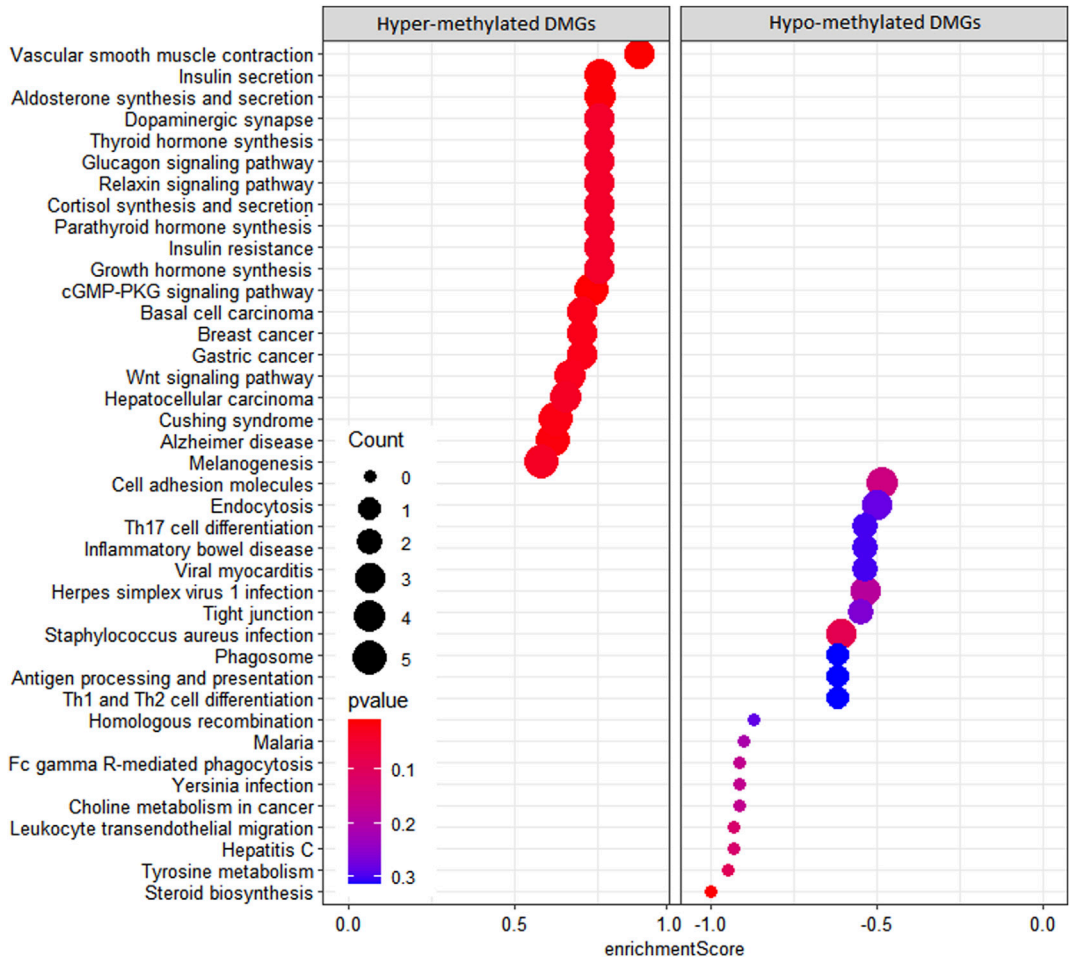


FIGURE 4 | Dot plot of KEGG pathways enriched by hypermethylated and hypomethylated DMGs.

in decreasing order were signal transduction, ion binding, biosynthetic process, plasma membrane, cellular protein modification process, enzyme binding, catalytic process, cellular nitrogen compound, cellular function, and transferase activity.

In the KEGG pathway enrichment analysis, 142 pathways were found for 367 DMGs at a p-value <0.1, out of which 40 most enriched pathways were represented by hypomethylated and hypermethylated DMGs (20 each) as shown in the form of a dot plot (**Figure 4**). The network of pathways showed the 14 most enriched DMGs involved in 20 enriched pathways (**Figure 5**). Phospholipase C beta 3 (PLCB3):102413748 was found to be connected to 13 pathways, cAMP-responsive element-binding protein 3 like 1 (CREB3L1):102406078 and (CREB3):102395032 both connected to 12 pathways, frizzled-1-like (LOC102414897):102414897 connected to eight pathways, Wnt family member 5A (WNT5A): 102403744 connected to eight pathways, and growth arrest and DNA damage-inducible alpha (GADD45A):102401934 connected to four pathways, which were the most enriched DMGs in the top networks (**Figure 5**).

RNA-Seq Data Analysis and Identified DEGs

In all 12 RNA-Seq libraries, a total of ~358 million PE reads were found with an average of ~15 million PE reads, which were aligned with an average of 75.65% (**Supplementary Table S1**). Furthermore, a total of 21,028 DERs were extracted from the SCM group with respect to the control group, out of which 8,408 and 12,620 DERs were downregulated and upregulated, respectively. While annotating these DERs with the buffalo RefSeq annotation file, 4,778 DEGs were extracted, out of which 2,908 and 1,870 DEGs were downregulated and upregulated, respectively.

Association of DNA Methylation with Gene Expression

A comparative study of DMGs from MeDIP-Seq analysis and DEGs from RNA-Seq analysis showed a total of 67 DMGs to have differential methylation along with differential expression. Their level of methylation along with the level of expression is shown in **Supplementary Figure S4**. Out of 67 DMGs, 33 (49.25%) were found to have gene expression as per methylation; i.e., 25 DMGs were negatively correlated with gene expression having methylation

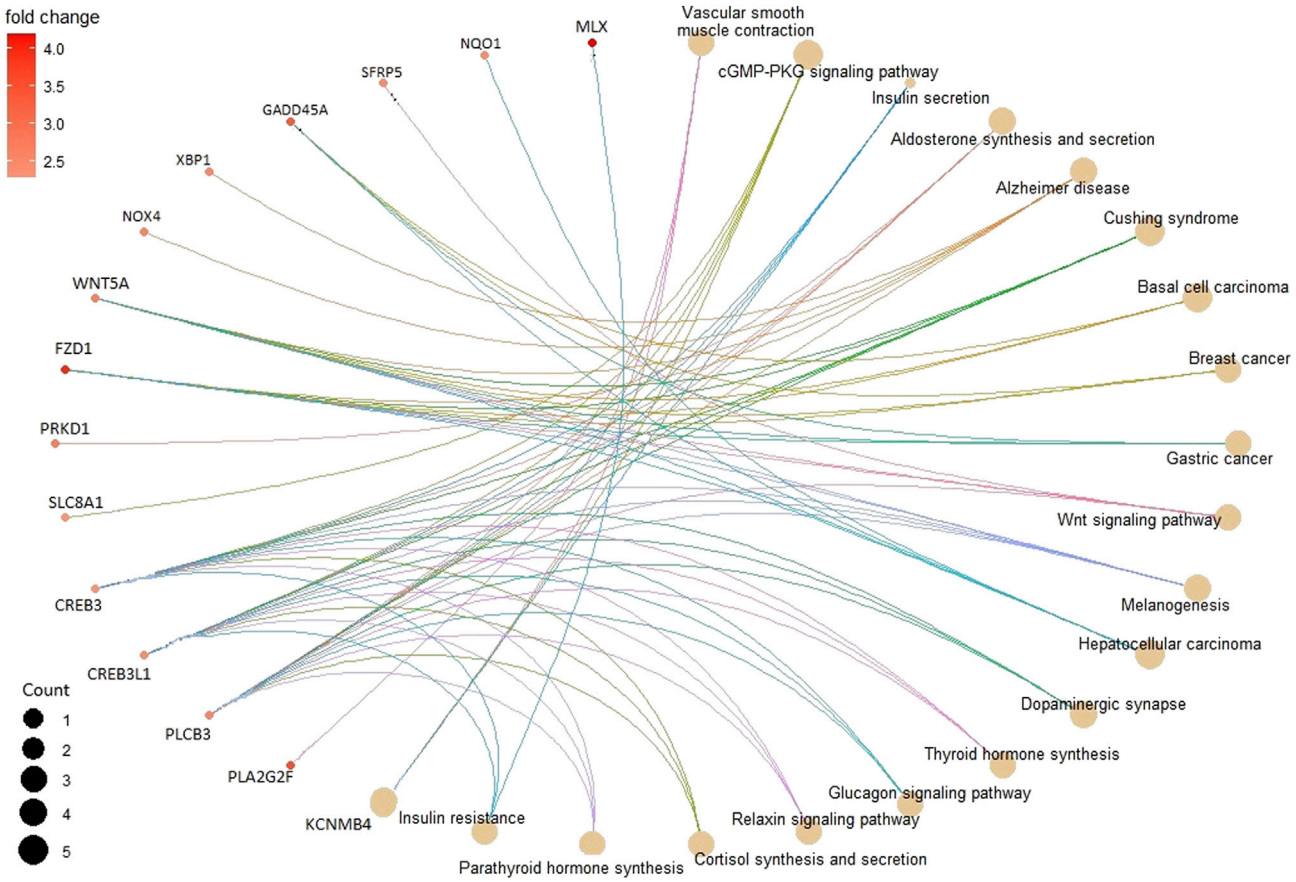


FIGURE 5 | Pathway network showing the 14 most enriched DMGs (in smaller circles) connected to 20 enriched pathways (in bigger brown circles) depicting their relation.

TABLE 1 | Hypomethylated and hypermethylated DMGs and their expression due to methylation in different genomic regions.

DNA methylation/gene expression	Exon	Promoter	Intron	TTS
Hypomethylated/downregulated	2	9	0	4
Hypermethylated/upregulated	2	2	4	8
Hypomethylated/upregulated	3	9	3	1
Hypermethylated/downregulated	5	10	0	5

in promoter and TTS regions and the remaining eight DMGs were positively correlated with gene expression having methylation in genic regions (exon and intron) (**Tables 1** and **2**).

Association of DNA Methylation with TEs and TAGs

A total of 3,377 TEs were found within DMRs, out of which 949 TEs belonged to introns, 52 (44 hypomethylated and 8 hypermethylated) to exons, 336 (241 hypomethylated and 95 hypermethylated) to promoters, 155 (109 hypomethylated and 46 hypermethylated) to TTS, and 1,896 to intergenic regions. These TEs belonged to eight major classes and their included subclasses along with their frequencies (**Table 3**) such as (1) DNA 1,022; (2) ERV 445; (3) integrated virus 39; (4) interspersed repeat 35; (5) LTR 570; (6) multicopy genes 6; (7) non-LTR 2,204; and (8) simple 5. The largest class found was non-LTR, and the largest subclass found was SINE, followed by LTR, Gypsy, hAT, RTE, and ERV3.

A total of 132 TAGs were extracted, out of which 100 and 32 TAGs were hypomethylated and hypermethylated, respectively (**Table 4**). SINE TEs were the most frequent, followed by Gypsy in hypomethylated TAGs. L1 and Gypsy TEs were more frequent in hypermethylated TAGs. In case of some TAGs, more than one and different type of TEs were found associated with a single TAG.

Association of DNA Methylation with miRNAs and their Target mRNAs

A total of eight methylation-regulated miRNAs were identified (**Supplementary Table S3**), out of which six miRNAs were targeting 44 mRNAs (transcribed from DEGs). The mode of action of miRNAs showed that 29 mRNAs were targeted by cleaving and 16 mRNAs were targeted by binding (**Table 5**). A total of five

TABLE 2 | Coordinated relation of DMGs with DNA methylation and gene expression according to genomic regions.

DNA methylation/gene expression	Gene IDs and their protein product
Hypo/Down in exon	espin-like (ESPNL), probable phospholipid-transporting ATPase IA-like (LOC102398473)
Hypo/Up in promoter	MAF bZIP transcription factor K (MAFK), selectin E (SELE), cryptochrome circadian clock 1 (CRY1), plastin 3 (PLS3), StAR-related lipid transfer domain containing 10 (STARD10), zinc finger protein 684 (ZNF684), calcium voltage-gated channel auxiliary subunit beta 2 (CACNB2), AT-rich interaction domain 5A (ARID5A), proline rich 15 like (PRR15L)
Hypo/Up in TTS	phosphatidylinositol-4-phosphate 5-kinase type 1 beta (PIP5K1B)
Hyper/Down in promoter	mediator complex subunit 25 (MED25), biogenesis of lysosomal organelles complex 1 subunit 6 (BLOC1S6), cAMP responsive element binding protein 3 like 1 (CREB3L1), FRY microtubule binding protein (FRY), protein kinase C-binding protein 1-like (LOC102401961), VPS11, CORVET/HOPS core subunit (VPS11), colony stimulating factor 2 receptor beta common subunit (CSF2RB), granulocyte-macrophage colony-stimulating factor receptor subunit alpha-like (LOC102408349), C3 and PZP like, alpha-2-macroglobulin domain containing 8 (CPAMD8), Myb like, SWIRM and MPN domains 1 (MYSM1)
Hyper/Down in TTS	TBC1 domain family member 9B (TBC1D9B), uncharacterized (LOC102390914), ARP8 actin-related protein 8 homolog (ACTR8), zinc finger and BTB domain containing 20 (ZBTB20), DDB1 and CUL4 associated factor 6 (DCAF6)
Hyper/Up in exon	Glycosyltransferase 1 domain containing 1 (GLT1D1)
Hyper/Up in intron	Uncharacterized (LOC102400551), uncharacterized (LOC102406144)

TABLE 3 | TEs in DMRs categorized into classes and their subclasses along with their frequencies.

S. No	Class and number of TEs	Subclasses of TEs and their frequencies
1	DNA: 1,022	DNA-68, Academ-11, Crypton-17, Dada-8, Enspm/CACTA-91, Ginger-7, Harbinger-64, hAT-306, Helitron-87, IS3EU-4, ISL2EU-14, Kolok-16, Mariner-129, Merlin-4, MuDR-82, P-11, Novosib-4, PiggyBac-19, Polinton-44, Sola-24, Transib-7, Zator-2, Zisupton-3
2	ERV: 445	ERV-17, ERV1-152, ERV2-29, ERV3-246, ERV4-1
3	Integrated virus: 39	Caulimovirus-29, DNAV-10
4	Interspersed repeat: 35	-
5	LTR: 570	LTR-13, BEL-24, Copia-160, DIRS-19, Gypsy-354
6	Multicopy gene: 6	Multicopy_gene-4, rRNA-2
7	Non-LTR: 2,204	Non-LTR-33, L1-517, L2-30, R1-13, R2-1, R4-5, Rex1-4, Loa-6, NeSL-1, CR1-198, Crack-5, CRE-4, Daphne-20, Hero-5, I-6, Ingi-1, Jockey-14, Kiri-6, Nimb-11, Outcast-3, Penelope-28, RTE-249, RETX-10, SINE-994, Tad1-8, Tx1-29, Vingi-3
8	Simple: 5	Sat-5

hypomethylation-regulated miRNAs were found positively correlated with 28 downregulated target DEGs (due to upregulation of miRNA) and one hypermethylation-regulated miRNA was positively correlated with one upregulated target DEG (due to downregulation of miRNA). The remaining 15 target DEGs did not show a positive correlation with methylation in miRNA genes.

Association of DNA Methylation with lncRNA Genes and their Target mRNAs

A total 140 methylated lncRNA genes were found transcribing 284 lncRNAs, out of which nine lncRNAs from seven lncRNA genes were found targeting 209 mRNAs (transcribed from DEGs) encoding 126 proteins, 2 tRNAs, and 2 snRNAs (**Table 6**). While analyzing targets of methylation-regulated miRNAs in methylation-regulated lncRNAs, it was found that only one methylated miRNA (bta-miR-12022) was sequestering the activity of six other methylated lncRNAs by binding with them (**Supplementary Table S4**). These lncRNAs were found to have methylation in the genic regions (exon and intron).

Development of the Web Resource, BSCM2TDb

The BSCM2TDb web resource has four main tabs, namely, Home, Statistics, Data, and Team (**Figure 6**). The *Home* page has a brief introduction about the database. The navigation key *Statistics* included a pie chart, showing the proportion of all included data, i.e., 7,900 DMRs, 370 DMGs, 208 DMG-KEGG pathways, 8 DM-miRNAs, 138 DM-lncRNAs, 3,377 DM-TEs, 131 DM-TAGs, 4,638 DEGs, and 64 DM-DEGs. The *Data* page is the main analyses result page that provides the options in the drop-down menu to navigate to the complete table of selected option. The *Team* page included the team member's name and link to the profile page of each member.

DISCUSSION

SCM is an inflammatory disease in water buffaloes that causes major losses to the dairy industry. The present study is the first genome-wide DNA methylation (MeDIP-Seq) study to compare an SCM group with a healthy group to understand the regulation of genes involving DNA methylation during host–pathogen

TABLE 4 | Hypomethylated and hypermethylated TAGs along with the frequencies of TE subclasses.

Class and number of TEs in TAGs	Hypomethylated TAGs	Hypermethylated TAGs
DNA: 52	DNA-5, P-1, Crypton-1, Dada-1, CACTA-6, Harbinger-3, hAT-15, Helitron-5, ISL2EU-2, Mariner-3, MuDR-4, PiggyBac-1, Novosib-1, Polinton-3, Sola-1	CACTA-2, Harbinger-4, hAT-4, Helitron-1, ISL2EU-1, Mariner-1, MuDR-1
ERV: 23	ERV-1, ERV1-6, ERV2-4, ERV3-9	ERV1-4
Integrated virus: 1	Caulimovirus-1	0
Interspersed repeat: 2	2	0
LTR: 37	LTR-1, BEL-1, Copia-5, DIRS-2, Gypsy-20	Copia-2, Gypsy-6
Multicopy gene: 2	Multicopy_gene-2	0
Non-LTR: 82	Non-LTR-1, L1-16, L2-1, R1-1, R2-1, RTE-7, CR1-3, Jockey-1, Kiri-1, Nimb-1, SINE-44, Tx1-3	CR1-5, Rex1-1, I-2, L1-6, L2-2, Outcast-1, Penelope-1, RTE-2, SINE-2

TABLE 5 | Target DEGs of miRNAs transcribed from DMRs along with their encoded proteins.

miRNAs	Log_2FC of DMRs	Target DEGs	Log_2FC of DEGs	Product of DEGs
bta-mir-2285cq	−2.16	ACTR3	−22.82	ARP3, actin related protein 3 homolog
bta-mir-12022	−2.37	ALG2	−6.53	ALG2, alpha-1,3/1,6-mannosyltransferase
bta-mir-2285cq	−2.16	ATP8A1	−6.45	ATPase phospholipid transporting 8A1
bta-mir-12022	−2.37	ATXN1	−6.18	ataxin 1
bta-mir-12022	−2.37	C7	−4.72	complement C7
bta-mir-12022	−2.37	DIS3L2	−4.41	DIS3 like 3′-5′ exoribonuclease 2
bta-mir-12022	−2.37	DLGAP4	−4.14	DLG-associated protein 4
bta-mir-12063	−2.61	DPY19L3	−4.05	dpy-19 like C-mannosyltransferase 3
bta-mir-12022	−2.37	ECM1	−4.017	extracellular matrix protein 1
bta-mir-2285cq	−2.16	EI24	−3.97	EI24, autophagy-associated transmembrane protein
bta-mir-12022	−2.37	EIF4E3	−3.93	eukaryotic translation initiation factor 4E family member 3
bta-mir-12063	−2.61	EPAS1	−3.8	endothelial PAS domain protein 1
bta-mir-2285cq	−2.16	ERAP1	−3.61	endoplasmic reticulum aminopeptidase 1
bta-mir-12063	−2.61	FAM227B	−3.53	family with sequence similarity 227 member B
bta-mir-12022	−2.37	FBXW7	−3.49	F-box and WD repeat domain containing 7
bta-mir-2285cq	−2.16	FOXK1	−3.35	forkhead box K1
bta-mir-2285cq	−2.16	GIMAP1	−3.22	GTPase, IMAP family member 1
bta-mir-12022	−2.37	GLYR1	−3.15	glyoxylate reductase 1 homolog
bta-mir-12063	−2.61	IKBKB	−3.05	inhibitor of nuclear factor kappa B kinase subunit beta
bta-miR-10161–5p	2.35	ITPR2	−2.85	inositol 1,4,5-trisphosphate receptor type 2
bta-mir-2285cq	−2.16	KIAA0232	−2.84	KIAA0232 ortholog
bta-miR-10161–5p	2.35	LOC102412044	−2.82	L-lactate dehydrogenase A chain
bta-mir-2285cq	−2.16	LOC102415248	−2.79	killer cell lectin-like receptor subfamily I member 1
bta-mir-2285cq	−2.16	LRP11	−2.58	LDL receptor related protein 11
bta-mir-11986	−2.76	LSMEM1	−2.37	leucine rich single-pass membrane protein 1
bta-miR-126–5p	−2.08	MAPKAP1	−2.28	mitogen-activated protein kinase associated protein 1
bta-mir-12063	−2.61	NAP1L1	−1.63	nucleosome assembly protein 1 like 1
bta-mir-12063	−2.61	NEXN	−1.43	nexilin F-actin binding protein
bta-mir-2285cq	−2.16	PLXNA2	1.54	plexin A2
bta-mir-12022	−2.37	PON1	1.75	paraoxonase 1
bta-mir-12063	−2.61	PPHLN1	1.9	periphilin 1
bta-mir-12022	−2.37	PPIL4	2.23	peptidylprolyl isomerase like 4
bta-mir-12022	−2.37	RCN1	2.56	reticulocalbin 1
bta-mir-12063	−2.61	RMC1	2.9	regulator of MON1-CCZ1
bta-mir-12063	−2.61	RRP1B	2.9	ribosomal RNA processing 1B
bta-mir-11986	−2.76	SCARB1	2.97	scavenger receptor class B member 1
bta-mir-12063	−2.61	SENP3	3.32	SUMO specific peptidase 3
bta-mir-2285cq	−2.16	SIN3A	3.68	SIN3 transcription regulator family member A
bta-mir-12063	−2.61	SYNE3	4.05	spectrin repeat containing nuclear envelope family member 3
bta-mir-12063	−2.61	TFEC	4.76	transcription factor EC
bta-mir-12022	−2.37	TNRC6A	5.31	trinucleotide repeat containing 6A
bta-mir-12063	−2.61	TNRC6B	5.64	trinucleotide repeat containing 6B
bta-miR-10161-5p	2.35	USP34	6.06	ubiquitin specific peptidase 34
bta-mir-12022	−2.37	ZRANB1	7.31	zinc finger RANBP2-type containing 1

TABLE 6 | Target DEGs of lncRNAs transcribed from methylated genes.

Methylated lncRNA genes		Target DEGs
LOC112583939, LOC112584670, LOC112585162, LOC112585197, LOC102413993, LOC112585597, LOC102408241	Proteins	PRELID3B, MICU3, LOC112577670, CABCOCO1, SENP6, RBM38, FAM219B, GXYLT2, DNAJC13, RIF1, BBS10, LOC112584770, LOC102415513, THNSL1, LOC102405919, CDK19, PTHLH, LOC102399155, FKBP3, STRN3, SEC11C, CENPU, PAXBP1, ATP5PF, BTG3, PFN2, UBE2G2, NCKAP1, CD302, EPC2, SUZ12, PPM1D, SAP30, HPGD, PLGRKT, PLEKHA5, AMN1, LOC102406990, PTGES3, SLC25A16, PCBD1, LOC102399263, CNIH4, TMEM262, TIPRL, HSD17B7, ZNF644, NEXN, SRSF11, SPATA6, ATPAF1, RRAGC, PI4K2B, ZNRF2, SRI, SLC25A46, FAM174A, RNF130, MPC1, QKI, CD24, SLIRP, FAM214A, PIGB, TPM1, AP3S1, RPS27A, VRK2, GEN1, GPR180, NDFIP2, PAN3, WFDC2, SYS1, COMMD7, HACD1, TRDMT1, IMPAD1, NSMCE2, FDX1, FAM76B, AASDHPPT, ISCU, FGF2, CBFB, EIF5, FAM177A1, RCN2, BTBD1, CMC1, LOC102402381, TXNL1, C22H18orf54, MBD2, RNF138, UBE2D1, KDELR2, NTAN1, DEXI, TTC14, KPNA4, SLC35B3, AGPS, RHOT1, ETNK1, FAM118B, RIIAD1, PTGR2, RPL31, PCSK7, STX2, KLC1, G2E3, PXK, FAM210A, DMD, RAB5A, ERP44, IVNS1ABP, SH2D1B, CCDC18, MKLN1, MAPKAP1, ZNF26, PPP1R37, GDNF
	snRNAs	LOC112578276, LOC112581316
	tRNAs	TRNAN-GUU, TRNAY-AUA

[1]*https://github.com/agordon/fastx_toolkit.*
[2]*https://www.ncbi.nlm.nih.gov/assembly/GCF_003121395.1/.*

interaction specific to SCM in buffaloes. MeDIP-Seq is a popular technique for the extraction of genome-wide DNA methylation, which has been used in numerous SCM studies on mammals including bovine (Song et al., 2016; Chen et al., 2019). A high correlation was found among MeDIP-Seq libraries, which is due to the sample taken from the buffaloes of the same genotype. The average methylation in healthy buffaloes was higher than that in SCM-infected buffaloes, which suggests that the upregulation and activation of certain genes may be required for the development of SCM. DNA methylation was found mostly in the intergenic region followed by intronic region in both healthy control and SCM groups, which is in agreement with other studies as well (Song et al., 2016; Fang et al., 2017; Wang et al., 2020). The DNA methylation level was significantly higher in the gene body than in upstream and downstream regions of genes; similar results were found by Wang et al. (2020).

We observed more hypomethylation in the SCM group than in the control group, which is in concordance with the results found in neutrophils of *E. coli* mastitis cows (Ju et al., 2020) and further suggests that the upregulation of certain genes may be required for the development of SCM during host–pathogen interaction. The 3,950 DMRs identified were annotated to 370 DMGs, and most of the DMGs were enriched in the promoter region, which is in concordance with earlier reports (Irizarry et al., 2009; Song et al., 2016). Some DMGs were found related to the immune system like colony-stimulating factor 2 receptor beta common subunit (CSF2RB) and granulocyte-macrophage colony-stimulating factor receptor subunit alpha-like (LOC102408349), which are cytokines considered essential for the survival, proliferation, and differentiation of blood cells such as granulocyte and macrophages as reported by Jeong et al. (2014). Additionally, antigen WC1.1-like (LOC102391350) expresses on subsets of $CD4^{-}CD8^{-}$ gamma delta T lymphocytes (Wijngaard et al., 1994), and V-set domain-containing T-cell activation inhibitor 1 (VTCN1) negatively regulates T-cell-mediated immune response. The DLA class II histocompatibility antigen DR-1 beta chain-like (LOC102389065) involves in the adaptive immune response. From results of the KEGG pathway analysis, interestingly, it was found that the *S. aureus* infection pathway was found significantly activated due to hypomethylation, suggesting that a major causal pathogen for SCM in this study could also be *S. aureus* and that there is a possible regulatory role of DMGs in the host response to *S. aureus*-induced mastitis, which is in agreement with the previous study involving SCM in cows (Wang et al., 2020). Additionally, important pathways activated due to hypomethylation related to immunity were Th17 cell differentiation and antigen processing and presentation, along with activation of pathways related to defense response such as Fc gamma R-mediated phagocytosis, phagosome, and leukocyte transendothelial migration. There is an important role of Th17 cell differentiation in the immune regulation of T cells (Sordillo, 2011) and *S. aureus* mastitis (Zhao et al., 2015; Wang et al., 2020). In the present study, Th17 cell differentiation pathway activation in response to SCM is further confirmation of the significance of Th17 cells in host immune response and regulation mediated by DNA methylation during SCM. The activated KEGG pathway related to cell adhesion molecules confirms previous studies involving the role of methylation in cell adhesion, which influences immune cell function during host–pathogen interaction (Li et al., 2018; Wang et al., 2020).

Furthermore, DNA methylation was compared with transcriptome data in the present study to understand the effect of DNA methylation on gene expression in response to SCM in buffaloes. A total of 4,778 significant DEGs were extracted specific to SCM, out of which 67 genes were also found differentially methylated along with differentially

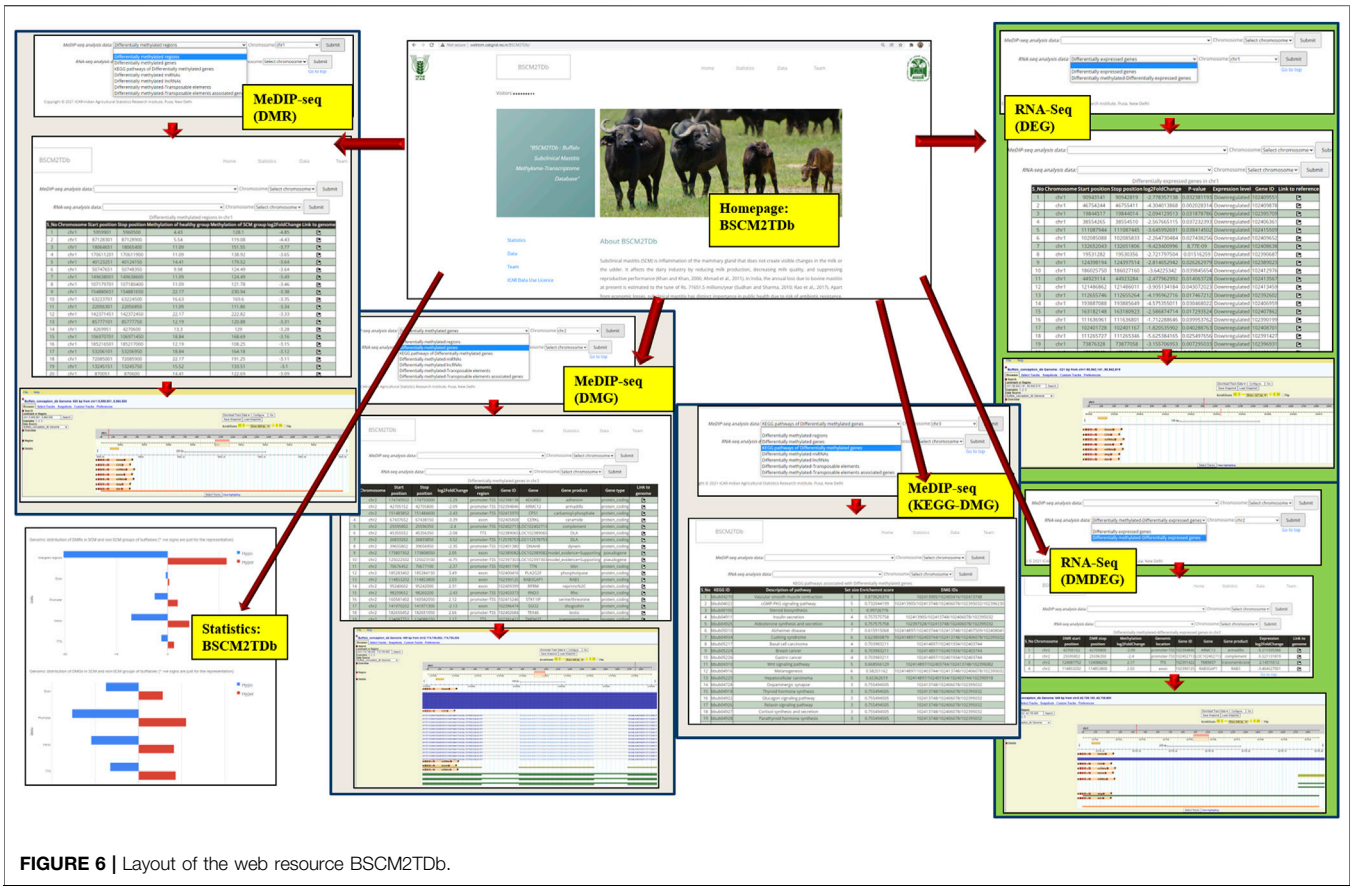


FIGURE 6 | Layout of the web resource BSCM2TDb.

expressed, suggesting that DNA methylation could be one of the epigenetic regulatory mechanisms of gene expression during SCM development. Among these genes, 73% DMGs were negatively correlated with gene expression in the promoter region while the rest were positively correlated with gene expression, which is in agreement with the previous studies (Selamat et al., 2012; Song et al., 2016), suggesting dynamic regulation due to DNA methylation in response to SCM. Some of the important DMGs with correlated differential gene expression were MED25, BLOC1S6, CREB3L1, FRY, protein kinase C-binding protein 1-like, VPS11, CSF2RB, granulocyte-macrophage colony-stimulating factor receptor subunit alpha-like (LOC102408349), C3 and PZP like, alpha-2-macroglobulin domain containing 8 (CPAMD8), and MYSM1, which could be involved in the development of SCM during host–pathogen interaction in buffaloes. Out of these genes, CSF2RB, LOC102408349, C3 and PZP like, and CPAMD8 were found to be downregulated due to hypermethylation in the promoter region during the development of SCM by suppressing immune response in buffaloes, which are known to be involved in immune response.

Additionally, a large number of TEs were found in DMRs in response to SCM in the present study. Apart from TEs in intergenic and intronic regions, there were higher hypomethylated TEs in the promoter region too, which is in agreement with the previous studies involving DNA methylation in cattle tissues (Zhou et al., 2020). The most common class of TEs in DMRs and TAGs was of the SINE family. A large number of TEs in differentially methylated intergenic and intronic regions suggests a role of DNA methylation in the stability of the genome (Adelson et al., 2009; Iwasaki et al., 2015; Czech and Hannon, 2016) and more hypomethylated TEs in response to SCM suggests more activation of TEs during the host–pathogen interaction during SCM. Similarly, probable differentially methylated TAGs were also found more hypomethylated in response to SCM.

Interestingly, we also studied the DNA methylation within miRNA and lncRNA transcribing genes to understand the indirect regulation of gene expression through DNA methylation at the posttranscriptional level. These methylation-regulated lncRNAs and miRNAs were found to target mRNAs transcribed through DEGs specific to SCM, suggesting that the DNA methylation is affecting the expression of genes not only directly at the transcription level but also indirectly at the posttranscriptional level during host–pathogen interaction specific to SCM in buffaloes, which is in agreement of with the findings of Lim et al. (2017) and Saripalli et al. (2020) that the correlation between DNA methylation and gene expression is nonlinear and complex. Similar findings were found by Ju et al. (2020) in the case of *E. coli*-infected mastitis cows. It is also reported that miRNA plays an important regulatory role in immune and inflammatory

responses to mastitis in cows (Ju et al., 2019) and DNA methylation plays a role in the regulation of around 10% miRNAs (Han et al., 2007).

All the results obtained from the present study were compiled and catalogued in the form of a web resource, BSCM2TDb. The present finding can not only be used for understanding the molecular regulation of mastitis disease but can also be used to identify candidate epigenetic markers related to the disease (Zheng et al., 2018). Such combined analysis of DNA methylome and transcriptome map has already been successfully used to identify candidate genes of mastitis disease susceptibility in cattle (Song et al., 2016). The unavailability of web resources related to mastitis disease obviates the development of BSCM2TDb, which catalogues the results obtained from this study at a single place for easy access. This will be useful to the bovine scientific community to be utilized in further studies and research.

CONCLUSION

The study is based on lactating Indian Murrah water buffaloes from the ICAR-CIRB, India, farm for the incidence of SCM. Here, DNA methylation was compared with transcriptome data to understand its effect on gene expression in response to SCM. The genomic annotation of obtained peaks is abundant in the intergenic region. On comparison of methylation in the SCM group vs control group, out of the total 3,950 DMRs, 2,451 were hypomethylated, while 1,449 were hypermethylated. Out of these 3,950 DMRs, 370 DMGs (169 hypermethylated and 201 and hypomethylated) were extracted. While annotating the 21,028 DERs with buffalo RefSeq annotation data, 4,778 DEGs (2,908 downregulated and 1,870 upregulated) were obtained. The KEGG pathway enrichment analysis revealed 142 pathways for 367 DMGs. A comparison of DMGs from MeDIP-Seq analysis and DEGs from RNA-Seq analysis shows that 67 DMGs have differential expression as well. Furthermore, 33 DMGs had gene expression as per methylation. Analyses also resulted in 3,377 TEs within DMRs and 132 TAGs. The CSF2RB, LOC102408349, C3 and PZP like, and CPAMD8 genes, which are known to involved in immune response, were found to be downregulated due to hypermethylation in the promoter region in our study. The present study is the first genome-wide DNA methylation study specific to SCM in buffaloes of the Murrah breed, aiming to understand the role of epigenetic regulation involving DNA methylation of genes involved in host–pathogen interaction during SCM in buffaloes. Interestingly, the present study also sheds a brief light to the role of DNA methylation in indirect regulation of SCM-specific mRNAs at the posttranscriptional level by methylation-regulated miRNAs and lncRNAs. All this information has been catalogued at one place in the BSCM2TDb, which may be of immense use to buffalo researchers in the endeavor of mastitis control and management for higher milk production.

DATA AVAILABILITY STATEMENT

The following data of whole-genome DNA methylation and transcriptome data of buffalo are submitted in the NCBI repository with BioProject ID PRJNA739886; MeDIP-Seq SRA IDs SRR14879252, SRR14879253, SRR14879254, SRR14879255, SRR14879256, SRR14879257, SRR14879258, SRR14879262, SRR14879272, SRR14879273, SRR14879274; RAN-Seq SRA IDs SRR14879259, SRR14879260, SRR14879261, SRR14879263, SRR14879264, SRR14879265, SRR14879266, SRR14879267, SRR14879268, SRR14879269, SRR14879270, SRR14879271.

ETHICS STATEMENT

The animal study was reviewed and approved by the Institute Animal Ethics Committee of the ICAR-Central Institute for Research on Buffaloes (ICAR-CIRB), Hisar.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: VN, DK, MI, SJ, AaB, AR, AgB, and TD. Performed the experiments: VN, MI, SJ, DK, AgB, CS, SK, RS, MS, SP, and RK. Analyzed the data: VN, KS, MI, SJ, AaB, and AgB. Contributed reagents/materials/analysis tools: VN, KS, MI, SJ, DK, and AaB. Drafted the manuscript: VN, KS, MI, SJ, DK, and AaB. Edited the manuscript: DK, AR, VN, and TD. All authors read and approved the manuscript.

FUNDING

This work was supported by the Indian Council of Agricultural Research, Ministry of Agriculture and Farmers' Welfare, Govt. of India, who provided financial assistance in the form of a CABin grant (F. no. Agril. Edn.4-1/2013-A&P), as well as Advanced Super Computing Hub for Omics Knowledge in Agriculture (ASHOKA) facility at ICAR-IASRI, New Delhi, India.

ACKNOWLEDGMENTS

Authors are thankful to the Indian Council of Agricultural Research, Ministry of Agriculture and Farmers' Welfare, Govt. of India, for providing financial assistance in the form of a CABin grant (F. no. Agril. Edn.4–1/2013-A&P) as well as to the Advanced Super Computing Hub for Omics Knowledge in Agriculture (ASHOKA) facility at ICAR-IASRI, New Delhi, India. The authors are very grateful to the Director ICAR-CIRB, Hisar; the Director ICAR-IASRI, New Delhi; and the Director ICAR-NRCE, Hisar, for providing the necessary facilities for this study. The authors also thankfully acknowledge LUVAS, Hisar.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.828292/full#supplementary-material

REFERENCES

Adelson, D. L., Raison, J. M., and Edgar, R. C. (2009). Characterization and Distribution of Retrotransposons and Simple Sequence Repeats in the Bovine Genome. *Proc. Natl. Acad. Sci.* 106 (31), 12855–12860. doi:10.1073/pnas.0901282106

AHDB (2016). *Agriculture and Horticulture Development Board Annual Report and Accounts 2016/17*. United Kingdom: APS Group. HC 193, SG/2017/99.

Ahmad, M. D., Muhammad, K., and Anjum, A. A. (2011). Prevalence of Sub Clinical Mastitis in Dairy Buffaloes of Punjab, Pakistan M. A. Ali. *J. Anim. Plant Sci.* 21 (3), 477–480. Available at: https://www.researchgate.net/publication/267785662_Prevalence_of_sub_clinical_mastitis_in_dairy_buffaloes_of_Punjab_Pakistan#:~:text=Six%20hundred%20lactating%20dairy%20buffaloes%20from%20four%20districts,Overall%20prevalence%20of%20subclinical%20mastitis%20was%2044%25%20%28264%2F600%29

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* 215, 403–410.

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Ball, M. P., Billy Li, J., Gao, Y., Lee, J.-H., LeProust, E. M., Park, I.-H., et al. (2009). Erratum: Corrigendum: Targeted and Genome-Scale Strategies Reveal Gene-Body Methylation Signatures in Human Cells. *Nat. Biotechnol.* 27, 485. doi:10.1038/nbt0509-485b

Bauer, A. W., Kirby, W. M., Sherris, J. C., and Turck, M. (1966). Antibiotic Susceptibility Testing by a Standardized Single Disk Method. *Am. J. Clin. Pathol.* 45 (4), 493–496. doi:10.1093/ajcp/45.4_ts.493

Basic Animal Husbandry Statistics (2019). *Department of Animal Husbandry and Dairying, Govt of India, Krishi Bhawan, New Delhi, India*. Available at: https://dahd.nic.in/circulars/basic-animal-husbandry-statistics-2019

Benić, M., Maæešiæ, N., Maćešić, N., Cvetnić, L., Habrun, B., Cvetnić, Ž., et al. (2018). Bovine Mastitis: a Persistent and Evolving Problem Requiring Novel Approaches for its Control - a Review. *Vet. Arhiv* 88, 535–557. doi:10.24099/vet.arhiv.0116

Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., et al. (2017). A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cel Rep.* 18 (3), 762–776. doi:10.1016/j.celrep.2016.12.063

Chen, J., Wu, Y., Sun, Y., Dong, X., Wang, Z., Zhang, Z., et al. (2019). Bacterial Lipopolysaccharide Induced Alterations of Genome-wide DNA Methylation and Promoter Methylation of Lactation-Related Genes in Bovine Mammary Epithelial Cells. *Toxins* 11, 298. doi:10.3390/toxins11050298

Conesa, A., and Götz, S. (2008). Blast2GO: a Comprehensive Suite for Functional Analysis in Plant Genomics. *Int. J. Plant Genom* 2008 (3), 1℃12. doi:10.1155/2008/619832

Czech, B., and Hannon, G. J. (2016). One Loop to Rule Them All: the Ping-Pong Cycle and piRNA-Guided Silencing. *Trends Biochem. Sci.* 41 (4), 324–337. doi:10.1016/j.tibs.2015.12.008

Dai, X., Zhuang, Z., and Zhao, P. X. (2018). psRNATarget: a Plant Small RNA Target Analysis Server (2017 Release). *Nucleic Acids Res.* 46 (W1), W49–W54. doi:10.1093/nar/gky316

De Vliegher, S., Fox, L. K., Piepers, S., McDougall, S., and Barkema, H. W. (2012). Invited Review: Mastitis in Dairy Heifers: Nature of the Disease, Potential Impact, Prevention, and Control. *J. Dairy Sci.* 95, 1025–1040. doi:10.3168/jds.2010-4074

Down, T. A., Rakyan, V. K., Turner, D. J., Flicek, P., Li, H., Kulesha, E., et al. (2008). A Bayesian Deconvolution Strategy for Immunoprecipitation-Based DNA Methylome Analysis. *Nat. Biotechnol.* 26 (7), 779–785. doi:10.1038/nbt1414

Dufour, S., Dohoo, I. R., Barkema, H. W., DesCôteaux, L., Devries, T. J., Reyher, K. K., .Roy, J.-P., and Scholl, D. T. (2012). Epidemiology of Coagulase-Negative Staphylococci Intramammary Infection in Dairy Cattle and the Effect of Bacteriological Culture Misclassification. *J. Dairy Sci.* 95 (6), 3110–3124. doi:10.3168/jds.2011-5164

Fagiolo, A., and Lai, O. (2007). Mastitis in buffalo. *Ital. J. Anim. Sci.* 6, 200–206. doi:10.4081/ijas.2007.s2.200

Fang, X., Zhao, Z., Yu, H., Li, G., Jiang, P., Yang, Y., et al. (2017). Comparative Genome-wide Methylation Analysis of Longissimus Dorsi Muscles between Japanese Black (Wagyu) and Chinese Red Steppes Cattle. *PLoS One* 12, e0182492. doi:10.1371/journal.pone.0182492

Feber, A., Wilson, G. A., Zhang, L., Presneau, N., Idowu, B., Down, T. A., et al. (2011). Comparative Methylome Analysis of Benign and Malignant Peripheral Nerve Sheath Tumors. *Genome Res.* 21 (4), 515–524. doi:10.1101/gr.109678.110

Geary, U., Lopez-Villalobos, N., O'Brien, B., Garrick, D. J., and Shalloo, L. (2013). Examining the Impact of Mastitis on the Profitability of the Irish Dairy Industry. *Irish J. Agric. Food Res.* 52, 135–149.

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 (7), 644–652. doi:10.1038/nbt.1883

Haas, B. J., and Papanicolaou, A. (2019). TransDecoder 5.5.0. Available at: https://github.com/TransDecoder/TransDecoder/wiki (Accessed July 21, 2021).

Han, L., Witmer, P. D. W., Casey, E., Valle, D., and Sukumar, S. (2007). DNA Methylation Regulates microRNA Expression. *Cancer Biol. Ther.* 6, 1290–1294. doi:10.4161/cbt.6.8.4486

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., et al. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cel* 38, 576–589. doi:10.1016/j.molcel.2010.05.004

Hoque, M. N., Das, Z. C., Talukder, A. K., Alam, M. S., and Rahman, A. N. M. A. (2015). Different Screening Tests and Milk Somatic Cell Count for the Prevalence of Subclinical Bovine Mastitis in Bangladesh. *Trop. Anim. Health Prod.* 47, 79–86. doi:10.1007/s11250-014-0688-0

Hoque, M. N., Istiaq, A., Rahman, M. S., Islam, M. R., Anwar, A., Siddiki, A. M. A. M. Z., et al. (2020). Microbiome Dynamics and Genomic Determinants of Bovine Mastitis. *Genomics* 112 (6), 5188–5203. doi:10.1016/j.ygeno.2020.09.039

Huda, A., and Jordan, I. K. (2009). Analysis of Transposable Element Sequences Using CENSOR and RepeatMasker. *Methods Mol. Biol.* 537, 323–336. doi:10.1007/978-1-59745-251-9_16

Ibeagha-Awemu, E. M., and Zhao, X. (2015). Epigenetic marks: Regulators of Livestock Phenotypes and Conceivable Sources of Missing Variation in Livestock Improvement Programs. *Front. Genet.* 6, 302. doi:10.3389/fgene.2015.00302

Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., et al. (2009). The Human colon Cancer Methylome Shows Similar Hypo- and Hypermethylation at Conserved Tissue-specific CpG Island Shores. *Nat. Genet.* 41, 178–186. doi:10.1038/ng.298

Iwasaki, Y. W., Siomi, M. C., and Siomi, H. (2015). PIWI-interacting RNA: its Biogenesis and Functions. *Annu. Rev. Biochem.* 84, 405–433. doi:10.1146/annurev-biochem-060614-034258

Jeong, W., Kim, J., Bazer, F. W., and Song, G. (2014). Proliferation-stimulating Effect of colony Stimulating Factor 2 on Porcine Trophectoderm Cells Is Mediated by Activation of Phosphatidylinositol 3-kinase and Extracellular Signal-Regulated Kinase 1/2 Mitogen-Activated Protein Kinase. *PLoS One* 9 (2), e88731. doi:10.1371/journal.pone.0088731

Ju, Z., Jiang, Q., Liu, G., Wang, X., Luo, G., Zhang, Y., et al. (2019). Solexa Sequencing and Custom microRNA Chip Reveal Repertoire of microRNAs in Mammary Gland of Bovine Suffering from Natural Infectious Mastitis. *Anim. Genet.* 49, 3–18. doi:10.1111/age.12628

Ju, Z., Jiang, Q., Wang, J., Wang, X., Yang, C., Sun, Y., et al. (2020). Genome-wide Methylation and Transcriptome of Blood Neutrophils Reveal the Roles of DNA Methylation in Affecting Transcription of Protein-Coding Genes and miRNAs in E. Coli-Infected Mastitis Cows. *BMC Genomics* 21, 102. doi:10.1186/s12864-020-6526-z

Khan, M. Z., and Khan, A. (2006). Basic Facts of Mastitis in Dairy Animals: A Review Pakistan. *Vet. J.* 26 (4), 204–208. Available at: https://www.semanticscholar.org/paper/BASIC-FACTS-OF-MASTITIS-IN-DAIRY-ANIMALS%3A-A-REVIEW-Khan-Khan/7299286f3ad2bf3a4c3704e14cdd7b9744fbdb96.

Kirkpatrick, M. A., and Olson, J. D. (2015). "Somatic Cell Counts at First Test: More Than a Number," in Proceedings NMC Annual Meeting, Memphis, Tennessee, February 2015, 53–56.

Langmead, B., and Salzberg, S. L. (2012). Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* 9 (4), 357–359. doi:10.1038/nmeth.1923

Lev Maor, G., Yearim, A., and Ast, G. (2015). The Alternative Role of DNA Methylation in Splicing Regulation. *Trends Genet.* 31, 274–280. doi:10.1016/j.tig.2015.03.002

Li, N., Ye, M., Li, Y., Yan, Z., Butcher, L. M., Sun, J., et al. (2010). Whole Genome DNA Methylation Analysis Based on High Throughput Sequencing Technology. *Methods* 52, 203–212. doi:10.1016/j.ymeth.2010.04.009

Li, D., Zhang, B., Xing, X., and Wang, T. (2015a). Combining MeDIP-Seq and MRE-Seq to Investigate Genome-wide CpG Methylation. *Methods* 72, 29–40. doi:10.1016/j.ymeth.2014.10.032

Li, J., Ma, W., Zeng, P., Wang, J., Geng, B., Yang, J., et al. (2015b). LncTar: a Tool for Predicting the RNA Targets of Long Noncoding RNAs. *Brief Bioinform* 16 (5), 806–812. doi:10.1093/bib/bbu048

Li, J., Hsu, H.-C., Mountz, J. D., and Allen, J. G. (2018). Unmasking Fucosylation: from Cell Adhesion to Immune System Regulation and Diseases. *Cel Chem. Biol.* 25, 499–512. doi:10.1016/j.chembiol.2018.02.005

Lim, Y. C., Li, J., Ni, Y., Liang, Q., Zhang, J., Yeo, G. S. H., et al. (2017). A Complex Association between DNA Methylation and Gene Expression in Human Placenta at First and Third Trimesters. *PLoS One* 12 (7), e0181155. doi:10.1371/journal.pone.0181155

Lou, S., Lee, H.-M., Qin, H., Li, J.-W., Gao, Z., Liu, X., et al. (2014). Whole-genome Bisulfite Sequencing of Multiple Individuals Reveals Complementary Roles of Promoter and Gene Body Methylation in Transcriptional Regulation. *Genome Biol.* 15, 408. doi:10.1186/s13059-014-0408-0

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8

Niazi, U., Geyer, K. K., Vickers, M. J., Hoffmann, K. F., and Swain, M. T. (2016). DISMISS: Detection of Stranded Methylation in MeDIP-Seq Data. *BMC Bioinformatics* 17 (1), 295. doi:10.1186/s12859-016-1158-7

Ott, S. L. (1999). Cost of Herd Level Production Losses Associated with Subclinical Mastitis in U.S. Dairy Cows. *NMC Annu. Meet. Proc.* 199, 152.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A., and Manke, T. (2014). deepTools: a Flexible Platform for Exploring Deep-Sequencing Data. *Nucleic Acids Res.* 42 (Web Server issue), W187–W191. doi:10.1093/nar/gku365

Rao, D. B. K., Reddy, N. K., Rangaswamy, A., Gangiah, M., Raghavan, S. K., Srinivas, M., et al. (2017). Comparative Evaluation of Oral Supplementation of Uncoated and Coated Trisodium Citrate in Subclinical Mastitis in Cows. *J.Bio.Innov* 6 (4), 616–623. Available at: https://www.jbino.com/docs/Issue04_13_2017.pdf

Reynolds, L. P., Ward, A. K., and Caton, J. S. (2017). *Epigenetics and Developmental Programming in Ruminants: Long-Term Impacts on Growth and Development, Biology of Domestic Animals*. Boca: CRC Press, 85–121. doi:10.1201/9781315152080-5

Ruegg, P. L., and Erskine, R. J. (2015). "Mammary Gland Health," in *Large Animal Internal Medicine*. Editor BP Smith. 5th Edition (St. Louis: Elsevier).

Ruzov, A., Tsenkina, Y., Serio, A., Dudnakova, T., Fletcher, J., Bai, Y., et al. (2011). Lineage-specific Distribution of High Levels of Genomic. *Cell Res* 21, 1332–1342. doi:10.1038/cr.2011.113

Saripalli, G., Sharma, C., Gautam, T., Singh, K., Jain, N., Prasad, P., et al. (2020). Complex Relationship between DNA Methylation and Gene Expression Due to Lr28 in Wheat-Leaf Rust Pathosystem. *Mol. Biol. Rep.* 47, 1339–1360. doi:10.1007/s11033-019-05236-1

Schalm, O. W., Carroll, E. J., and Jain, N. C. (1971). *Bovine Mastitis. 1971*. Philadelphia: Lea & Febiger.

Selamat, S. A., Chung, B. S., Girard, L., Zhang, W., Zhang, Y., Campan, M., et al. (2012). Genome-scale Analysis of DNA Methylation in Lung Adenocarcinoma and Integration with mRNA Expression. *Genome Res.* 22, 1197–1211. doi:10.1101/gr.132662.111

Shen, L., Shao, N.-Y., Liu, X., Maze, I., Feng, J., and Nestler, E. J. (2013). diffReps: Detecting Differential Chromatin Modification Sites from ChIP-Seq Data with Biological Replicates. *PLoS One* 8 (6), e65598. doi:10.1371/journal.pone.0065598

Sinha, M. K., Thombare, N. N., and Mondal, B. (2014). Subclinical Mastitis in Dairy Animals: Incidence, Economics, and Predisposing Factors. *Scientific World J.* 2014, 1–4. doi:10.1155/2014/523984

Song, M., He, Y., Zhou, H., Zhang, Y., Li, X., and Yu, Y. (2016). Combined Analysis of DNA Methylome and Transcriptome Reveal Novel Candidate Genes with Susceptibility to Bovine *Staphylococcus aureus* Subclinical Mastitis. *Sci. Rep.* 6, 29390. doi:10.1038/srep29390

Sordillo, L. M. (2011). New Concepts in the Causes and Control of Mastitis. *J. Mammary Gland Biol. Neoplasia* 16, 271–273. doi:10.1007/s10911-011-9239-8

Sudhan, N. A., and Sharma, N. (2010). *Mastitis-An Important Production Disease of Dairy Animals*. SMVS Dairy Year Book, 72–88.

Sun, X., Li, J., Fu, L., Jiang, J., Zhao, J., and Wang, G. (2019). The Epigenetic Modification in Mammals under Heat Stress. *World J. Vet. Sci.* 1, 1005Available at: http://www.medtextpublications.com/open-access/the-epigenetic-modification-in-mammals-undernbspheat-stress-236.pdf

Suzuki, M. M., and Bird, A. (2008). DNA Methylation Landscapes: Provocative Insights from Epigenomics. *Nat. Rev. Genet.* 9, 465–476. doi:10.1038/nrg2341

Taiwo, O., Wilson, G. A., Morris, T., Seisenberger, S., Reik, W., Pearce, D., et al. (2012). Methylome Analysis Using MeDIP-Seq with Low DNA Concentrations. *Nat. Protoc.* 7 (4), 617–636. doi:10.1038/nprot.2012.012

Wang, M., Liang, Y., Ibeagha-Awemu, E. M., Li, M., Zhang, H., Chen, Z., et al. (2020). Genome-Wide DNA Methylation Analysis of Mammary Gland Tissues from Chinese Holstein Cows with *Staphylococcus aureus* Induced Mastitis. *Front. Genet.* 11, 550515. doi:10.3389/fgene.2020.550515

Wijngaard, P. L., MacHugh, N. D., Metzelaar, M. J., Romberg, S., Bensaid, A., Pepin, L., et al. (1994). Members of the Novel WC1 Gene Family Are Differentially Expressed on Subsets of Bovine CD4-CD8- Gamma delta T Lymphocytes. *J. Immunol.* 152 (7), 3476–3482.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137. doi:10.1186/gb-2008-9-9-r137

Zhao, Y., Zhou, M., Gao, Y., Liu, H., Yang, W., Yue, J., et al. (2015). Shifted T Helper Cell Polarization in a Murine *Staphylococcus aureus* Mastitis Model. *PLoS One* 10, e0134797. doi:10.1371/journal.pone.0134797

Zheng, Y., Huang, Q., Ding, Z., Liu, T., Xue, C., Sang, X., et al. (2018). Genome-wide DNA Methylation Analysis Identifies Candidate Epigenetic Markers and Drivers of Hepatocellular Carcinoma. *Brief Bioinform* 19 (1), bbw094. doi:10.1093/bib/bbw094

Zhou, Y., Liu, S., Hu, Y., Fang, L., Gao, Y., Xia, H., et al. (2020). Comparative Whole Genome DNA Methylation Profiling across Cattle Tissues Reveals Global and Tissue-specific Methylation Patterns. *BMC Biol.* 18, 85. doi:10.1186/s12915-020-00793-5

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

The handling editor declared a past collaboration with the authors.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 23 May 2022
doi: 10.3389/fgene.2022.852165

RNA-Seq Analysis of the Growth Hormone Transgenic Female Triploid Atlantic Salmon (*Salmo salar*) Hepatic Transcriptome Reveals Broad Temperature-Mediated Effects on Metabolism and Other Biological Processes

Eric H. Ignatz[1], Tiago S. Hori[2], Surendra Kumar[1], Tillmann J. Benfey[3], Laura M. Braden[4,5], C. Dawn Runighan[4], Jillian D. Westcott[6] and Matthew L. Rise[1]**

[1]Department of Ocean Sciences, Memorial University of Newfoundland and Labrador, St. John's, NL, Canada, [2]Atlantic Aqua Farms Ltd., Charlottetown, PE, Canada, [3]Department of Biology, University of New Brunswick, Fredericton, NB, Canada, [4]AquaBounty Canada, Inc., Souris, PE, Canada, [5]Department of Pathology and Microbiology, Atlantic Veterinary College, University of Prince Edward Island, Charlottetown, PE, Canada, [6]Fisheries and Marine Institute, Memorial University of Newfoundland and Labrador, St. John's, NL, Canada

OPEN ACCESS

Edited by:
Suxu Tan,
Michigan State University,
United States

Reviewed by:
Yi Zhou,
Hunan Normal University, China
Mir Asif Iquebal,
Indian Council of Agricultural Research, India

***Correspondence:**
Eric H. Ignatz
ehignatz@mun.ca
Matthew L. Rise
mrise@mun.ca

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 10 January 2022
Accepted: 02 May 2022
Published: 23 May 2022

Citation:
Ignatz EH, Hori TS, Kumar S, Benfey TJ, Braden LM, Runighan CD, Westcott JD and Rise ML (2022) RNA-Seq Analysis of the Growth Hormone Transgenic Female Triploid Atlantic Salmon (Salmo salar) Hepatic Transcriptome Reveals Broad Temperature-Mediated Effects on Metabolism and Other Biological Processes. Front. Genet. 13:852165. doi: 10.3389/fgene.2022.852165

This study examined the impact of rearing temperature (10.5, 13.5 or 16.5°C) on the hepatic transcriptome of AquAdvantage Salmon (growth hormone transgenic female triploid Atlantic salmon) at an average weight of 800 g. Six stranded PE libraries were Illumina-sequenced from each temperature group, resulting in an average of over 100 M raw reads per individual fish. RNA-sequencing (RNA-seq) results showed the greatest difference in the number of differentially expressed transcripts (1750 DETs), as revealed by both DESeq2 and edgeR ($q < 0.05$; fold-change > |1.5|), was between the 10.5 and 16.5°C temperature groups. In contrast, 172 and 52 DETs were found in the 10.5 vs. 13.5°C and the 13.5 vs. 16.5°C comparisons, respectively. Considering the DETs between the 10.5 and 16.5°C groups, 282 enriched gene ontology (GO) terms were identified ($q < 0.05$), including "response to stress", "immune system process", "lipid metabolic process", "oxidation-reduction process", and "cholesterol metabolic process", suggesting elevated temperature elicited broad effects on multiple biological systems. Pathway analysis using ClueGO showed additional impacts on amino acid and lipid metabolism. There was a significant positive correlation between RNA-seq and real-time quantitative polymerase chain reaction (RT-qPCR) results for 8 of 9 metabolic-related transcripts tested. RT-qPCR results also correlated to changes in fillet tissue composition previously reported in these salmon (e.g., methionine and lysine concentrations positively correlated with *hsp90ab1* transcript expression), suggesting that rearing temperature played a significant role in mediating metabolic/biosynthetic pathways of AquAdvantage Salmon. Many transcripts related to lipid/fatty acid metabolism (e.g., *elovl2*, *fabpi*, *hacd2*, *mgll*, *s27a2*, *thrsp*) were downregulated at 16.5°C compared to both other temperature groups. Additionally,

enrichment of stress-, apoptosis- and catabolism-relevant GO terms at 16.5°C suggests that this temperature may not be ideal for commercial production when using freshwater recirculating aquaculture systems (RAS). This study relates phenotypic responses to transcript-specific findings and therefore aids in the determination of an optimal rearing temperature for AquAdvantage Salmon. With approval to grow and sell AquAdvantage Salmon in the United States and Canada, the novel insights provided by this research can help industry expansion by promoting optimal physiological performance and health.

Keywords: Atlantic salmon, metabolism, RNA-seq, transcriptomics, transgenic, triploid

INTRODUCTION

Global aquaculture production has grown 527% from 1990 to 2018 (FAO, 2020) and is now responsible for producing over half of the world's seafood since the capture fisheries sector is unable to sustainably expand further to meet the increasing demand for food fish consumption (Cai and Leung, 2017; FAO, 2018). However, fish farming itself must be sustainable, and the industry must be proactive in developing new tools to maximize production and limit environmental impact.

Genetic modification is one option available that has sizeable potential to advance aquaculture and help meet the growing demand for safe, nutritious seafood. An example is the AquAdvantage Salmon (AAS), a growth hormone transgenic female triploid Atlantic salmon (*Salmo salar*). These reproductively sterile fish grow at approximately twice the rate of their non-transgenic siblings (Cook et al., 2000; Ganga et al., 2015). With lower food conversion ratios (FCRs), AAS also require less food to reach the same target weight (Tibbetts et al., 2013), further advancing their economic potential compared with conventional farmed Atlantic salmon. AquAdvantage Salmon are currently approved for production and sale within the United States and Canada, with regulatory approval obtained in 2015 and 2016, respectively.

Triploids may benefit the aquaculture industry, as they offer an effective, albeit not 100% guaranteed, option for reproductive sterility and genetic containment (Benfey, 2016). While conventional salmonids are pseudotetraploid due to incomplete restoration of diploidy following a whole-genome duplication event in the salmonid lineage (Lien et al., 2016), Atlantic salmon will be referred to as diploid, and where the second polar body is retained as triploid, to improve the paper's readability. Growth performance between ploidies varies in salmonids, but there is a general trend that diploids grow faster during early development, while triploids catch up or surpass diploids in weight gain at harvest as they avoid the diversion of energy toward sexual development (Chiasson et al., 2009; Weber et al., 2014; Nuez-Ortín et al., 2017). However, it has been suggested that triploid salmonids have a lower optimum temperature for routine metabolism, feed intake and growth than diploids (Atkins and Benfey, 2008; Sambraus et al., 2017b; Sambraus et al.,2018), but other studies have found no such effect (Benfey et al., 1997; Bowden et al., 2018) or contradictory results (Ellis et al., 2013).

Further, recent results raised welfare concerns for the commercial production of conventional triploid Atlantic salmon, where reduced survival, higher levels of emaciation, and lower quality processing scores have been reported in Norway (Madaro et al., 2021). Therefore, it is of high importance for producers of triploid salmon to understand well how their fish will respond to varying environmental conditions during their production cycle. This information can then be used to develop production strategies that maximize growth potential while limiting stress and other negative consequences in recirculating aquaculture systems (RAS).

While recommendations have been made for the optimal rearing temperature of conventional diploid Atlantic salmon (Handeland et al., 2008; Hevrøy et al., 2013) and non-transgenic triploid Atlantic salmon (Sambraus et al., 2017a) in seawater, until recently, no information was available in this regard for AAS reared in freshwater RAS. It is now known what effect rearing temperature (i.e., 10.5, 13.5, and 16.5°C) has on the growth performance, nutrient utilization and innate antiviral response of AAS (Ignatz et al., 2020a; Ignatz et al., 2020b). Results showed that at 16.5°C, FCR was higher, dietary lipid deposition was diverted more to the viscera than to muscle, fillet yields were lower, and ω3 fatty acid deposition was less efficient, but weight gain was higher and faster than AAS reared at 10.5 and 13.5°C (Ignatz et al., 2020b). Further, when intraperitoneally injected with either polyriboinosinic polyribocytidylic acid [pIC; a synthetic double-stranded RNA (dsRNA) analog that elicits a potent antiviral-like response] or an equal volume of sterile phosphate-buffered saline (PBS), higher induction of target antiviral biomarker transcripts was generally found at 10.5°C than either 13.5 or 16.5°C (Ignatz et al., 2020a). Collectively, this suggests that rearing AAS at 16.5°C is not optimal for commercial production and that lower temperatures (i.e., 10.5 or 13.5°C) may be more suitable. However, until the current study, nothing was known about how temperature regulates the broader expression of liver transcripts in AAS.

The field of genomics offers powerful tools that can be used to assess different production models and help reveal the overall impact of changes in rearing conditions. In the past, microarrays have shown differences in hepatic gene expression between fast and slow-growing families of transgenic triploid Atlantic salmon (Xu et al., 2013). Multiple microarray platforms have also indicated changes in hepatic gene expression between transgenic coho salmon (*Oncorhynchus kisutch*) fed different ration levels compared to non-transgenic

control fish (Rise et al., 2006). More recently, RNA-sequencing (RNA-seq) has emerged as a popular method allowing for the entire transcriptome to be assessed rather than a predefined number of transcripts (a potential limiting factor with microarrays). Several studies have examined the impact of elevated temperature on various tissue/organ transcriptomes in multiple fish species using RNA-seq (e.g., Tomalty et al., 2015; Tan et al., 2016; Veilleux et al., 2018; Shi et al., 2019; Bowen et al., 2020; Pan et al., 2021; Quan et al., 2021; Yang et al., 2021; Zhou et al., 2021). However, most of these studies measured responses following acute exposure (i.e., hours to days) to thermal extremes and there is limited information on transcriptomic signatures of fish reared long-term (i.e., months) within typical thermal ranges observed in aquaculture.

As part of the objective to determine the optimal rearing temperature for the commercial production of AAS, RNA-seq was used to compare the hepatic transcriptomes of AAS reared at either 10.5, 13.5 or 16.5°C at an average weight of 800 g. This is the first transcriptomic study to examine the impact of temperature on this transgenic line of salmon and provides novel insight into the mechanisms behind the phenotypic differences previously described in these fish. Not only does this research elucidate hepatic gene expression responses observed in AAS, but the results generated shed new light on the molecular pathways involved in how salmonids respond to differences in thermal conditions.

MATERIALS AND METHODS

Experimental Animals and Rearing Conditions

A full description of the production and rearing conditions of fish used in the current study can be found in Ignatz et al. (2020b). Briefly, all fish were from St. John River stock and were hatched and reared at AquaBounty Canada (PE, Canada). AAS were produced from a specific transgenic line (termed EO-1α) containing a single copy of a gene construct (opAFP-GHc2) consisting of an ocean pout (*Macrozoarces americanus*) antifreeze protein promoter and the coding sequence for the Chinook salmon (*Oncorhynchus tshawytscha*) growth hormone gene (Yaskowiak et al., 2007, Yaskowiak et al., 2006). A single homozygous transgenic sex-reversed neomale (i.e., functionally masculinized genetic female) was crossed with three non-transgenic females, producing all-female transgenic offspring. The neomale parent was originally produced through oral delivery of 17α-methyltestosterone for a limited period during juvenile rearing. All fertilized eggs were pooled together before subjection to hydrostatic pressure shock to induce triploidy, making the fish reproductively sterile (AquaBounty Technologies, Inc., 2010; Benfey, 2016).

AAS were grown from first feeding fry (starting weight of ~0.51 g) in 1.5 m^3 tanks at 10.5°C, 13.5°C, and 16.5°C (±0.5°C) in triplicate. De-gassed and oxygenated freshwater from a well (<1 g/L salinity) was supplied to each tank in a RAS (10%–20% make-up daily), and dissolved oxygen was measured daily and maintained at >9.0 mg/L. Fish were fed commercial diets throughout the day via automatic feeders with continuous light provided. Passive integrated transponder (PIT) tags (Avid Canada Corp., Calgary, AB, Canada) were previously inserted into the peritoneal cavities of each fish when they averaged 400 g. At an average weight of 800 g, liver samples were collected to determine the overall impact of temperature on the hepatic transcriptome of AAS. Samplings took place approximately 10, 11, and 12 months after the onset of the experiment in the 16.5, 13.5, and 10.5°C treatments, respectively.

Sample Collection

Fish were taken off feed at least 24 h in advance of sampling, and fish were euthanized by an overdose (0.4 g/L) of anesthetic (MS-222; Syndel Canada, Nanaimo, BC, Canada), buffered with an equal mass of sodium bicarbonate. As salmon at each rearing temperature reached an average weight of 800 g, fish were sampled by convenience, in that the first salmon found that fell within 700–900 g were sampled. Liver samples, taken from the most distal portion of the posterior lobe, were collected using standard aseptic techniques. Samples were placed in 1 ml of RNA*later*™ (Sigma-Aldrich, Oakville, ON, Canada) kept on ice, then stored at 4°C overnight. The RNA*later* was subsequently removed, and samples were stored at −70°C until RNA extraction. Twenty-one liver samples were collected per rearing temperature, but only two samples per tank (n = 6 per temperature treatment) were chosen for subsequent analyses. The sample list was narrowed down to only fish that fit within ±1 standard deviation from the mean weight calculated after sampling and, if more than two fish per tank remained, then two were randomly chosen.

Fillet samples were collected concurrently during each sampling period from the same fish sampled for liver transcript expression. Fillets were maintained on ice in plastic bags and subsequently stored at −70°C until samples were shipped for compositional analysis. A full description of how fillet composition was analyzed and the results of those analyses can be found in Ignatz et al. (2020b).

RNA Extraction and Purification

RNA extractions were performed using the RNeasy® Mini Kit (Qiagen, Mississauga, ON, Canada), and the TURBO DNA-free™ Kit (Invitrogen, Burlington, ON, Canada) was utilized for DNase digestion in solution. All procedures were completed according to the manufacturers' instructions, with small adjustments made based on guidelines found in the Qiagen Purification of Total RNA Using the RNeasy® Fibrous Tissue Mini Kit Protocol. RNA quantity and quality were first assessed using spectrophotometry and 0.7% agarose gel electrophoresis, respectively. All samples showed evidence of high-quality RNA (A260/280 & A260/230 ratios > 2.0, with distinct 18S/28S ribosomal RNA bands) before shipment to the Centre d'expertise et de services Génome Québec (Montréal, QC, Canada) for further analysis. RNA Integrity

TABLE 1 | Trimming and alignment metrics of each biological replicate RNA-seq library.

	Library name	SRA[a] accession	Raw reads #	Surviving reads # (%)	Aligned reads (%)	Alternative alignments (%)	% Coverage	% Exonic rate
16.5°C	S1_16.5	SRR16105563	112,257,950	112,227,600 (99.97)	108,739,280 (96.89)	43,960,352 (40.43)	30.40	86.26
	S2_16.5	SRR16115861	114,484,598	114,440,190 (99.96)	110,338,704 (96.42)	35,997,993 (32.63)	32.03	85.43
	S3_16.5	SRR16122595	101,880,276	101,848,592 (99.97)	97,969,194 (96.19)	30,641,463 (31.28)	28.69	85.52
	S4_16.5	SRR16958631	107,258,770	107,234,866 (99.98)	103,748,543 (96.75)	29,735,650 (28.66)	30.54	86.03
	S5_16.5	SRR16958615	117,304,540	117,277,318 (99.98)	113,299,533 (96.61)	39,245,504 (34.64)	32.36	86.36
	S6_16.5	SRR16975862	98,692,524	98,668,078 (99.98)	95,693,153 (96.99)	31,917,863 (33.35)	27.78	86.34
13.5°C	S1_13.5	SRR16988317	105,027,464	105,000,658 (99.97)	101,581,042 (96.74)	33,599,049 (33.08)	28.46	86.15
	S2_13.5	SRR17038623	74,841,184	74,821,020 (99.97)	72,261,894 (96.58)	20,308,874 (28.11)	20.51	85.17
	S3_13.5	SRR17481977	109,834,502	109,803,350 (99.97)	106,465,002 (96.96)	32,170,009 (30.22)	30.64	86.62
	S4_13.5	SRR17503995	90,177,906	90,155,004 (99.98)	87,224,875 (96.75)	29,904,438 (34.28)	24.55	86.53
	S5_13.5	SRR17481976	115,909,552	115,859,460 (99.96)	111,534,501 (96.27)	35,826,986 (32.12)	30.88	86.18
	S6_13.5	SRR17500070	105,921,688	105,894,066 (99.97)	102,319,865 (96.63)	32,417,461 (31.68)	28.55	85.75
10.5°C	S1_10.5	SRR16097286	99,113,996	99,077,664 (99.96)	95,533,239 (96.42)	24,867,589 (26.03)	28.39	86.84
	S2_10.5	SRR17446820	92,641,680	92,565,332 (99.92)	88,928,239 (96.07)	25,834,830 (29.05)	26.18	85.41
	S3_10.5	SRR17500116	94,647,236	94,620,554 (99.97)	91,837,075 (97.06)	24,089,272 (26.23)	26.81	87.02
	S4_10.5	SRR17469864	100,106,994	100,071,212 (99.96)	96,974,288 (96.91)	26,257,994 (27.08)	28.24	86.65
	S5_10.5	SRR17504015	93,838,306	93,807,530 (99.97)	90,983,847 (96.99)	26,985,379 (29.66)	26.65	86.46
	S6_10.5	SRR17500112	106,725,990	106,685,144 (99.96)	102,643,077 (96.21)	28,060,332 (27.34)	30.58	87.37

[a]*Sequence read archive.*

Numbers (RIN), assessed using a Bioanalyzer 2100 (Agilent, Santa Clara, CA, United States), of all RNAs used for sequencing and qPCR were found to be ≥8.9.

Sequencing, Read Alignment and Annotation

Library construction and sequencing were performed at Génome Québec CES, with bioinformatics support provided by the Canadian Centre for Computational Genomics (C3G; Montréal, QC, Canada) using their GenPipes next-generation sequencing data processing framework (Bourgey et al., 2019). A total of 18 stranded PE100 libraries were created from the six RNA samples per rearing temperature (**Table 1**) using the NEBNext Ultra kit. RNA sequencing was completed using three lanes (six libraries per lane) of a flow cell in an Illumina® HiSeq4000 sequencing platform. Trimming of the resulting reads were performed using Trimmomatic (v.0.36) (Bolger et al., 2014). Reads were trimmed from the 3′ end with a Phred score cut-off of 30. Illumina sequencing adapters were removed from the reads, and all reads were required to have a length of at least 32 bp. The filtered reads were then aligned to a reference genome (*Salmo salar* assembly ICSASG_v2 taken from NCBI) using STAR (v2.5.3a) (Dobin et al., 2013), creating Binary Alignment Map (BAM) files. All BAM files from the same sample were merged into one single BAM file using Picard (v.2.9.0) (Picard Tools, 2019). Wiggle tracks format (wig) files were generated from the merged file using bedGraphToBigWig (ucsc-userApps v.346) (Kent et al., 2010). Cufflinks (v.2.2.1) (Roberts et al., 2011) was used to assemble aligned RNA-Seq reads into transcripts and estimate their abundance as fragments per kilobase of exon per million fragments mapped (FPKM). Individual count data were extracted by multiplying each Cuffnorm sample count table by the library size factors to de-normalize the counts using a method as in Myers et al. (2021). Once transcripts were assembled and their corresponding FPKM estimated, they were annotated with a known reference set of transcripts obtained from the Ensembl database (Accession GCA_000233375). The transcript list was also annotated with putative *Homo sapiens* ortholog RefSeq mRNA accession identifiers based on lowest BLASTx score with an Expect (E) value cut-off <1e-5 to better assist functional annotation, which was able to successfully annotate ~79% of the assembled transcriptome. Typically, non-model species' gene ontology (GO) annotations are assigned by orthology and lack experimental evidence. Human annotations were added in this study as they are more reliable and allowed for further downstream analyses to be conducted (i.e., GOrilla, ClueGO).

Identification of Differentially Expressed Transcripts and GO Term Enrichment Analysis

Utilizing the transformed read counts obtained from Cuffnorm, differentially expressed transcripts (DETs) were found using the online Galaxy (Afgan et al., 2018) platform. DESeq2 (Love et al., 2014) under Galaxy Version 2.11.40.2 was used with outlier and independent filtering to test pairwise comparisons of rearing temperature first. Then, edgeR (Robinson et al., 2009) under Galaxy Version 3.20.7.2, without filtering lowly expressed genes, was used to perform the exact same comparisons. As count data exist as compositions, this presents a challenge to normalizing results and comparing differential expression (Quinn et al., 2018). DESeq2 and edgeR analyses were compared to obtain higher confidence as they use the same reasonable assumption that most genes will not be differentially expressed between treatments but

TABLE 2 | RT-qPCR primers.

Gene name (Symbol) (GenBank accession number)	Nucleotide sequence (5′-3′)	Efficiency[a] (%)	r^2	Amplicon size (bp)	Source
Elongation of very long chain fatty acids protein 2 (*elovl2*) (TC91192)[b]	F: CGGGTACAAAATGTGCTGGT R: TCTGTTTGCCGATAGCCATT	100.9	0.983	145	Morais et al. (2009)
Fatty acid-binding protein, intestinal-like (*fabpi*) (XM_014206454)	F: CCTGGGCGTACAGTTTGACT R: TATAGCTCTGTACTAGCTCTC CTCC	100.9	0.998	164	This study
Glutathione S-transferase kappa 1 (*gstk1*) (XM_014154124)	F: GGAGTGGACATCAGCATCAGT R: ACACCTTATGATGCCAGAGGAA	92.8	0.960	197	This study
3-hydroxyacyl-CoA dehydratase 2 (*hacd2*) (XM_014165367)	F: CAGACCGGAGCTCTTCTGG R: TGTCTTCATTCTGGACCTCTCG	104.1	0.975	145	This study
Hepatocyte growth factor a (*hgfa*) (NM_001140139)	F: CAGACGGGGACAAGATGCC R: CCGCGAAGAAGATAAACGCA	110.2	0.999	104	This study
Heat shock protein 90 alpha family class B member 1 (*hsp90ab1*) (NM_001123532)	F: AGCCTCACGTTTTTCCAATCG R: TGCGTTGCCCACCATTAACT	92.4	0.978	150	This study
Monoglyceride lipase (*mgll*) (NM_001140001)	F: CTTGTCAGTAATCTTTGGACCCCTA R: ATGCCTCTGTGAATTGCGCT	107.7	0.997	192	This study
Very long-chain acyl-CoA synthetase (*s27a2*, alias *slc27a2*) (NM_001141797)	F: TTCACCCAGAAGCATAGGAGC R: GGCCAAACTGGCAAAAAGGA	100.3	0.999	75	This study
Thyroid hormone responsive (*thrsp*) (XM_014163692)	F: CTACCGGGAACAGCCAGAAA R: GCATAGTGTTGGACTCGGCA	106.9	0.995	74	This study
Elongation factor 1 alpha (*ef1a*) (NM_001141909)[c]	F: GTGGAGACTGGAACCCTGAA R: CTTGACGGACACGTTCTTGA	97.4	0.997	155	Jones et al. (2007)
Polyadenylate-binding protein 1 (*pabpc1*) (EG908498)[c]	F: TGACCGTCTCGGGTTTTTAG R: CCAAGGTGGATGAAGCTGTT	99.2	0.999	108	Caballero-Solares et al. (2017)
60S ribosomal protein 32 (*rpl32*) (BT043656)[d]	F: AGGCGGTTTAAGGGTCAGAT R: TCGAGCTCCTTGATGTTGTG	99.6	0.995	119	Xue et al. (2015)

[a]*Amplification efficiencies were calculated using a 6-point 1:5 dilution series starting with an equal mixture of cDNA from every sample.*
[b]*Atlantic salmon Gene Index identification.*
[c]*Normalizer genes chosen for this study.*
[d]*Normalizer genes tested, but ultimately not chosen for this study.*

apply different approaches to estimate dispersion (Hall et al., 2021). By using both programs, more robust observations can be made between comparisons. Once each DET list was obtained from both DESeq2 and edgeR, they were filtered by their false discovery rate (FDR) (q-value) to keep only transcripts with values less than 0.05, as well as fold-change values above |1.5|. The matched pairwise comparisons by DESeq2 and edgeR were then compared, and only filtered DETs found by both programs were ultimately used for GO term enrichment analysis (GTEA). Comparisons between different gene lists were accomplished through VENNY (v.2.1) (Oliveros, 2007).

With the transcriptome annotated with *Homo sapiens* putative orthologs, the concordant lists of DETs found by DESeq2 and edgeR were used as input for the GOrilla platform (Eden et al., 2009; Eden et al., 2007), with both target and background lists present in the analysis. Only GO terms with q-values below 0.05 were considered significant. Pathway analysis was performed using the ClueGO (v.2.5.7) (Bindea et al., 2009) Cytoscape (v.3.7.2) plugin, illustrating enriched GO (Biological Process, Molecular Function, Cellular Component; v.08.05.2020) and Reactome pathway (v.08.05.2020) terms among annotated DETs. Analyses were performed using a two-sided hypergeometric test after a Bonferroni step-down p-value adjustment. Networks were also designed using a kappa statistic threshold of 0.4 and a medium specificity level, and only including network terms with $p < 0.05$.

RT-qPCR Validation

Nine transcripts were chosen for RT-qPCR validation of RNA-seq results based on their GO terms. Selected GO terms were lipid/fatty acid metabolism (*elovl2*, *fabpi*, *hacd2*, *mgll*, *s27a2*, *thrsp*), protein metabolism (*hgfa*, *hsp90ab1*) and oxidoreductase activity (*gstk1*). These areas of focus were selected based on past interest in nutrient utilization (Tibbetts et al., 2013; Ganga et al., 2015; Ignatz et al., 2020b) and oxidation reduction processes of growth hormone transgenic Atlantic salmon (Xu et al., 2013). All transcripts chosen for validation were also found significantly differentially expressed ($q < 0.01$) by both DESeq2 and edgeR, with fold-changes between the 10.5°C and 16.5°C treatments > |2.0|. One target gene primer pair, and all normalizer transcript primers, were designed from previous studies (Jones et al., 2007; Morais et al., 2009; Xue et al., 2015; Caballero-Solares et al., 2017), while all remaining target transcript primers were designed using NCBI's Primer-Blast (Ye et al., 2012). cDNA was synthesized from 1 µg of purified RNA (the same extracted RNA that was sent for sequencing) using the iScript™ cDNA synthesis kit (Bio-Rad; Saint-Laurent, QC, Canada) following the manufacturer's instructions. To assess potential gDNA contamination, no reverse transcriptase (no-RT) controls were also performed using pooled RNA of randomized samples ($n = 9$ per pool). Every primer set was quality-tested using a cDNA pool derived from equal quantity input from each sample to determine amplification efficiencies (Pfaffl, 2001) with

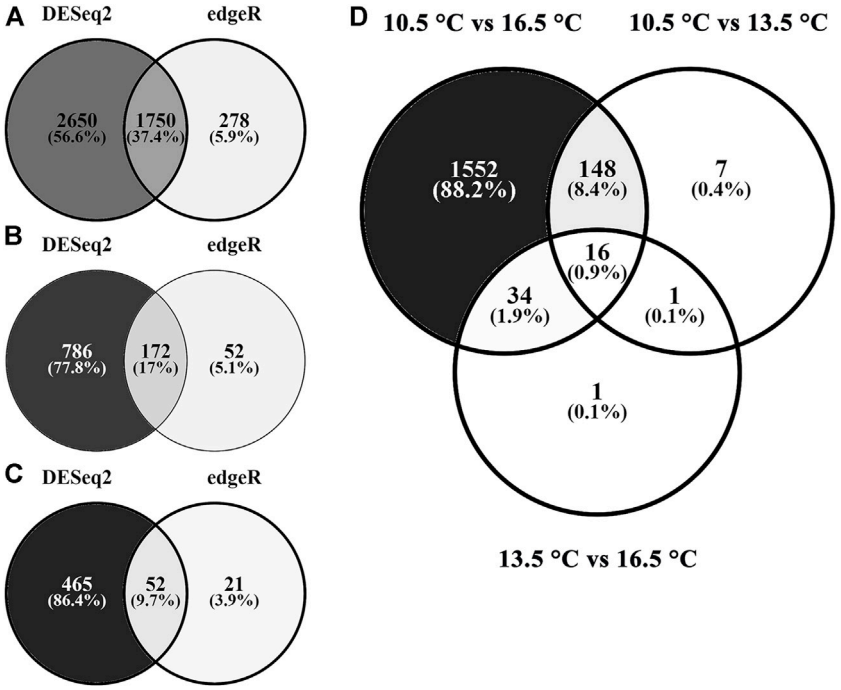


FIGURE 1 | Differential transcript expression in response to rearing temperature. Venn diagrams depict the number of shared and separate differentially expressed (both up- and down-regulated; FDR-adjusted $p < 0.05$) transcripts identified by either DESeq2 (left), edgeR (right), or those in common (middle). Temperature comparisons are presented as **(A)** 10.5°C vs. 16.5°C, **(B)** 10.5°C vs. 13.5°C, **(C)** 13.5°C vs. 16.5°C, and **(D)** differentially expressed transcripts shared between both DESeq2 and edgeR for each pairwise comparison.

a 5-fold, 6-point dilution series in triplicate. Gel electrophoresis also confirmed single products and amplicon size (74–192 bp) for each primer pair. A single peak was found in each dissociation curve with no evidence of primer dimers, no gDNA contamination was found in no-RT controls, and amplification efficiencies ranged between 92 and 110%. All information related to the primers that were utilized in this study is shown in **Table 2**. Based on amplification results from QC testing, cDNA was diluted in nuclease-free water (Invitrogen) at 1:5 or 1:25 concentrations before RT-qPCR analysis for lower and higher expressed transcripts, respectively.

qPCR amplification was performed using a SsoAdvanced™ Universal SYBR® Green Supermix qPCR kit (Bio-Rad) per the manufacturer's instructions with a CFX-384 thermal cycler (Bio-Rad). The master mix was loaded using an Aurora automated plate dispenser with VERSAwar 10.v.1.2.48 software, and samples were plated in triplicate using an electronic Sartorius Picus® pipette. The total reaction volume was 11 µl, comprised of: 5 µl SYBR® Green Supermix, 4 µl nuclease-free water, 0.5 µl forward primer (10 µM), 0.5 µl reverse primer (10 µM), and 1 µl cDNA template (2 or 10 ng input RNA, depending on dilution concentration). PCR amplification and cycling were performed using the same programs described in Ignatz et al. (2020a). Each gene was run on a single plate, which included no-template controls (NTC).

Raw expression profiles were imported into qbase+ (Biogazelle; Gent, Belgium) (Hellemans et al., 2007), and technical replicates falling outside of ±0.5 cycle threshold (CT) from two close replicates were removed. A total of three normalizer genes were tested (*ef1a*, *pabpc1*, *rpl32*); however, only *ef1a* and *pabpc1* were chosen as the final study normalizers as that combination showed the highest stability (geNorm M value and coefficient of variation of 0.438 and 0.154, respectively) (Vandesompele et al., 2002). Normalized relative quantities (NRQs) (Vandesompele et al., 2002) were calculated using amplification efficiencies for each primer pair (**Table 2**), which were then $\log_2$ transformed in Microsoft Excel.

Statistical Procedures

Statistical comparison of treatment means was performed by one-way ANOVA, testing the influence of temperature on NRQs. Tukey's post-hoc tests were performed where significant differences were detected. Linear regression, conducted in Excel, was used to model the relationship between fold-change values calculated from RNA-sequencing and those from the RT-qPCR validation study. Generation of hierarchical clustering, correlogram and principal component analysis (PCA) diagrams, together with ANOVA testing, were performed in R (version 3.6.3) (R Studio Team, 2015). Statistical differences were considered significant at $p < 0.05$.

Hierarchical clustering heatmaps applied standardized FPKM values, calculated using the following equation:

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

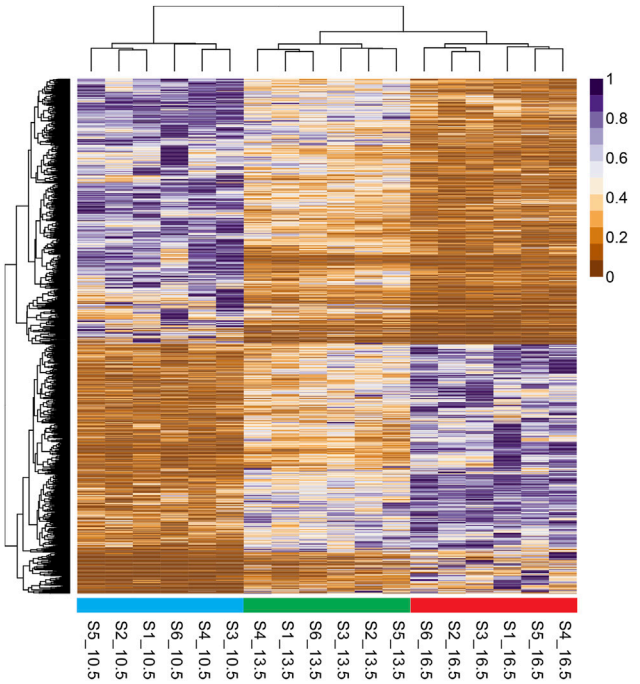


FIGURE 2 | Heatmap and hierarchical clustering of differentially expressed transcripts (1759 DETs) found in comparison of all three rearing temperature treatments. RNA samples used in sequencing are labeled on the base of the figure, with standardized fragments per kilobase of transcript per million mapped reads (FPKM) values for each DET clustered along the vertical axis.

where x_i is a specific FPKM value for a given sample and x is the range of FPKM values across samples for a given transcript used to calculate an individual standardized value (z_i).

RESULTS

Differentially Expressed Transcripts

On average, more than 100 M raw reads per sample were generated from sequencing (**Table 1**). For all samples, >96% of those reads survived the trimming process and were aligned to the reference *Salmo salar* genome. The final read counts were then pairwise compared between each rearing temperature treatment (**Figures 1A–C**) using both DESeq2 and edgeR. In all comparisons, DESeq2 found more DETs than the edgeR program (both analyzed with significance determined at q-value < 0.05, fold change ≥ |1.50|). Full DET lists with DESeq2, edgeR and concordant results can be found in **Supplementary Table S1** (10.5 vs. 16.5°C), **Supplementary Table S2** (10.5 vs. 13.5°C) and **Supplementary Table S3** (13.5 vs. 16.5°C). Only concordant DET lists for each comparison were used in subsequent analyses. Overlap between pairwise lists is shown in **Figure 1D**. Overall, the highest number of DETs was found in the comparison between the 10.5 and 16.5°C treatments (1750 DETs), followed by the 10.5 vs. 13.5°C (172 DETs) and the 13.5 vs. 16.5°C (52 DETs) assessments. Only 16 DETs were found in all three pairwise tests (i.e., found significantly different among all three temperature groups).

A heatmap with hierarchical clustering of standardized FPKM values of all 1759 DETs found in this study is shown in **Figure 2**. All three temperature groups clustered separately from one another; however, the 13.5 and 16.5°C samples share the same initial branching point.

Gene Ontology Term Enrichment Analysis

Taking the DETs that were found significant by both DESeq2 and edgeR, GTEA was performed on each pairwise list. The full results from each of those comparisons can be found in **Supplementary Table S4** (10.5 vs. 16.5°C), **Supplementary Table S5** (10.5 vs. 13.5°C), and **Supplementary Table S6** (13.5 vs. 16.5°C). Due to the 10.5 vs. 16.5°C comparison containing the highest number of DETs, we focused on these results. A total of 225 Biological Process, 24 Molecular Function, and 33 Cellular Component GO terms were enriched in the annotated DET list between the 10.5 and 16.5°C groups. **Figure 3** illustrates selected significantly enriched (q-value < 0.05) GO terms from this specific assessment. Notably, GO terms such as "response to stress", "immune system process", "lipid metabolic process", "oxidation-reduction process", and "cholesterol metabolic process" were all enriched, suggesting that elevated temperature elicited broad effects on multiple biological systems in the liver tissues of AquAdvantage Salmon.

Pathway Analysis

Figure 4 shows the results of pathway analyses conducted using both GO and Reactome terms from the annotated DET list comparing the 10.5 and 16.5°C treatments. In **Figure 4A**, all the enriched terms downregulated at 10.5°C (i.e., upregulated at 16.5°C) are shown. These terms were grouped into five general categories: catabolism, stress and apoptosis, RNA binding, extracellular region, and hormone response. In contrast, **Figure 4B** shows all the upregulated terms at 10.5°C (i.e., downregulated at 16.5°C). These terms mostly fit into four general categories: sterol and fatty acid metabolism, cell regulation, response to stimulus, and oxidoreductase activity.

Examining the transcripts associated with the GO term "fatty acid metabolic process" found during this analysis, **Figure 5** shows that many of these transcripts were upregulated at 10.5°C when compared to 16.5°C. Furthermore, fish reared at 13.5°C had an intermediate response for these transcripts, but overall, their clustering showed a closer relationship to the 10.5°C group than 16.5°C. Further, the relationships between GO terms "apoptotic process" (**Supplementary Figure S1**) and "oxidoreductase activity" (**Supplementary Figure S2**) revealed that at elevated temperature (i.e., 16.5°C), salmon expressed more transcripts related to apoptosis and less transcripts associated with oxygen reduction processes compared to the lower two temperature groups (particularly 10.5°C).

RT-qPCR Validation

Figures 6A–I show the expression of nine metabolic-related transcripts chosen for RT-qPCR validation. All but one of the transcripts demonstrated significantly correlated fold-change

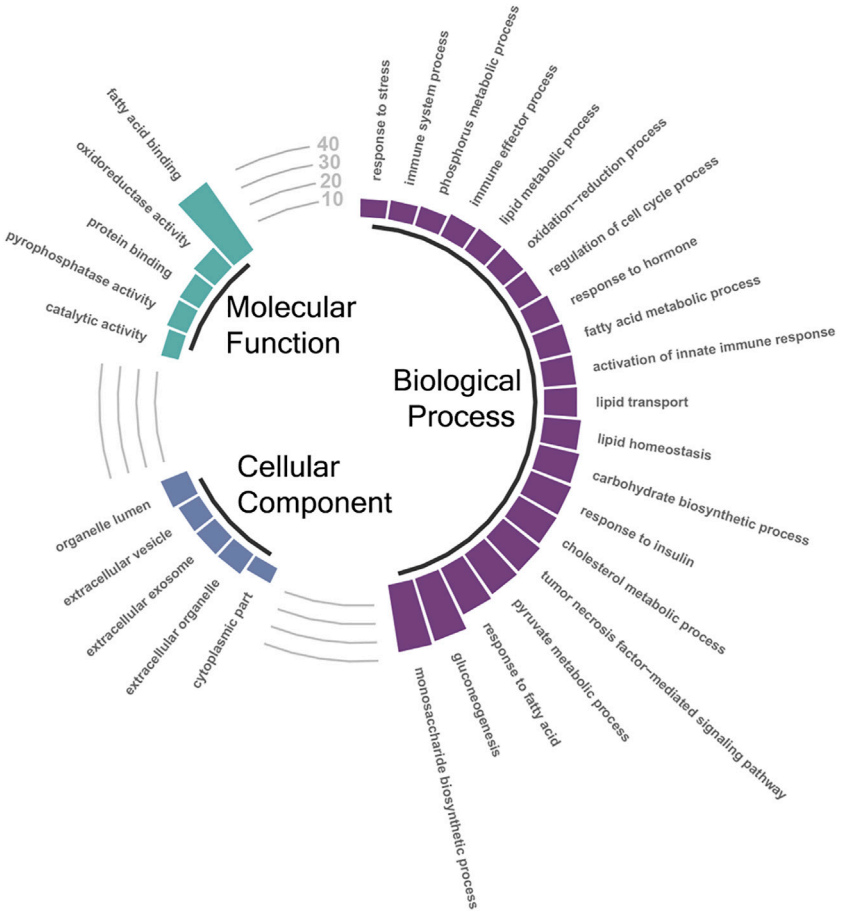


FIGURE 3 | Selected significantly enriched (FDR-adjusted $p < 0.05$) gene ontology (GO) terms associated with significantly differentially expressed transcripts (DETs; both up- and down-regulated) between the 10.5°C and 16.5°C groups. The height of the bars represents the percentage of DETs found in the current study compared to total number of transcripts assigned to that specific GO term. A full list of enriched GO terms identified between the 10.5°C and 16.5°C groups can be found in **Supplementary Table S4**.

differences between the 10.5 and 16.5°C temperature groups between RT-qPCR and RNA-seq results, with *gstk1* being the one exception. This result is further visualized in **Figure 7**, where an overall highly significant ($p < 0.001$) positive correlation ($R^2 = 0.96$) between both methods measuring transcript expression was found. Therefore, we concluded that the RNA-seq results of this study were successfully validated.

All transcripts associated with fatty acid/lipid metabolism (**Figures 6A–F**) demonstrated significant upregulation at 10.5 and 13.5°C compared to 16.5°C. In five of these six cases (*elovl2*, *fabpi*, *hacd2*, *mgll*, *thrsp*), transcript expression did not significantly differ between the 10.5 and 13.5°C temperature groups. However, the expression of *s27a2* showed significant differences between each treatment, indicating an inverse relationship with temperature. Although lacking statistical significance, this trend was also evident in four of the other transcripts (*elovl2*, *fabpi*, *hacd2*, *thrsp*). For the two remaining transcripts that successfully validated that are related to protein metabolism (**Figures 6G,H**), opposing patterns emerged. Expression of *hgfa* exhibited an inverse relationship to rearing temperature, with significant differences between all three treatments. In contrast, *hsp90ab1* was expressed more highly at both 13.5 and 16.5°C compared to 10.5°C. While the one transcript related to oxidoreductase activity (*gstk1*; **Figure 6I**) failed to successfully validate, it was still clear that several fish reared at 16.5°C exhibited lower expression of this transcript compared to the other two temperature groups.

Multivariate Analyses

Figure 8A shows the degree of correlation among the expression profiles of all nine target transcripts involved in RT-qPCR validation along with fillet composition [i.e., docosahexaenoic acid (DHA), eicosapentaenoic acid (EPA), total omega-3 fatty acids (Σ ω3), lysine, methionine, histidine, phosphorus, calcium concentrations], and morphometric [i.e., Fulton's condition factor (K), viscerosomatic index (VSI), hepatosomatic index (HSI)] measurements of all fish from this study. While we detected numerous significant relationships between variables, some are particularly noteworthy. For example, DHA concentrations in

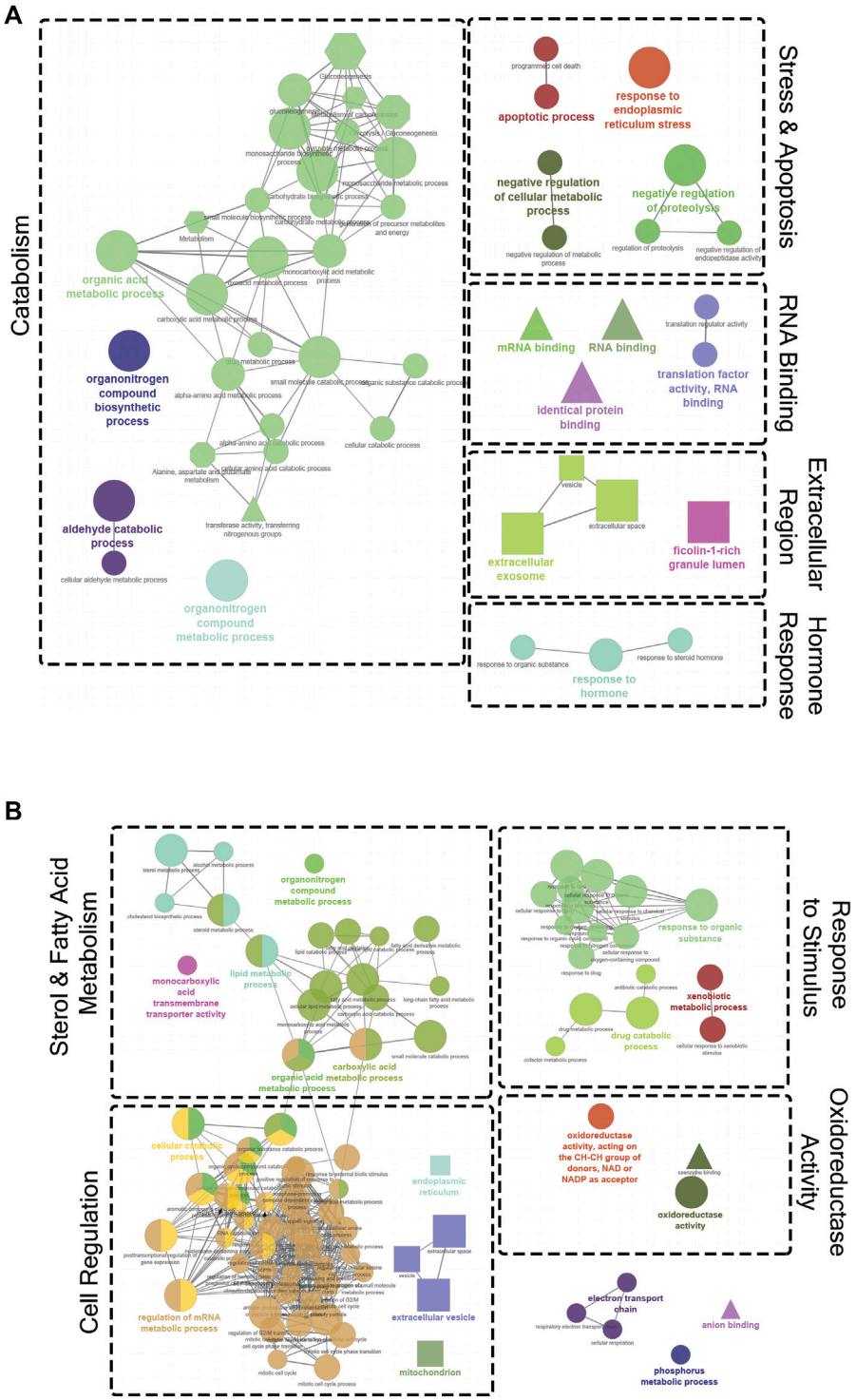
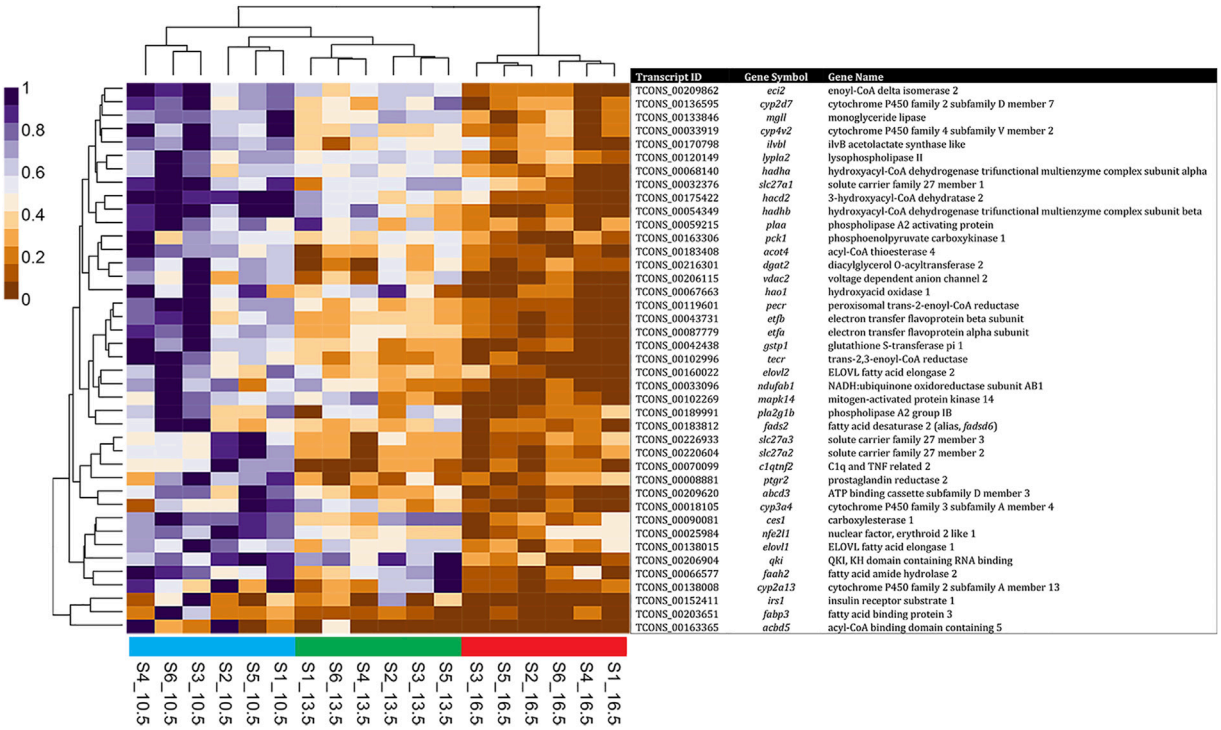


FIGURE 4 | Gene ontology (GO) term enrichment and pathway network analysis of significantly differentially expressed transcripts (DETs) between the 10.5°C and 16.5°C groups. **(A)** Downregulated DETs at 10.5°C (659 DETs). **(B)** Upregulated DETs at 10.5°C (718 DETs). Only significant ($p < 0.05$) network terms were included in the visualization. The diameter of each node indicates the level of significance for that specific term, with a larger node corresponding to a lower p-value. Node colour signifies related processes that share similar DETs. The shape of the node reflects the database where the term originated (circle, GO Biological Process; triangle, GO Molecular Function; square, GO Cellular Component; hexagon, Reactome). Dotted borders and labels were used to group related clusters and highlight general themes.



Transcript ID	Gene Symbol	Gene Name
TCONS_00209862	*eci2*	enoyl-CoA delta isomerase 2
TCONS_00136595	*cyp2d7*	cytochrome P450 family 2 subfamily D member 7
TCONS_00133846	*mgll*	monoglyceride lipase
TCONS_00033919	*cyp4v2*	cytochrome P450 family 4 subfamily V member 2
TCONS_00170798	*ilvbl*	ilvB acetolactate synthase like
TCONS_00120149	*lypla2*	lysophospholipase II
TCONS_00068140	*hadha*	hydroxyacyl-CoA dehydrogenase trifunctional multienzyme complex subunit alpha
TCONS_00032376	*slc27a1*	solute carrier family 27 member 1
TCONS_00175422	*hacd2*	3-hydroxyacyl-CoA dehydratase 2
TCONS_00054349	*hadhb*	hydroxyacyl-CoA dehydrogenase trifunctional multienzyme complex subunit beta
TCONS_00059215	*plaa*	phospholipase A2 activating protein
TCONS_00163306	*pck1*	phosphoenolpyruvate carboxylkinase 1
TCONS_00183408	*acot4*	acyl-CoA thioesterase 4
TCONS_00216301	*dgat2*	diacylglycerol O-acyltransferase 2
TCONS_00206115	*vdac2*	voltage dependent anion channel 2
TCONS_00067663	*hao1*	hydroxyacid oxidase 1
TCONS_00119601	*pecr*	peroxisomal trans-2-enoyl-CoA reductase
TCONS_00043731	*etfb*	electron transfer flavoprotein beta subunit
TCONS_00087779	*etfa*	electron transfer flavoprotein alpha subunit
TCONS_00042438	*gstp1*	glutathione S-transferase pi 1
TCONS_00102996	*tecr*	trans-2,3-enoyl-CoA reductase
TCONS_00160022	*elovl2*	ELOVL fatty acid elongase 2
TCONS_00033096	*ndufab1*	NADH:ubiquinone oxidoreductase subunit AB1
TCONS_00102269	*mapk14*	mitogen-activated protein kinase 14
TCONS_00189991	*pla2g1b*	phospholipase A2 group IB
TCONS_00183812	*fads2*	fatty acid desaturase 2 (alias, *fadsd6*)
TCONS_00226933	*slc27a3*	solute carrier family 27 member 3
TCONS_00220604	*slc27a2*	solute carrier family 27 member 2
TCONS_00070099	*c1qtnf2*	C1q and TNF related 2
TCONS_00008881	*ptgr2*	prostaglandin reductase 2
TCONS_00209620	*abcd3*	ATP binding cassette subfamily D member 3
TCONS_00018105	*cyp3a4*	cytochrome P450 family 3 subfamily A member 4
TCONS_00090081	*ces1*	carboxylesterase 1
TCONS_00025984	*nfe2l1*	nuclear factor, erythroid 2 like 1
TCONS_00138015	*elovl1*	ELOVL fatty acid elongase 1
TCONS_00206904	*qki*	QKI, KH domain containing RNA binding
TCONS_00066577	*faah2*	fatty acid amide hydrolase 2
TCONS_00138008	*cyp2a13*	cytochrome P450 family 2 subfamily A member 13
TCONS_00152411	*irs1*	insulin receptor substrate 1
TCONS_00203651	*fabp3*	fatty acid binding protein 3
TCONS_00163365	*acbd5*	acyl-CoA binding domain containing 5

FIGURE 5 | Heatmap and hierarchical clustering of non-redundant differentially expressed transcripts (41 DETs) found in comparison of the 10.5°C and 16.5°C groups associated with GO: 0006631—Fatty Acid Metabolic Process. RNA samples used in sequencing are labeled on the base of the figure, with standardized fragments per kilobase of transcript per million mapped reads (FPKM) values for each DET clustered along the vertical axis.

the fillet were negatively correlated with expression of *thrsp*, *hacd2*, and *fabpi*. Numerous transcripts (i.e., *elovl2*, *thrsp*, *s27a2*, *hacd2*, *hgfa*, *fabpi*) also exhibited a negative relationship to lysine and/or methionine fillet concentrations. In contrast, *hsp90ab1* showed a positive correlation with both lysine and methionine concentrations. The PCA seen in **Figure 8B** illustrates that each temperature group clustered separately in multivariate space, with the 13.5°C aggregated between the other two treatments. The eigenvectors of the target transcripts influenced Dimension 1 (Dim1) the most (i.e., the top five contributing variables to Dim1 were *hacd2*, *thrsp*, *hgfa*, *fabpi*, and *s27a2*), while phenotypic data contributed more so to Dimension 2 (Dim2) (i.e., the top five contributing variables to Dim2 were EPA, Σ ω3, DHA, HSI, and P). Altogether, this information helped elucidate which factors are most relevant in helping differentiate temperature treatments in this study and allowed for transcriptional responses to be related to phenotypic measurements.

DISCUSSION

The aim of this study was to compare the transcriptomic responses in the liver of AAS reared at either 10.5, 13.5 or 16.5°C using RNA-seq. Comparing results from DESeq2 and edgeR differential expression analyses, the greatest number of DETs was found when comparing the 10.5 and 16.5°C temperature groups. That comparison showed that temperature elicits broad effects on numerous biological systems in the livers of AAS, many related to metabolism, as evidenced by finding 282 enriched GO terms. Pathway analysis further illustrated that at elevated temperature, catabolism-, stress- and apoptosis-related transcripts were upregulated in comparison to 10.5°C. In contrast, expression of oxidoreductase activity- and lipid/fatty acid metabolism-relevant transcripts was downregulated at 16.5°C. Our results were validated by qPCR, with eight out of nine metabolic-related transcripts tested positively correlated to the RNA-seq data.

It was important to size-match salmon from the separate rearing temperatures for this study, as previous findings have shown that life stage significantly impacts hepatic transcriptomic responses in diploid and triploid Atlantic salmon (Odei et al., 2020). Also, as evidenced by the results of Ignatz et al. (2020b), several phenotypic differences were observed when comparing AAS reared at these three temperatures as they grew from first feeding fry up to 1,500 g. The fact that these fish were all reproductively sterile, monosex and genetically related also limited variation, allowing for a better assessment of temperature-mediated effects.

A stringent approach was taken when assessing DETs, whereby only transcripts found differentially expressed by both DESeq2 and edgeR were included in downstream GTEA and pathway analyses. In this study, more DETs were found by

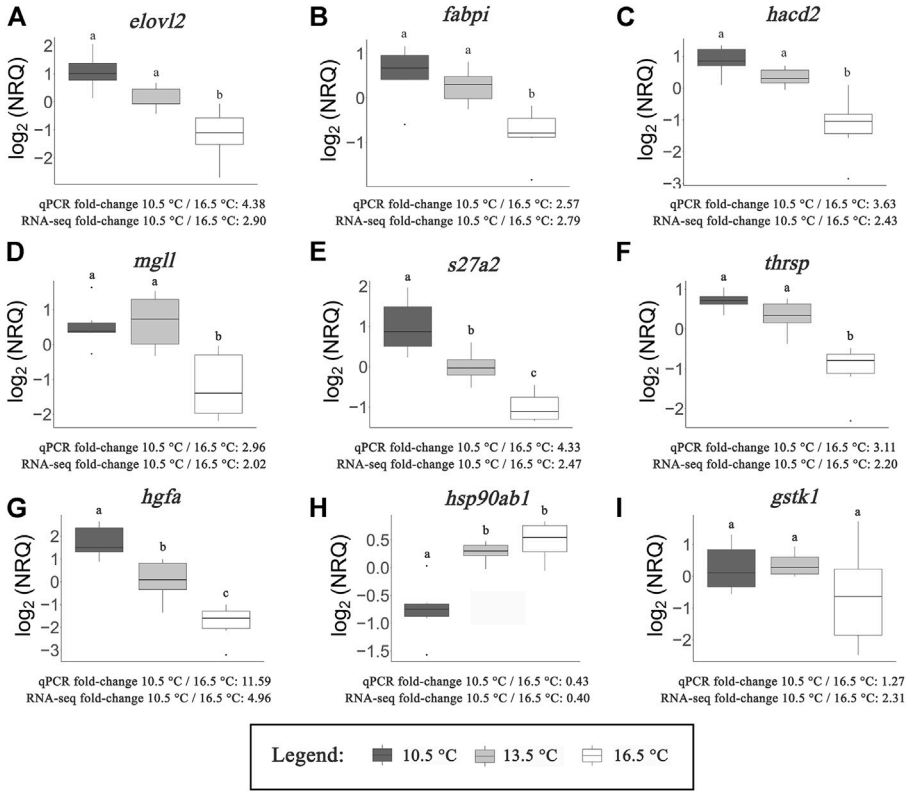


FIGURE 6 | qPCR results of targeted transcripts involved in fatty acid/lipid metabolism **(A–F)**, protein metabolism **(G,H)** and oxidoreductase activity **(I)**. Normalized relative quantities (NRQs) are provided $\log_2$ transformed. Letters represent significant differences ($p < 0.05$) between rearing temperatures as determined by one-way ANOVA and subsequent Tukey's post-hoc tests. Average fold-change values calculated from qPCR results and RNA-seq data (average between DESeq2 and edgeR) are provided in comparison of the 10.5°C and 16.5°C treatments.

DESeq2 compared to edgeR, a general trend that has been previously reported (Polinski et al., 2016; Hall et al., 2021). In Tomalty et al. (2015), juvenile diploid Chinook salmon (*Oncorhynchus tshawytscha*) were held in 12°C and then exposed to higher temperatures (i.e., 15, 18, 21, or 25°C) for 3 h followed by a 1 h recovery at 12°C. These authors reported a smaller number of DETs when comparing 12 vs. 15°C (9 DETs) or even 12 vs. 18°C (120 DETs) in gill tissue. These data differ from the results presented herein and could be explained by the difference in tissue sampled (i.e., gill vs. liver), the difference in study species, and/or the possibility that chronic exposure (i.e., after 10–12 months) to a different temperature may lead to more extensive changes in the transcriptome compared to acute thermal challenges. The latter point is further supported, wherein only 243 and 88 DE genes were identified in the liver of juvenile Atlantic salmon exposed to 23°C for 6 and 24 h, respectively compared to fish held at 13°C (Shi et al., 2019). This shows that exposure to even more extreme temperatures for short durations does not impact the transcriptome as much as shown in the current study between 10.5 and 16.5°C. However, this result could also potentially be influenced by life stage of the fish.

In 2.5 kg Atlantic salmon reared in saltwater, triploids exhibited a greater basal cortisol response when exposed to gradual increases in temperature up to 18°C compared to diploids (Sambraus et al., 2018). Further, cortisol levels increased at 18°C as compared to lower rearing temperatures (i.e., 9–15°C) in triploids (Sambraus et al., 2018). These results reflect that at 18°C, triploid Atlantic salmon experience chronic stress, a feature in agreement with the current study with AAS reared at 16.5°C as reflected in the GO term enrichment and pathway analyses. It is noteworthy that differences in basal cortisol concentrations of AAS were not detected between 10.5, 13.5, and 16.5°C at 800 g previously (Ignatz et al., 2020a). However, several classic indicators of heat stress were either upregulated [e.g., *hsp90ab1*, *serpinh1* (alias *hsp47*, encoding Serpin H1)] or downregulated [e.g., mitochondrial uncoupling protein 2 (*ucp2*), cold-inducible RNA-binding protein (*cirbp*), peroxiredoxin 6 (*prdx6*)] at 16.5°C compared to 10.5°C (**Supplementary Table S1**). These markers align with previous studies that investigated expression differences and DNA methylation dynamics in the liver between non-transgenic Atlantic salmon reared at 12 and 20°C (Beemelmanns et al., 2021a; Beemelmanns et al., 2021b; Beemelmanns et al., 2021c). Several of these genes have also been identified as differentially expressed during heat stress in other salmonids across a variety of tissues (Jeffries et al., 2012; Jeffries et al., 2014; Tomalty et al., 2015; Tan et al., 2016; Shi et al., 2019; Quan et al., 2021). Further study is recommended to parse out potential separate effects of

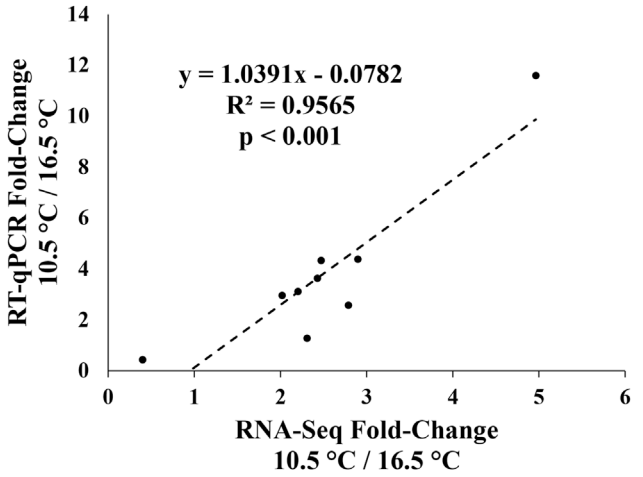


FIGURE 7 | Regression plot of transcript expression fold-changes between the 10.5°C and 16.5°C treatments calculated from the RNA-seq dataset (average between DESeq2 and edgeR) and normalized relative quantity (NRQ) values from the qPCR-validation.

triploidy and growth hormone transgenesis on thermally responsive transcripts of Atlantic salmon.

In both juvenile Atlantic salmon and brook charr (*Salvelinus fontinalis*), triploids had lower erythrocyte levels of several heat shock proteins, including HSP90, compared to diploids at several acclimation temperatures (i.e., 5, 15, and 25°C) (Saranyan et al., 2017). Lower basal levels of these proteins were hypothesized to limit triploid salmonids in their ability to prevent protein damage or denaturation at high temperatures (Saranyan et al., 2017). In the current study, expression of *hsp90ab1* was upregulated at 16.5 and 13.5°C compared to 10.5°C. It may be that even at 13.5°C, AAS are attempting to compensate for their lower basal protein synthesis by upregulating expression of *hsp90ab1* to prevent protein degradation, but further study is warranted. Similarly, with significantly higher expression of *serpinh1* at 16.5 vs. 10.5°C, it has been suggested that Atlantic salmon attempt to synthesize and stabilize collagen molecules to maintain hepatocyte structures at high temperatures (Beemelmanns et al., 2021b; Beemelmanns et al., 2021c). Additionally, Serpin H1 is thought to help neutralize reactive oxygen species (ROS) during heat stress as evidenced by increased transcript expression observed in rainbow trout (*Oncorhynchus mykiss*) (Wang et al., 2016). Thus, by upregulating *serpinh1* at 16.5°C, AAS are likely attempting to reestablish cellular homeostasis by eliminating ROS.

In contrast, downregulating *ucp2* at 16.5°C might help reduce mitochondrial ROS formation by increasing mitochondrial coupling (Laskowski et al., 2016). In fact, studies have shown that long-term acclimation to 20°C in diploid Atlantic salmon significantly reduces mitochondrial ROS production when exposed to higher temperatures compared to salmon acclimated to 12°C (Gerber et al., 2021; Gerber et al., 2020). The lower expression of *cirbp* at 16.5°C compared to 10.5°C is also indicative of heat stress, as this transcript is routinely downregulated at elevated temperature but upregulated at lower temperatures in salmonids and other vertebrates (Zhong and Huang, 2017; Akbarzadeh et al., 2018; Beemelmanns et al., 2021b; Beemelmanns et al., 2021c). This chaperone-encoding mRNA is involved in cell proliferation, survival and apoptosis and is also hypoxia sensitive (Zhong and Huang, 2017). Further, *prdx6*, which is involved in phospholipid homeostasis and lipid peroxidation repair (Fisher, 2011; Arevalo and Vázquez-Medina, 2018), was also a key downregulated transcript at 16.5°C. It has also been noted that in the liver of large (1.5–2.0 kg) immature Atlantic salmon held at 19°C for

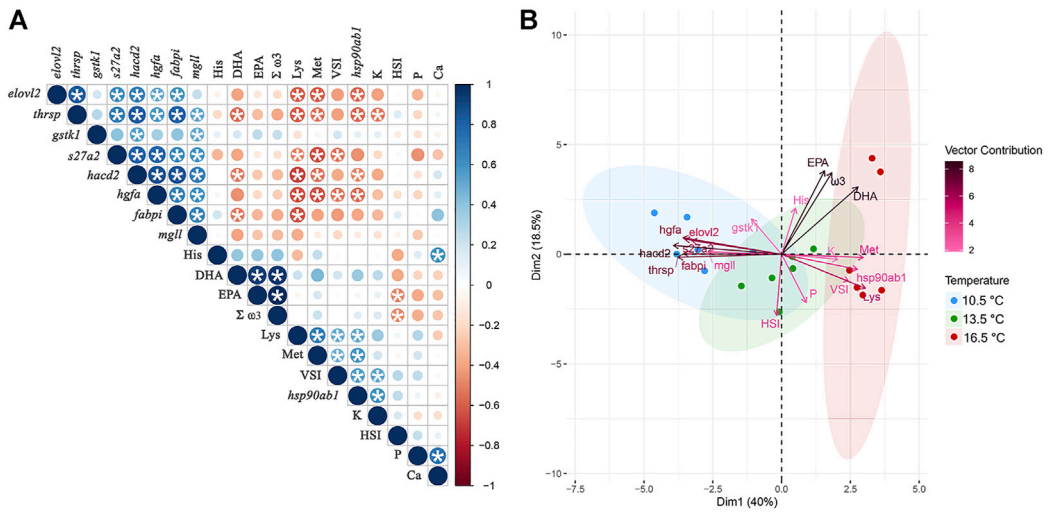


FIGURE 8 | (A) Correlogram of expression profiles for all transcripts used for RT-qPCR validation alongside fillet fatty acid (ω3, total omega-3; DHA, docosahexaenoic acid; EPA, eicosapentaenoic acid), amino acid (Lys, lysine; Met, methionine; His, histidine), and mineral (P, phosphorus; Ca, calcium) compositional results and phenotypic (HSI, hepatosomatic index; VSI, viscerosomatic index; K, condition factor) measurement data amongst all fish. The coloured scale shows degree of correlation ranging from $r = -1.0$ (red) to $r = 1.0$ (blue). The size of the coloured circle indicates its significance level and circles containing asterisks are significant at $p < 0.05$. **(B)** Principal component analysis (PCA) of the same multivariate dataset with distinction between rearing temperature treatments based on their phenotypic characteristics and transcript expression. Variable eigenvectors are coloured based on their contribution to Dimension (Dim) 1 and 2 of the PCA plot.

45 days, that transcripts associated with protection against oxidative stress become downregulated compared to salmon held at 13°C (Olsvik et al., 2013). This feature is also found comparing the 10.5 and 16.5°C groups in the current study. Altogether, this suggests that AAS reared at this higher temperature may struggle with regulating oxidative stress responses.

As female triploids reduce investment in sexual maturation, this prevents the reallocation of nutrients from muscle tissue toward gonad development (Manor et al., 2014). Thus, triploid Atlantic salmon have different fillet lipid/fatty acid profiles compared to diploids (Murray et al., 2018; Everson et al., 2021). Previous findings from this experiment have also shown that rearing temperature influenced deposition rates and retention efficiencies of ω3 fatty acids (Ignatz et al., 2020b). While expression of transcripts related to fatty acid metabolism (i.e., *hacd2*, *fabpi*, *thrsp*) were negatively correlated with ω3 fatty acid composition (significant with DHA composition comparisons) within the fillet of these salmon, this likely reflects the ongoing shift in nutrient utilization of fish at 16.5°C. Although fillet ω3 fatty acids did not differ on a compositional level between rearing temperature groups at 800 g, deposition rates of total ω3, DHA, and EPA were found significantly lower during the 800–1,500 g growth period in fish reared at 16.5°C compared to the two lower temperature treatments (Ignatz et al., 2020b). Therefore, the changes in gene expression found in the current study (i.e., downregulation of fatty acid metabolism genes at 16.5°C compared to 10.5 and 13.5°C) were likely indicative of the start of this change in nutrient allocation. Ultimately, this can lead to adverse effects on final product composition and quality, as lower production and deposition of ω3 fatty acids in particular would not be ideal for the marketability of AAS. Thus, this study helps to define the optimal rearing temperature of AAS to avoid these adverse effects.

Both GTEA and pathway analysis highlighted that lipid/fatty acid metabolism was negatively impacted at 16.5°C. Similar findings have been reported in other transcriptomic studies at elevated temperature in diploid salmonids (Jeffries et al., 2012; Tomalty et al., 2015; Shi et al., 2019; Bowen et al., 2020; Beemelmanns et al., 2021c). Changes in lipid metabolism are relevant as regulation of cholesterol and phospholipids help maintain cell membrane rigidity during changes in temperature; as temperature increases, membrane structure becomes more fluid which can ultimately lead to cell death (Hazel, 1979; Fodor et al., 1995; Crockett, 1998; Farkas et al., 2001; Liu et al., 2019). "Cholesterol metabolic process" was an enriched GO term between the 10.5 and 16.5°C temperature groups, which further supports this notion. Similarly, with the downregulation of transcripts associated with the GO term "oxidoreductase activity", this potentially offers AAS at 16.5°C less protection against oxidative stress. Several studies in diploid salmonids have shown similar results either following a heat shock or after chronic exposure to elevated temperatures (Olsvik et al., 2013; Nakano et al., 2014; Beemelmanns et al., 2021b; Beemelmanns et al., 2021c). Additionally, as catabolism, stress and apoptosis were all identified as general themes of upregulated pathways at 16.5°C, this suggests that it is not ideal for AAS to be reared at this high of a temperature.

Potentially due to differences in cellular dimensions and decreased cellular surface-to-volume ratio (Piferrer et al., 2009), triploids are also more prone to skeletal deformities and the development of cataracts (Sambraus et al., 2017b; Fraser et al., 2020; Jagiełło et al., 2021), which has led to the recommendation to increase dietary phosphorus and histidine for triploids to address each of those respective issues (Burke et al., 2010; Fjelldal et al., 2015; Taylor et al., 2015; Smedley et al., 2016). Therefore, these nutrients were important considerations in both past (Ignatz et al., 2020b) and present studies. In the current study, no correlations were found between phosphorus and histidine concentrations in the fillet with any of the target transcripts that were qPCR-validated. However, "phosphorus metabolic process" was a significantly enriched GO term that was found between the 10.5 and 16.5°C groups. Additionally, levels of methionine and/or lysine in the fillet were negatively correlated with 6 out of 7 fatty acid metabolism relevant transcripts. A previous recommendation was made that AAS may require additional methionine at temperatures >10.5°C (Ignatz et al., 2020b). It may be possible that especially at 16.5°C, the dietary requirements of AAS were not being fully met, which may have negatively impacted their metabolism of numerous nutrients.

No differences in immune responses between triploid and diploid salmonids during both bacterial and parasitic infections have been reported (Weber et al., 2013; Frenzl et al., 2014; Chalmers et al., 2016, Chalmers et al., 2017). In contrast, some evidence suggests that triploid Atlantic salmon may mount a more effective immune response against viral agents than diploids (Herath et al., 2017; Moore et al., 2017; Brown et al., 2021). In a subset of fish also assessed at 800 g from the same overarching experiment as the current study, the innate antiviral immune response was found to be significantly impacted by rearing temperature (Ignatz et al., 2020a). Significantly higher fold-change values of several immune biomarkers (i.e., *rsad2*, *isg15a*, *ifng*) were found in AAS reared at 10.5°C vs. 16.5°C comparing fish at 24 h post-injection with either PBS or pIC (Ignatz et al., 2020a). It was suggested that salmon reared at 10.5°C ultimately had a more robust innate antiviral response than both other temperature treatments (Ignatz et al., 2020a). From the current RNA-seq results, several GO terms including "immune system process", "immune effector process", and "activation of the innate immune response" were enriched when comparing DETs between the 10.5 and 16.5°C groups. This illustrates that temperature significantly influenced basal immune-relevant transcript expression in AAS liver, with higher expression levels at 10.5 than 16.5°C. This could mean that AAS reared at 10.5°C would be more prepared to respond to pathogens and mount a more effective immune response than if they are reared at elevated temperature as previously suggested (Ignatz et al., 2020a), but this requires further testing. While temperature did not influence basal expression of immune-relevant genes in Ignatz et al. (2020a), these results differ from what has been shown in basal expression in liver and head kidney of post-smolt diploid Atlantic salmon reared at 12 and 20°C where expression was generally found higher at elevated temperature (Zanuzzo et al., 2020; Beemelmanns et al., 2021b).

CONCLUSION

Overall, the results from this study indicate that it is not optimal to rear AAS at 16.5°C due to increases in expression of transcripts related to stress, apoptosis and catabolism. Moreover, negative regulation of lipid/fatty acid metabolism is not ideal especially as it has previously been noted that deposition rates of ω3 fatty acids were lower at 16.5°C compared to 10.5 and 13.5°C (Ignatz et al., 2020b). This is the first RNA-seq study to investigate the impact of rearing temperature on the hepatic transcriptome of AAS and helps inform AquaBounty's husbandry practices using their land-based freshwater RAS. This is important as commercialization of AquAdvantage Salmon is ongoing in the United States and Canada, and these results can be used to optimize the physiological performance of these fish. The current results are also informative as AAS appear to have responded similarly to conventional salmonids when exposed to elevated temperatures. Future work could consider optimizing the diet for AAS reared at 10.5 and/or 13.5°C and using transcriptomics to help determine best practices for their commercial production.

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 **Table 1**.

ETHICS STATEMENT

The animal study was reviewed and approved by the Animal Care Committees of both AquaBounty Canada (Animal Care Protocol ABC-ACC-013) and Memorial University of Newfoundland (17-03-JW). The animals in this study were handled and cared for in accordance with the Canadian Council on Animal Care's Guidelines on the Care and Use of Fish in Research, Teaching, and Testing (Canadian Council on Animal Care, 2005). Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

EHI assisted in study design and implementation, as well as performed and analyzed all differential expression analysis, including GO term enrichment and pathway analyses. EHI also took the lead role of writing this manuscript. TSH was involved in study design and editing of the manuscript, as well as assisted with interpretation of sequencing results and performed transcriptome annotation with putative orthologs. SK assisted in data analysis and interpretation, in addition to manuscript editing. LMB, TJB, and CDR all contributed to study design and assisted in editing of the manuscript. JDW and MLR supervised and provided comments on all steps of the research and manuscript preparation. All authors read and approved the final manuscript.

FUNDING

Financial support of this project was provided through the Prince Edward Island Department of Agriculture and Fisheries—Aquaculture Technology Program, under the project title "Temperature-Dependent Gene Expression Study in AquAdvantage® Salmon", awarded to AquaBounty Canada. Funding from a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (2020-04519) to MLR also contributed to this project. SK's salary was funded by the Ocean Frontier Institute through an award from the Canada First Research Excellence Fund, Publishing fees were paid by Memorial University Libraries' Open Access Author Fund.

ACKNOWLEDGMENTS

Assistance from Damien MacDonald, Amanda MacLean-Franciosa, Jonathan MacInnis, Jonathan Veinot, Lee Pagnutti, and Mauricio Rojas with fish husbandry and sample collection was greatly appreciated. Training and use of equipment from Dr. Sara Purcell and Dr. Mark Fast in the Hoplite Laboratory of the Atlantic Veterinary College was also extremely valued. Lastly, thank you to the teams at Génome Québec and C3G, in particular Eloi Mercier and François Lefebvre, for their assistance with sequencing and bioinformatics analysis. C3G is a Genomics Technology Platform (GTP) supported by the Canadian Government through Genome Canada.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.852165/full#supplementary-material

REFERENCES

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., et al. (2018). The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2018 Update. *Nucleic Acids Res.* 46, W537–W544. doi:10.1093/nar/gky379

Akbarzadeh, A., Günther, O. P., Houde, A. L., Li, S., Ming, T. J., Jeffries, K. M., et al. (2018). Developing Specific Molecular Biomarkers for Thermal Stress in Salmonids. *BMC Genomics* 19, Article 749. doi:10.1186/s12864-018-5108-9

AquaBounty Technologies, Inc. (2010). *Environmental Assessment for AquAdvantage® Salmon*.

Arevalo, J., and Vázquez-Medina, J. (2018). The Role of Peroxiredoxin 6 in Cell Signaling. *Antioxidants* 7, 172. doi:10.3390/antiox7120172

Atkins, M. E., and Benfey, T. J. (2008). Effect of Acclimation Temperature on Routine Metabolic Rate in Triploid Salmonids. *Comp. Biochem. Physiology Part A Mol. Integr. Physiology* 149, 157–161. doi:10.1016/j.cbpa.2007.11.004

Beemelmanns, A., Ribas, L., Anastasiadi, D., Moraleda-Prados, J., Zanuzzo, F. S., Rise, M. L., et al. (2021a). DNA Methylation Dynamics in Atlantic Salmon (*Salmo salar*) Challenged with High Temperature and Moderate Hypoxia. *Front. Mar. Sci.* 7, Article 604878. doi:10.3389/fmars.2020.604878

Beemelmanns, A., Zanuzzo, F. S., Sandrelli, R. M., Rise, M. L., and Gamperl, A. K. (2021b). The Atlantic Salmon's Stress- and Immune-Related Transcriptional Responses to Moderate Hypoxia, an Incremental Temperature Increase, and These Challenges Combined. *G3 Genes|Genomes|Genetics* 11, Article jkab102. doi:10.1093/g3journal/jkab102

Beemelmanns, A., Zanuzzo, F. S., Xue, X., Sandrelli, R. M., Rise, M. L., and Gamperl, A. K. (2021c). The Transcriptomic Responses of Atlantic Salmon (*Salmo salar*) to High Temperature Stress Alone, and in Combination with Moderate Hypoxia. *BMC Genomics* 22, Article 261. doi:10.1186/s12864-021-07464-x

Benfey, T. J. (2016). Effectiveness of Triploidy as a Management Tool for Reproductive Containment of Farmed Fish: Atlantic Salmon (*Salmo salar*) as a Case Study. *Rev. Aquacult* 8, 264–282. doi:10.1111/raq.12092

Benfey, T. J., Mccabe, L. E., and Pepin, P. (1997). Critical Thermal Maxima of Diploid and Triploid Brook Charr, *Salvelinus fontinalis*. *Environ. Biol. Fishes* 49, 259–264. doi:10.1023/a:1007361231325

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a Cytoscape Plug-In to Decipher Functionally Grouped Gene Ontology and Pathway Annotation Networks. *Bioinformatics* 25, 1091–1093. doi:10.1093/bioinformatics/btp101

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170

Bourgey, M., Dali, R., Eveleigh, R., Chen, K. C., Letourneau, L., Fillon, J., et al. (2019). GenPipes: an Open-Source Framework for Distributed and Scalable Genomic Analyses. *GigaScience* 8, Article giz037. doi:10.1093/gigascience/giz037

Bowden, A. J., Andrewartha, S. J., Elliott, N. G., Frappell, P. B., and Clark, T. D. (2018). Negligible Differences in Metabolism and Thermal Tolerance between Diploid and Triploid Atlantic Salmon (*Salmo salar* L.). *J. Exp. Biol.* 221, Article jeb166975. doi:10.1242/jeb.166975

Bowen, L., von Biela, V. R., McCormick, S. D., Regish, A. M., Waters, S. C., Durbin-Johnson, B., et al. (2020). Transcriptomic Response to Elevated Water Temperatures in Adult Migrating Yukon River Chinook Salmon (*Oncorhynchus tshawytscha*). *Conserv. Physiol.* 8, Article coaa084. doi:10.1093/conphys/coaa084

Brown, R., Moore, L., Mani, A., Patel, S., and Salinas, I. (2021). Effects of Ploidy and Salmonid Alphavirus Infection on the Skin and Gill Microbiome of Atlantic Salmon (*Salmo salar*). *PLOS ONE* 16, Article e0243684. doi:10.1371/journal.pone.0243684

Burke, H. A., Sacobie, C. F. D., Lall, S. P., and Benfey, T. J. (2010). The Effect of Triploidy on Juvenile Atlantic Salmon (*Salmo salar*) Response to Varying Levels of Dietary Phosphorus. *Aquaculture* 306, 295–301. doi:10.1016/j.aquaculture.2010.05.002

Caballero-Solares, A., Hall, J. R., Xue, X., Eslamloo, K., Taylor, R. G., Parrish, C. C., et al. (2017). The Dietary Replacement of Marine Ingredients by Terrestrial Animal and Plant Alternatives Modulates the Antiviral Immune Response of Atlantic Salmon (*Salmo salar*). *Fish Shellfish Immunol.* 64, 24–38. doi:10.1016/j.fsi.2017.02.040

Cai, J., and Leung, P. (2017). *Short-term Projection of Global Fish Demand and Supply Gaps*. Rome: Food and Agriculture Organization of the United Nations. FAO Fisheries and Aquaculture Technical Paper no. 607.

Canadian Council on Animal Care (2005). Canadian Council on Animal Care Guidelines on: the Care and Use of Fish in Research, Teaching and Testing. *Can. Counc. Animal Care*, 1–94.

Chalmers, L., Taylor, J. F., Roy, W., Preston, A. C., Migaud, H., and Adams, A. (2017). A Comparison of Disease Susceptibility and Innate Immune Response between Diploid and Triploid Atlantic Salmon (*Salmo salar*) Siblings Following Experimental Infection with *Neoparamoeba pararurans*, Causative Agent of Amoebic Gill Disease. *Parasitology* 144, 1229–1242. doi:10.1017/S0031182017000622

Chalmers, L., Thompson, K. D., Taylor, J. F., Black, S., Migaud, H., North, B., et al. (2016). A Comparison of the Response of Diploid and Triploid Atlantic Salmon (*Salmo salar*) Siblings to a Commercial Furunculosis Vaccine and Subsequent Experimental Infection with Aeromonas salmonicida. *Fish Shellfish Immunol.* 57, 301–308. doi:10.1016/j.fsi.2016.08.049

Chiasson, M. A., Pelletier, C. S., and Benfey, T. J. (2009). Triploidy and Full-Sib Family Effects on Survival and Growth in Juvenile Arctic Charr (*Salvelinus alpinus*). *Aquaculture* 289, 244–252. doi:10.1016/j.aquaculture.2009.01.010

Cook, J. T., McNiven, M. A., Richardson, G. F., and Sutterlin, A. M. (2000). Growth Rate, Body Composition and Feed Digestibility/conversion of Growth-Enhanced Transgenic Atlantic Salmon (*Salmo salar*). *Aquaculture* 188, 15–32. doi:10.1016/s0044-8486(00)00331-8

Crockett, E. L. (1998). Cholesterol Function in Plasma Membranes from Ectotherms: Membrane-specific Roles in Adaptation to Temperature. *Am. Zool.* 38, 291–304. doi:10.1093/icb/38.2.291

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* 29, 15–21. doi:10.1093/bioinformatics/bts635

Eden, E., Lipson, D., Yogev, S., and Yakhini, Z. (2007). Discovering Motifs in Ranked Lists of DNA Sequences. *PLoS Comput. Biol.* 3, e39–0522. doi:10.1371/journal.pcbi.0030039

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a Tool for Discovery and Visualization of Enriched GO Terms in Ranked Gene Lists. *BMC Bioinforma.* 10, Article 48. doi:10.1186/1471-2105-10-48

Ellis, L. E., Sacobie, C. F. D., Kieffer, J. D., and Benfey, T. J. (2013). The Effects of Dissolved Oxygen and Triploidy on Critical Thermal Maximum in Brook Charr (*Salvelinus fontinalis*). *Comp. Biochem. Physiology Part A Mol. Integr. Physiology* 166, 426–433. doi:10.1016/j.cbpa.2013.07.026

Everson, J. L., Weber, G. M., Manor, M. L., Tou, J. C., and Kenney, P. B. (2021). Polyploidy Affects Fillet Yield, Composition, and Fatty Acid Profile in Two-Year Old, Female Rainbow Trout, *Oncorhynchus mykiss*. *Aquaculture* 531, 735873. doi:10.1016/j.aquaculture.2020.735873

FAO (2018). *The State of World Fisheries and Aquaculture 2018 - Meeting the Sustainable Development Goals. (No. Licence: CC BY-NC-SA 3.0 IGO*. Rome: Food and Agriculture Organization of the United Nations.

FAO (2020). The State of World Fisheries and Aquaculture 2020. Rome: Food and Agriculture Organization of the United Nations. Sustainability in action. doi:10.4060/ca9229en

Farkas, T., Fodor, E., Kitajka, K., and Halver, J. E. (2001). Response of Fish Membranes to Environmental Temperature. *Aquac. Res.* 32, 645–655. doi:10.1046/j.1365-2109.2001.00600.x

Fisher, A. B. (2011). Peroxiredoxin 6: A Bifunctional Enzyme with Glutathione Peroxidase and Phospholipase A2 Activities. *Antioxidants Redox Signal.* 15, 831–844. doi:10.1089/ars.2010.3412

Fjelldal, P. G., Hansen, T. J., Lock, E.-J., Wargelius, A., Fraser, T. W. K., Sambraus, F., et al. (2015). Increased Dietary Phosphorous Prevents Vertebral Deformities in Triploid Atlantic Salmon (*Salmo salar* L.). *Aquacult Nutr.* 22, 72–90. doi:10.1111/anu.12238

Fodor, E., Jones, R. H., Buda, C., Kitajka, K., Dey, I., and Farkas, T. (1995). Molecular Architecture and Biophysical Properties of Phospholipids during Thermal Adaptation in Fish: An Experimental and Model Study. *Lipids* 30, 1119–1126. doi:10.1007/BF02536612

Fraser, T. W. K., Hansen, T. J., Sambraus, F., and Fjelldal, P. G. (2020). Vertebral Deformities in Interspecific Diploid and Triploid Salmonid Hybrids. *J. Fish. Biol.* 98, 1059–1070. doi:10.1111/jfb.14353

Frenzl, B., Migaud, H., Fjelldal, P. G., Shinn, A. P., Taylor, J. F., Richards, R. H., et al. (2014). Triploid and Diploid Atlantic Salmon Show Similar Susceptibility to Infection with *Salmon lice Lepeophtheirus Salmonis*. *Pest. Manag. Sci.* 70, 982–988. doi:10.1002/ps.3639

Ganga, R., Tibbetts, S. M., Wall, C. L., Plouffe, D. A., Bryenton, M. D., Peters, A. R., et al. (2015). Influence of Feeding a High Plant Protein Diet on Growth and

Nutrient Utilization to Combined 'all-Fish' Growth-Hormone Transgenic Diploid and Triploid Atlantic Salmon (*Salmo salar* L.). *Aquaculture* 446, 272–282. doi:10.1016/j.aquaculture.2015.05.010

Gerber, L., Clow, K. A., and Gamperl, A. K. (2021). Acclimation to Warm Temperatures Has Important Implications for Mitochondrial Function in Atlantic Salmon (*Salmo salar*). *J. Exp. Biol.* 224, Article jeb236257. doi:10.1242/jeb.236257

Gerber, L., Clow, K. A., Mark, F. C., and Gamperl, A. K. (2020). Improved Mitochondrial Function in Salmon (*Salmo salar*) Following High Temperature Acclimation Suggests that There Are Cracks in the Proverbial 'Ceiling'. *Sci. Rep.* 10, Article 21636. doi:10.1038/s41598-020-78519-4

Hall, J. R., Lehnert, S. J., Gonzalez, E., Kumar, S., Hanlon, J. M., Morris, C. J., et al. (2021). Snow Crab (*Chionoecetes opilio*) Hepatopancreas Transcriptome: Identification and Testing of Candidate Molecular Biomarkers of Seismic Survey Impact. *Fish. Res.* 234, 105794. doi:10.1016/j.fishres.2020.105794

Handeland, S. O., Imsland, A. K., and Stefansson, S. O. (2008). The Effect of Temperature and Fish Size on Growth, Feed Intake, Food Conversion Efficiency and Stomach Evacuation Rate of Atlantic Salmon Post-smolts. *Aquaculture* 283, 36–42. doi:10.1016/j.aquaculture.2008.06.042

Hazel, J. R. (1979). Influence of Thermal Acclimation on Membrane Lipid Composition of Rainbow Trout Liver. *Am. J. Physiology-Regulatory, Integr. Comp. Physiology* 236, R91–R101. doi:10.1152/ajpregu.1979.236.1.R91

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., and Vandesompele, J. (2007). qBase Relative Quantification Framework and Software for Management and Automated Analysis of Real-Time Quantitative PCR Data. *Genome Biol.* 8, R19. doi:10.1186/gb-2007-8-2-r19

Herath, T. K., Ashby, A. J., Jayasuriya, N. S., Bron, J. E., Taylor, J. F., Adams, A., et al. (2017). Impact of Salmonid Alphavirus Infection in Diploid and Triploid Atlantic Salmon (*Salmo salar* L.) Fry. *PLOS ONE* 12, Article e0179192. doi:10.1371/journal.pone.0179192

Hevrøy, E. M., Hunskår, C., de Gelder, S., Shimizu, M., Waagbø, R., Breck, O., et al. (2013). GH-IGF System Regulation of Attenuated Muscle Growth and Lipolysis in Atlantic Salmon Reared at Elevated Sea Temperatures. *J. Comp. Physiol. B* 183, 243–259. doi:10.1007/s00360-012-0704-5

Ignatz, E. H., Braden, L. M., Benfey, T. J., Caballero-Solares, A., Hori, T. S., Runighan, C. D., et al. (2020a). Impact of Rearing Temperature on the Innate Antiviral Immune Response of Growth Hormone Transgenic Female Triploid Atlantic Salmon (*Salmo salar*). *Fish Shellfish Immunol.* 97, 656–668. doi:10.1016/j.fsi.2019.12.081

Ignatz, E. H., Dumas, A., Benfey, T. J., Hori, T. S., Braden, L. M., Runighan, C. D., et al. (2020b). Growth Performance and Nutrient Utilization of Growth Hormone Transgenic Female Triploid Atlantic Salmon (*Salmo salar*) Reared at Three Temperatures in a Land-Based Freshwater Recirculating Aquaculture System (RAS). *Aquaculture* 519, 734896. doi:10.1016/j.aquaculture.2019.734896

Jagiełło, K., Polonis, M., and Ocalewicz, K. (2021). Incidence of Skeletal Deformities in Induced Triploid Rainbow Trout *Oncorhynchus mykiss* (Walbaum, 1792). *Oceanol. Hydrobiological Stud.* 50, 150–159. doi:10.2478/oandhs-2021-0014

Jeffries, K. M., Hinch, S. G., Sierocinski, T., Clark, T. D., Eliason, E. J., Donaldson, M. R., et al. (2012). Consequences of High Temperatures and Premature Mortality on the Transcriptome and Blood Physiology of Wild Adult Sockeye Salmon (*Oncorhynchus nerka*). *Ecol. Evol.* 2, 1747–1764. doi:10.1002/ece3.274

Jeffries, K. M., Hinch, S. G., Sierocinski, T., Pavlidis, P., and Miller, K. M. (2014). Transcriptomic Responses to High Water Temperature in Two Species of Pacific Salmon. *Evol. Appl.* 7, 286–300. doi:10.1111/eva.12119

Jones, S., Fast, M., Johnson, S., and Groman, D. (2007). Differential Rejection of Salmon Lice by Pink and Chum Salmon: Disease Consequences and Expression of Proinflammatory Genes. *Dis. Aquat. Org.* 75, 229–238. doi:10.3354/dao075229

Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S., and Karolchik, D. (2010). BigWig and BigBed: Enabling Browsing of Large Distributed Datasets. *Bioinformatics* 26, 2204–2207. doi:10.1093/bioinformatics/btq351

Laskowski, M., Augustynek, B., Kulawiak, B., Koprowski, P., Bednarczyk, P., Jarmuszkiewicz, W., et al. (2016). What Do We Not Know about Mitochondrial Potassium Channels? *Biochimica Biophysica Acta (BBA) - Bioenergetics* 1857, 1247–1257. doi:10.1016/j.bbabio.2016.03.007

Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., et al. (2016). The Atlantic Salmon Genome Provides Insights into Rediploidization. *Nature* 533, 200–205. doi:10.1038/nature17164

Liu, C., Dong, S., Zhou, Y., Shi, K., Pan, Z., Sun, D., et al. (2019). Temperature-dependent Fatty Acid Composition Change of Phospholipid in Steelhead Trout (*Oncorhynchus mykiss*) Tissues. *J. Ocean. Univ. China* 18, 519–527. doi:10.1007/s11802-019-3775-z

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, Article 550. doi:10.1186/s13059-014-0550-8

Madaro, A., Kjøglum, S., Hansen, T., Fjelldal, P. G., and Stien, L. H. (2021). A Comparison of Triploid and Diploid Atlantic Salmon (*Salmo salar*) Performance and Welfare under Commercial Farming Conditions in Norway. *J. Appl. Aquac.*, 1–15. doi:10.1080/10454438.2021.1916671

Manor, M. L., Weber, G. M., Cleveland, B. M., and Kenney, P. B. (2014). Effects of Feeding Level and Sexual Maturation on Fatty Acid Composition of Energy Stores in Diploid and Triploid Rainbow Trout (*Oncorhynchus mykiss*). *Aquaculture* 418-419, 17–25. doi:10.1016/j.aquaculture.2013.09.023

Moore, L. J., Nilsen, T. O., Jarungsriapisit, J., Fjelldal, P. G., Stefansson, S. O., Taranger, G. L., et al. (2017). Triploid Atlantic Salmon (*Salmo salar* L.) Post-smolts Accumulate Prevalence More Slowly Than Diploid Salmon Following Bath Challenge with Salmonid Alphavirus Subtype 3. *PLOS ONE* 12, Article e0175468. doi:10.1371/journal.pone.0175468

Morais, S., Monroig, O., Zheng, X., Leaver, M. J., and Tocher, D. R. (2009). Highly Unsaturated Fatty Acid Synthesis in Atlantic Salmon: Characterization of ELOVL5- and ELOVL2-like Elongases. *Mar. Biotechnol.* 11, 627–639. doi:10.1007/s10126-009-9179-0

Murray, D. S., Kainz, M. J., Hebberecht, L., Sales, K. R., Hindar, K., and Gage, M. J. G. (2018). Comparisons of Reproductive Function and Fatty Acid Fillet Quality between Triploid and Diploid Farm Atlantic Salmon (*Salmo salar*). *R. Soc. Open Sci.* 5, 180493. doi:10.1098/rsos.180493

Myers, R. A., Ortel, T. L., Waldrop, A., Dave, S., Ginsburg, G. S., and Voora, D. (2021). Aspirin Effects on Platelet Gene Expression Are Associated with a Paradoxical, Increase in Platelet Function. *Brit J. Clin. Pharma* 88, 2074–2083. doi:10.1111/bcp.15127

Nakano, T., Kameda, M., Shoji, Y., Hayashi, S., Yamaguchi, T., and Sato, M. (2014). Effect of Severe Environmental Thermal Stress on Redox State in Salmon. *Redox Biol.* 2, 772–776. doi:10.1016/j.redox.2014.05.007

Nuez-Ortín, W. G., Carter, C. G., Wilson, R., Cooke, I. R., Amoroso, G., Cobcroft, J. M., et al. (2017). Triploid Atlantic Salmon Shows Similar Performance, Fatty Acid Composition and Proteome Response to Diploids during Early Freshwater Rearing. *Comp. Biochem. Physiology Part D Genomics Proteomics* 22, 67–77. doi:10.1016/j.cbd.2017.02.005

Odei, D. K., Hagen, Ø., Peruzzi, S., Falk-Petersen, I.-B., and Fernandes, J. M. O. (2020). Transcriptome Sequencing and Histology Reveal Dosage Compensation in the Liver of Triploid Pre-smolt Atlantic Salmon. *Sci. Rep.* 10, Article 16836. doi:10.1038/s41598-020-73814-6

Oliveros, J. C. (2007). An Interactive Tool for Comparing Lists with Venn Diagrams [WWW Document]. VENNY. Available at: http://bioinfogp.cnb.csic.es/tools/venny/index.html.

Olsvik, P. A., Vikeså, V., Lie, K. K., and Hevrøy, E. M. (2013). Transcriptional Responses to Temperature and Low Oxygen Stress in Atlantic Salmon Studied with Next-Generation Sequencing Technology. *BMC Genomics* 14, 817. doi:10.1186/1471-2164-14-817

Pan, Y., Zhao, X., Li, D., Gao, T., and Song, N. (2021). Transcriptome Analysis Provides the First Insight into the Molecular Basis of Temperature Plasticity in Banggai Cardinalfish, *Pterapogon kauderni*. *Comp. Biochem. Physiology Part D Genomics Proteomics* 40, Article 100909. doi:10.1016/j.cbd.2021.100909

Pfaffl, M. W. (2001). A New Mathematical Model for Relative Quantification in Real-Time RT-PCR. *Nucleic Acids Res.* 29, e45. doi:10.1093/nar/29.9.e45

Picard Tools (2019). WWW Document. Available at: http://broadinstitute.github.io/picard.

Piferrer, F., Beaumont, A., Falguière, J.-C., Flajšhans, M., Haffray, P., and Colombo, L. (2009). Polyploid Fish and Shellfish: Production, Biology and Applications to Aquaculture for Performance Improvement and Genetic Containment. *Aquaculture* 293, 125–156. doi:10.1016/j.aquaculture.2009.04.036

Polinski, M. P., Bradshaw, J. C., Inkpen, S. M., Richard, J., Fritsvold, C., Poppe, T. T., et al. (2016). *De Novo* assembly of Sockeye Salmon Kidney Transcriptomes Reveal a Limited Early Response to Piscine Reovirus with or without Infectious Hematopoietic Necrosis Virus Superinfection. *BMC Genomics* 17, Article 848. doi:10.1186/s12864-016-3196-y

Quan, J., Kang, Y., Luo, Z., Zhao, G., Li, L., and Liu, Z. (2021). Integrated Analysis of the Responses of a circRNA-miRNA-mRNA ceRNA Network to Heat Stress in Rainbow Trout (*Oncorhynchus mykiss*) Liver. *BMC Genomics* 22, Article 48. doi:10.1186/s12864-020-07335-x

Quinn, T. P., Erb, I., Richardson, M. F., and Crowley, T. M. (2018). Understanding Sequencing Data as Compositions: an Outlook and Review. *Bioinformatics* 34, 2870–2878. doi:10.1093/bioinformatics/bty175

R Studio Team (2015). RStudio: Integrated Development for R [WWW Document]. Available at: http://www.rstudio.com/.

Rise, M. L., Douglas, S. E., Sakhrani, D., Williams, J., Ewart, K. V., Rise, M., et al. (2006). Multiple Microarray Platforms Utilized for Hepatic Gene Expression Profiling of GH Transgenic Coho Salmon with and without Ration Restriction. *J. Mol. Endocrinol.* 37, 259–282. doi:10.1677/jme.1.02031

Roberts, A., Pimentel, H., Trapnell, C., and Pachter, L. (2011). Identification of Novel Transcripts in Annotated Genomes Using RNA-Seq. *Bioinformatics* 27, 2325–2329. doi:10.1093/bioinformatics/btr355

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2009). edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* 26, 139–140. doi:10.1093/bioinformatics/btp616

Sambraus, F., Fjelldal, P. G., Remø, S. C., Hevrøy, E. M., Nilsen, T. O., Thorsen, A., et al. (2017a). Water Temperature and Dietary Histidine Affect Cataract Formation in Atlantic Salmon (*Salmo salar* L.) Diploid and Triploid Yearling Smolt. *J. Fish. Dis.* 40, 1195–1212. doi:10.1111/jfd.12594

Sambraus, F., Olsen, R. E., Remen, M., Hansen, T. J., Torgersen, T., and Fjelldal, P. G. (2017b). Water Temperature and Oxygen: The Effect of Triploidy on Performance and Metabolism in Farmed Atlantic Salmon (*Salmo salar* L.) Post-smolts. *Aquaculture* 473, 1–12. doi:10.1016/j.aquaculture.2017.01.024

Sambraus, F., Remen, M., Olsen, R., Hansen, T., Waagbø, R., Torgersen, T., et al. (2018). Changes in Water Temperature and Oxygen: the Effect of Triploidy on Performance and Metabolism in Large Farmed Atlantic Salmon. *Aquacult. Environ. Interact.* 10, 157–172. doi:10.3354/aei00260

Saranyan, P. V., Ross, N. W., and Benfey, T. J. (2017). Erythrocyte Heat Shock Protein Responses to Chronic (*In Vivo*) and Acute (*In Vitro*) Temperature Challenge in Diploid and Triploid Salmonids. *Comp. Biochem. Physiology Part A Mol. Integr. Physiology* 206, 95–104. doi:10.1016/j.cbpa.2017.01.007

Shi, K.-P., Dong, S.-L., Zhou, Y.-G., Li, Y., Gao, Q.-F., and Sun, D.-J. (2019). RNA-seq Reveals Temporal Differences in the Transcriptome Response to Acute Heat Stress in the Atlantic Salmon (*Salmo salar*). *Comp. Biochem. Physiology Part D Genomics Proteomics* 30, 169–178. doi:10.1016/j.cbd.2018.12.011

Smedley, M. A., Clokie, B. G. J., Migaud, H., Campbell, P., Walton, J., Hunter, D., et al. (2016). Dietary Phosphorous and Protein Supplementation Enhances Seawater Growth and Reduces Severity of Vertebral Malformation in Triploid Atlantic Salmon (*Salmo salar* L.). *Aquaculture* 451, 357–368. doi:10.1016/j.aquaculture.2015.10.001

Tan, E., Kinoshita, S., Suzuki, Y., Ineno, T., Tamaki, K., Kera, A., et al. (2016). Different Gene Expression Profiles between Normal and Thermally Selected Strains of Rainbow Trout, *Oncorhynchus mykiss*, as Revealed by Comprehensive Transcriptome Analysis. *Gene* 576, 637–643. doi:10.1016/j.gene.2015.10.028

Taylor, J. F., Waagbø, R., Diez-Padrisa, M., Campbell, P., Walton, J., Hunter, D., et al. (2015). Adult Triploid Atlantic Salmon (*Salmo salar*) Have Higher Dietary Histidine Requirements to Prevent Cataract Development in Seawater. *Aquacult Nutr.* 21, 18–32. doi:10.1111/anu.12130

Tibbetts, S. M., Wall, C. L., Barbosa-Solomieu, V., Bryenton, M. D., Plouffe, D. A., Buchanan, J. T., et al. (2013). Effects of Combined 'all-Fish' Growth Hormone Transgenics and Triploidy on Growth and Nutrient Utilization of Atlantic Salmon (*Salmo salar* L.) Fed a Practical Grower Diet of Known Composition. *Aquaculture* 406 (407), 141–152. doi:10.1016/j.aquaculture.2013.05.005

Tomalty, K. M. H., Meek, M. H., Stephens, M. R., Rincón, G., Fangue, N. A., May, B. P., et al. (2015). Transcriptional Response to Acute Thermal Exposure in Juvenile Chinook Salmon Determined by RNAseq. *G3 Genes, GENOMES, Genet.* 5, 1335–1349. doi:10.1534/g3.115.017699

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., et al. (2002). Accurate Normalization of Real-Time Quantitative RT-PCR Data by Geometric Averaging of Multiple Internal Control Genes. *Genome Biol.* 3 (1), research0034. doi:10.1186/gb-2002-3-7-research0034

Veilleux, H. D., Ryu, T., Donelson, J. M., Ravasi, T., and Munday, P. L. (2018). Molecular Response to Extreme Summer Temperatures Differs between Two Genetically Differentiated Populations of a Coral Reef Fish. *Front. Mar. Sci.* 5, Article 349. doi:10.3389/fmars.2018.00349

Wang, Y., Liu, Z., Li, Z., Shi, H., Kang, Y., Wang, J., et al. (2016). Effects of Heat Stress on Respiratory Burst, Oxidative Damage and *SERPINH1* (*HSP47*) mRNA Expression in Rainbow Trout *Oncorhynchus mykiss*. *Fish. Physiol. Biochem.* 42, 701–710. doi:10.1007/s10695-015-0170-6

Weber, G. M., Hostuttler, M. A., Cleveland, B. M., and Leeds, T. D. (2014). Growth Performance Comparison of Intercross-Triploid, Induced Triploid, and Diploid Rainbow Trout. *Aquaculture* 433, 85–93. doi:10.1016/j.aquaculture.2014.06.003

Weber, G. M., Wiens, G. D., Welch, T. J., Hostuttler, M. A., and Leeds, T. D. (2013). Comparison of Disease Resistance between Diploid, Induced-Triploid, and Intercross-Triploid Rainbow Trout Including Trout Selected for Resistance to *Flavobacterium psychrophilum*. *Aquaculture* 410-411, 66–71. doi:10.1016/j.aquaculture.2013.06.014

Xu, Q., Feng, C. Y., Hori, T. S., Plouffe, D. A., Buchanan, J. T., and Rise, M. L. (2013). Family-specific Differences in Growth Rate and Hepatic Gene Expression in Juvenile Triploid Growth Hormone (GH) Transgenic Atlantic Salmon (*Salmo salar*). *Comp. Biochem. Physiology Part D Genomics Proteomics* 8, 317–333. doi:10.1016/j.cbd.2013.09.002

Xue, X., Hixson, S. M., Hori, T. S., Booman, M., Parrish, C. C., Anderson, D. M., et al. (2015). Atlantic Salmon (*Salmo salar*) Liver Transcriptome Response to Diets Containing *Camelina sativa* Products. *Comp. Biochem. Physiology Part D Genomics Proteomics* 14, 1–15. doi:10.1016/j.cbd.2015.01.005

Yang, T., Zhang, Y., Meng, W., Zhong, X., Shan, Y., and Gao, T. (2021). Comparative Transcriptomic Analysis Brings New Insights into the Response to Acute Temperature Acclimation in Burbot (*Lota Lota Lota*). *Aquac. Rep.* 20, 100657. doi:10.1016/j.aqrep.2021.100657

Yaskowiak, E. S., Shears, M. A., Agarwal-Mawal, A., and Fletcher, G. L. (2006). Characterization and Multi-Generational Stability of the Growth Hormone Transgene (EO-1α) Responsible for Enhanced Growth Rates in Atlantic Salmon. *Transgenic Res.* 15, 465–480. doi:10.1007/s11248-006-0020-5

Yaskowiak, E. S., Shears, M. A., Agarwal-Mawal, A., and Fletcher, G. L. (2007). Characterization and Multi-Generational Stability of the Growth Hormone Transgene (EO-1α) Responsible for Enhanced Growth Rates in Atlantic Salmon. *Transgenic Res.* 16, 253–259. doi:10.1007/s11248-006-0020-510.1007/s11248-006-9059-6

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T. L. (2012). Primer-BLAST: A Tool to Design Target-specific Primers for Polymerase Chain Reaction. *BMC Bioinforma.* 13, Article 134. doi:10.1186/1471-2105-13-134

Zanuzzo, F. S., Beemelmanns, A., Hall, J. R., Rise, M. L., and Gamperl, A. K. (2020). The Innate Immune Response of Atlantic Salmon (*Salmo salar*) Is Not Negatively Affected by High Temperature and Moderate Hypoxia. *Front. Immunol.* 11, Article 1009. doi:10.3389/fimmu.2020.01009

Zhong, P., and Huang, H. (2017). Recent Progress in the Research of Cold-Inducible RNA-Binding Protein. *Future Sci. OA* 3, FSO246. doi:10.4155/fsoa-2017-0077

Zhou, C.-Q., Ka, W., Yuan, W.-K., and Wang, J.-L. (2021). The Effect of Acute Heat Stress on the Innate Immune Function of Rainbow Trout Based on the Transcriptome. *J. Therm. Biol.* 96, Article 102834. doi:10.1016/j.jtherbio.2021.102834

Conflict of Interest: Authors EHI, LMB and CDR were all employed by AquaBounty Canada (ABC) during the time of the experimentation. EHI performed all data analysis under the direct supervision of JDW and MLR when he was no longer employed by ABC. ABC had no role in the decision to publish this manuscript. Atlantic Aqua Farms was also devoid of influence on this article.

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.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 08 June 2022
doi: 10.3389/fgene.2022.901838

Genome-Wide Characterization and Comprehensive Analysis of NAC Transcription Factor Family in *Nelumbo nucifera*

Heyun Song[1,2], *Yanling Liu*[1,3], *Gangqiang Dong*[4], *Minghua Zhang*[1,2], *Yuxin Wang*[1,2], *Jia Xin*[1,2], *Yanyan Su*[4], *Heng Sun*[1,3]* *and Mei Yang*[1,3]*

[1]*Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, China,* [2]*University of Chinese Academy of Sciences, Beijing, China,* [3]*Aquatic Plant Research Center, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, China,* [4]*Amway (China) Botanical R&D Centre, Wuxi, China*

NAC (NAM, ATAF, and CUC) is a ubiquitously expressed plant-specific transcription factor (TF) family which is involved in the regulation of various biological processes. However, a systematic characterization of *NAC* gene family is yet to be reported in lotus. Here, 82 *NnNAC* genes which included five predicted membrane-bound NAC proteins were identified in the lotus genome. Phylogenetic analysis revealed seven-subfamily clusters (I–VII) of NnNAC proteins, with homologous gene pairs displaying similar conserved motifs and gene structure characteristics. Transactivation assay of NnNAC proteins revealed an extensive transcriptional activation capacity which is mediated by the highly divergent C-terminal activation domain (AD). Expression analysis of *NnNAC* genes in lotus tissues showed high transcript levels in root, stamen, petal and seed coat. In addition, 30 and 29 differentially expressed *NnNAC* candidate genes putatively involved in lotus seed development and response to complete submergence stress, respectively, were identified. Overall, our study provides potentially useful candidate gene resources for future molecular breeding of lotus varieties with novel agronomic traits.

Keywords: lotus, NAC transcription factor, transactivation, development, stress

OPEN ACCESS

Edited by:
Turgay Unver,
FicusBio, Turkey

Reviewed by:
Ming Li,
Hubei University, China
Peitao Lü,
Fujian Agriculture and Forestry University, China

***Correspondence:**
Heng Sun
sunheng@wbgcas.cn
Mei Yang
yangmei815815@wbgcas.cn

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *22 March 2022*
Accepted: *17 May 2022*
Published: *08 June 2022*

Citation:
Song H, Liu Y, Dong G, Zhang M, Wang Y, Xin J, Su Y, Sun H and Yang M (2022) Genome-Wide Characterization and Comprehensive Analysis of NAC Transcription Factor Family in Nelumbo nucifera. Front. Genet. 13:901838. doi: 10.3389/fgene.2022.901838

INTRODUCTION

Transcription factors (TFs) are master regulators of gene expression. NAC is one of the largest and most characterized plant-specific TF superfamily with 117 and 152 members in the dicotyledonous *Arabidopsis thaliana* and *Glycine max*, respectively, and 151 members in the monocotyledonous *Oryza sativa* (Nuruzzaman et al., 2010; Dung et al., 2011; Yan et al., 2017; Li et al., 2018a). NAC protein generally harbors a highly-conserved N-terminal DNA-binding domain (BD) and a highly variable C-terminal transcriptional regulatory region (TRR) (Puranik et al., 2012). The DNA binding of NAC TFs is determined by the presence of NAC domains, which are classified into five subdomains designated, A to E. The highly conserved C and D subdomains are necessary for DNA interaction, subdomain A is necessary for protein dimerization or heterodimerization, whereas the highly variable subdomains B and E determine the functional diversity of NAC proteins (Puranik et al., 2012). The highly divergent C-terminal TRRs are responsible for transcriptional regulation (Christiansen and Gregersen, 2014; Lindemose et al., 2014). Most NAC TFs are located in the cell nucleus; however, some membrane-associated NAC proteins show subcellular localization at the

endoplasmic reticulum or plasma membrane, which could be attributed to the α-helical transmembrane (TM) motif in the C-terminus (Seo et al., 2010; Li et al., 2016). Membrane-bound *NAC* genes have previously been reported in other species, such as *A. thaliana* (*Arabidopsis*), *O. sativa* (rice), and *S. lycopersicum* (tomato) with 13, 5, and 13 gene members, respectively (Kim et al., 2010a; Bhattacharjee et al., 2017). Membrane-bound *NAC* genes have important regulatory functions in stress responses and are strictly post-translationally regulated under specific conditions (Dung et al., 2011).

The first reported *NAC* gene developed no shoot apical meristem (SAM) in Petunia embryos and showed an equiformal phenotype in SAM formation in *Arabidopsis* mutants (Souer et al., 1996; Takada et al., 2001). The NAC TF family has been associated with various biological processes, including tissue development, hormone response, organogenesis, secondary cell wall (SCW) biosynthesis, and stress response (Zhong et al., 2006; Yamaguchi et al., 2010; Nakano et al., 2015; Yuan et al., 2020). In addition to regulating these processes, numerous NAC TFs have been shown to modulate seed development in some plant species. For example, three seed-size related *NAC* genes, *ONAC020*, *ONAC023*, and *ONAC026* with varied expression levels among rice accessions were highly expressed during seed development (Mathew et al., 2016). Similarly, *ONAC025*, *ONAC127*, and *ONAC129* were found to be involved in seed development by modulating rice grain filling during reproductive period (Mathew et al., 2020; Ren et al., 2021). Plants constantly encounter diverse biotic and abiotic stresses throughout their lives, which cause adverse effects on their growth and development. Increasing evidence suggests that NAC proteins play pivotal roles in abiotic stress tolerance (Puranik et al., 2012; Shao et al., 2015; Yuan et al., 2020; Singh et al., 2021). In *Arabidopsis*, *ANAC019*, *ANAC055*, and *ANAC072* are transcriptional activators of drought stress response, and their overexpression improved drought tolerance in transgenic plants (Tran et al., 2004). Similarly, a *NAC SlJUB1* gene, was shown to enhance tomato drought tolerance by activating the expression of several stress related genes, such as *SlDREB1*, *SlDREB*, and *SlDELLA* (Thirumalaikumar et al., 2018). Due to their roles in regulating important biological processes, comprehensive identification of stress related *NAC* genes is thus crucial for studying stress response mechanisms in plants.

Lotus (*Nelumbo nucifera* Gaertn.) is an old domesticated perennial wetland plant in the family Nelumbonaceae, which contains a single genus, *Nelumbo*, with two extant species: *N. nucifera* Gaertn. and *N. lutea* Pers. (Ming et al., 2013; Sun et al., 2021). The lotus seeds, rhizomes, and flowers have versatile uses, including as popular vegetable and ornamental plant with diverse medicinal properties (Sun et al., 2020). Lotus is popularly cultivated in Asia, and it is mainly categorized as seed-, rhizome-, and flower-lotus based on agronomic traits and distinct uses (Yang et al., 2015). Seed-lotus is predominantly cultivated for its edible lotus seeds, which are rich in proteins, vitamins, minerals, essential amino acids, and other numerous health-promoting bioactive components (Li et al., 2018b; Sun et al., 2021). Rhizome-lotus is produced for its edible underground stems, while flower-lotus is popular for its diversely colored and shaped flowers (Yang et al., 2015). TFs, including MYB, WRKY, bHLH, ERF, and bZIP have previously been reported to play crucial roles in lotus growth and development (Chen et al., 2013; Deng et al., 2016; Li et al., 2018b; Li et al., 2019; Sun et al., 2020). For example, Deng et al. (2016) demonstrated the *NnMYB5*-mediated transcriptional activation of anthocyanin synthesis, and its overexpression induced higher anthocyanin accumulation in immature transgenic *Arabidopsis* seeds and flower stalks. In addition, *NnANT* and *NnAP2* have been reported as potential negative regulators of lotus seed size and development (Li et al., 2018b). Moreover, the lotus *NAC* genes have been implicated in root browning under anaerobic stress and the formation of adventitious roots (Cheng et al., 2019; Min et al., 2019). However, despite previous reports, comprehensive mapping of genome-wide NAC TF family, which could potentially improve our knowledge on the gene family function, distribution, and evolution in sacred lotus is still lacking.

The recently improved sacred lotus genome assembly by Shi et al. (2020) provides an excellent opportunity to investigate the genome-wide distribution and evolution of *NAC* gene family in lotus. This study aimed to conduct a comprehensive analysis to clarify sequence features, phylogenesis, genome synteny, expression patterns, subcellular localization, and transcriptional activation capacity of *NAC* gene family members in sacred lotus. The study also integrated the analysis of possible roles of *NnNAC* genes in lotus seed development and response to complete submergence stress. Our results will not only improve the understanding on *NnNAC* gene family function, genome distribution, and evolution, but also provides vital candidate genes for future molecular breeding in lotus.

MATERIALS AND METHODS

Identification and Phylogenetic Analysis of *NAC* Genes in the Sacred Lotus Genome

To identify *NnNAC* genes, we retrieved the gene annotation gff3 file containing all the predicted protein sequences from lotus reference genome (*N.* Gaertn.) (Li et al., 2021a). The *NnNAC* genes were predicted using PlantTFDB 4.0 (http://planttfdb.cbi.pku.edu.cn/), and their complete amino acid sequences were confirmed with Clustalx v.1.81 and MEME software (http://meme-suite.org/index.html). Predicted molecular weights (Mw) and isoelectric points (pI) of NnNAC proteins were calculated using ExPASy software (http://web.expasy.org/protparam/). Membrane-bound NnNAC members were predicted using the TMHMM server v.2.0 (http://www.cbs.dtu.dk/services/TMHMM/).

Multiple sequence alignments of the identified 82 NnNAC proteins were carried out using Clustalx v.1.81 with default settings. Phylogenetic tree was constructed in MEGA7 using the Neighbor-Joining algorithm, the evolutionary distances were computed with the Poisson correction method, and tree nodes were evaluated with 1,000 bootstrap replications (Kumar et al., 2016).

Gene Structure, Conserved Motifs, Promoter, Chromosomal Location and Genome Scale Syntenic Analysis of *NnNAC* Genes

To obtain the intron and exon structures, the coding and genomic sequences of *NnNAC* genes were analyzed with the Gene Structure Display Server (GSDS v.2.0) online program (http://gsds.cbi.pku.edu.cn/). Conserved motifs were identified using the MEME online tool (http://meme-suite.org/index.html) with default settings. For promoter sequence analysis, a 2-kb length sequence upstream of the start codon of each gene was downloaded for *cis*-elements prediction using PlantCARE program (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al., 2002). Distribution of *NnNAC* genes within the lotus chromosomes was visualized using TBtools software (Chen et al., 2020). *NAC* genes were sequentially named from chromosome (Chr) 1 to 8 with numerical digits in the order of their physical location from top to bottom of the chromosomes.

Genome sequences of *A. thaliana* and *O. sativa* were downloaded from Phytozome v13 (https://phytozome-next.jgi.doe.gov/). Genome scale syntenic analysis between the three plant species was conducted by MCScan X, and TBtools software was used to visualize their relationship (Wang et al., 2012; Chen et al., 2020).

Expression Profiling of *NnNAC* Genes

To investigate the expression patterns of *NnNAC* genes in lotus tissues, the transcriptome data corresponding to gene expression abundance in various tissues were obtained from Nelumbo Genome Database (http://nelumbo.biocloud.net/nelumbo/home). Public transcriptome datasets on seed development and response to complete submergence stress were retrieved from the National Center for Biotechnology Information (NCBI) with accession numbers, SRP127765 and PRJNA723672, respectively (Li et al., 2018b; Deng et al., 2022). We performed quality control of the downloaded transcriptome data sets with FastQC and Trimmomatics programs (Bolger et al., 2014; Wingett and Andrews, 2018). Subsequently, gene expression levels were quantified by fragments per kilobase of transcript per million mapped fragments (FPKM) using StringTie software (Pertea et al., 2015). Gene expression was visualized using TBtools software. DESeq2 R package v.1.10.1 was used for pairwise comparisons to identify differentially expressed genes (DEGs) with filter the criteria, Fold Change (FC) ≥ 2 and False Discovery Rate (FDR) < 0.05 (Love et al., 2014).

qRT-PCR Analysis

High-quality RNAs were reverse transcribed to cDNA using TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix (Lot#M31212, Beijing TransGene Biotech, Beijing, China). qRT-PCR analysis was carried out on a StepOnePlus™ Real-time PCR System (Applied Biosystems, United States). The relative gene expression level was calculated and normalized using *NnACTIN* (Gene ID NNU_24864) used as the internal standard. All primers used for qRT-PCR were listed in **Supplementary Table S1**.

Subcellular Localization of *NnNAC* Genes

ProtComp v9.0 (http://linux1.softberry.com/berry.phtml) was used to obtain the subcellular localization predictions of NnNAC proteins. Coding regions of *NnNAC* genes were cloned into entry vectors (pDONR221 Zeo) using high fidelity primers (**Supplementary Table S1**) following the BP-clonase kit instruction manual (Lot#2335893, Invitrogen by Thermo Fisher Scientific, United States). Transformed plasmids were then cloned into PMDC43 vectors with LR-clonase according to the manufacturer's instructions. These plasmids were subsequently inserted into the *Agrobacterium tumefaciens* cv. GV3101 using electric shock method. Young leaves of tobacco plants (4–6 weeks-old) were selected for transformation assay according to the methods of Kokkirala et al. (2010). A confocal laser scanning microscope (Leica TCS SP2; Leica microsystems, Wetzlar, Germany) was used to photograph the agroinfiltrated leaves 48 h after infiltration.

Transcriptional Activation Analysis

Transactivation capacity of NnNAC proteins was assessed using yeast transactivation assay (Sun et al., 2018). The full-length coding sequences of *NnNAC* genes and the truncated *NnNAC-AD* and *NnNAC-BD* forms were independently cloned in pGBKT7 vectors. The resulting plasmids were subsequently transformed into Y2H gold strains according to manufacturer's instructions (Cat.No.630489, Clontech) then plated on SD growth mediums without Trp, but with X-α-Gal (SD-Trp + X-α-Gal) to identify the transactivation activity of NnNAC proteins. SD mediums were photographed after 36 h incubation. Primers used for this assay were listed in **Supplementary Table S1**.

RESULTS

Genome-Wide Identification of *NAC* Genes in the Lotus Genome and Their Gene Structural and Conserved Motifs Analyses

A total of 82 complete non-redundant *NnNAC* genes were identified in the lotus genome (*N. nucifera* Gaertn.) from PlantTFDB (**Supplementary Table S2**). Multiple sequence alignment revealed the presence of conserved NAC domain and the five (A–E) subdomains among all predicted NnNAC proteins (**Figures 1A**, **2B**). Five putative membrane-bound NAC proteins each harboring a single predicted TM at the C-terminal were identified, including NnNAC6, NnNAC12, NnNAC23, NnNAC27, and NnNAC44 (**Figure 1A**; **Supplementary Figure S1**; **Supplementary Table S2**). Analysis of NnNAC protein features revealed that their amino acid residue, molecular weight (Mw), and protein isoelectric point (pI) ranged from 152 to 683, 7.63 to 77.94 KDa, and 4.48 to 9.92, respectively (**Figure 1B**; **Supplementary Table S3**).

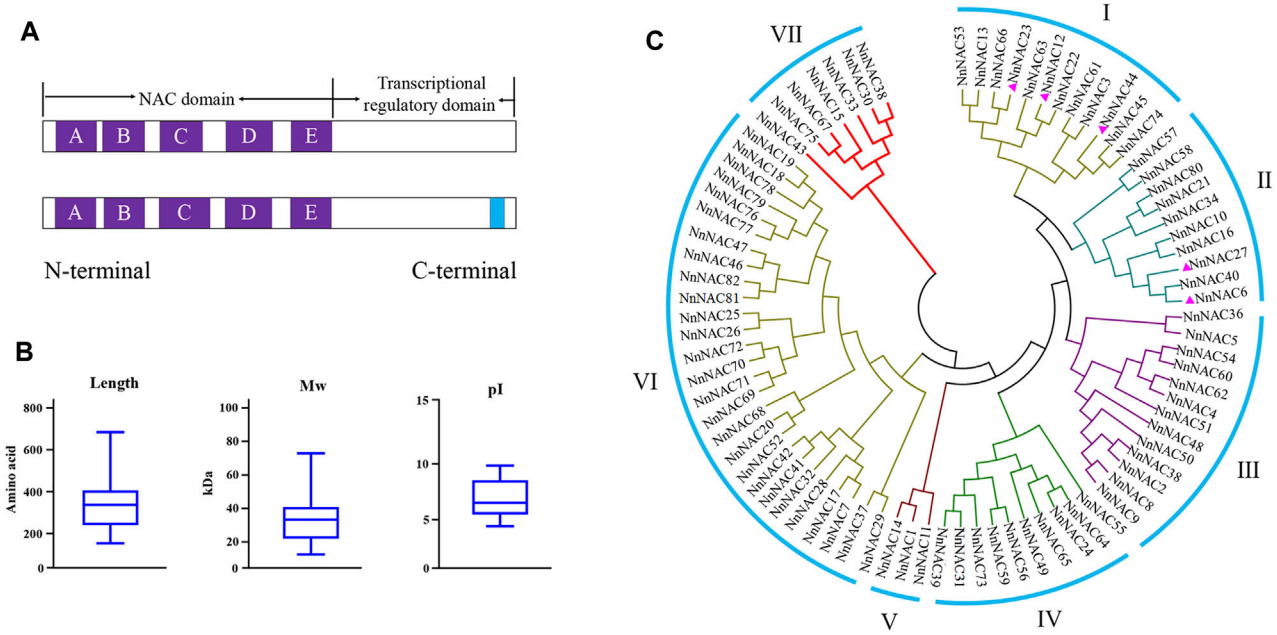


FIGURE 1 | NAC domains, protein sequence features and phylogenetic analysis in lotus. **(A)** Schematic representation of the highly conserved NnNAC key domain and subdomains A–E (purple squares) at the N-terminal. The blue square represents the transmembrane motif (TM). **(B)** Box plots showing statistical distribution of amino acid sequence length, molecular weight (Mw) and isoelectric point (pI) of NnNACs. **(C)** Phylogenetic clustering of 82 identified lotus NAC proteins. The purple triangle represents NnNAC protein harboring the TM motif.

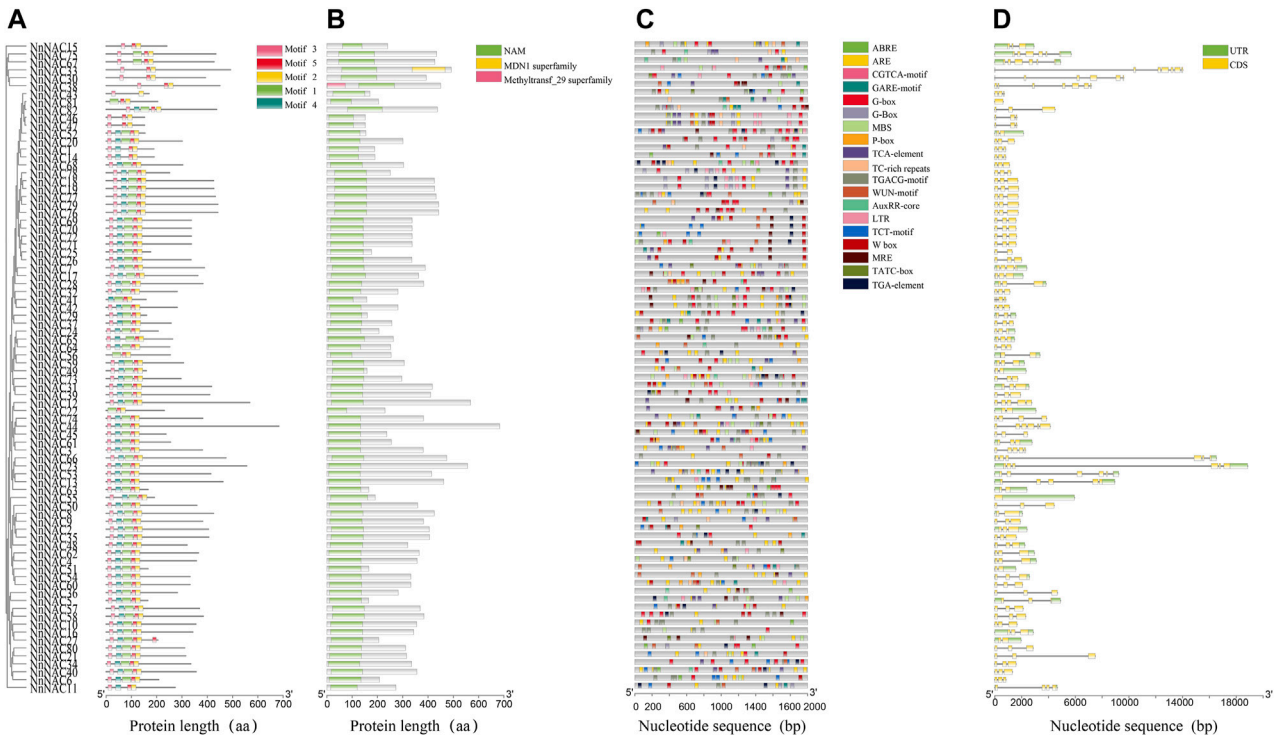


FIGURE 2 | Conserved motifs, *cis*-elements and *NnNAC* gene structure analyses. **(A)** Clusters and conserved motifs of NnNAC proteins. **(B)** Screening of conserved NAC domain in lotus. **(C)** Identification of *cis*-regulatory elements in the 2-kb upstream promoter regions of *NnNAC* genes. **(D)** The lotus *NAC* gene structures.

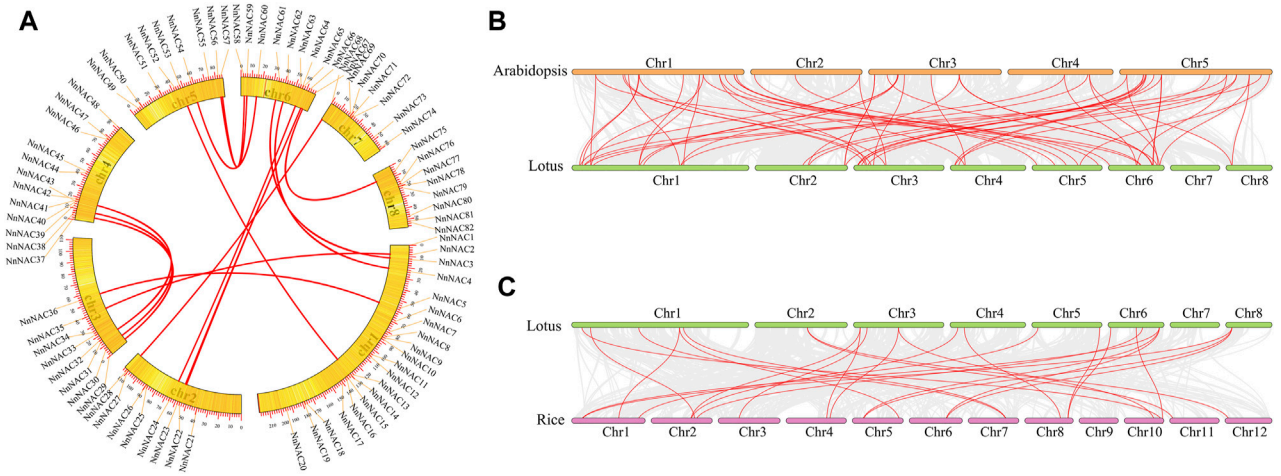


FIGURE 3 | Genome scale synteny analysis of lotus, *Arabidopsis*, and rice *NAC* genes. **(A)** Syntenic relationship of *NnNAC* genes. All 82 *NnNAC* genes are marked according to their chromosome distribution in the lotus genome and syntenic gene pairs are connected with red lines. **(B)** Syntenic pairs of *NAC* genes between *Arabidopsis* and lotus. **(C)** Syntenic pairs of *NAC* genes between lotus and rice. The lotus, *Arabidopsis*, and rice chromosomes are shown in orange, green, and purple bars, respectively. Syntenic gene pairs are linked with red lines.

To understand the evolutionary relationships among the *NnNAC* gene family, a phylogenetic tree was constructed using the 82 identified protein sequences. All NnNAC proteins were clustered into seven subfamilies (I–VII) (**Figure 1C**). Notably, the five predicted membrane-bound proteins were exclusively present in I and II subfamily clusters (**Figure 1C**). Conserved motif analysis identified five conserved protein motifs, with motif 2 and motif 5 as the most conserved (**Figure 2A**). Interestingly, NnNACs within the same phylogenetic cluster tended to harbor similar conserved motifs, while paired NnNACs, such as NnNAC1 and NnNAC14, NnNAC30 and NnNAC38, NnNAC46 and NnNAC47, and NnNAC67 and NnNAC75 shared equiformal conserved motifs (**Figure 2A**).

Gene structure variability due to different exon and intron combinations is useful for understanding the diverse gene functions and genome evolution. To explore the structural diversity of *NnNAC* genes, the exon/intron arrangements were analyzed. The numbers of exons among *NnNACs* genes varied from one to seven, with most genes (64.2%) having three exons (**Supplementary Figure S2**). *NnNAC38* had seven exons which was the highest number observed, while *NnNAC55* and *NnNAC81* only harbored one exon each (**Figure 2D**). Notably, genes sharing the same phylogenetic cluster showed similar conserved structural characteristics (**Figure 2**).

Chromosomal Location and Syntenic Analysis of *NAC* Genes in Lotus, *Arabidopsis* and Rice

Chromosomal localization revealed uneven distribution of *NnNAC* genes within the lotus genome with each of the eight lotus chromosomes harboring at least one gene. The highest gene density was observed in Chr 1 with 20 *NnNACs*, representing 24% of total genes identified, while only seven *NnNAC* genes were anchored on Chr 7. In addition, 8, 8, 12, 9, 10, and 8 *NnNAC* genes were anchored on Chr 2, 3, 4, 5, 6, and 8, respectively (**Figure 3A**; **Supplementary Figure S3**). Interestingly, some *NnNACs* indicated potential evidence of gene duplication events in the lotus genome (**Supplementary Figure S3**). For example, *NnNAC18* and its homolog *NnNAC19* located within a 0.77-Mb interval on Chr 1 shared about 99.8% identity (**Supplementary Figure S4A**). Similarly, four *NnNAC69*, *NnNAC70*, *NnNAC71*, and *NnNAC72* anchored within a 0.41-Mb chromosomal region on Chr 7 shared about 98.8% amino acid sequence identity (**Supplementary Figure S4B**). In addition, 16 pairs of segmentally duplicated *NnNAC* genes were identified by intragenomic synteny analysis. Highly conserved amino acid sequences were observed among some collinear gene pairs, such as *NnNAC54* and its paralog *NnNAC60* located on Chr 6 which shared about 87.5% amino acid sequence identity, and *NnNAC67* on Chr 6 which shared about 88.5% identity with *NnNAC75* anchored on Chr 8 (**Supplementary Figures S4C,D**).

Genome synteny analysis is an effective method for inferring genome evolutionary history. The evolution and collinearity of *NAC* gene family was explored using the complete genome sequences of sacred lotus, *Arabidopsis* and rice. As a result, a total of 59 orthologous gene pairs sharing high similarities were identified between *Arabidopsis* to lotus (**Figure 3B**). In contrast, only 26 orthologous gene pairs were identified between rice to lotus genomes (**Figure 3C**). These results suggest that lotus, an ancient true dicot, is more likely to be evolutionarily closer to dicots than monocots.

Screening of *NAC* Gene *Cis*-Elements in Lotus

To study the transcriptional regulation of *NnNAC* genes, *cis*-elements were identified using the PLANTCARE database. As a

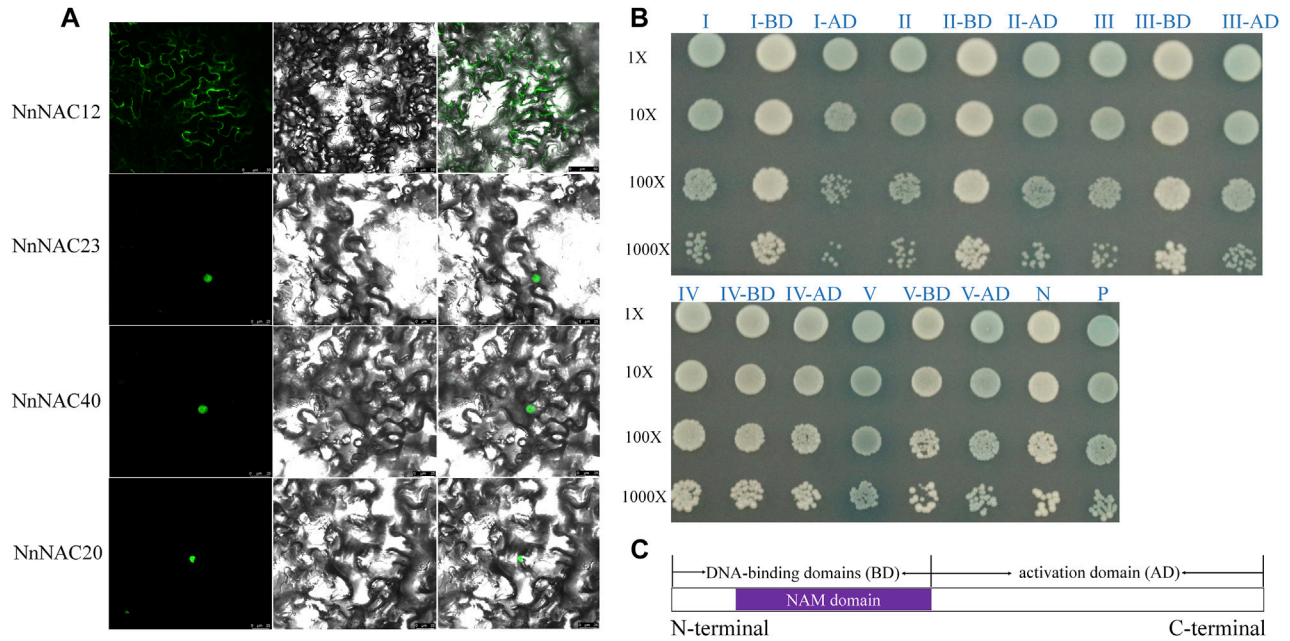


FIGURE 4 | Subcellular localization and transcriptional activation assay of NnNAC proteins. **(A)** Subcellular localization of four NnNAC proteins in tobacco leaves. **(B)** Transactivation assay of NnNAC proteins in the Y2H system. I: NnNAC12; II: NnNAC23; III: NnNAC20; IV: NnNAC52; V: NnNAC40; AD and BD represent NnNAC protein with AD and BD domains, respectively; P: positive control (pGBKT7-53); N: negative control. **(C)** Schematic representation of NnNAC protein truncations.

result, numerous *cis*-regulatory elements linked to stress responses, development, and phytohormone responses were detected in the promoter regions of *NnNAC* genes. Of the 82 gene promoters screened, ARE element, CGTCA-motif, TGACG-motif, G-box, ABRE, and W box were the six most commonly identified *cis*-elements with percentage frequencies of 89, 84.1, 82.9, 76.8, 71.9, and 65.9, respectively (**Figure 2C**; **Supplementary Figure S5**). The W box and ABRE are both associated with drought and salt stress responses, while TGACG-motif and CGTCA-motif are involved in phytohormone response and development. These results suggest the likely roles of *NnNAC* genes as crucial stress response and development regulators in lotus.

Subcellular Localization and Transcriptional Activation Assay of NnNAC Proteins

Bioinformatic studies have predicted that about 74 (~90%) of the identified NnNAC proteins were nuclear while others were extracellularly located (**Supplementary Table S3**). Three predicted nuclear located *NnNAC* genes, including *NnNAC20*, *NnNAC23*, and *NnNAC40* and a membrane located *NnNAC12* gene were selected for subcellular localization to test the prediction accuracy results. Microscopic observations showed nuclear localization of NnNAC20, NnNAC23 and NnNAC40, while the GFP signals of NnNAC12 were observed in plasma membranes (**Figure 4A**).

Transcriptional activation assay is useful for determining the regulatory capacity of TFs. Transactivation capacities of NnNAC proteins were determined using the yeast two-hybrid system (Y2H). Full-length cDNA of five selected NAC proteins, including NnNAC12, NnNAC20, NnNAC23, NnNAC40, and NnNAC52 were individually cloned into GAL4DB vector. As a result, blue colonies indicating transcriptional activation activity were observed in all tested NnNAC proteins except NnNAC52 (**Figure 4B**). To investigate the transactivation domain, we further truncated the NnNAC proteins into conserved N-terminal DNA-binding domains (BD) and high divergent C-terminal activation domain (AD) regions (**Figure 4C**). Expectedly, transactivation activity was observed for the AD domain and not for BD domain (**Figure 4B**).

Expression Profiling of *NnNAC* Genes in Lotus Tissues

Published transcriptome data of leaf, petiole, petal, receptacle (immature and mature), stamen (immature and mature), carpel (unpollinated and pollinated), seed coat (6, 12, and 18 days after pollination, DAP), root and rhizome were obtained and used to investigate the tissue expression patterns of *NnNAC* genes (**Figure 5A**). Overall, 64 (~78%) of the 82 *NnNAC* genes were expressed in at least one lotus tissue, whereas 14 (~17%) were expressed in all tested tissues (FPKM ≥ 1). Low transcript abundance of FPKM < 5 in all tested tissues was observed in 42 (~51%) *NnNAC* genes. Conversely, the expressions of *NnNAC25*, *NnNAC43*, *NnNAC46*, *NnNAC47*, *NnNAC71*, and *NnNAC72* were not detected in all analyzed tissues (**Supplementary Table S4**). Notably, highly expressed *NnNAC* genes with FPKM ≥ 20 were identified in root, stamen, petal, and seed coat (**Figure 5B**; **Supplementary Table S4**). *NnNAC* genes,

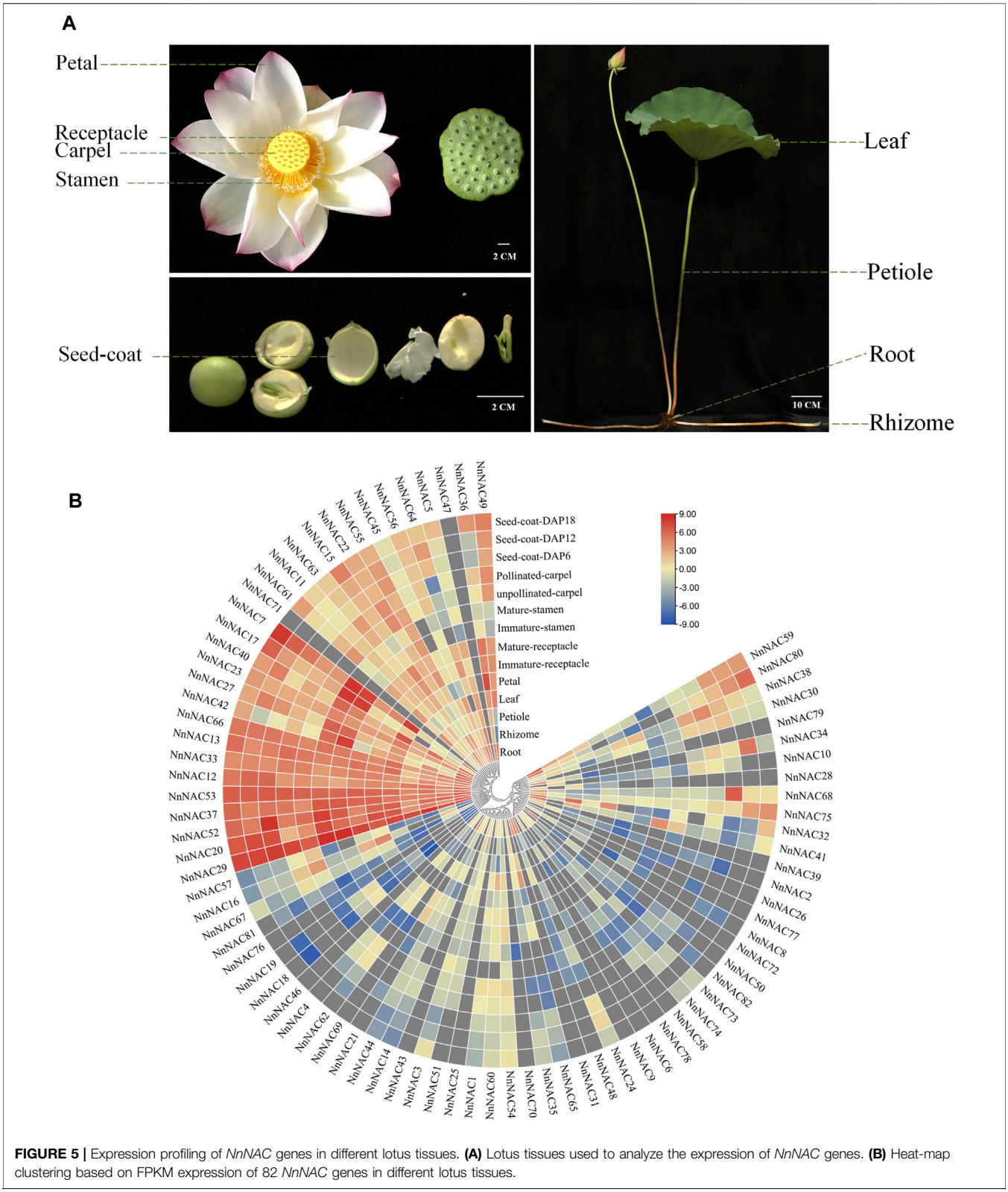


FIGURE 5 | Expression profiling of *NnNAC* genes in different lotus tissues. **(A)** Lotus tissues used to analyze the expression of *NnNAC* genes. **(B)** Heat-map clustering based on FPKM expression of 82 *NnNAC* genes in different lotus tissues.

including *NnNAC24*, *NnNAC31*, *NnNAC41*, *NnNAC48*, and *NnNAC50* exhibited tissue-specific expression in roots, while *NnNAC27* and *NnNAC42* were only expressed in root and stamen. In contrast, *NnNAC20*, *NnNAC29*, *NnNAC37*, *NnNAC52*, *NnNAC53*, and *NnNAC66* were expressed in all tested tissues. Interestingly, phylogenetically close *NnNAC*

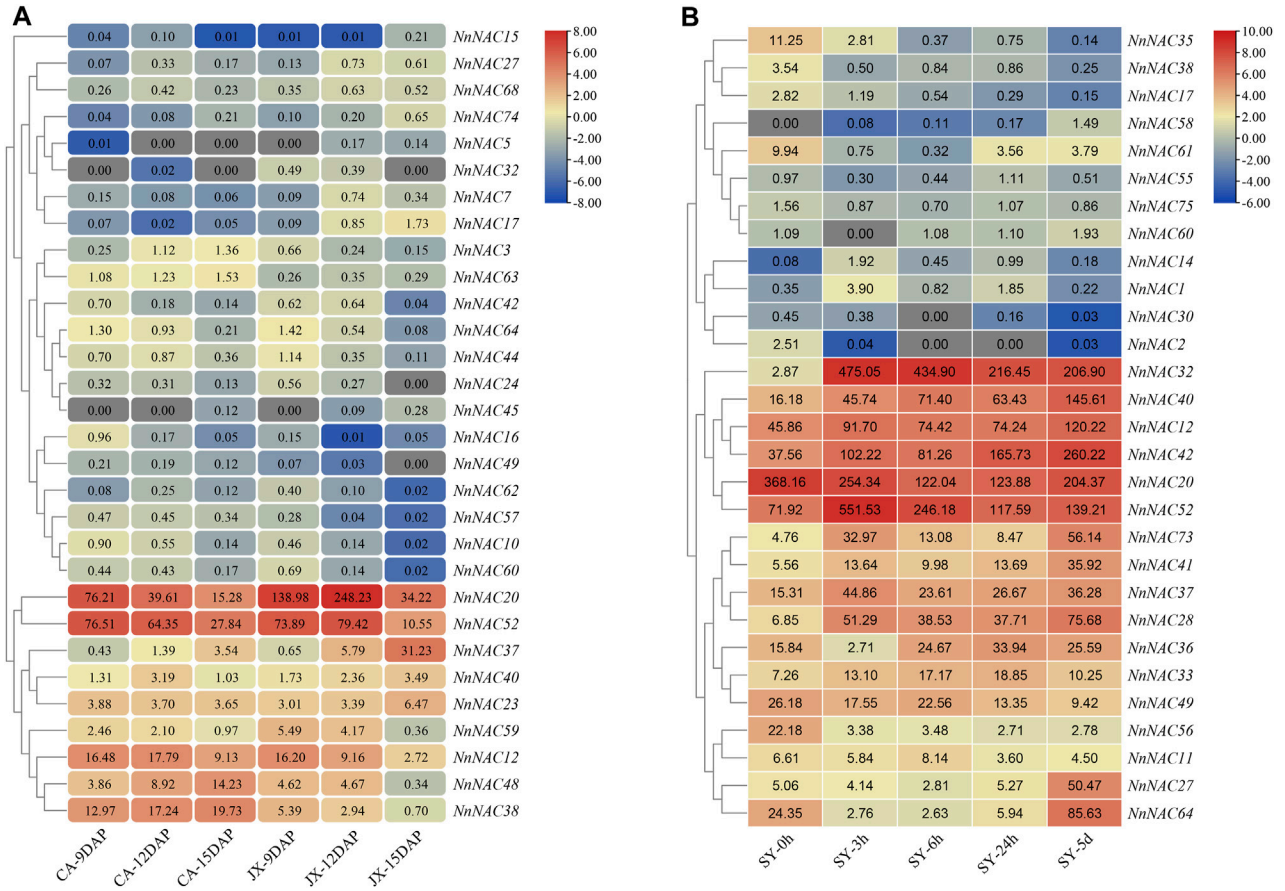


FIGURE 6 | Expression analysis of *NnNAC* genes during lotus seed development and response to complete submergence. **(A)** Heat-map showing the expression patterns of 30 *NnNAC* genes involved in the lotus seed development (FPKM values). **(B)** Heat-map showing the expression patterns of 29 complete submergence stress responsive *NnNAC* genes in lotus (FPKM values). "SY" represent complete submergence stress treatment.

genes, including *NnNAC24/64*, *NnNAC31/39*, and *NnNAC44/45* exhibited comparable expression patterns. In addition, duplicated gene pairs, such as *NnNAC18/19* and *NnNAC69/70/71/72* also shared similar expression patterns.

Identification of *NnNAC* Genes Involved in Lotus Seed Development and Complete Submergence Stress Response

Multiple complex pathways respond to different plant signals in a coordinated manner that involves the recruitment of various transcription factors (Li et al., 2018b; Sun et al., 2020). The published transcriptome data of seed cotyledons at 9, 12 and 15 DAP in "China Antique" (CA) and "Jianxuan 17" (JX) were used to investigate the possible roles of *NnNACs* in lotus seed development. A total of 10 and 23 *NnNAC* DEGs were subsequently identified in CA and JX, respectively, while 20 *NnNAC* DEGs were identified between CA and JX. Overall, 30 *NnNAC* DEGs were identified during lotus seed development (**Supplementary Table S5**). Of these, 14 (46.7%) *NnNACs* were expressed in the cotyledon in at least one development stage (FPKM ≥ 1) (**Figure 6A**). Highly expressed *NnNAC* genes, including *NnNAC12*, *NnNAC20*, *NnNAC37*, *NnNAC38*, *NnNAC48*, and *NnNAC52* with FPKM abundance ≥10, and with varied expression patterns during lotus seed development were identified, suggesting their likely roles in the regulation of lotus seed development. The expression of *NnNAC38* showed a continuous increase from 9 to 15 DAP in CA, while it was downregulated in JX. In contrast, the expression profile of *NnNAC37* showed a continuous increase from 9 to 15 DAP in both CA and JX. Notably, preferential expression of *NnNAC20* and *NnNAC52* gene pairs over other *NnNACs* was observed, with the latter showing a continuous decrease in expression from 9 to 15 DAP in CA. In contrast, the expression of *NnNAC20* showed a continuous downregulation in CA, whereas it was upregulated in JX from 9 to 12 DAP. Four *NnNAC* genes, including *NnNAC23*, *NnNAC37*, *NnNAC38*, and *NnNAC40* were randomly selected to evaluate the reliability of RNA-Seq by qRT-PCR. The results showed comparable expression changes with of transcriptome data (**Supplementary Figure S6**). For example, qRT-PCR and FPKM expression results showed that the levels of *NnNAC37* were persistently elevated from 9 to 15 DAP in both CA and JX. *NAC* genes have previously been implicated in the regulation of various stress signaling pathways. As a perennial aquatic plant,

lotus experience frequent submergence stress during its life cycle due to flooding (Deng et al., 2022). To determine potential submergence responsive *NnNACs*, the published transcriptome of completely submerged "Qiu Xing" at different time intervals (0, 3, 6, and 24 h and 5 days) was analyzed. As a result, a total of 29 *NnNAC* DEGs were identified (**Figure 6B**; **Supplementary Table S6**). Seventeen highly expressed *NnNAC* genes with FPKM ≥ 10, and exhibiting two different expression patterns were identified. Of these genes, nine (~52.9%), including *NnNAC12*, *NnNAC28*, *NnNAC32*, *NnNAC37*, *NnNAC40*, *NnNAC41*, *NnNAC42*, *NnNAC52*, and *NnNAC73* were all significantly upregulated with transcript fold increase ranging from 2 to 165.5 after 3 h complete submergence treatment. For example, a 165.5-fold increase in the expression level of *NnNAC32* compared to that of control was observed after 3 h complete submergence, suggesting that it could be a crucial complete submergence stress responsive regulator in lotus. On the contrary, four *NnNAC* genes, *NnNAC20*, *NnNAC35*, *NnNAC49*, and *NnNAC56* showed decreased expression profiles after complete submergence treatment, suggesting their unlikely involvement in the lotus submergence stress response. Taken together, these results provide important possible candidate NAC regulators for studying the mechanisms of seed development and response to complete submergence in lotus.

DISCUSSION

TFs function as important switches of transcriptional networks that accurately regulate gene expression (Jiang et al., 2016; Droge-Laser et al., 2018; Kumar et al., 2021). NAC TF family has been identified as one of the largest plant-specific regulators involved in diverse biological processes of numerous plant species, such as *Arabidopsis*, rice, tomato, and maize (Christiansen and Gregersen, 2014; Su et al., 2015; Singh et al., 2021). However, despite their crucial roles and the availability of complete genome sequence, a comprehensive characterization of *NAC* gene family has not been conducted in the aquatic lotus plant.

Identification and Sequence Characteristics of *NAC* Gene Family in Lotus

Previous studies have reported that members of the NAC TF family are widely distributed in lower and higher plants, such as moss with 31 members, and *Arabidopsis*, rice, soybean, wheat, tomato, grape, and cucumber with 117, 151, 269, 263, 101, 70, and 83 members, respectively (Nuruzzaman et al., 2010; Xu et al., 2014; Ma et al., 2018; Singh et al., 2021). Our study identified for the first time 82 *NnNAC* genes in the published genome of sacred lotus, and this number is very close to that of cucumber. As a basal eudicot, lotus experienced a whole genome duplication (WGD) event about 18 million years ago (Ming et al., 2013; Wang et al., 2013). Consequently, genes associated with signal transduction might have been preferentially retained during WGD to form the present NAC TF family (De Smet et al., 2013). Despite wide variations in gene length, predicted protein molecular weight, isoelectric point, and exons/introns organization among the identified *NnNAC* genes, they remain relatively conserved with about 64% of the genes having three exons (**Figure 2D**; **Supplementary Figure S2**). Similar observations were also made in *Arabidopsis*, rice, tomato, tobacco, and maize (Nuruzzaman et al., 2010; Peng et al., 2015; Su et al., 2015; Li et al., 2018a; Li et al., 2021b; Singh et al., 2021). Homologous *NnNAC* gene pairs predominantly shared higher degree of similarities in predicted protein features, gene structures, and conserved protein motifs, which is consistent with previous findings that duplicate genes derived from a common ancestor evolve independently at the same rate with few variations (Zhang and Wang, 2005; Wang et al., 2010; Rinerson et al., 2015). Membrane-bound NAC TFs have been identified in various plant species, providing direct evidence of their involvement in many biological processes (De Clercq et al., 2013; Li et al., 2016; Bhattacharjee et al., 2017). Here, we identified five membrane-bound NAC TFs which showed over-representation in subfamilies I and II in the lotus genome (**Figure 1C**; **Supplementary Figure S1**). All the five genes belonged to the classical membrane-bound NAC TFs, harboring a single α-helical TM located at the C-terminal, which is consistent with the membrane-bound NACs identified in *Arabidopsis* and rice (Kim et al., 2007; Ng et al., 2013). Interestingly, a divergent form of NAC protein with two TM located in front of the NAC domain has been reported in tomato and cotton, indicating evolution and functional differentiation (Olsen et al., 2005; Bhattacharjee et al., 2017; Sun et al., 2018). Studies have shown that membrane-bound NAC TFs could be activated by specific signal on the cell membrane (Kim et al., 2007; Seo et al., 2008). In this study, a predicted NAC protein with TM, NnNAC23, was localized in the nucleus, suggesting that this gene might be regulated by post-translation and being cut off the transmembrane domain.

Functional and Evolutionary Relationship Assessment of *NnNACs* Using Gene Expression Patterns

Plenty of studies have reported the ubiquitous distribution of NAC TFs in the plant kingdom and their association with various biological processes (Kim et al., 2010b; Nakano et al., 2015; Yuan et al., 2020). The presence of *cis*-elements is crucial for the expression of *NAC* TF genes and their functions. Here, we comprehensively identified numerous *cis*-elements with various functions in the promoter region of *NnNAC* genes, suggesting their likely involvement in various biological processes, such as stress responses and regulation of lotus development. For example, classical stress-responsive *cis*-elements, including LTREs, MYB and DREB (Tran et al., 2004; Nakashima et al., 2012; Yan et al., 2017; Yuan et al., 2020), were identified in the promoter region of *NnNAC* genes. These results provide a basis for future studies on the transcriptional regulation mechanisms and functions of *NnNAC* genes.

The expression patterns of a gene family can be used to predict their functions and evolutionary relationships (Wang et al., 2013). Gene expression pattern showed multiple expression of

NnNAC genes in different lotus tissues. For example, tissue-specific expression in roots was observed in five *NnNAC* genes, including *NnNAC24*, *NnNAC31*, *NnNAC41*, *NnNAC48*, and *NnNAC50*, suggesting their potential association with lotus root development. Conversely, several *NnNAC* genes, such as *NnNAC20*, *NnNAC29*, *NnNAC37*, *NnNAC52*, *NnNAC53*, and *NnNAC66* were expressed in all tested tissues, thus they could be essential for various developmental processes in the aquatic lotus (**Figure 5B**). Overall, members of the *NnNAC* gene family displayed evidence of functional redundancy and evolutionary diversity in the lotus genome. Although varied expression patterns were observed among *NnNACs*, most homologous gene pairs shared similar expression patterns, for example, *NnNAC31* and *NnNAC39* were uniquely expressed in the root, *NnNAC29* and *NnNAC37* were constitutively expressed in all tested tissues, while *NnNAC18*/*19* and *NnNAC25*/*26* were not expressed in any of the tissues tested (**Figure 5B**). Interestingly, similar observations were made for *Arabidopsis* and cotton *NAC* genes (Jensen et al., 2010; Sun et al., 2018). Taken together, our results indicated that *NAC* gene family might have undergone expansion through gene duplication during evolution, resulting in functional redundancy of gene members.

Identification of Putative Candidate *NnNAC* Genes Involved in Lotus Seed Development and Complete Submergence Stress Response

Lotus seed has important commercial, nutritional and medicinal values, and improving its yield and quality is one of the breeding targets today (Sun et al., 2021). Despite previous studies focused on lotus seed development, little is still known about whether *NnNAC* genes are associated with this process (Wang et al., 2016; Li et al., 2018b). In rice, *ONAC020*, *ONAC023*, and *ONAC026* showed high expression level during seed development, and their sequence variations were correlated with seed size (Mathew et al., 2016). Herein, 30 *NnNAC* DEGs during lotus seed development were identified for the first time, with six genes, including *NnNAC12*, *NnNAC20*, *NnNAC37*, *NnNAC38*, *NnNAC48*, and *NnNAC52* being preferentially expressed, suggesting their potential involvement in the regulation of lotus seed development (**Figure 6A**).

Efforts have been made to clarify the role of NAC TFs in responses to diverse stress, such as drought, salinity and flooding (Puranik et al., 2012; Shao et al., 2015; Yuan et al., 2020; Singh et al., 2021). Lotus is an aquatic plant that experiences frequent submergence stress due to sudden frequent flooding events. Twenty-nine *NnNAC* DEGs associated with submergence stress response were identified in the lotus genome, of which 17 genes showed high expression levels. For example, *NnNAC32*, an orthologous of *Arabidopsis* stress response *AtNAP* (At1g69490) gene (Sakuraba et al., 2015; Seok et al., 2017), was significantly upregulated under complete submergence treatment, which suggest its possible regulatory role in lotus submergence stress response (**Supplementary Figure S7**). This result might further confirm the functional differentiation of *NAC* gene family. In addition, two lotus homologs, *NnNAC20* and *NnNAC52*, sharing ~46% amino acid sequence identity, and orthologous to the *Arabidopsis* stress related *ANAC2* (At1g01720) NAC TF gene, were identified (Lahiri et al., 2021). However, *NnNAC20* and *NnNAC52* showed different expression patterns under complete submergence, suggesting they could be diverging functionally. Notably, thirteen common *NnNAC* genes were differentially expressed during seed development and stress response, indicating they might play multiple biological functions in lotus. This systematic characterization of lotus *NAC* genes in will pave way for future functional analysis to clarify the roles of these candidate genes in lotus.

CONCLUSION

Our study identified 82 *NnNAC* genes for the first time in the sacred lotus genome, and conducted a comprehensive sequence characterization, gene structure, expression profiling and transcriptional activation. In addition, *NnNAC* genes involved in lotus seed development and response to complete submergence stress were highlighted. In future, functional characterization of candidate *NnNAC* genes identified in this study is warranted. Our results provide potentially valuable *NnNAC* gene resources for further lotus genetic improvement.

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

HnS conceived, designed, and analyzed the data. YL, MZ, and GD performed experiments. HnS and HgS wrote the manuscript. HgS and MY supervised the research. YW, JX, and YS revised the manuscript. All authors have read and approved the manuscript.

FUNDING

This project was supported by the National Natural Science Foundation of China (Grant No. 31872136), the Biological Resources Program CAS (Grant No. KFJ-BRP-007-009), and the Hubei Provincial Natural Science Foundation of China (Grant No. 2020CFB484).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.901838/full#supplementary-material

Supplementary Figure S1 | Structures of membrane-bound NnNAC proteins in lotus.

Supplementary Figure S2 | Statistical summary of exon numbers in *NnNAC* genes.

Supplementary Figure S3 | Chromosome localization of *NnNAC* genes in lotus.

Supplementary Figure S4 | Comparison of NAC amino acid sequence similarities.

Supplementary Figure S5 | *Cis*-element analysis in the promoters of *NAC* genes in lotus.

Supplementary Figure S6 | Validation of gene expression through qRT-PCR.

Supplementary Figure S7 | Phylogenetic relationship between NnNAC proteins and AtNAC proteins in *Arabidopsis*.

Supplementary Table S1 | Primer sequences used for vector construction and qRT-PCR validation.

Supplementary Table S2 | Protein sequences of *NnNAC* genes.

Supplementary Table S3 | Statistical analyses of sequence length, molecular weight, isoelectric point, and predicted subcellar localizations for NAC proteins in lotus.

Supplementary Table S4 | The expression of *NnNAC* genes in different organs and different tissues in Lotus (FPKM values).

Supplementary Table S5 | The expression of *NnNAC* DEGs involved in lotus seed development (FPKM values).

Supplementary Table S6 | The expression of *NnNAC* DEGs involved in completely submerged stress response (FPKM values).

REFERENCES

Bhattacharjee, P., Das, R., Mandal, A., and Kundu, P. (2017). Functional Characterization of Tomato Membrane-Bound NAC Transcription Factors. *Plant Mol. Biol.* 93, 511–532. doi:10.1007/s11103-016-0579-z

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170

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, S., Xiang, Y., Deng, J., Liu, Y., and Li, S. (2013). Simultaneous Analysis of Anthocyanin and Non-Anthocyanin Flavonoid in Various Tissues of Different lotus (*Nelumbo*) Cultivars by HPLC-DAD-ESI-MSn. *PLoS One* 8, e62291. doi:10.1371/journal.pone.0062291

Cheng, L. B., Han, Y. Y., Liu, H. Y., Jiang, R. Z., and Li, S. Y. (2019). Transcriptomic Analysis Reveals Ethylene's Regulation Involved in Adventitious Roots Formation in lotus (*Nelumbo nucifera* Gaertn.). *Acta Physiol. Plant* 41, 97. doi:10.1007/s11738-019-2895-9

Christiansen, M. W., and Gregersen, P. L. (2014). Members of the Barley NAC Transcription Factor Gene Family Show Differential Co-Regulation with Senescence-Associated Genes during Senescence of Flag Leaves. *J. Exp. Bot.* 65, 4009–4022. doi:10.1093/jxb/eru046

De Clercq, I., Vermeirssen, V., Van Aken, O., Vandepoele, K., Murcha, M. W., Law, S. R., et al. (2013). The Membrane-Bound NAC Transcription Factor ANAC013 Functions in Mitochondrial Retrograde Regulation of the Oxidative Stress Response in *Arabidopsis*. *Plant Cell* 25, 3472–3490. doi:10.1105/tpc.113.117168

De Smet, R., Adams, K. L., Vandepoele, K., Van Montagu, M. C. E., Maere, S., and Van de Peer, Y. (2013). Convergent Gene Loss Following Gene and Genome Duplications Creates Single-Copy Families in Flowering Plants. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2898–2903. doi:10.1073/pnas.1300127110

Deng, J., Li, M., Huang, L., Yang, M., and Yang, P. (2016). Genome-Wide Analysis of the *R2R3 MYB* Subfamily Genes in Lotus (*Nelumbo Nucifera*). *Plant Mol. Biol. Rep.* 34, 1016–1026. doi:10.1007/s11105-016-0981-3

Deng, X., Yang, D., Sun, H., Liu, J., Song, H., Xiong, Y., et al. (2022). Time-course Analysis and Transcriptomic Identification of Key Response Strategies to Complete Submergence in *Nelumbo Nucifera*. *Hortic. Res.* 9, uhac001. doi:10.1093/hr/uhac001

Droge-Laser, W., Snoek, B. L., Snel, B., and Weiste, C. (2018). The *Arabidopsis* bZIP Transcription Factor Family-An Update. *Curr. Opin. Plant Biol.* 45, 36–49. doi:10.1016/j.pbi.2018.05.001

Jensen, M. K., Kjaersgaard, T., Nielsen, M. M., Galberg, P., Petersen, K., O'Shea, C., et al. (2010). The *Arabidopsis Thaliana* NAC Transcription Factor Family: Structure-Function Relationships and Determinants of ANAC019 Stress Signaling. *Biochem. J.* 426, 183–196. doi:10.1042/BJ20091234

Jiang, J., Ma, S., Ye, N., Jiang, M., Cao, J., and Zhang, J. (2016). WRKY Transcription Factors in Plant Responses to Stresses. *J. Integr. Plant Biol.* 59, 86–101. doi:10.1111/jipb.12513

Kim, S.-G., Lee, S., Ryu, J., and Park, C.-M. (2010a). Probing Protein Structural Requirements for Activation of Membrane-Bound NAC Transcription Factors in *Arabidopsis* and Rice. *Plant Sci.* 178, 239–244. doi:10.1016/j.plantsci.2009.12.007

Kim, S.-G., Lee, S., Seo, P. J., Kim, S.-K., Kim, J.-K., and Park, C.-M. (2010b). Genome-Scale Screening and Molecular Characterization of Membrane-Bound Transcription Factors in *Arabidopsis* and Rice. *Genomics* 95, 56–65. doi:10.1016/j.ygeno.2009.09.003

Kim, S.-Y., Kim, S.-G., Kim, Y.-S., Seo, P. J., Bae, M., Yoon, H.-K., et al. (2007). Exploring Membrane-Associated NAC Transcription Factors in *Arabidopsis*: Implications for Membrane Biology in Genome Regulation. *Nucleic Acids Res.* 35, 203–213. doi:10.1093/nar/gkl1068

Kokkirala, V. R., Yonggang, P., Abbagani, S., Zhu, Z., and Umate, P. (2010). Subcellular Localization of Proteins of *Oryza Sativa L.* In the Model Tobacco and Tomato Plants. *Plant Signal Behav.* 5, 1336–1341. doi:10.4161/psb.5.11.13318

Kumar, R., Das, S., Mishra, M., Choudhury, D. R., Sharma, K., Kumari, A., et al. (2021). Emerging Roles of NAC Transcription Factor in Medicinal Plants: Progress and Prospects. *3 Biotech.* 11, 1–14. doi:10.1007/s13205-021-02970-x

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi:10.1093/molbev/msw054

Lahiri, A., Zhou, L., He, P., and Datta, A. (2021). Detecting Drought Regulators Using Stochastic Inference in Bayesian Networks. *PLos One* 16, e0255486. doi:10.1371/journal.pone.0255486

Le, D. T., Nishiyama, R., Watanabe, Y., Mochida, K., Yamaguchi-Shinozaki, K., Shinozaki, K., et al. (2011). Genome-Wide Survey and Expression Analysis of the Plant-Specific NAC Transcription Factor Family in Soybean during Development and Dehydration Stress. *DNA Res.* 18, 263–276. doi:10.1093/dnares/dsr015

Lescot, M., Dehais, P., Thijs, G., Marchal, K., Moreau, Y., Peer, Y. V., et al. (2002). PlantCARE, a Database of Plant *Cis*-Acting Regulatory Elements and a Portal to Tools for In Silico Analysis of Promoter Sequences. *Nucleic Acids Res.* 30, 325–327. doi:10.1093/nar/30.1.325

Li, H., Yang, X., Zhang, Y., Gao, Z., Liang, Y., Chen, J., et al. (2021a). Nelumbo Genome Database, an Integrative Resource for Gene Expression and Variants of *Nelumbo N. Sci. Data* 8, 38. doi:10.1038/s41597-021-00828-8

Li, J., Shi, T., Huang, L., He, D., Nyong'A, T. M., and Yang, P. (2018b). Systematic Transcriptomic Analysis Provides Insights into lotus (*Nelumbo Nucifera*) Seed Development. *Plant Growth Regul.* 86, 339–350. doi:10.1007/s10725-018-0433-1

Li, J., Xiong, Y., Li, Y., Ye, S., Yin, Q., Gao, S., et al. (2019). Comprehensive Analysis and Functional Studies of WRKY Transcription Factors in *Nelumbo Nucifera*. *Int. J. Mol. Sci.* 20, 5006. doi:10.3390/ijms20205006

Li, S., Wang, N., Ji, D., Xue, Z., Yu, Y., Jiang, Y., et al. (2016). Evolutionary and Functional Analysis of Membrane-Bound NAC Transcription Factor Genes in Soybean. *Plant Physiol.* 172, 1804–1820. doi:10.1104/pp.17.0001810.1104/pp.16.01132

Li, W., Li, X., Chao, J., Zhang, Z., Wang, W., and Guo, Y. (2018a). NAC Family Transcription Factors in Tobacco and Their Potential Role in Regulating Leaf Senescence. *Front. Plant Sci.* 9, 1900. doi:10.3389/fpls.2018.01900

Li, W., Zeng, Y., Yin, F., Wei, R., and Mao, X. (2021b). Genome-Wide Identification and Comprehensive Analysis of the NAC Transcription

Factor Family in Sunflower during Salt and Drought Stress. *Sci. Rep.* 11, 19865. doi:10.1038/s41598-021-98107-4

Lindemose, S., Jensen, M. K., de Velde, J. V., O'Shea, C., Heyndrickx, K. S., Workman, C. T., et al. (2014). A DNA-Binding-Site Landscape and Regulatory Network Analysis for NAC Transcription Factors in *Arabidopsis Thaliana*. *Nucleic Acids Res.* 42, 7681–7693. doi:10.1093/nar/gku502

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8

Ma, X., Zhang, Y., Turečková, V., Xue, G.-P., Fernie, A. R., Mueller-Roeber, B., et al. (2018). The NAC Transcription Factor SlNAP2 Regulates Leaf Senescence and Fruit Yield in Tomato. *Plant Physiol.* 177, 1286–1302. doi:10.1104/pp.18.00292

Mathew, I. E., Das, S., Mahto, A., and Agarwal, P. (2016). Three Rice NAC Transcription Factors Heteromerize and Are Associated with Seed Size. *Front. Plant Sci.* 7, 1638. doi:10.3389/fpls.2016.01638

Mathew, I. E., Priyadarshini, R., Mahto, A., Jaiswal, P., Parida, S. K., and Agarwal, P. (2020). SUPER STARCHY1/ONAC025 Participates in Rice Grain Filling. *Plant Direct* 4, e00249. doi:10.1002/pld3.249

Min, T., Niu, L.-F., Xie, J., Yi, Y., Wang, L.-M., Ai, Y.-W., et al. (2019). Effects of Vacuum Packaging on NAC Gene Expression in Fresh-Cut Lotus Root. *J. Amer. Soc. Hort. Sci.* 145, 36–44. doi:10.21273/JASHS04806-19

Ming, R., Vanburen, R., Liu, Y., Yang, M., Han, Y., Li, L. T., et al. (2013). Genome of the Long-Living Sacred lotus (*Nelumbo Nucifera* Gaertn.). *Genome Biol.* 14, R41. doi:10.1186/gb-2013-14-5-r41

Nakano, Y., Yamaguchi, M., Endo, H., Rejab, N. A., and Ohtani, M. (2015). NAC-MYB-Based Transcriptional Regulation of Secondary Cell Wall Biosynthesis in Land Plants. *Front. Plant Sci.* 6, 288. doi:10.3389/fpls.2015.00288

Nakashima, K., Takasaki, H., Mizoi, J., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2012). NAC Transcription Factors in Plant Abiotic Stress Responses. *Biochem. Biophys. Acta* 1819, 97–103. doi:10.1016/j.bbagrm.2011.10.005

Ng, S., Ivanova, A., Duncan, O., Law, S. R., Van Aken, O., De Clercq, I., et al. (2013). A Membrane-Bound NAC Transcription Factor, *ANAC017*, Mediates Mitochondrial Retrograde Signaling in *Arabidopsis*. *Plant Cell* 25, 3450–3471. doi:10.1105/tpc.113.113985

Nuruzzaman, M., Manimekalai, R., Sharoni, A. M., Satoh, K., Kondoh, H., Ooka, H., et al. (2010). Genome-Wide Analysis of NAC Transcription Factor Family in Rice. *Gene* 465, 30–44. doi:10.1016/j.gene.2010.06.008

Olsen, A. N., Ernst, H. A., Leggio, L. L., and Skriver, K. (2005). NAC Transcription Factors: Structurally Distinct, Functionally Diverse. *Trends Plant Sci.* 10, 79–87. doi:10.1016/j.tplants.2004.12.010

Peng, X., Zhao, Y., Li, X., Wu, M., Chai, W., Sheng, L., et al. (2015). Genome-Wide Identification, Classification and Analysis of NAC Type Gene Family in Maize. *J. Genet.* 94, 377–390. doi:10.1007/s12041-015-0526-9

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

Puranik, S., Sahu, P. P., Srivastava, P. S., and Prasad, M. (2012). NAC Proteins: Regulation and Role in Stress Tolerance. *Trends Plant Sci.* 17, 369–381. doi:10.1016/j.tplants.2012.02.004

Ren, Y., Huang, Z., Jiang, H., Wang, Z., Wu, F., Xiong, Y., et al. (2021). A Heat Stress Responsive NAC Transcription Factor Heterodimer Plays Key Roles in Rice Grain Filling. *J. Exp. Bot.* 72, 2947–2964. doi:10.1093/jxb/erab027

Rinerson, C. I., Rabara, R. C., Tripathi, P., Shen, Q. J., and Rushton, P. J. (2015). The Evolution of WRKY Transcription Factors. *BMC Plant Biol.* 15, 66. doi:10.1186/s12870-015-0456-y

Sakuraba, Y., Kim, Y.-S., Han, S.-H., Lee, B.-D., and Paek, N.-C. (2015). The Arabidopsis Transcription Factor NAC016 Promotes Drought Stress Responses by Repressing AREB1 Transcription through a Trifurcate Feed-Forward Regulatory Loop Involving NAP. *Plant Cell* 27, 1771–1787. doi:10.1105/tpc.15.00222

Seo, P. J., Kim, M. J., Park, J.-Y., Kim, S.-Y., Jeon, J., Lee, Y.-H., et al. (2010). Cold Activation of a Plasma Membrane-Tethered NAC Transcription Factor Induces a Pathogen Resistance Response in *Arabidopsis*. *Plant J.* 61, 661–671. doi:10.1111/j.1365-313X.2009.04091.x

Seo, P. J., Kim, S.-G., and Park, C.-M. (2008). Membrane-Bound Transcription Factors in Plants. *Trends Plant Sci.* 13, 550–556. doi:10.1016/j.tplants.2008.06.008

Seok, H.-Y., Woo, D.-H., Nguyen, L. V., Tran, H. T., Tarte, V. N., Mehdi, S. M. M., et al. (2017). *Arabidopsis AtNAP* Functions as a Negative Regulator via Repression of *AREB1* in Salt Stress Response. *Planta* 245, 329–341. doi:10.1007/s00425-016-2609-0

Shao, H., Wang, H., and Tang, X. (2015). NAC Transcription Factors in Plant Multiple Abiotic Stress Responses: Progress and Prospects. *Front. Plant Sci.* 6, 902. doi:10.3389/fpls.2015.00902

Shi, T., Rahmani, R. S., Gugger, P. F., Wang, M., Li, H., Zhang, Y., et al. (2020). Distinct Expression and Methylation Patterns for Genes with Different Fates Following a Single Whole-Genome Duplication in Flowering Plant. *Mol. Biol. Evol.* 37, 2394–2413. doi:10.1093/molbev/msaa105

Singh, S., Koyama, H., Bhati, K. K. K., and Alok, A. (2021). Correction to: The Biotechnological Importance of the Plant-Specific NAC Transcription Factor Family in Crop Improvement. *J. Plant Res.* 134, 643. doi:10.1007/s10265-021-01281-9

Souer, E., van Houwelingen, A., Kloos, D., Mol, J., and Koes, R. (1996). The *No Apical Meristem* Gene of Petunia Is Required for Pattern Formation in Embryos and Flowers and Is Expressed at Meristem and Primordia Boundaries. *Cell* 85, 159–170. doi:10.1016/S0092-8674(00)81093-4

Su, H., Zhang, S., Yin, Y., Zhu, D., and Han, L. (2015). Genome-Wide Analysis of NAM-ATAF1,2-CUC2 Transcription Factor Family in *Solanum Lycopersicum*. *J. Plant Biochem. Biotechnol.* 24, 176–183. doi:10.1007/s13562-014-0255-9

Sun, H., Hu, M., Li, J., Chen, L., Li, M., Zhang, S., et al. (2018). Comprehensive Analysis of NAC Transcription Factors Uncovers Their Roles during Fiber Development and Stress Response in Cotton. *BMC Plant Biol.* 18, 150. doi:10.1186/s12870-018-1367-5

Sun, H., Li, J., Song, H., Yang, D., Deng, X., Liu, J., et al. (2020). Comprehensive Analysis of AGPase Genes Uncovers Their Potential Roles in Starch Biosynthesis in Lotus Seed. *BMC Plant Biol.* 20, 457. doi:10.1186/s12870-020-02666-z

Sun, H., Liu, Y., Ma, J., Wang, Y., Song, H., Li, J., et al. (2021). Transcriptome Analysis Provides Strategies for Postharvest lotus Seeds Preservation. *Postharvest Biol. Technol.* 179, 111583. doi:10.1016/j.postharvbio.2021.111583

Takada, S., Hibara, K., Ishida, T., and Tasaka, M. (2001). The *CUP-SHAPED COTYLEDON1* Gene of *Arabidopsis* Regulates Shoot Apical Meristem Formation. *Development* 128, 1127–1135. doi:10.1242/dev.128.7.1127

Thirumalaikumar, V. P., Devkar, V., Mehterov, N., Ali, S., Ozgur, R., Turkan, I., et al. (2018). NAC Transcription Factor JUNGBRUNNEN1 Enhances Drought Tolerance in Tomato. *Plant Biotechnol. J.* 16, 354–366. doi:10.1111/pbi.12776

Tran, L.-S. P., Nakashima, K., Sakuma, Y., Simpson, S. D., Fujita, Y., Maruyama, K., et al. (2004). Isolation and Functional Analysis of *Arabidopsis* Stress-Inducible NAC Transcription Factors that Bind to a Drought-Responsive *Cis*-Element in the Early Responsive to Dehydration Stress 1 Promoter. *Plant Cell* 16, 2481–2498. doi:10.1105/tpc.104.022699

Wang, K., Guo, W., Yang, Z., Hu, Y., Zhang, W., Zhou, B., et al. (2010). Structure and Size Variations between 12 A and 12 D Homoeologous Chromosomes Based on High-Resolution Cytogenetic Map in Allotetraploid Cotton. *Chromosoma* 119, 255–266. doi:10.1007/s00412-009-0254-0

Wang, L., Fu, J., Li, M., Fragner, L., Weckwerth, W., and Yang, P. (2016). Metabolomic and Proteomic Profiles Reveal the Dynamics of Primary Metabolism during Seed Development of lotus (*Nelumbo nucifera*). *Front. Plant Sci.* 7, 750. doi:10.3389/fpls.2016.00750

Wang, Y., Fan, G., Liu, Y., Sun, F., Shi, C., Liu, X., et al. (2013). The Sacred lotus Genome Provides Insights into the Evolution of Flowering Plants. *Plant J.* 76, 557–567. doi:10.1111/tpj.12313

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: A Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity. *Nucleic Acids Res.* 40, e49. doi:10.1093/nar/gkr1293

Wingett, S. W., and Andrews, S. (2018). FastQ Screen: A Tool for Multi-Genome Mapping and Quality Control. *F1000Research* 7, 1338. doi:10.12688/f1000research.15931.2

Xu, B., Ohtani, M., Yamaguchi, M., Toyooka, K., Wakazaki, M., Sato, M., et al. (2014). Contribution of NAC Transcription Factors to Plant Adaptation to Land. *Science* 343, 1505–1508. doi:10.1126/science.1248417

Yamaguchi, M., Ohtani, M., Mitsuda, N., Kubo, M., Ohme-Takagi, M., Fukuda, H., et al. (2010). VND-INTERACTING2, a NAC Domain Transcription Factor, Negatively Regulates Xylem Vessel Formation in *Arabidopsis*. *Plant Cell* 22, 1249–1263. doi:10.1105/tpc.108.064048

Yan, H., Zhang, A., Ye, Y., Xu, B., Chen, J., He, X., et al. (2017). Genome-Wide Survey of Switchgrass NACs Family Provides New Insights into Motif and Structure Arrangements and Reveals Stress-Related and Tissue-Specific NACs. *Sci. Rep.* 7, 3056. doi:10.1038/s41598-017-03435-z

Yang, M., Zhu, L., Pan, C., Xu, L., Liu, Y., Ke, W., et al. (2015). Transcriptomic Analysis of the Regulation of Rhizome Formation in Temperate and Tropical Lotus (*Nelumbo Nucifera*). *Sci. Rep.* 5, 13059. doi:10.1038/srep13059

Yuan, C., Li, C., Lu, X., Zhao, X., Yan, C., Wang, J., et al. (2020). Comprehensive Genomic Characterization of NAC Transcription Factor Family and Their Response to Salt and Drought Stress in Peanut. *BMC Plant Biol.* 20, 454. doi:10.1186/s12870-020-02678-9

Zhang, Y., and Wang, L. (2005). The WRKY Transcription Factor Superfamily: Its Origin in Eukaryotes and Expansion in Plants. *BMC Evol. Biol.* 5, 1. doi:10.1186/1471-2148-5-1

Zhong, R., Demura, T., and Ye, Z.-H. (2006). SND1, a NAC Domain Transcription Factor, Is a Key Regulator of Secondary Wall Synthesis in Fibers of *Arabidopsis*. *Plant Cell* 18, 3158–3170. doi:10.1105/tpc.106.047399

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.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 14 June 2022
doi: 10.3389/fgene.2022.892674

Transcriptome Analysis Reveals that Exogenous Melatonin Confers *Lilium* Disease Resistance to *Botrytis elliptica*

Xuehua Xie[1,2,3,4,5,6], *Yu Han*[1,2,3,4,5,6], *Xi Yuan*[1,2,3,4,5,6], *Man Zhang*[1,2,3,4,5,6], *Ping Li*[1,2,3,4,5,6], *Aiqin Ding*[1,2,3,4,5,6], *Jia Wang*[1,2,3,4,5,6], *Tangren Cheng*[1,2,3,4,5,6] and *Qixiang Zhang*[1,2,3,4,5,6]*

[1]*Beijing Key Laboratory of Ornamental Plants Germplasm Innovation & Molecular Breeding, Beijing Forestry University, Beijing, China,* [2]*National Engineering Research Center for Floriculture, Beijing Forestry University, Beijing, China,* [3]*Beijing Laboratory of Urban and Rural Ecological Environment, Beijing Forestry University, Beijing, China,* [4]*Engineering Research Center of Landscape Environment of Ministry of Education, Beijing Forestry University, Beijing, China,* [5]*Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants of Ministry of Education, Beijing Forestry University, Beijing, China,* [6]*School of Landscape Architecture, Beijing Forestry University, Beijing, China*

OPEN ACCESS

Edited by:
Suxu Tan,
Michigan State University, United States

Reviewed by:
Neeti Sanan-Mishra, International Centre for Genetic Engineering and Biotechnology, India
Ashutosh Pandey, National Institute of Plant Genome Research (NIPGR), India

***Correspondence:**
Qixiang Zhang
zqxbjfu@126.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *09 March 2022*
Accepted: *19 May 2022*
Published: *14 June 2022*

Citation:
Xie X, Han Y, Yuan X, Zhang M, Li P, Ding A, Wang J, Cheng T and Zhang Q (2022) Transcriptome Analysis Reveals that Exogenous Melatonin Confers Lilium Disease Resistance to Botrytis elliptica. Front. Genet. 13:892674. doi: 10.3389/fgene.2022.892674

Leaf blight, caused by *Botrytis elliptica* (Berk.) Cooke, is a devastating disease that limits the production of *Lilium* in China and in other countries worldwide. Numerous studies have indicated that plants have evolved sophisticated and effective signal transduction and defense-related pathways in response to pathogen invasion. Recently, particular attention has been given to the action(s) of melatonin in plants in response to biotic stress, and the role of melatonin in plant–pathogen interactions has also been discussed. In this study, RNA-seq was applied to analyze the transcriptomic changes in *Lilium* leaves that were pre-treated and post-treated with melatonin after *B. elliptica* infection for 0, 12, 24, 36, and 72 h and then compare those changes with those of the control. Treatment with exogenous melatonin and infection with *B. elliptica* caused differential expression of a large number of genes in *Lilium* leaves. KEGG pathway analysis showed that, after melatonin treatment, the defense-related DEGs were mainly enriched in plant–pathogen interactions, plant hormone signal transduction, MAPK signaling pathways, phenylpropanoid biosynthesis, and phenylalanine metabolism. RT–qPCR was used to verify the expression changes of 12 DEGs, the results of which were consistent with the RNA-seq analysis results. The expression of DEGs related to the MAPK pathway were significantly different between the MB group and the HB group, suggesting that, via the MAPK signaling cascade, melatonin may play a role in the disease resistance of *Lilium* to *B. elliptica*. This study provides a new perspective and information for molecular-based breeding of *Lilium* disease resistance.

Keywords: melatonin, disease resistance, *Botrytis elliptica*, transcriptome analysis, MAPK cascades

1 INTRODUCTION

Lilium is constantly affected by bacteria, viruses, and fungi, resulting in heavy economic losses during its growth. Specifically, leaf blight caused by *Botrytis elliptica* (Berk.) Cooke is a devastating disease that causes enormous losses to cut flower production and gardening applications under hot and heavy rainfall conditions. Understanding the molecular basis or resistance mechanism against pathogens is important for the sustainable production and breeding of new varieties. Studies on the mechanisms of plant disease resistance have shown that plants have evolved complex mechanisms to sense invading pathogens and employ proper responses accordingly downstream of effector-triggered immunity (ETI) or pathogen-triggered immunity (PTI) activation (Chisholm et al., 2006; Jones and Dangl, 2006; Boller and Felix, 2009; Dodds and Rathjen, 2010; Cook et al., 2015). In addition to three primary defense hormones, other plant hormones, such as abscisic acid (ABA), brassinosteroids (BRs), and cytokinins (CKs), are also vital resistance-related compounds that activate complex phytohormone signaling networks involved in plant defense against *Botrytis cinerea* (AbuQamar et al., 2017).

To date, the majority of related studies have focused the role of melatonin (N-acetyl-5-methoxytryptamine) since it was discovered in vascular plants in 1995 (Dubbels et al., 1995). Due to its wide-ranging functions in plants, as a pleiotropic signaling molecule, melatonin plays important roles in the regulation of plant growth; development; and defense against various environmental stresses, such as drought, salt, cold, heat, and heavy metals (Arnao and Hernández-Ruiz, 2014; Byeon and Back, 2014; Nawaz et al., 2015; Wei et al., 2015; Arnao and Hernandez-Ruiz, 2018; Gu et al., 2021; Tiwari et al., 2021). In addition, previous studies have suggested that exogenous melatonin improves plant resistance to pathogen infection and shows its efficacy against devastating fungal pathogens such as *Diplocarpon mali* in apple (Yin et al., 2013), *Phytophthora infestans* in potato (Zhang et al., 2017), *Podosphaera xanthii* in cucurbits (Mandal et al., 2018). Mycelial growth of phytopathogenic fungi (*Colletotrichum gloeosporioides* and *Colletotrichum acutatum*) which caused severe anthracnose in *Capsicum annum* were significantly reduced 76 and 71% after a 100 μM concentration of phytomelatonin treatment (Ali et al., 2021).

Melatonin is a powerful antioxidant that can remove excess ROS and RNS and indirectly activate both enzymatic and non-enzymatic antioxidant systems under biological stress, acting as a signal molecule for elevated defense response against bacterial infection. The application of melatonin has been found to enhance the production of SA, nitric oxide (NO) and hydrogen peroxide (H_2O_2) in plants infected with pathogens (Lee et al., 2014). Studies have implicated melatonin in the induction of plant resistance to pathogens alone or in conjunction with the defense-related hormones SA, NO and H_2O_2 (Shi et al., 2015a; Shi et al., 2015b; Qian et al., 2015), indirect crosstalk of melatonin-phytohormone will help the plants to cope with the pathogen attack and a working model of melatonin-mediated plant resistance has been proposed (Zhao et al., 2019; Tiwari et al., 2021). In addition, by affecting lignin and gossypol synthesis genes involved in the phenylpropanoid and gossypol pathways, melatonin enhances cotton immunity to *Verticillium wilt* (Li et al., 2019). An increasing number of researchers have implicated melatonin in the regulation of the immune signaling network and interactions between plants and pathogens (Shi et al., 2015a; Li et al., 2016; Wei Y. et al., 2018). Study of melatonin spray in alleviating powdery mildew infection in watermelon. Demonstrated that melatonin played a positive role in enhancing plant resistance by enhancing the expression of PTI and ETI response-related genes in watermelon (Mandal et al., 2018), however, the direct effect of melatonin on promoting PTI and ETI is not well explored. Researches show that signaling with respect to melatonin-mediated innate immunity in plants occurs through various *MAPKKK* kinases within MAPK signaling cascades (Lee and Back, 2016; Lee and Back, 2017). MAPK signaling cascades are important signaling modules and play important roles in plant growth; development; and adaptation to environmental stress, such as cold, drought, and to pathogens (Banerjee et al., 2020). Following pathogen infection, ROS and NO bursts activate MAPK signaling cascades to phosphorylate downstream targets, including transcription factors, and to promote the synthesis of defense-related enzymes and other antimicrobial compounds that in turn activate cellular responses (Frawley and Bayram, 2020; Zhang and Zhang, 2022). However, it is still unclear whether such a response of melatonin in disease resistance is universal across plant species, and the mechanism underlying melatonin-mediated disease resistance is unknown.

In this study, we conducted an RNA sequencing (RNA-seq) analysis of *Lilium* genes that were differentially expressed in response to melatonin and involved in resistance to *B. elliptica*. The results therefore reveal the metabolic mechanism underlying melatonin-mediated resistance of *Lilium* to *B. elliptica*.

2 MATERIALS AND METHODS

2.1 Plant Materials and Fungal Culture Conditions

The oriental hybrid *Lilium* cultivar "Sorbonne" was selected for this study. Bulbs were planted in pots (17 cm in diameter) containing peat and pearlite (1:1) as a growth substrate. The pots were placed in a greenhouse at Beijing Forestry University under a 12 h day/night photoperiod at 25/22°C. *B. elliptica* isolated from diseased *Lilium* leaves was grown on potato dextrose agar media for 7 days at 25°C. Mycelial discs (1 cm in diameter) were subsequently obtained with a leather punch for inoculation.

2.2 Treatment and Experimental Design

The roots of plants growing for 45 days were treated with 0, 0.02, 0.2, 2, or 20 mM melatonin (100 ml per plant) for 6 and 10 days (once every 2 days). Melatonin (Sigma–Aldrich, St. Louis, MO, United States) solutions were prepared by dissolving the soluble materials in ethanol followed by dilution with Milli-Q water.

After melatonin pretreatment, six leaves from the middle of plants were inoculated with *B. elliptica*, and the area of each lesion was measured at 36, 72 and 120 h after inoculation. Six plants were used per treatment, and this process was repeated three times. Based on the percentage of lesion area covering the total area of the *Lilium* leaves, the disease severity was assessed on the basis of a 5-point scale: 0, no visible symptoms; 1, symptoms <5%; 2, symptoms = 5%–25%; 3, symptoms = 25%–50%; 4, symptoms >50%. The disease index was calculated by assessing 36 leaves and then calculated with the following formula:

$$\text{Disease index} = \frac{\sum(\text{number of leaves for different grade} \times \text{grade})}{\text{Total assessed leaves} \times 7}$$

The experiment for transcriptome sequencing consisted of two parts: pre-treatment and post-treatment. In the pretreatment part, *Lilium* plants were treated with 100 ml of a 2 mM melatonin solution by irrigating the roots once every other day with the solution, for a total of five times. Control plants were treated with 0 mM melatonin, and then the leaves in the middle of the plants were inoculated with *B. elliptica*. The plants inoculated with *B. elliptica* were cultivated in a growth chamber at 25°C under a 16 h light/8 h dark photoperiod, and samples were obtained at five time points (0, 12, 24, 36, and 72 h) after inoculation from the inoculated sites for transcriptome sequencing. In the posttreatment part, *Lilium* plants were inoculated with *B. elliptica* for 36 h, and then the inoculated plants were treated with 100 ml of 2 and 0 mM melatonin solutions by irrigating the roots with those solutions. Leaf samples from the inoculated sites were obtained for transcriptome sequencing at 36 h after treatment. Each treatment included three biological replications, and each replication comprised 12 plants. All the samples used for transcriptome sequencing were frozen immediately in liquid nitrogen and stored at −80°C.

2.3 Enzyme Extraction and Activity Assays

Leaf samples from plants pre-treated with 2 and 0 mM melatonin were obtained at five time points (0, 12, 24, 36, 48, and 72 h) after inoculation with *B. elliptica*. The harvested samples were rapidly frozen in liquid nitrogen and stored at −80°C for the phenylalanine ammonia lyase (PAL) and catalase (CAT) assays. A PAL test kit (Nanjing Jiancheng Bioengineering Institute, A137-1-1) and CAT assay kit-visible light (Nanjing Jiancheng Bioengineering Institute, A007-1-1) were used for enzyme extraction and activity assays. The reaction mixture, operation process and enzyme activity were prepared, performed, and calculated, respectively, according to the instructions of the kits.

2.4 RNA Extraction, Library Preparation, and Illumina Sequencing

Total RNA was extracted using an RNAprep Pure Plant Kit (Tiangen, Beijing, China) as described by the manufacturer's protocol for each RNA-seq sample, and RNA quality was detected via 1% agarose gel electrophoresis. The RNA degradation, purity, concentration, and integrity were measured using a Nanodrop™ 2000 spectrophotometer (Thermo Fisher Scientific, United States), Qubit RNA Assay Kit with Qubit Fluorometer 2.0 (Life Technology, Carlsbad, CA, United States), and an Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, United States). Sequencing libraries were constructed using a NEB Next® Ultra™ RNA Library Prep Kit for Illumina® (NEB, Beverly, CA, United States). Library preparation sequencing was performed on an Illumina HiSeq 4000 platform, with 150 bp paired-end reads.

2.5 Sequence Assembly and Annotation

Quality control analysis was conducted for raw reads obtained from the sequencing data to identify high-quality sequencing data and clean reads. The Trinity v2.4.0 platform (with the parameters K-mer = 25 and group pair distance = 250) was used to stitch and assemble clean reads (Grabherr et al., 2011). Redundant sequences were removed from the unigene sequences with TGICL software, and sequences with lengths as long as possible were assembled. Fragments per kilobase of exon per million mapped fragments (FPKM) were used to standardize the read counts of each gene. BLASTX searches were performed based on the information housed in the following databases: the National Center for Biotechnology Information (NCBI) nonredundant (Nr) protein sequence database (e-value = $1e^{-5}$), the NCBI nonredundant nucleotide (Nt) sequence database (e-value = $1e^{-5}$), the SwissProt manually annotated and reviewed protein sequence database (e-value = $1e^{-5}$), the Protein family (Pfam) database (e-value = 0.01), the EuKaryotic Orthologous Groups (KOG)/Clusters of Orthologous Genes (COG) database (e-value = $1e^{-3}$), the Gene Ontology (GO) database (standardized classification for gene function; e-value = $1e^{-6}$), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (gene product functions and metabolic pathways; e-value = $1e^{-10}$).

2.6 Differentially Expressed Gene Analysis

DEG analysis was performed using the DESeq R package. To evaluate the genes that were significantly expressed between two samples, genes with an adjusted p value (q-value) < 0.05 and an $|\log_2(\text{fold change})\ \text{FPKM}| > 2$ were defined as differentially expressed. The GOseq R package (1.10.0) was used for GO enrichment analysis, with a false discovery rate (FDR) < 0.01, and KOBAS (v2.0.12) was used for KEGG pathway enrichment analysis, with FDR of <0.01 (http://kobas.cbi.pku.edu.cn/). Venn diagrams, bubble maps, and heatmaps were constructed using the online data analysis platform Omicstudio tools (https://www.omicstudio.cn/tool/6).

2.7 Quantitative Real-Time PCR Analysis

To validate the RNA-seq results of the gene expression levels reflected by the FPKM values, RT–qPCR was performed using a 7500 Real-Time PCR System (Applied Biosystems, CA, United States) and a SYBR R Premix Ex Taq™ Kit (TaKaRa, Tokyo, Japan). Total RNA was extracted and reverse-transcribed into first-strand cDNA using a PrimeScript™ RT Reagent Kit

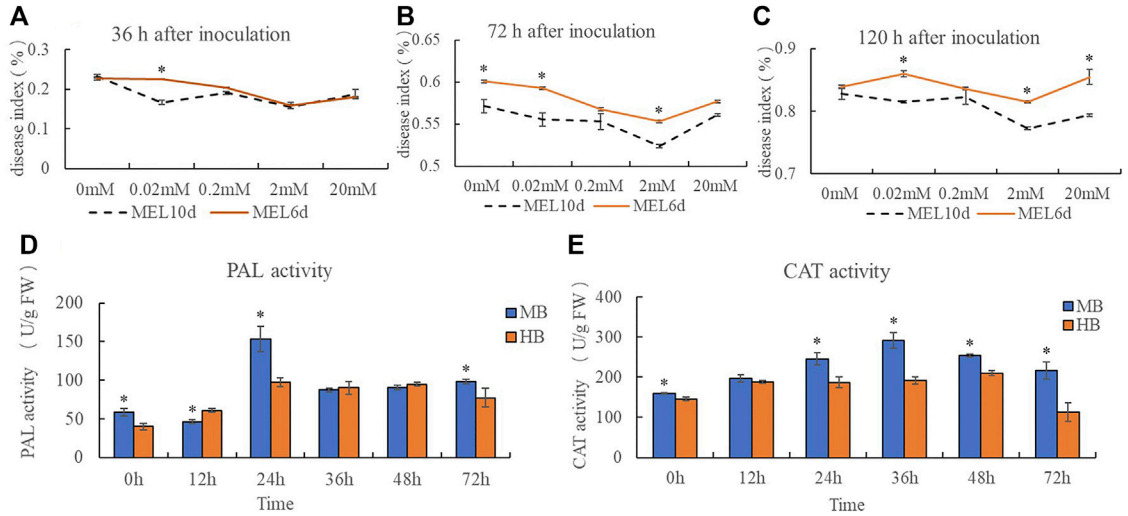


FIGURE 1 | Effect of exogenous melatonin pre-treatment on *Lilium* resistance to *B. elliptica*. **(A)** Effect of melatonin concentration on disease index at 36 h for plants pretreated with 0, 0.02, 0.2, 2 or 20 mM melatonin. MEL6d represents plants were pre-treatment with melatonin for 6 days (once every 2 days), MEL10d represents plants were pre-treatment with melatonin for 10 days (once every 2 days). **(B)** Effect of melatonin concentration on disease index at 72 h for plants pretreated with 0, 0.02, 0.2, 2 or 20 mM melatonin. MEL6d represents plants were pre-treatment with melatonin for 6 days (once every 2 days), MEL10d represents plants were pre-treatment with melatonin for 10 days (once every 2 days). **(C)** Effect of melatonin concentration on disease index at 120 h for plants pretreated with 0, 0.02, 0.2, 2 or 20 mM melatonin. MEL6d represents plants were pre-treatment with melatonin for 6 days (once every 2 days), MEL10d represents plants were pre-treatment with melatonin for 10 days (once every 2 days). **(D)** Effect of exogenous melatonin pre-treatment on Phenylalanine ammonia lyase (PAL) activity of *Lilium* after inoculation. MB represents plants were pre-treatment with 2 mM melatonin and then inoculated with *B. elliptica*; HB represents plants were pre-treatment with Milli-Q water [ethanol/water (v/v) D 1/10,000], and then inoculated with *B. elliptica*. **(E)** Effect of exogenous melatonin pre-treatment on catalase (CAT) activity of *Lilium* after inoculation. MB represents plants were pre-treatment with 2 mM melatonin and then inoculated with *B. elliptica*; HB represents plants were pre-treatment with Milli-Q water [ethanol/water (v/v) D 1/10,000], and then inoculated with *B. elliptica*. Data represent mean ± SE of triplicate assays. The line charts were generated based on IBM SPSS Statistics 20. The "*" represents the significant differences.

with gDNA Eraser (Takara Bio, Inc., Shiga, Japan) following the manufacturer's protocol. RT–qPCR was performed on a 20 ml reaction mixture including 2 µl of first-strand cDNA, 0.6 µl of forward primer, 0.6 µl of reverse primer, 10 µl of SYBR Premix Ex Taq and 6.8 µl of sterile distilled water under the following reaction conditions: 95°C for 3 min, followed by 40 cycles of 95°C for 10 s, 55°C for 15 s and 72°C for 15 s. The *Lilium* eukaryotic elongation factor 1 (EF1, KJ543461) gene was used as a reference gene for normalization (Liu et al., 2016). Experiments were performed for three independent biological replicates and three technical replicates. The $2^{-\Delta\Delta Ct}$ method was used to calculate the relative expression levels of the selected transcripts (Livak and Schmittgen, 2001). Detailed information on the primer sequences used for RT–qPCR is listed in **Supplementary Table S6**. The correlation coefficients between the RT–qPCR results and FPKM value were analyzed using SPSS 22.0 software.

3 RESULTS

3.1 Exogenous Melatonin Enhanced *Lilium* Resistance to *B. elliptica*

A disease index was calculated for plants inoculated with *B. elliptica* after treatment with 0, 0.02, 0.2, 2, or 20 mM melatonin. The results indicate that pre-treatment with exogenous melatonin improved the resistance of *Lilium* plants to *B. elliptica*. The addition of 0.02, 0.2, 2 or 20 mM melatonin alleviated blotch damage to varying degrees compared with that of plants pre-treated with 0 mM melatonin (**Figures 1A–C**). The 2 mM melatonin concentration was used for subsequent analyses. Compared with plants treated with melatonin for 6 days, plants treated with melatonin for 10 days presented a lower disease index. PAL activity and CAT activity first decreased but then increased during the process of pathogen infection. Moreover, the PAL activity of plants pre-treated with 2 mM melatonin was significantly higher than that of the plants in the control group at 0 and 24 h after inoculation (**Figure 1D**), and the CAT activity of plants pre-treated with 2 mM melatonin was significantly higher than that of plants in the control group at 24, 36 and 72 h after inoculation (**Figure 1E**), suggesting that melatonin might act as a regulator to enhance *Lilium* resistance against *B. elliptica*.

3.2 Summary of Transcriptome Sequencing Data From the *Lilium* Hybrid Cultivar "Sorbonne"

Thirty six samples for transcriptome sequencing were obtained from pre-treatment plants at five time points after inoculation (MB 0 h, MB 12 h, MB 24 h, MB 36 h, MB 72 h and HB 0 h, HB 12 h, HB 24 h, HB 36 h, HB 72 h) and post-treatment plants (BEM, BEH), three biological replicates were performed for each treatment. Approximately 298.72 Gb of data from 36 samples was

obtained after filtering and quality control measures were performed in this study. The total number of clean reads for each sample ranged from 65,403,890 to 86,281,814, and the clean read percentage was between 98.02 and 98.45%. All the above data and the sequencing quality, represented by the Q20 value, Q30 percentage and GC content for each sample, are shown in **Supplementary Table S1**. The Q30 percentage was greater than 95%, and the GC content was between 45.65% and 45.88%. The clean reads of each sample were mapped to the reference sequence, and the alignment proportion of each sample was greater than 78%. The quality of the sequencing data met the requirements for assembly. After assembly and the removal of redundancy were performed, 220,215 unigenes were ultimately generated with a total length, average length, N50 and GC content of 187,808,990 bp, 852 bp, 1,428 bp and 46.78%, respectively (**Supplementary Table S2**).

Then, the annotation information of the unigenes was checked via the Nr, Nt, SwissProt, KOG, KEGG, GO, and Pfam databases (**Supplementary Table S3**). A total of 122,909 (55.81%) unigenes were annotated in the public databases, and the numbers of annotated unigenes in the seven functional databases were 103,159 (Nr: 46.84%), 48,154 (Nt: 21.87%), 71,635 (SwissProt: 32.53%), 81,270 (KOG: 36.90%), 78,538 (KEGG: 35.66%), 77,926 (GO: 35.39%), and 88,964 (Pfam: 40.40%). The unigenes that matched sequences from the genome of oil palm (*Elaeis guineensis*) accounted for 21.37% of the total annotations, followed by date (*Phoenix dactylifera*), *Asparagus officinalis*, pineapple (*Ananas comosus*) and others.

Unigenes annotated in the Nr database were classified into three major functional categories [biological processes (BPs), cellular component (CCs) and molecular functions (MFs)] after GO mapping and functional characterization (**Supplementary Figure S2**). The top terms with the most annotated genes in the three functional categories were metabolic process (15,042 genes, BP), cell (13,875 genes, CC), and catalytic activity (20,251 genes, MF). We next identified the biological pathways via the KEGG database associated with the annotated sequences, and all the unigenes were assigned to 136 pathways and five KEGG categories based on the pathway hierarchy. A total of 91,104 coding DNA sequences (CDSs) were detected using TransDecoder. It also revealed 20,806 simple sequence repeats (SSRs) distributed among 17,875 unigenes and predicted 3,013 unigenes encoding transcription factors. All the assembled contigs and data can be found in the NCBI BioProject database under accession number PRJNA799047.

3.3 Differential Expression Analysis of the *Lilium* Hybrid "Sorbonne" After Treatment With Exogenous Melatonin and Inoculation With *B. elliptica*

To investigate role of melatonin on the resistance to gray mold of *Lilium*, comparisons were conducted between the melatonin pretreatment group (MB) and the control group (HB) and between the melatonin posttreatment group (BEM) and the control group (BEH) during different infection stages (MB-HB-0 h, MB-HB-12 h, MB-HB-24 h, MB-HB-36 h, MB-HB-72 h, BEM vs. BEH) as shown in **Figure 2**. **Figure 2A** displayed upregulated and downregulated DEGs of MB group vs. HB group at five points in time after inoculation, The results showed that the number of upregulated and downregulated DEGs increased significantly at 36 and 72 h after inoculation in the MB vs. HB comparison group. **Figure 2B** illustrated the number of DEGs overlapping in MB group vs. HB group at five points in time after inoculation, The unoverlapped region represents the unique differential genes in each comparison, The results showed that unique DEGs mainly appeared at 36 and 72 h after inoculation in the MB vs. HB comparison group, suggesting that exogenous melatonin may play a major role in the induction of *Lilium* resistance to *B. elliptica* at 36 and 72 h after inoculation. 58 DEGs were upregulated and 126 DEGs were downregulated in the BEM vs. BEH comparison group (**Figure 2C**), 15,282 unique DEGs and 11,973 unique DEGs began to present in BEM and BEH group, respectively (**Figure 2D**).

To investigate the DEGs expressed in *Lilium* in response to *B. elliptica* infection in both the MB group and the HB group, the data in the gene expression libraries corresponding to the five time points were organized into eight pairwise comparisons (12 vs. 0 h, 24 vs. 12 h, 36 vs. 24 h, 72 vs. 36 h; and 12 vs. 0 h, 24 vs. 0 h, 36 vs. 0 h, 72 vs. 0 h) to identify the genes that were differentially expressed during different infection stages as shown in **Figure 3**. Comprising four points in time with 0 h, the number of upregulated and downregulated DEGs increased significantly in 36 vs. 0 h comparison (**Figure 3A**), the highest number of unique DEGs was found in 36 vs. 0 h comparison in both the MB group and the HB group (**Figures 3B,C**). Comprising four points in time with previous points, the number of upregulated and downregulated DEGs of MB group increased significantly in 36 vs. 24 h comparison, while the number of DEGs of HB group was observed in 72 vs. 36 h comparison (**Figure 3D**), the highest number of unique DEGs was found in 36 vs. 24 h comparison of MB group and 72 vs. 36 h comparison of HB group, respectively (**Figures 3E,F**), suggesting that exogenous melatonin may accelerate the response of *Lilium* against *B. elliptica*.

To investigate global changes in expression patterns, we performed k-means clustering of the gene expression profiles for orthologous genes in the MB group and the HB group during pathogen infection. We classified the different expression modes into 12 clusters that showed distinctly different expression patterns (**Supplementary Figure S1**).

3.4 Functional Analysis of DEGs in *Lilium* After Treatment With Exogenous Melatonin and Inoculation With *B. elliptica*

DEGs were enriched and analyzed at different time points after pathogen infection with different treatments. GO terms with significant enrichment of DEGs were identified, and GO functional classification analysis of the DEGs between MB group and the HB group (MB-HB-0h, MB-HB-12h, MB-HB-24h, MB-HB-36h, MB-HB-72h, BEM vs. BEH) at five time different points (MB12h vs. MB0h, MB24h vs. MB12h, MB36h vs. MB24h, MB72h vs. MB36h and HB12h vs. HB0h, HB24h vs.

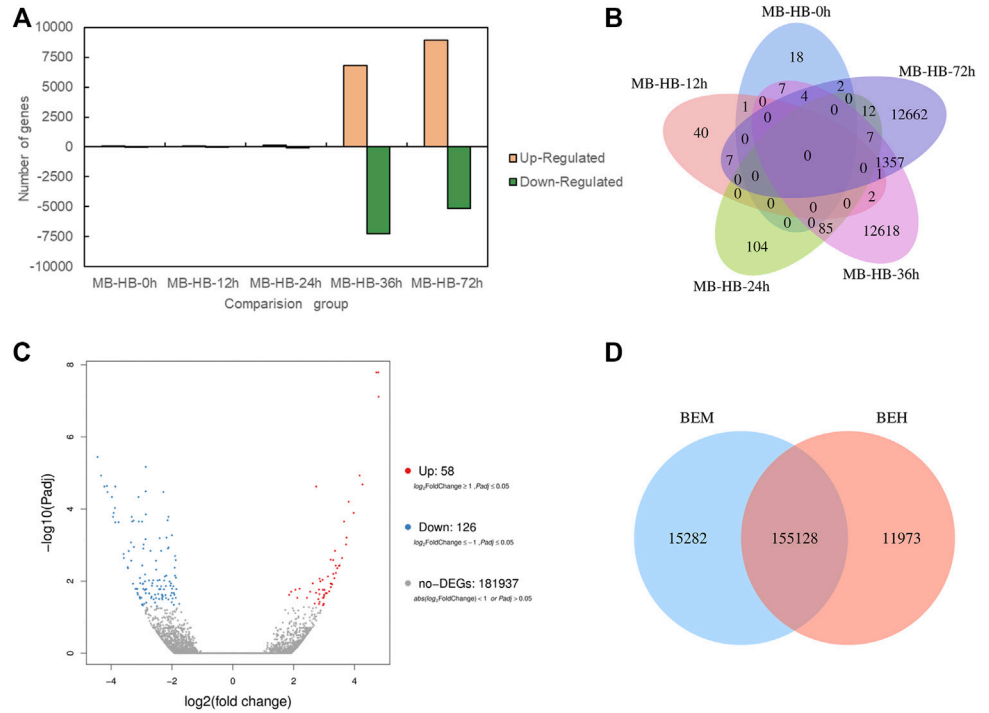


FIGURE 2 | Differentially expressed genes statistics between melatonin treatment group and control group. **(A)** Upregulated and down regulated DEGs (*p*adj < 0.05) statistics of MB group vs. HB group at five points in time after inoculation. MB represents plants were pre-treatment with melatonin and then inoculated with *B. elliptica*; HB represents plants were pre-treatment with Milli-Q water [ethanol/water (v/v) D 1/10,000], and then inoculated with *B. elliptica*. **(B)** Venn diagram of the number of DEGs (*p*adj < 0.05) between MB group vs. HB group. MB represents plants were pre-treatment with melatonin and then inoculated with *B. elliptica*; HB represents plants were pre-treatment with Milli-Q water [ethanol/water (v/v) D 1/10,000], and then inoculated with *B. elliptica*. **(C)** Volcano plot of DEGs (*p*adj < 0.05) between the comparison BEM vs. BEH, BEM represents plants were inoculated with *B. elliptica* for 36 h and then treated with 100 ml melatonin (2 mM/L) for 36 h; BEH represents plants were inoculated with *B. elliptica* for 36 h and then treated with 100 ml Milli-Q water [ethanol/water (v/v). D 1/10,000] for 36 h. **(D)** Venn diagram illustrate the number of DEGs (*p*adj < 0.05) between the comparison BEM vs. BEH, BEM represents plants were inoculated with *B. elliptica* for 36 h and then treated with 100 ml melatonin (2 mM/L) for 36 h; BEH represents plants were inoculated with *B. elliptica* 36 h and then treated with 100 ml Milli-Q water [ethanol/water (v/v) D 1/10,000] for 36 h.

HB0h, HB36h vs. HB0h, HB72h vs. HB0h) was performed. In the MB vs. HB comparison group (**Supplementary Table S4**), the results showed that at 0–24 h after pathogen inoculation, DEGs were significantly enriched in only 11 GO terms: cell wall organization or biogenesis, cellular component organization or biogenesis, structural constituent of cell wall, structural molecular activity, cytidine deaminase activity, and isocitrate lyase activity. At 36 and 72 h after inoculation, DEGs were significantly enriched in 92 GO terms, including carbohydrate metabolic process, chloroplast organization, cutin biosynthetic process, defense response, photosynthesis, peroxiredoxin activity, oxidoreductase activity, and thylakoid membrane. In the BEM vs. BEH comparison group, 16 GO terms were significantly enriched, and three of them were associated with MFs, namely, diacylglycerol O-acyltransferase activity, acylglycerol O-acyltransferase activity, and sequence-specific DNA binding. Thirteen were associated with BPs, including cell response to injury, migration of metal ions, and response to light intensity and ultraviolet light. At different time points, the results showed that, among 301 significantly enriched GO terms in the HB group, 144 (47.8%) were associated with BPs, and 25 (8.3%) were associated with CCs. Among 234 significantly enriched GO terms in the MB vs. HB comparison group, 132 (43.9%) were associated with MFs in the MB group, and 95 (40.6%) were associated with BPs. Sixty (25.6%) were associated with CCs; 79 (33.8%), MFs.

KEGG pathway cluster analysis was conducted between the MB group and the HB group at the five time points to explore the main metabolic pathways in which the DEGs were involved (**Supplementary Table S5**). The results of KEGG enrichment analysis of the MB vs. HB comparison group showed that 59 metabolic pathways were significantly enriched, which mainly involved energy metabolism, carbohydrate metabolism, amino acid metabolism and synthesis of secondary metabolites. In the BEM vs. BEH comparison group, the DEGs were significantly enriched in 13 metabolic pathways, including circadian rhythm plant (Ko04712); phenylpropanoid biosynthesis (Ko00360); and cutin, suberin and wax biosynthesis (Ko00073). A possible explanation is that exogenous melatonin treatment induced significant changes in the basal metabolism of *Lilium* during pathogen infection, and activity involving 10 metabolic pathways was detected at 0, 12 and 24 h. Activity of the other 49 metabolic pathways was detected at 36 and 72 h after inoculation, which indicated that, within 0–24 h after inoculation, the pathogen was

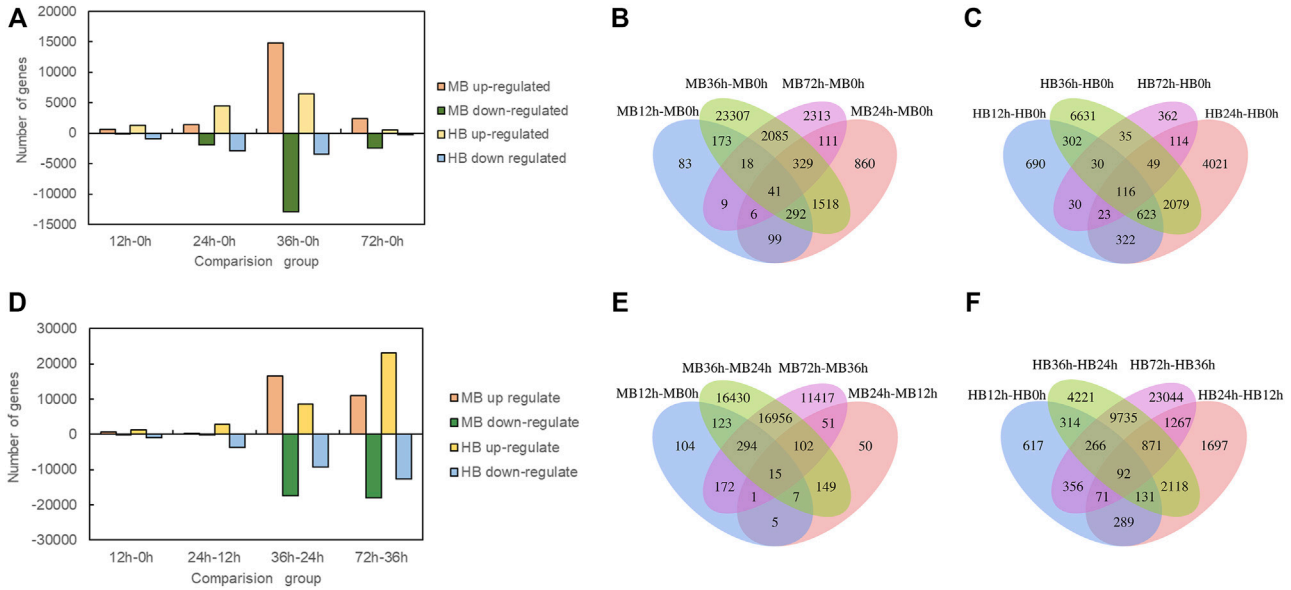


FIGURE 3 | Differentially expressed genes statistics at different time point's comparison. **(A)** Number of up regulated and down regulated DEGs (*padj* < 0.05) statistics for four comparisons (12 vs. 0 h; 24 vs. 0 h; 36 vs. 0 h; 72 vs. 0 h) in MB group and HB group. MB represents plants were pre-treatment with melatonin and then inoculated with *B. elliptica*; HB represents plants were pre-treatment with Milli-Q water [ethanol/water (v/v) D 1/10,000], and then inoculated with *B. elliptica*. **(B)** Venn diagram illustrate the number of DEGs in four comparisons (12 vs. 0 h; 24 vs. 0 h; 36 vs. 0 h; 72 vs. 0 h) overlapping in MB group, MB represents plants were pre-treatment with melatonin and then inoculated with *B. elliptica*; **(C)** Venn diagram illustrate the number of DEGs in four comparisons (12 vs. 0 h; 24 vs. 0 h; 36 vs. 0 h; 72 vs. 0 h) overlapping in HB group, HB represents plants were pre-treatment with Milli-Q water [ethanol/water (v/v) D 1/10,000], and then inoculated with *B. elliptica*; **(D)** Number of up regulated and down regulated DEGs (*padj* < 0.05) statistics for four comparisons (12 vs. 0 h; 24 vs. 12 h; 36 vs. 24 h; 72 vs. 36 h) in MB group and HB group. MB represents plants were pre-treatment with melatonin and then inoculated with *B. elliptica*; HB represents plants were pre-treatment with Milli-Q water [ethanol/water (v/v) D 1/10,000], and then inoculated with *B. elliptica*. **(E)** Venn diagram illustrate the number of DEGs in four comparisons (12 vs. 0 h; 24 vs. 12 h; 36 vs. 24 h; 72 vs. 36 h) overlapping in MB group, MB represents plants were pre-treatment with melatonin and then inoculated with *B. elliptica*. **(F)** Venn diagram illustrate the number of DEGs in four comparisons (12 vs. 0 h; 24 vs. 12 h; 36 vs. 24 h; 72 vs. 36 h) overlapping in HB group, HB represents plants were pre-treatment with Milli-Q water [ethanol/water (v/v) D 1/10,000], and then inoculated with *B. elliptica*.

in the colonization stage, did not completely invade the *Lilium* leaves, and did not induce the defense response in *Lilium* leaves. Melatonin may play a role in the response of *Lilium* to pathogen infection and in pathways related to plant disease resistance, such as the peroxisome pathway (Ko04146); MAPK signaling pathway-plant (Ko04016); the phenylpropanoid biosynthesis pathway (Ko00940); plant hormone signal transduction (Ko04075); the phenylalanine metabolism pathway (Ko00360); the cutin, suberin and wax biosynthesis pathway (Ko00073); and fatty acid metabolism pathway (Ko01212). Activity in these pathways was detected within 36 and 72 h after inoculation.

At different points in time, it was found that among all the significantly enriched pathways, the DEGs were mapped to 67 metabolic pathways in the MB group, while the DEGs were mapped to 72 metabolic pathways in the HB group. Significantly enriched pathways related to plant disease resistance mainly included MAPK signaling pathway-plant (Ko04146); phenylpropanoid biosynthesis (Ko00940); phenylalanine metabolism (Ko00360); terpenoid backbone biosynthesis (Ko00900); plant–pathogen interactions (Ko04626); plant hormone signal transduction (Ko04075); cutin, suberin and wax biosynthesis (Ko00073); etc. These pathways might play a key role in *Lilium* resistance to *B. elliptica*.

3.5 Transcriptome Analysis of DEGs Involved in the Plant–Pathogen Interaction Pathway, Phenylpropanoid Biosynthesis Pathway, and SA and JA Signaling Pathways During *B. elliptica* Infection

On the basis of the KEGG pathway enrichment data for the 5 time point comparisons after *B. elliptica* inoculation, DEGs that were associated with disease resistance to *B. elliptica* and involved in the phenylpropanoid biosynthesis pathway, plant–pathogen interaction pathway, and plant hormone signal transduction pathway were selected. The gene expression data at the 5 time points is shown in **Figure 4**. Twenty-five unigenes involved in the phenylpropanoid biosynthesis pathway, including *peroxidase* (*POD*), *caffeic acid 3-O-methyltransferase* (*COMT*), *cinnamyl alcohol dehydrogenase* (*CAD*), *naringenin 3-dioxygenase* (*F3H*), *coumaroyl quinate 3′-monooxygenase* (*C3′H*), *chalcone synthase* (*CHS*), *4-coumarate: CoA ligase* (*4CL*), *PAL*, *cinnamate 4-hydroxylase* (*C4H*), *flavonol synthase* (*FLS*), bifunctional dihydroflavonol 4-reductase (*DFR*), *shikimate O-hydroxycinnamoyl transferase* (*HCT*), *flavonoid 3′-monooxygenase* (*F3′H*), and *chalcone isomerase* (*CHI*), were

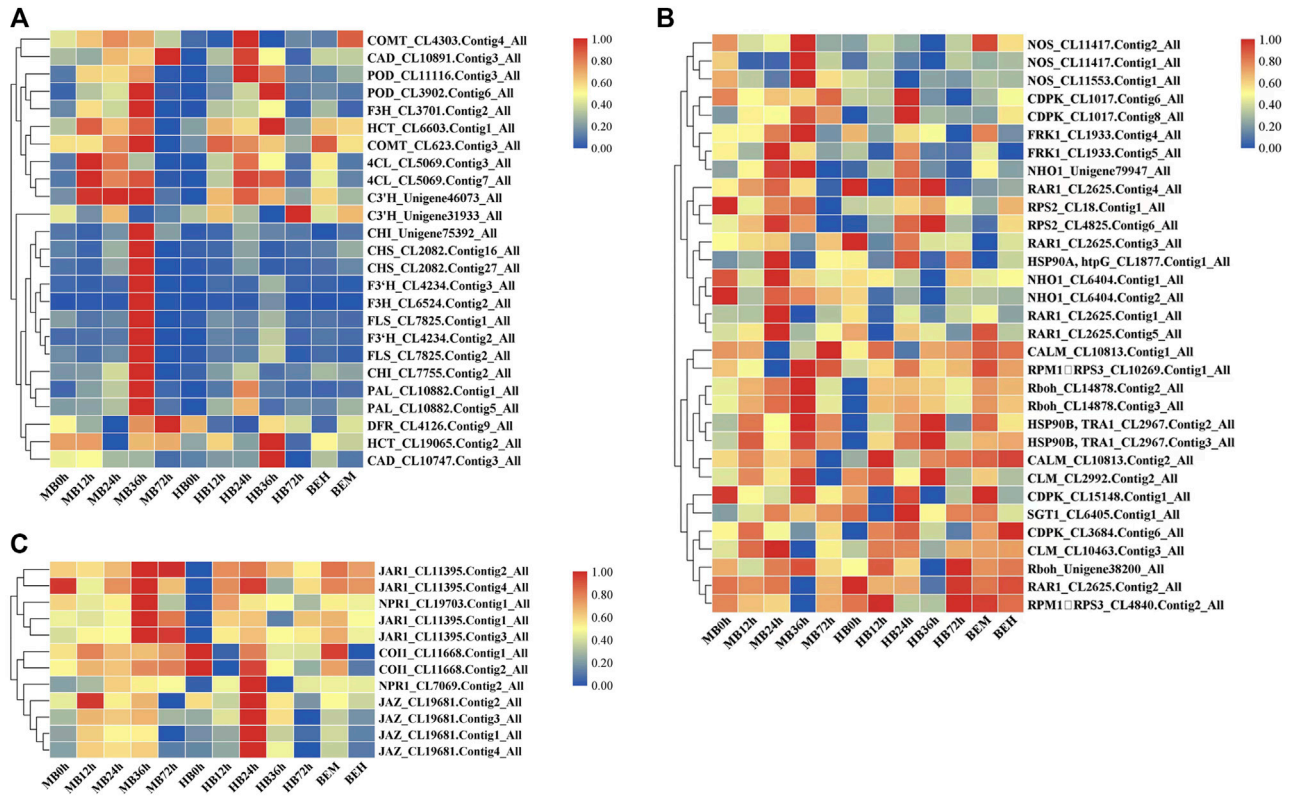


FIGURE 4 | Differential gene expression profiles involved in plant-pathogen interaction pathway, phenylpropanoid biosynthesis pathway, SA and JA signaling pathway during *B. elliptica* infection. **(A)** Gene expression profiles in plant-pathogen interaction pathway. **(B)** Gene expression profiles in phenylpropanoid biosynthesis pathway. **(C)** Gene expression profiles in SA and JA signaling pathway. MB represents plants were pre-treatment with melatonin and then inoculated with *B. elliptica*; HB represents plants were pre-treatment with Milli-Q water [ethanol/water (v/v) D 1/10,000], and then inoculated with *B. elliptica*. BEM represents plants were inoculated with *B. elliptica* for 36 h and then treated with 100 ml melatonin (2 mM/L) for 36 h; BEH represents plants were inoculated with *B. elliptica* for 36 h and then treated with 100 ml Milli-Q water [ethanol/water (v/v) D 1/10,000] for 36 h.

selected, and most of them were upregulated at 36 h after *B. elliptica* inoculation in the MB group (**Figure 4A**). Thirty-six unigenes involved in the phenylpropanoid biosynthesis pathway, including *respiratory burst oxidase* (*Rboh*), *calcium-dependent protein kinase* (*CDPK*), *nitric-oxide synthase* (*NOS*), and *LRR receptor-like serine* (*FLS2*), were found to be differentially expressed during *B. elliptica* infection in both the MB group and the HB group (**Figure 4C**). The plant hormones SA and JA play a major role in disease resistance signaling (Yang et al., 2015). We identified significant expression of the NPR1 and JAR1 genes at 36 h after *B. elliptica* inoculation in the MB group. The F-box protein *coronatine insensitive 1* (*COI1*), working together with *jasmonate zim* (*JAZ*) domain-containing transcriptional repressor proteins, is a key regulator of the JA signaling pathway (Pieterse et al., 2012). The expression of the *COI1* and *JAZ* genes was upregulated at 24 h in the HB group, and spraying exogenous melatonin on infected plants increased *COI1* and *JAZ* gene expression **Figure 4B**. Detailed information on the DEGs from the KEGG pathway enrichment analysis is listed in **Supplementary Table S6**.

3.6 Role of MAPK Signaling Cascades in the *Lilium* Defense Response Against *B. elliptica*

The findings of DEGs involved in MAPK signaling cascades prompted our interests. Plant MAPK signaling cascades play indispensable roles in plant defense against external pathogen attack. Many members of the MAPK signaling cascade pathways are related to plant disease resistance, including defense gene activation, reactive oxygen species (ROS) generation, stomatal closure, phytoalexin biosynthesis, cell wall strengthening, and hypersensitive response (HR)-related cell death (Meng and Zhang, 2013), There may be crosstalk between each MAPK signaling cascade pathway and each member. DMGs in the MAPK signaling cascade were widely involved in nearly all stages of *B. elliptica* infection. DMGs including *MEKK1*, *MKK4/5*, *MKK3*, *MKK2*, *MPK3*, *MPK4*, *MPK6*, and *MPK1/2* were enriched in the MAPK pathway and were differentially expressed at different time points (**Figure 5**). Exogenous melatonin treatment increased MEKK1 gene expression. *MKK4/5* had the highest expressional level at 24 h respond to *B. elliptica* infection in both MB group and HB group, expression

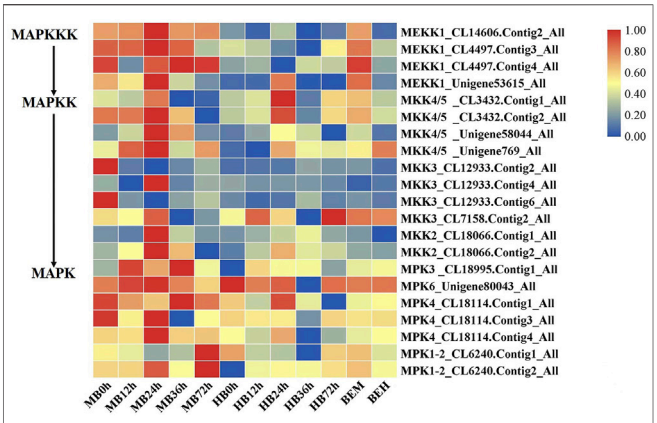


FIGURE 5 | Differential gene expression profiles involved in MAPK Signaling pathway. MB represents plants were pre-treatment with melatonin and then inoculated with *B. elliptica*; HB represents plants were pre-treatment with Milli-Q water [ethanol/water (v/v) D 1/10,000], and then inoculated with *B. elliptica*. BEM represents plants were inoculated with *B. elliptica* for 36 h and then treated with 100 ml melatonin (2 mM/L) for 36 h; BEH represents plants were inoculated with *B. elliptica* for 36 h and then treated with 100 ml Milli-Q water [ethanol/water (v/v) D 1/10,000] for 36 h.

of MKK3 showed no significant difference in HB group, but exogenous melatonin treatment improved its expression at 0 and 24 h, and exogenous melatonin treatment increased MKK2 gene expression at 24 h after *B. elliptica* infection. The expression of MPK3 was significantly higher in the MB group than that in the HB group, *MPK4*, *MPK6*, and *MPK1/2* were differentially expressed during the *B. elliptica* infection process, and exogenous melatonin treatment increased gene expression at different time, respectively.

To confirm the reproducibility of the transcriptome data, 12 defense-related DEGs associated with the MAPK signaling pathway, plant–pathogen interaction pathway and plant hormone signal transduction pathway from the RNA-seq data were evaluated via qRT-PCR at 0, 12, 24, 36, and 72 h after infection according to their FPKM value. As shown in **Supplementary Figure S3**, the results of the qRT–PCR analyses were generally in accordance with the gene expression profile data from the transcriptome during the infection process. These results indicate good correlations between the transcription profiles and the RNA-seq data.

4 DISCUSSION

4.1 Diverse Mechanisms of Plant Resistance to Pathogen Infection

Studies on the response of *Lilium* to *B. elliptica* infection have revealed that the plant defense response to pathogens is a complex biological process involving various changes at the cellular structure, physiological, biochemical and molecular levels. Studies on the molecular basis underlying resistance to *B. elliptica* have been performed, and on the basis of transcriptomic data, key genes involved in the jasmonate signaling pathway have been shown to play roles in certain plant species (*Lilium regale*) in defense against *B. elliptica*, whereas salicylic acid (SA) and ethylene (ET) were found to not be involved (Cui et al., 2018). The *Lilium* mRNA transcriptome revealed targets of miRNAs involved in metabolic processes (Gao et al., 2018), and lre-miR159a was shown to positively regulate resistance to *B. elliptica* and activate a defense response (Gao et al., 2020). In our study, the plant hormone signal transduction pathway and the phenylpropanoid and flavonoid pathways participated in the response of *Lilium* plants to pathogens at all stages of infection, according to KEGG analysis of different infection time points (**Supplementary Table S5**). This was similar to the findings reported by Nan Chai, in which Metabolic and transcriptomic analysis of *Lilium* plants infected with *B. elliptica* showed that differentially expressed genes (DEGs) and differentially accumulated metabolites (DAMs) were enriched in the phenylpropanoid and flavonoid pathways. Moreover, weighted gene coexpression network analysis (WGCNA) indicated that jasmonic acid (JA), SA, brassinolide (BR), and calcium ions (Ca^{2+}) are important for defense against *B. elliptica* in *Lilium* (Chai et al., 2021). In addition, other plant hormones, including butyric acid (BA), BRs, and CKs, also play important roles in plant defense and immune responses (Glazebrook, 2005; Yang et al., 2015; Arnao and Hernandez-Ruiz, 2018). Recently, as reported with respect to the interactions between other hosts and pathogens, interactions between melatonin and plant hormones including ET, JA, SA, and ABA have been documented during pathogen infection (Zhao et al., 2021), but more research is needed for an in-depth understanding of the crosstalk between melatonin and plant hormone signaling pathways.

4.2 MAPK Cascade Is Involved in Plant Defense Induced by Melatonin

The MAPK signaling cascade pathway participates in plant disease resistance by amplifying and transmitting foreign signals through phosphorylation cascades involving MAPK kinase kinase (MAPKKK, MEKK), MAPKK kinase (MAPKK, MEK), and MAPK proteins (Xu and Zhang, 2015; Thulasi Devendrakumar et al., 2018). There are many complex MAPK cascades involved in plant defense responses, such as *MEKK1-MKK4/MKK5-MPK3/MPK6* and *MEKK1-MKK1/MKK2-MPK4* cascades in *Arabidopsis thaliana* (Pecher et al., 2014; Peng et al., 2018), In our study, Exogenous melatonin treatment increased MEKK1 gene expression during the *B. elliptica* infection process. *MKK4/5* had the highest expressional level at 24 h respond to *B. elliptica* infection in both MB group and HB group, The expression of MPK3and MPK6 were significantly higher at 12, 24 and 36 h in MB group than that in HB group, We speculate that melatonin may play an important regulatory role via *MEKK1-MKK4/MKK5-MPK3/MPK6* in *Lilium* disease resistance to *B. elliptica*. A MAPK signaling cascade triggering melatonin-induced defense mechanisms via *OXI1/MAPKKK3–MAKK4/5/7/9–MAPK3/6* cascades has been elucidated, MAPKKK3 and oxidative signal-inducible 1 (OXI1) kinases play roles in triggering the melatonin-induced defense signaling pathway in Arabidopsis mutants,

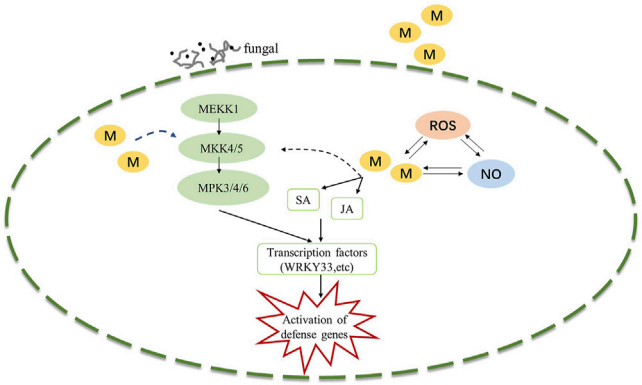


FIGURE 6 | Potential model illustrating the melatonin induced defense signaling in *Lilium* against *B. elliptica*. *B. elliptica* infection or exogenous melatonin application induced activation of MAPK cascades and NO accumulation, which in turn activates defense related genes. Abbreviation: ROS, reactive oxygen species; MAPK, Mitogen-activated protein kinase; NO, nitric oxide; SA, salicylic acid; JA, jasmonic acid; M, melatonin.

and MKK4/5/7/9-MPK3/6 cascades are responsible for melatonin-mediated innate immunity in MKK knockout Arabidopsis mutants (Lee et al., 2014; Lee and Back, 2016; Lee and Back, 2017). In our study, the DEGs involved in the MAPK pathway (*MEKK1, MKK4/5, MKK3, MKK2, MPK3, MPK4, MPK6, MPK1/2*) were differentially expressed between the MB group and the HB group according to the results of the qRT–PCR analyses and the RNA-seq data (**Supplementary Figure S2**), suggesting that melatonin was involved in the MAPK pathway and played a role in the resistance of *Lilium* to *B. elliptica*. According to the experimental results and those of previous studies, we speculate that melatonin-mediated stimulation of the MAPK kinase cascade reaction (*MAPKKK1-MAPKK4/5-MAPK3/6*) secondarily activates the transcription of the defense response (**Figure 6**). *MPK3*/*MPK6* phosphorylates *WRKY* transcription factors including *WRKY22*, *WRKY23*, *WRKY29*, *WRKY46* and *WRKY53*, which mediate the pathogen induced plant defense response (Yoo et al., 2014; Li et al., 2021; Xie et al., 2021). The study found that LrWRKYs may be important regulators involved in the biotic stress responses of lilies, the effects on plant immunity may result from the regulation of the SA-/JA-dependent signaling pathway (Cui et al., 2018; Fu et al., 2022). The mechanism of the MAPK signaling cascade-mediated plant defense response to pathogens has been gradually elucidated; this cascade mainly regulates transcriptional activation of defense genes, synthesis of plant antitoxins, cell wall thickening, hypersensitivity, stomatal closure, the production of endogenous hormones and ROS. However, the direct link between melatonin and the MAPK signaling cascades is unclear. Wei J. et al. (2018) found that the first phytomelatonin receptor (CAND2/PMTR1) in *A. thaliana*, CAND2/PMTR1, was induced by melatonin and acted on the α-subunit of heterotrimeric G proteins, which then activated NADPH oxidase (NOX) to produce H_2O_2 and promoted Ca^{2+} influx and K^+ outflow, leading to stomatal closure (Wei J. et al., 2018). Nevertheless, does melatonin depend on the CAND2/PMTR1 receptor to convert extracellular signals into intracellular signals, activating intracellular signal transmission and regulating plant biological activities ? Answers to this question have not been reported. Many MAPK components are activated by melatonin (Zhao et al., 2021), but their specific pathways need to be further elucidated, which is also one of the key research goals in the future.

In conclusion, treatment with exogenous melatonin and infection with *B. elliptica* caused the differential expression of a large number of genes in *Lilium*. KEGG analysis showed that defense-related DEGs were mainly enriched in the plant–pathogen interaction pathway, plant hormone signal transduction, MAPK signaling pathway-plant, phenylpropanoid biosynthesis, and phenylalanine metabolism after melatonin treatment. The DEGs related to the MAPK pathway were significantly different between the MB group and the HB group, suggesting that melatonin may play a role in the disease resistance of *Lilium* to *B. elliptica*. Thus, plant resistance to fungi through the MAPK signaling cascade mediated by melatonin provides a new direction for fungal disease research.

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

XX, YH, TC, and QZ conceived and designed the experiments. XX, AD, and JW performed the experiments. XX, XY, MZ, and PL analyzed the data. XX and YH wrote the manuscript. All authors have read and approved the published version of the manuscript.

FUNDING

This research was supported by Special Fund for Beijing Common Construction Project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.892674/full#supplementary-material

Supplementary Figure S1 | Analysis of gene expression trend at 5 time point in melatonin pre-treatment group **(A)** and control group **(B)**.

Supplementary Figure S2 | GO functional classification of differentially expressed genes.

Supplementary Figure S3 | The qRT-PCR result and FPKM value of gene.

Supplementary Table S1 | Quality statistics for transcriptome sequencing (RNA-seq) data.

Supplementary Table S2 | The summary statistics of the assembled transcripts and unigenes.

Supplementary Table S3 | The success rate of functional annotation in transcriptome unigenes.

Supplementary Table S4 | Statistics of DEGs for GO enrichment analysis in different comparison group.

Supplementary Table S5 | Statistics of DEGS for KEGG pathway enrichment analysis in different comparison group.

Supplementary Table S6 | The sequence information of primers for RT-qPCR.

REFERENCES

AbuQamar, S., Moustafa, K., and Tran, L. S. (2017). Mechanisms and Strategies of Plant Defense Against Botrytis Cinerea. *Crit. Rev. Biotechnol.* 37 (2), 262–274. doi:10.1080/07388551.2016.1271767

Ali, M., Tumbeh Lamin-Samu, A., Muhammad, I., Farghal, M., Khattak, A. M., Jan, I., et al. (2021). Melatonin Mitigates the Infection of Colletotrichum Gloeosporioides via Modulation of the Chitinase Gene and Antioxidant Activity in Capsicum Annuum L. *Antioxidants* 10 (1), 7. doi:10.3390/antiox10010007

Arnao, M. B., and Hernández-Ruiz, J. (2014). Melatonin: Plant Growth Regulator and/or Biostimulator During Stress? *Trends Plant Sci.* 19 (12), 789–797. doi:10.1016/j.tplants.2014.07.006

Arnao, M. B., and Hernández-Ruiz, J. (2018). Melatonin and its Relationship to Plant Hormones. *Ann. Bot.* 121 (2), 195–207. doi:10.1093/aob/mcx114

Banerjee, G., Singh, D., and Sinha, A. K. (2020). Plant Cell Cycle Regulators: Mitogen-Activated Protein Kinase, a New Regulating Switch? *Plant Sci.* 301, 110660–660. doi:10.1016/j.plantsci.2020.110660

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

Byeon, Y., and Back, K. (2014). An Increase in Melatonin in Transgenic Rice Causes Pleiotropic Phenotypes, Including Enhanced Seedling Growth, Delayed Flowering, and Low Grain Yield. *J. Pineal Res.* 56 (4), 408–414. doi:10.1111/jpi.12129

Chai, N., Xu, J., Zuo, R., Sun, Z., Cheng, Y., Sui, S., et al. (2021). Metabolic and Transcriptomic Profiling of Lilium Leaves Infected with Botrytis Elliptica Reveals Different Stages of Plant Defense Mechanisms. *Front. Plant Sci.* 12, 730620. doi:10.3389/fpls.2021.730620

Chisholm, S. T., Coaker, G., Day, B., and Staskawicz, B. J. (2006). Host-Microbe Interactions: Shaping the Evolution of the Plant Immune Response. *Cell* 124 (4), 803–814. doi:10.1016/j.cell.2006.02.008

Cook, D. E., Mesarich, C. H., and Thomma, B. P. H. J. (2015). Understanding Plant Immunity as a Surveillance System to Detect Invasion. *Annu. Rev. Phytopathol.* 53, 541–563. doi:10.1146/annurev-phyto-080614-120114

Cui, Q., Liu, Q., Gao, X., Yan, X., and Jia, G.-X. (2018). Transcriptome-Based Identification of Genes Related to Resistance against Botrytis Elliptica in Lilium Regale. *Can. J. Plant Sci.* 98 (5), 1058–1071. doi:10.1139/CJPS-2017-0254

Dodds, P. N., and Rathjen, J. P. (2010). Plant Immunity: Towards an Integrated View of Plant-Pathogen Interactions. *Nat. Rev. Genet.* 11 (8), 539–548. doi:10.1038/nrg2812

Dubbels, R., Reiter, R. J., Klenke, E., Goebel, A., Schnakenberg, E., Ehlers, C., et al. (1995). Melatonin in Edible Plants Identified by Radioimmunoassay and by High Performance Liquid Chromatography-Mass Spectrometry. *J. Pineal Res.* 18 (1), 28–31. doi:10.1111/j.1600-079x.1995.tb00136.x

Frawley, D., and Bayram, Ö. (2020). The Pheromone Response Module, a Mitogen-Activated Protein Kinase Pathway Implicated in the Regulation of Fungal Development, Secondary Metabolism and Pathogenicity. *Fungal Genet. Biol.* 144, 103469. doi:10.1016/j.fgb.2020.103469

Fu, Y., Li, J., Wu, H., Jiang, S., Zhu, Y., Liu, C., et al. (2022). Analyses of Botrytis Cinerea-Responsive LrWRKY Genes from Lilium Regale Reveal Distinct Roles of Two LrWRKY Transcription Factors in Mediating Responses to *B. cinerea*. *Plant Cell Rep.* 41 (4), 995–1012. doi:10.1007/s00299-022-02833-6

Gao, X., Cui, Q., Cao, Q. Z., Zhao, Y. Q., Liu, Q., He, H. B., et al. (2018). Evaluation of Resistance to Botrytis Elliptica in Lilium Hybrid Cultivars. *Plant Physiol. Biochem.* 123, 392–399. doi:10.1016/j.plaphy.2017.12.025

Gao, X., Zhang, Q., Zhao, Y. Q., Yang, J., He, H. B., and Jia, G. X. (2020). The lre-miR159a- LrGAMYB Pathway Mediates Resistance to Grey Mould Infection in Lilium Regale. *Mol. Plant Pathol.* 21 (6), 749–760. doi:10.1111/mpp.12923

Glazebrook, J. (2005). Contrasting Mechanisms of Defense Against Biotrophic and Necrotrophic Pathogens. *Annu. Rev. Phytopathol.* 43 (1), 205–227. doi:10.1146/annurev.phyto.43.040204.135923

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 (7), 644–652. doi:10.1038/nbt.1883

Gu, Q., Wang, C., Xiao, Q., Chen, Z., and Han, Y. (2021). Melatonin Confers Plant Cadmium Tolerance: An Update. *Int. J. Mol. Sci.* 22 (21), 11704. doi:10.3390/ijms222111704

Jones, J. D. G., and Dangl, J. L. (2006). The Plant Immune System. *Nature* 444 (7117), 323–329. doi:10.1038/nature05286

Lee, H. Y., and Back, K. (2016). Mitogen-Activated Protein Kinase Pathways are Required for Melatonin-Mediated Defense Responses in Plants. *J. Pineal Res.* 60 (3), 327–335. doi:10.1111/jpi.12314

Lee, H. Y., and Back, K. (2017). Melatonin is Required for H2O2- and NO-Mediated Defense Signaling Through MAPKKK3 and OXI1 in *Arabidopsis thaliana*. *J. Pineal Res.* 62 (2), e12379. doi:10.1111/jpi.12379

Lee, H. Y., Byeon, Y., and Back, K. (2014). Melatonin as a Signal Molecule Triggering Defense Responses Against Pathogen Attack in Arabidopsis and Tobacco. *J. Pineal Res.* 57 (3), 262–268. doi:10.1111/jpi.12165

Li, H., He, J., Yang, X., Li, X., Luo, D., Wei, C., et al. (2016). Glutathione-Dependent Induction of Local and Systemic Defense Against Oxidative Stress by Exogenous Melatonin in Cucumber (*Cucumis sativus* L.). *J. Pineal Res.* 60 (2), 206–216. doi:10.1111/jpi.12304

Li, C., He, Q., Zhang, F., Yu, J., Li, C., Zhao, T., et al. (2019). Melatonin Enhances Cotton Immunity to Verticillium Wilt via Manipulating Lignin and Gossypol Biosynthesis. *Plant J.* 100 (4), 784–800. doi:10.1111/tpj.14477

Li, N., Yang, Z., Li, J., Xie, W., Qin, X., Kang, Y., et al. (2021). Two VQ Proteins are Substrates of the OsMPKK6-OsMPK4 Cascade in Rice Defense Against Bacterial Blight. *Rice* 14 (1), 39. doi:10.1186/s12284-021-00483-y

Liu, Q., Wei, C., Zhang, M.-F., and Jia, G.-X. (2016). Evaluation of Putative Reference Genes for Quantitative Real-Time PCR Normalization in Lilium Regale During Development and Under Stress. *PeerJ* 4 (3), e1837. doi:10.7717/peerj.1837

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25 (4), 402–408. doi:10.1006/meth.2001.1262

Mandal, M. K., Suren, H., Ward, B., Boroujerdi, A., and Kousik, C. (2018). Differential Roles of Melatonin in Plant-Host Resistance and Pathogen Suppression in Cucurbits. *J. Pineal Res.* 65 (3), e12505. doi:10.1111/jpi.12505

Meng, X., and Zhang, S. (2013). MAPK Cascades in Plant Disease Resistance Signaling. *Annu. Rev. Phytopathol.* 51, 245–266. doi:10.1146/annurev-phyto-082712-102314

Nawaz, M. A., Huang, Y., Bie, Z., Ahmed, W., Reiter, R. J., Niu, M., et al. (2015). Melatonin: Current Status and Future Perspectives in Plant Science. *Front. Plant Sci.* 6, 1230. doi:10.3389/fpls.2015.01230

Pecher, P., Eschen-Lippold, L., Herklotz, S., Kuhle, K., Naumann, K., Bethke, G., et al. (2014). The *Arabidopsis thaliana* Mitogen-Activated Protein Kinases MPK3 and MPK6 Target a Subclass of 'VQ-Motif'-Containing Proteins to Regulate Immune Responses. *New Phytol.* 203 (2), 592–606. doi:10.1111/nph.12817

Peng, Y., van Wersch, R., and Zhang, Y. (2018). Convergent and Divergent Signaling in PAMP-Triggered Immunity and Effector-Triggered Immunity. *Mol. Plant Microbe Interact.* 31 (4), 403–409. doi:10.1094/mpmi-06-17-0145-cr

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 (1), 489–521. doi:10.1146/annurev-cellbio-092910-154055

Qian, Y., Tan, D.-X., Reiter, R. J., and Shi, H. (2015). Comparative Metabolomic Analysis Highlights the Involvement of Sugars and Glycerol in Melatonin-Mediated Innate Immunity Against Bacterial Pathogen in Arabidopsis. *Sci. Rep.* 5, 15815. doi:10.1038/srep15815

Shi, H., Chen, Y., Tan, D.-X., Reiter, R. J., Chan, Z., and He, C. (2015a). Melatonin Induces Nitric Oxide and the Potential Mechanisms Relate to Innate Immunity Against Bacterial Pathogen Infection in Arabidopsis. *J. Pineal Res.* 59 (1), 102–108. doi:10.1111/jpi.12244

Shi, H., Qian, Y., Tan, D. X., Reiter, R. J., and He, C. (2015b). Melatonin Induces the Transcripts of CBF/DREB1s and Their Involvement in Both Abiotic and Biotic Stresses in Arabidopsis. *J. Pineal Res.* 59 (3), 334–342. doi:10.1111/jpi.12262

Thulasi Devendrakumar, K., Li, X., and Zhang, Y. (2018). MAP Kinase Signalling: Interplays Between Plant PAMP- and Effector-Triggered Immunity. *Cell. Mol. Life Sci.* 75 (16), 2981–2989. doi:10.1007/s00018-018-2839-3

Tiwari, R. K., Lal, M. K., Kumar, R., Mangal, V., Altaf, M. A., Sharma, S., et al. (2021). Insight into Melatonin-Mediated Response and Signaling in the Regulation of Plant Defense Under Biotic Stress. *Plant Mol. Biol.*. [Epub Ahead of Print] doi:10.1007/s11103-021-01202-3

Wei, W., Li, Q.-T., Chu, Y.-N., Reiter, R. J., Yu, X.-M., Zhu, D.-H., et al. (2015). Melatonin Enhances Plant Growth and Abiotic Stress Tolerance in Soybean Plants. *J. Exp. Bot.* 66 (3), 695–707. doi:10.1093/jxb/eru392

Wei J, J., Li, D.-X., Zhang, J.-R., Shan, C., Rengel, Z., Song, Z.-B., et al. (2018). Phytomelatonin Receptor PMTR1-Mediated Signaling Regulates Stomatal Closure in *Arabidopsis thaliana*. *J. Pineal Res.* 65 (2), e12500. doi:10.1111/jpi.12500

Wei Y, Y., Chang, Y., Zeng, H., Liu, G., He, C., and Shi, H. (2018). RAV Transcription Factors Are Essential for Disease Resistance Against Cassava Bacterial Blight via Activation of Melatonin Biosynthesis Genes. *J. Pineal Res.* 64 (1), e12454. doi:10.1111/jpi.12454

Xie, W., Ke, Y., Cao, J., Wang, S., and Yuan, M. (2021). Knock Out of Transcription Factor WRKY53 Thickens Sclerenchyma Cell Walls, Confers Bacterial Blight Resistance. *Plant Physiol.* 187 (3), 1746–1761. doi:10.1093/plphys/kiab400

Xu, J., and Zhang, S. (2015). Mitogen-Activated Protein Kinase Cascades in Signaling Plant Growth and Development. *Trends Plant Sci.* 20 (1), 56–64. doi:10.1016/j.tplants.2014.10.001

Yang, Y. X., Ahammed, G. J., Wu, C., Fan, S. Y., and Zhou, Y. H. (2015). Crosstalk Among Jasmonate, Salicylate and Ethylene Signaling Pathways in Plant Disease and Immune Responses. *Curr. Protein Pept. Sci.* 16 (5), 450–461. doi:10.2174/1389203716666150330141638

Yin, L., Wang, P., Li, M., Ke, X., Li, C., Liang, D., et al. (2013). Exogenous Melatonin Improves Malus Resistance to Marssonina Apple Blotch. *J. Pineal Res.* 54 (4), 426–434. doi:10.1111/jpi.12038

Yoo, S. J., Kim, S.-H., Kim, M.-J., Ryu, C.-M., Kim, Y. C., Cho, B. H., et al. (2014). Involvement of the OsMKK4-OsMPK1 Cascade and its Downstream Transcription Factor OsWRKY53 in the Wounding Response in Rice. *Plant Pathol. J.* 30 (2), 168–177. doi:10.5423/ppj.oa.10.2013.0106

Zhang, M., and Zhang, S. (2022). Mitogen-Activated Protein Kinase Cascades in Plant Signaling. *Integr. Plant Biol.*. [Epub Ahead of Print]. doi:10.1111/jipb.13215

Zhang, S., Zheng, X., Reiter, R. J., Feng, S., Wang, Y., Liu, S., et al. (2017). Melatonin Attenuates Potato Late Blight by Disrupting Cell Growth, Stress Tolerance, Fungicide Susceptibility and Homeostasis of Gene Expression in Phytophthora Infestans. *Front. Plant Sci.* 8, 1993. doi:10.3389/fpls.2017.01993

Zhao, L., Chen, L., Gu, P., Zhan, X., Zhang, Y., Hou, C., et al. (2019). Exogenous Application of Melatonin Improves Plant Resistance to Virus Infection. *Plant Pathol.* 68 (7), 1287–1295. doi:10.1111/ppa.13057

Zhao, D., Wang, H., Chen, S., Yu, D., and Reiter, R. J. (2021). Phytomelatonin: An Emerging Regulator of Plant Biotic Stress Resistance. *Trends Plant Sci.* 26 (1), 70–82. doi:10.1016/j.tplants.2020.08.009

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 Genetics

TYPE Original Research
PUBLISHED 02 September 2022
DOI 10.3389/fgene.2022.932731

OPEN ACCESS

EDITED BY
Suxu Tan,
Qingdao University, China

REVIEWED BY
Jun Tang,
Jiangsu Academy of Agricultural Sciences, China
Ajay Kumar Mahato,
Centre for DNA Fingerprinting and Diagnostics (CDFD), India

*CORRESPONDENCE
Xiaofang Xie,
xxf317@fafu.edu.cn
Yongxian Wen,
Wen9681@sina.com

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

RECEIVED 30 April 2022
ACCEPTED 11 August 2022
PUBLISHED 02 September 2022

CITATION
Yang T, Li C, Zhang H, Wang J, Xie X and Wen Y (2022), Genome-wide identification and expression analysis of the GRAS transcription in eggplant (Solanum melongena L.). *Front. Genet.* 13:932731. doi: 10.3389/fgene.2022.932731

Genome-wide identification and expression analysis of the GRAS transcription in eggplant (*Solanum melongena* L.)

Ting Yang[1], Cheng Li[2], Hui Zhang[2], Jingyu Wang[2], Xiaofang Xie[1,3]* and Yongxian Wen[2,4]*

[1]Fujian Key Laboratory of Crop Breeding By Design, Fujian Agriculture and Forestry University, Fuzhou, China, [2]Institute of Statistics and Applications, Fujian Agriculture and Forestry University, Fuzhou, China, [3]College of Life Sciences, Fujian Agriculture & Forestry University, Fuzhou, China, [4]College of Computer and Information Science, Fujian Agriculture and Forestry University, Fuzhou, China

GRAS proteins are plant-specific transcription factors and play important roles in plant growth, development, and stress responses. In this study, a total of 48 *GRAS* genes in the eggplant (*S. melongena*) genome were identified. These genes were distributed on 11 chromosomes unevenly, with amino acid lengths ranging from 417 to 841 aa. A total of 48 GRAS proteins were divided into 13 subgroups based on the maximum likelihood (ML) model. The gene structure showed that 60.42% (29/48) of *SmGRAS*s did not contain any introns. Nine pairs of *SmGRAS* appeared to have a collinear relationship, and all of them belonged to segmental duplication. Four types of cis-acting elements, namely, light response, growth and development, hormone response, and stress response, were identified by a cis-acting element predictive analysis. The expression pattern analysis based on the RNA-seq data of eggplant indicated that *SmGRAS*s were expressed differently in various tissues and responded specifically to cold stress. In addition, five out of ten selected *SmGRAS*s (*SmGRAS2/28/32/41/44*) were upregulated under cold stress. These results provided a theoretical basis for further functional study of *GRAS* genes in eggplant.

KEYWORDS

eggplant, GRAS, phylogenetic analysis, expression pattern, cold stress

Introduction

The name of the *GRAS* gene family is derived from its first three identified members, namely, gibberellic acid insensitive (GAI) (Peng et al., 1997), repressor of GA1 (RGA) (Silverstone et al., 1998), and scarecrow (SCR) (Di Laurenzio et al., 1996). Members of the GRAS protein family have great differences in sequence length and structure. Typical GRAS proteins are generally composed of 400–770 amino acids, including a variable N-terminal sequence and a relatively conserved C-terminal sequence (Bolle, 2004). The conserved C-terminal of typical GRAS has five different sequence motifs: leucine heptad repeat I (LHRI), VHIID, leucine heptad repeat II (LHRII), PFYRE, and SAW (Pysh et al., 1999). It

was reported that GRAS proteins of rice, *Arabidopsis*, *Populus*, grape, and tomato were divided into 13 subfamilies: DELLA, HAM, LISCL, AtSCR, AtPAT1, DLT, AtSCL3, AtSCL4/7, AtLAS, AtSHR, Pt20, Os19, and Os4 based on phylogenetic analysis (Liu and Widmer, 2014).

Many studies have confirmed that GRAS proteins are important proteins in plant growth and development. For example, *AtLAS* controls the formation of the axillary meristem (Greb et al., 2003). *OsSCR* regulates asymmetric cell division (Kamiya et al., 2003). *AtSCL3* is a tissue-specific integration factor of the GA pathway, which can promote the division and elongation of *Arabidopsis* root cells (Heo et al., 2011). In addition, *GRAS* genes are also involved in plant responses to various stresses. *AtSCL14* interacts with Class II TGA to activate the detoxification system of the plant to reduce harm (Fode et al., 2008). In *Arabidopsis thaliana*, overexpression of *BnLAS* in *Brassica napus* leads to smaller stomatal opening, more wax deposition in leaves, and a lower water loss rate, indicating that *BnLAS* has potential applications in improving drought tolerance of plants (Yang, et al., 2011). In tomatoes, the transcript accumulation of *SlGRAS4* exhibited more than 250-fold change during cold stress compared to that in the control plants, which means that GRAS can respond positively to cold stress (Huang et al., 2015). In grapes, overexpression of *VaPAT1* gene, a member of the PAT1 subfamily, significantly increased plant resistance to cold by regulating jasmonic acid biosynthesis (Yuan et al., 2016; Wang et al., 2021). Torres-Galea et al. (2006) detected *AtSCL13* expression patterns in PAT1 branches of *Arabidopsis thaliana* and found that *AtSCL13* expression was induced under low temperature stress.

Eggplant (*Solanum melongena* L.) is a popular vegetable. It is cultivated all over the world, with the largest acreage in Asia. Eggplant is a kind of temperature-bias plant (Saini and Kaushik, 2019), and it is much more sensitive to low temperatures than other solanaceous vegetables (Wan et al., 2014), which is one of the main factors affecting the production of eggplant.

There are few studies on genes related to low-temperature tolerance in eggplants, which mainly focus on the analysis of genes related to low-temperature tolerance in *A. thaliana* and the mining of low-temperature tolerant genes in wild eggplant based on high-throughput sequencing technology. Wan et al. (2014) transferred *AtCOR15* and *AtCBF3* into the eggplant cultivar Sanyueqie and assessed their cold tolerance. The results showed that the expression of the exogenous *AtCBF3* and *AtCOR15A* could promote the cold adaptation process to protect eggplant plants from chilling stress (Wan et al., 2014). Zhou et al. (2020) isolated an ICE1-like gene (*SmICE1a*) from eggplant and functionally characterized its role in cold tolerance by overexpressing it in *A. thaliana*. The findings of this study indicated that the *SmICE1a* gene can be used to enhance cold tolerance in eggplant. A study using high-throughput sequencing technology to study the miRNA and its target genes of *Solanum aculeatissimum* at low temperatures obtained and verified nine significantly differentially expressed miRNAs and 12 targeted mRNAs (Yang et al., 2017). Therefore, excavating key cold resistance genes of eggplant is of significance for the cold resistance varieties' breeding. At present, the response mode of GRAS transcription factor family members of eggplant under low temperature stress is not clear.

In this study, a comprehensive investigation of the *GRAS* gene family, including physicochemical properties, gene structure, conserved motifs, phylogeny, chromosome location, collinearity, cis-acting elements, and gene expression patterns under low temperature stress, was performed based on the current eggplant genome sequence data. The information derived from this study would provide a solid foundation for the further functional investigation of *GRAS* genes in eggplant.

Materials and methods

Screening and domain identification of eggplant GRAS proteins

The protein sequences of *A. thaliana* 33) and *Oryza sativa* L. (50) GRAS members (Liu and Widmer, 2014) were downloaded from TAIR (http://www.Arabidopsis.org/) and Phytozome (http://www.phytozome.net/search.php). These protein sequences were used as the queries to identify the GRAS orthologs in eggplant using the BLASTP tool of SpudDB and the Eggplant Genome Database (http://eggplant-hq.cn). Proteins with more than 30% similarity to the query sequence and an E-value less than E^{-10} were selected. The domains for GRAS proteins were further confirmed using the Conserved Domain Database (CDD) of NCBI (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) and visualized by TBtools (Chen et al., 2020). The genes containing the GRAS domain were selected as the final candidates for eggplant *GRAS* genes and were renamed based on their physical position in the eggplant genome. The information of these genes was extracted from the eggplant database (http://eggplant-hq.cn), including gene IDs, physical position, gene sequence, and protein sequence. The molecular weight (MW), theoretical isoelectric point (pI), and protein length (PL) for the predicted SmGRAS proteins were calculated using the ExPASy online tool (http://prosite.expasy.org/) (Gasteiger et al., 2003).

Gene structure and conserved motif analysis of eggplant GRAS proteins

The *SmGRAS* gene structures were identified using the Gene Structure Display Server (http://gsds.cbi.pku.edu.cn) (Guo et al., 2007). The conserved motifs were identified using MEME software (http://meme-suite.org) (Bailey et al., 2009). The parameters were set at 10 motifs with an optimum motif width of 50–300 residues. The conserved motifs were further annotated with the CDD tool (Marchler et al., 2017).

TABLE 1 *GRAS* genes identified in eggplant.

No.	Gene name	Gene id	Location (bp)	Chr	MW (Da)	pI	PL (aa)
1	*SmGRAS1*	*Smechr0100956.1*	9096627..9597	1	74440.13	5.83	657
2	*SmGRAS2*	*Smechr0100957.1*	9100908..9588	1	74295.13	6.03	656
3	*SmGRAS3*	*Smechr0101655.1*	16353704..16356949	1	82853.66	5.64	766
4	*SmGRAS4*	*Smechr0102147.1*	23751511..23085	1	59071.26	6.58	525
5	*SmGRAS5*	*Smechr0102148.1*	23755015..23177	1	68002.17	5.1	598
6	*SmGRAS6*	*Smechr0102782.1*	44313493..44868	1	46955.22	6.29	417
7	*SmGRAS7*	*Smechr0103899.1*	101543583..101983	1	52423.91	6.2	467
8	*SmGRAS8*	*Smechr0202369.1*	68167819..68195	2	67484.03	5.04	616
9	*SmGRAS9*	*Smechr0202394.1*	68366257..68555	2	60491.63	5.58	548
10	*SmGRAS10*	*Smechr0202592.1*	70117709..70951	2	58773.7	5.74	518
11	*SmGRAS11*	*Smechr0202609.1*	70298129..70059	2	47501.12	5.46	421
12	*SmGRAS12*	*Smechr0203187.1*	74994635..74828	2	76998.66	5.25	697
13	*SmGRAS13*	*Smechr0302081.1*	80300515..80024	3	71252.27	6.26	652
14	*SmGRAS14*	*Smechr0302579.1*	86165898..86167872	3	58899.44	6.61	528
15	*SmGRAS15*	*Smechr0303662.1*	95960868..95962490	3	60658.13	5.41	541
16	*SmGRAS16*	*Smechr0401109.1*	48573914..48575963	4	55008.66	5.36	495
17	*SmGRAS17*	*Smechr0401258.1*	55717340..55718779	4	53181.96	5.96	465
18	*SmGRAS18*	*Smechr0401532.1*	64833269..64836967	4	62884.01	5.11	561
19	*SmGRAS19*	*Smechr0500353.1*	3950772..3952937	5	80739.03	5.69	720
20	*SmGRAS20*	*Smechr0500354.1*	3953971..3958015	5	81748.31	5.32	726
21	*SmGRAS21*	*Smechr0500468.1*	5367175..5374708	5	60210.61	5.49	541
22	*SmGRAS22*	*Smechr0501999.1*	71422121..71424489	5	63678.71	6.42	567
23	*SmGRAS23*	*Smechr0502680.1*	80486152..80487825	5	57594.82	5.44	501
24	*SmGRAS24*	*Smechr0502748.1*	81063283..81065782	5	52495.18	6.09	469
25	*SmGRAS25*	*Smechr0600059.1*	897922..900511	6	65616.54	5.6	579
26	*SmGRAS26*	*Smechr0601319.1*	66235763..66237871	6	65621.08	6.38	599
27	*SmGRAS27*	*Smechr0602914.1*	87323086..87326478	6	95884.91	5.88	841
28	*SmGRAS28*	*Smechr0603017.1*	88208662..88210722	6	72410.31	6.1	636
29	*SmGRAS29*	*Smechr0701498.1*	84937623..84938921	7	48011.8	5.25	433
30	*SmGRAS30*	*Smechr0701743.1*	92623237..92627075	7	60658.9	6.36	546
31	*SmGRAS31*	*Smechr0701943.1*	97245321..97247331	7	49218.94	5.53	434
32	*SmGRAS32*	*Smechr0702503.1*	104537045..104540305	7	64262.94	6.27	578
33	*SmGRAS33*	*Smechr0702636.1*	105493514..105496196	7	59959.92	5.27	538
34	*SmGRAS34*	*Smechr0702726.1*	106308917..106310185	7	47913.35	8.43	423
35	*SmGRAS35*	*Smechr0800451.1*	6768043..6770159	8	51558.29	5.45	450
36	*SmGRAS36*	*Smechr0802185.1*	82575983..82579091	8	76017.79	5.74	701
37	*SmGRAS37*	*Smechr0802313.1*	84095305..84098713	8	75327.93	5.71	680
38	*SmGRAS38*	*Smechr0900438.1*	6411669..6415271	9	86289.75	5.75	764
39	*SmGRAS39*	*Smechr0900855.1*	18793688..18795304	9	55612.27	5.08	484
40	*SmGRAS40*	*Smechr0901929.1*	79662612..79664345	9	55711.21	6.23	500
41	*SmGRAS41*	*Smechr1000504.1*	6691939..6694766	10	62131.32	5.73	556
42	*SmGRAS42*	*Smechr1001636.1*	64747643..64751256	10	86484.02	5.73	760
43	*SmGRAS43*	*Smechr1001829.1*	68156739..68160660	10	89747.97	5.92	826
44	*SmGRAS44*	*Smechr1100032.1*	470509..473623	11	59993	5.72	540
45	*SmGRAS45*	*Smechr1100163.1*	1902184..1903926	11	63608.38	4.95	581
46	*SmGRAS46*	*Smechr1100616.1*	8037435..8039120	11	55463.21	5.19	489
47	*SmGRAS47*	*Smechr1102566.1*	98372183..98374214	11	60852.15	6.65	546
48	*SmGRAS48*	*Smechr1102664.1*	99602433..99606100	11	83061.76	5.69	738

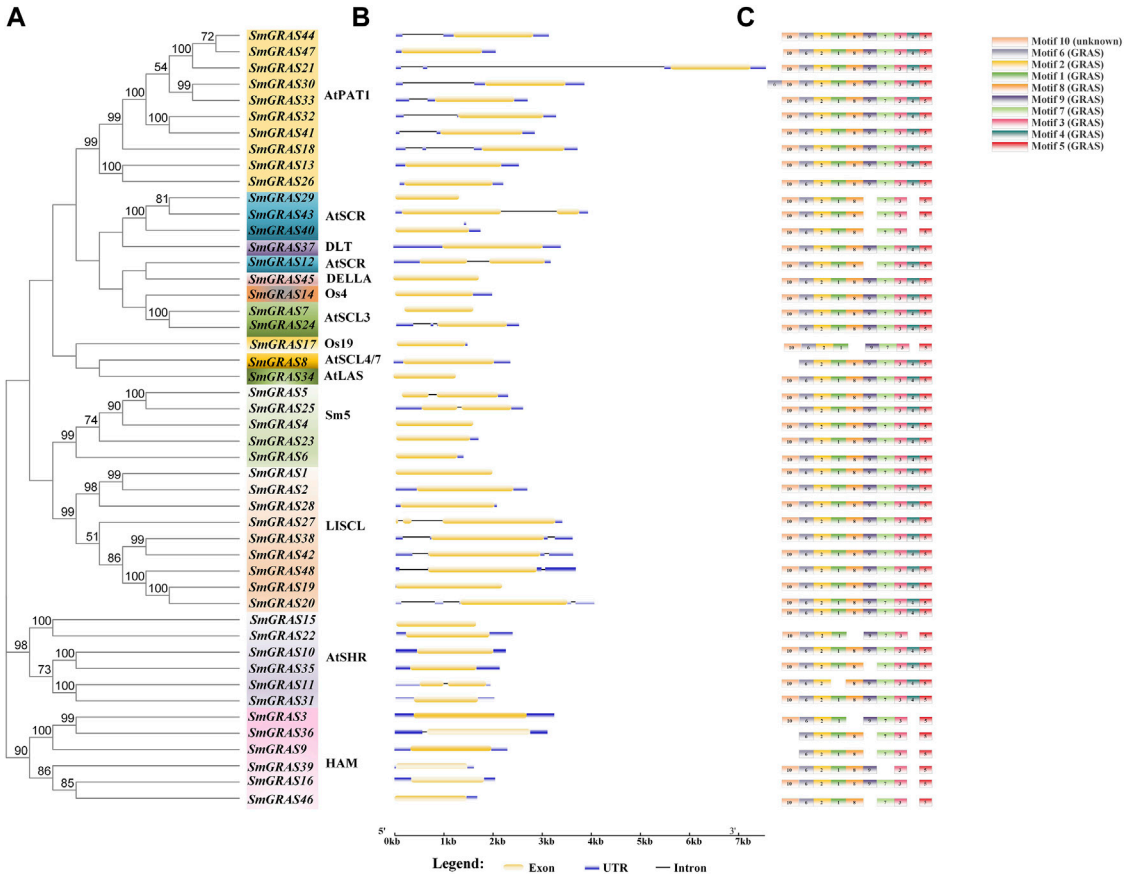


FIGURE 1
Phylogenetic tree, gene structure, and conserved motif of 48 GRAS transcription factors in eggplant. **(A)** Maximum likelihood (ML) tree generated by MEGA-X under the LG + G model with bootstrapping analysis (1,000 replicates). **(B)** Exon–intron structures of *SmGRASs* generated by the online software GSDS. **(C)** Conserved motifs of *SmGRASs* were further annotated with the CDD tool.

Phylogenetic analysis of GRAS proteins

The GRAS protein sequences of eggplant, *Arabidopsis*, and rice were aligned using the multiple sequence alignment tool ClustalX (Thompson et al., 1997). The phylogenetic tree of GRAS family proteins was generated using the MEGA-X maximum likelihood (ML) model (Kumar et al., 2018) with 1,000 bootstrap replicates. The SmGRASs were categorized based on taxonomic rules for the subfamily of *Arabidopsis* and rice GRAS protein sequences (Liu and Widmer, 2014).

Chromosomal location and collinearity analysis of *SmGRASs*

According to the annotation information of the eggplant genome downloaded from the Eggplant Genome Database, the physical positions of *SmGRASs* on chromosomes were obtained, and the chromosome mapping was performed using TBtools software (Chen et al., 2020). Collinearity of *SmGRASs* was analyzed using MCScanX software (Wang et al., 2012), and Circos (Krzywinski et al., 2009) software was used for visualization. To identify gene duplication, the parameters for the proportion of alignment and the similarity of the two sequences were set to be greater than 70%. Two duplicated genes with a distance of less than 100 kb on the same chromosome were defined as tandem duplicates (Yang et al., 2020), while any other duplicated gene pairs that did not meet the parameters of tandem duplicated genes were all defined as segmental duplicated genes.

Cis-acting element analysis of *SmGRASs*

According to the annotation information of the eggplant genome sequence, the 2 kb upstream sequence of *SmGRASs* was submitted to the PlantCARE website (http://bioinformatics.psb.ugent.be/webtools/plantcare/html) (Lescot et al., 2002) for cis-

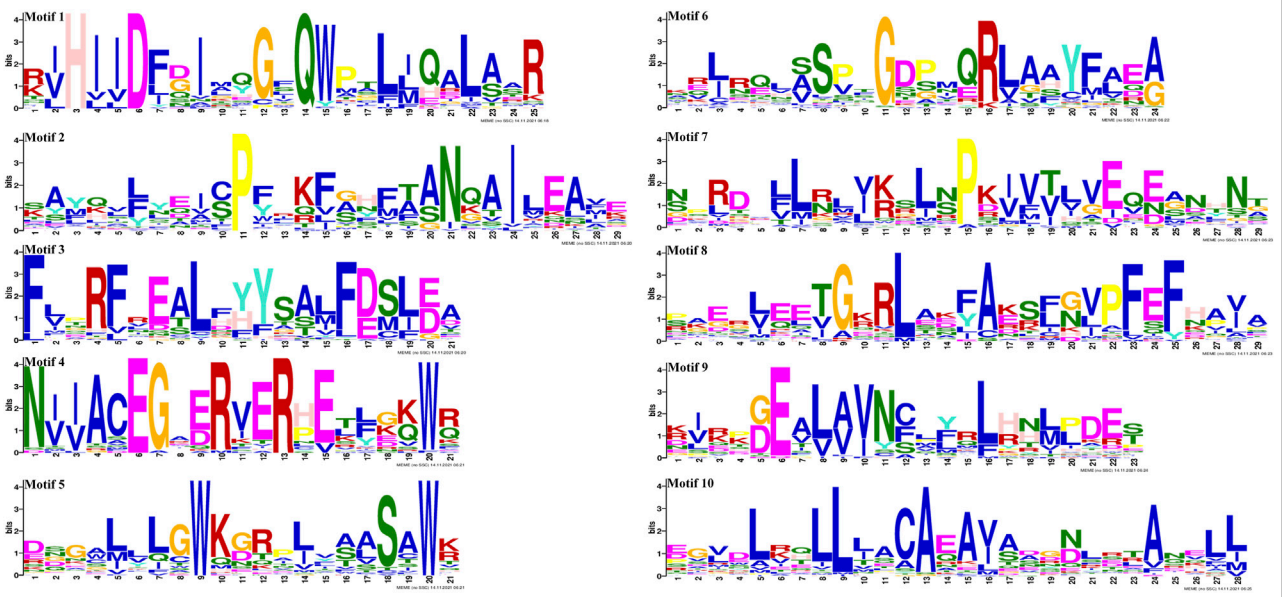


FIGURE 2
SmGRAS protein motifs confirmed using CDD of NCBI.

acting element prediction, and TBtools software (Chen et al., 2020) was used for visualization.

Expression pattern analysis of *SmGRASs*

According to the transcriptome data of eggplant (PRJNA328564 and PRJNA572318) published by NCBI, the FPKM (fragments per kilobase per million) value (log2 conversion) was used to analyze the expression of *GRAS* family genes under different tissues and low-temperature stress. Heatmaps of *SmGRAS* expression patterns were generated using the TBtools' Heatmap module (Chen et al., 2020).

Plant treatments and qRT-PCR analysis

The commercial species "Selected Purple Long Eggplant" was used for qRT-PCR. Eighteen plants were grown in a growth chamber at 28°C with a 16/8 h light/dark photoperiod until the fourth leaves were fully expanded. Then, nine plants of these seedlings were transferred to 4°C for 12 h. The seedlings grown under normal conditions were used as a control. The treated and control plantlets were collected 12 h after treatment and then stored at −80°C before RNA extraction.

The total RNA of the plantlets was extracted using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. The cDNA samples were then assessed by qRT-PCR using SYBR Premix ExTaq (Takara). Actin was used as an internal control gene. Three biological replicates (each containing three plants) and three technical replicates were measured for each treatment. The relative expression level of a gene was calculated according to the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). The primers used for qRT-PCR analysis are listed in Supplementary Table S1.

Results and analysis

Identification and physicochemical property analysis of the SmGRASs

A total of 48 SmGRASs containing GRAS domains were identified from the whole genome of eggplant and renamed from *SmGRAS1* to *SmGRAS48* based on their physical position on chromosomes (Table 1). The protein lengths of the 48 SmGRASs varied from 417 (SmGRAS6) to 841 (SmGRAS27) amino acids. The molecular weights ranged from 46,955.22 Da (SmGRAS6) to 95,884.91 Da (SmGRAS27). The theoretical isoelectric points (pI) of these SmGRASs varied from 4.95 (SmGRAS45) to 8.43 (SmGRAS34).

Gene structure and conserved motif analysis of *SmGRASs*

Gene structure analysis showed that up to 60.42% (29/48) of *SmGRASs* genes did not contain any introns (Figure 1B), and the

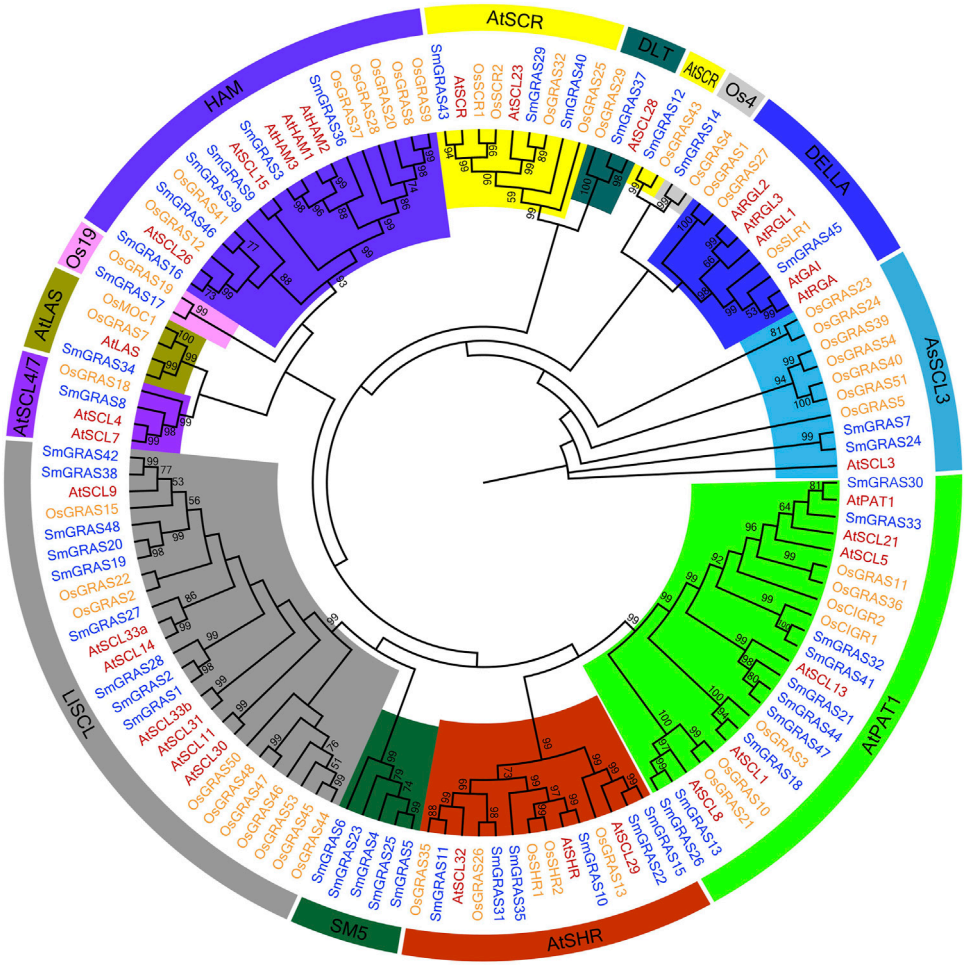


FIGURE 3
Combined phylogenetic analysis of GRAS proteins from *Arabidopsis,* rice, and eggplant. The rooted maximum likelihood tree was constructed from alignments of 131 GRAS protein sequences from *Arabidopsis* (33), rice (50), and eggplant (48) under the LG + G + F model with 1,000 bootstrap replications. Bootstrap values above 50% are shown. For each subfamily, bootstrap values are shown in different colors.

members clustered in the same subfamily had similar gene structures (Figure 1). Motif analysis (Figure 1C) indicated that members grouped in the same subfamily possessed similar conserved motif composition and sorting order, suggesting that members clustered in the same subfamily might share similar functions. All members contained motifs 1, 2, 3, 5, 6, 7, and 8, indicating that these motifs played an important role in the *GRAS* gene family. Among them, motif 5 is found in the C terminus of all *SmGRAS*s, indicating that the C terminus of the *SmGRAS* gene family is relatively conserved, which is consistent with previous studies (Pysh et al., 1999). Using the CDD tool, a total of nine motifs (motifs 1–9) were functionally annotated for the components of the conserved GRAS domain. Moreover, conserved motif sequence analysis found that motif 1 contained a VHIID fragment in the GRAS domain, and motif 5 contained a SAW fragment in the GRAS domain (Figure 2).

Phylogenetic analysis of GRAS proteins

The phylogenetic trees of GRAS proteins of *A. thaliana* (33), rice (50), and eggplant (48) were constructed by using the MEGA-X tool (Figure 3). Upon consultation with previous systematic classification rules of *Arabidopsis* and rice GRAS subfamilies (Liu and Widmer, 2014), the GRAS proteins of *Arabidopsis*, rice, and eggplant were divided into 13 subfamilies. Each subfamily contained SmGRAS members, and the number of SmGRAS members in different subfamilies varied greatly. Among them, PAT1 had the largest number of SmGRAS members (10),followed by LISCL (9). In contrast, LAS, OS4, DLT, and OS19 had the smallest subfamily, which contained only one member. The majority subfamilies contained the common members across *A. thaliana*, rice, and eggplant. However, the SmGRAS5 was a new subfamily identified

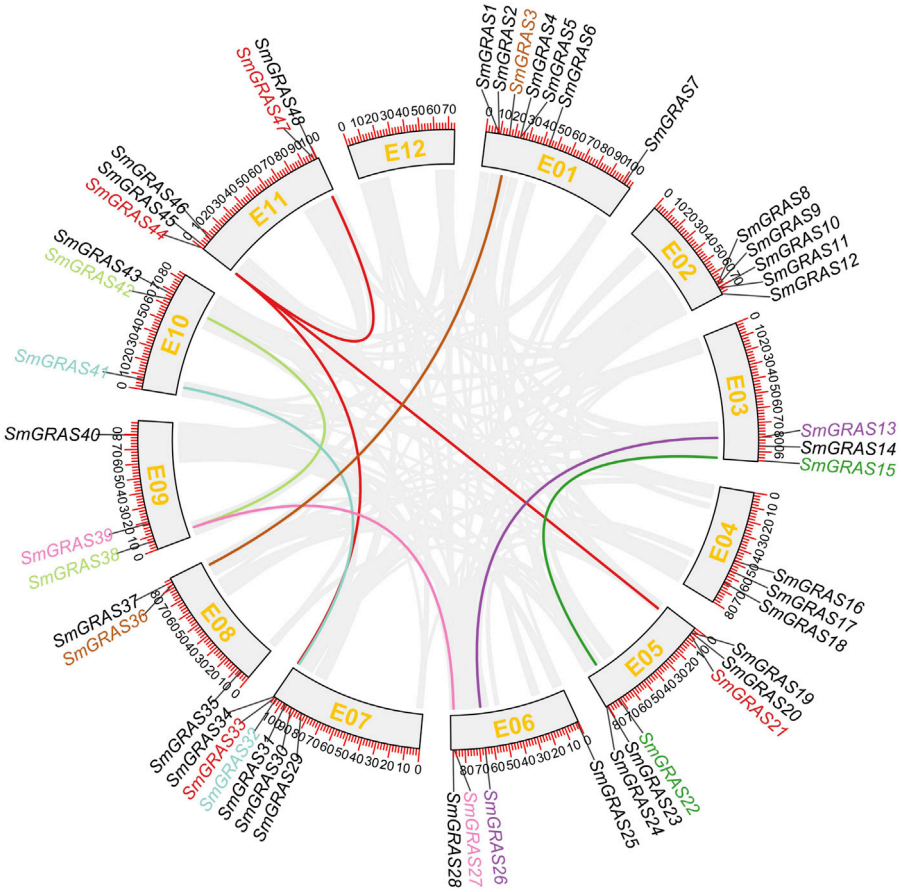


FIGURE 4
Collinearity analysis of SmGRASs. Scale bar on the chromosome shows the length of the chromosome (Mb). Genes with segmental duplication are shown in the same color.

in this study, which only contained GRAS members in eggplant. Moreover, OS4 and OS19 of GRAS members were only identified in rice and eggplant, and no members of these families were found in *Arabidopsis*. These results indicated that the GRAS protein subfamily might undergo differentiation between monocotyledons and dicotyledons to some extent.

Chromosomal location and collinearity analysis of SmGRASs

The 48 *SmGRASs* were unevenly distributed on 11 chromosomes of eggplant (Figure 4). The majority of *SmGRAS* members were distributed on chromosomes 1, 2, 5, 7, and 11, and chromosome 1 contained the largest number of *SmGRASs* (seven genes) while chromosomes 3, 4, 8, 9, and 10 all contained three SmGRAS members. Gene duplication events, including segmental duplication and tandem duplication (Cannon et al., 2004), were analyzed in this study. A total of nine *SmGRAS* gene pairs (3/36, 13/26, 15/2, 27/39, 32/41, 33/44, 30/42, 38/42, and 44/47) were identified (Figure 5). All of them were confirmed to be segmental duplication events.

Cis-acting elements of *SmGRAS*s

The 2 kb upstream CDS sequences were extracted from the promoter regions of 48 *SmGRASs*. The *cis*-acting elements in the promoter region of *SmGRASs* were predicted by the PlantCARE online tool (Figure 5). Four types of cis-acting elements were found in the promoter regions of *SmGRASs*. Each *SmGRAS* gene contains 5–19 photoresponsive cis-acting elements, which is the most abundant type. This is followed by hormone-associated cis-acting elements consisting of a combination of abscisic acid, auxin, gibberellin, methyl jasmonate, and salicylic acid responsive elements, which all *SmGRAS* members contain. However, there were few cis-acting elements related to growth and development and stress response, and the *SmGRAS* members

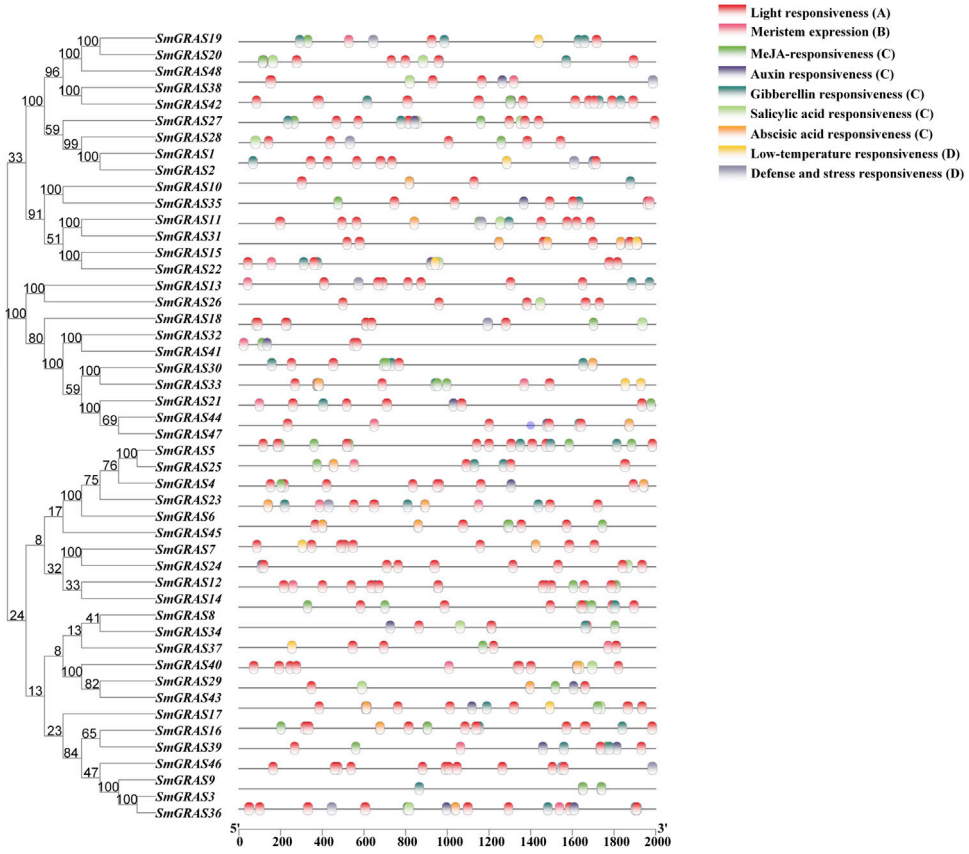


FIGURE 5
Cis-acting element analysis of *SmGRASs*. Note: **(A)** Photoresponsive cis-acting element; **(B)** growth- and development-related cis-acting element; **(C)** hormone response-related cis-acting elements; **(D)** cis-acting elements associated with stress.

containing these two types of cis-acting elements were 18 (*SmGRAS04/08/12/15/16/18/19/22/25/30/33/35/36/37/40/41/44/48*) and 22 (*SmGRAS01/05/06/09/11/15/19/20/22/25/26/27/28/31/36/37/39/40/41/43/45/48*), respectively.

Expression profiles of *SmGRASs*

To further explore the expression patterns of the *GRAS* genes, the transcript data of 16 tissues were obtained from the public genome database, including roots, stems, leaves, flowers, and fruits. A heatmap was generated based on the transcript data of 48 *SmGRAS* genes (Figure 6). As shown in Figure 6, some *SmGRASs* exhibited distinct tissue-specific expression patterns, while others were expressed throughout in the whole plant. A total of 48 genes were grouped into 3 groups (Figure 6). Group I contained 10 genes (*SmGRAS32/41/26/18/44/45/13/30/33/36*), which showed higher expression levels in the whole growth period, indicating that these genes might have a relatively important role in the whole growth stage of plants. In contrast, 24 genes in group II showed low or no expression in most tissues. A total of 14 genes (*SmGRAS1/2/7/10/29/9/20/48/3/8/24/38/21/27*) were included in group III, which were at moderate expression levels in most of the analyzed tissues.

Expression of *SmGRASs* in response to cold stress

To further investigate the expression patterns of *SmGRASs* in response to cold stress, the relative expression levels were measured based on the expression FPKM values (Figure 7). Under cold stress, the expression levels of different genes showed great variation, such as *SmGRAS2*, *SmGRAS7*, *SmGRAS8*, *SmGRAS16*, *SmGRAS24*, *SmGRAS29*, *SmGRAS32*, and *SmGRAS44*. Among them, *SmGRAS2*, *SmGRAS28*, *SmGRAS32*, *SmGRAS41*, and *SmGRAS44* showed significant upregulation, while the rest showed significant downregulation.

According to the transcriptome data (Figure 8), ten *SmGRAS* genes that exhibited significant change under cold stress were

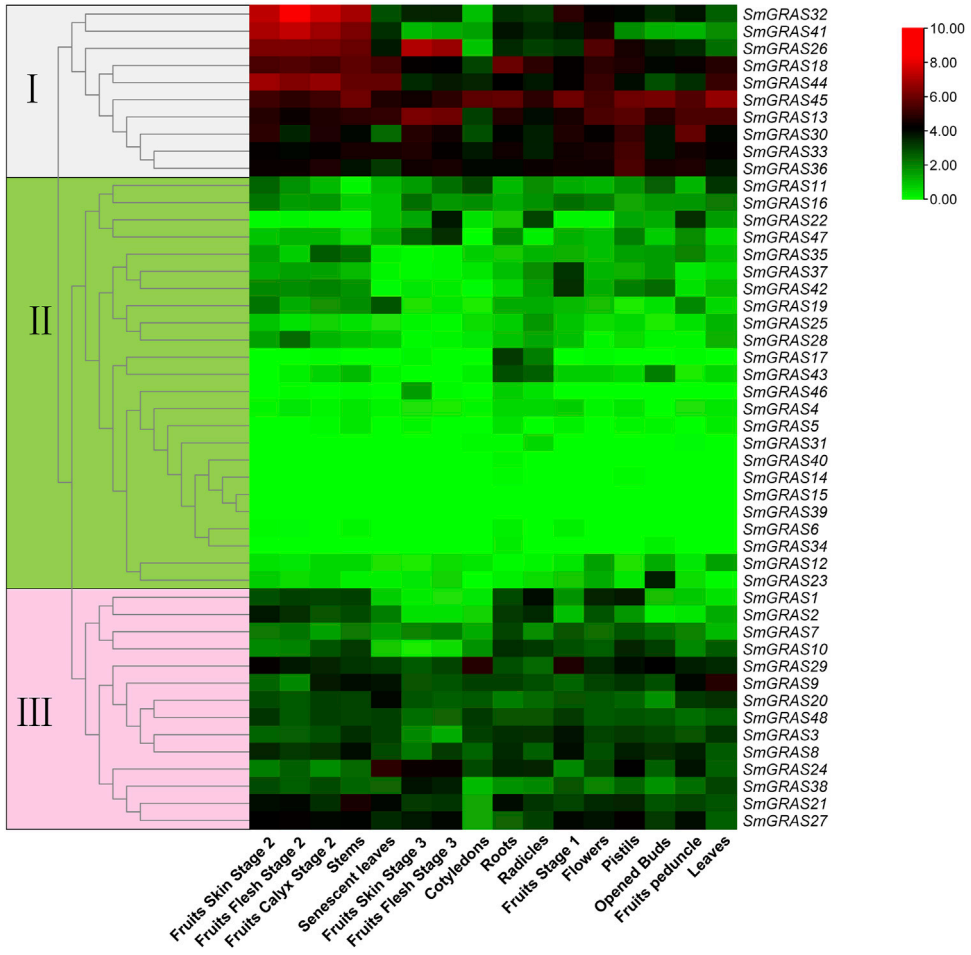


FIGURE 6
Expression analysis of Sm*GRAS*s in different tissues and different stages. The three colored boxes on the left of the picture symbol indicate the subfamily that the corresponding gene on the right belongs to.

selected for further qRT-PCR analysis (Figure 8). The results showed that five genes (*SmGRAS2*, *SmGRAS28*, *SmGRAS32*, *SmGRAS41*, and *SmGRAS44*) were upregulated and five genes (*SmGRAS7/8/16/24/29*) were downregulated. A similar gene expression pattern was found between the qRT-PCR and transcriptome analysis (Figure 7).

Discussion

In this study, a total of 48 *SmGRAS*s were identified from the eggplant genome, higher than that of *A. thaliana* (34) and lower than that of rice (60) and tomato(53) (Liu and Widmer, 2014; Huang et al., 2015). As we all know, the genome size of eggplant was 1.07 Gb (Wei et al., 2020), much higher than that of rice (420 Mb) (Goff et al., 2002), *A. thaliana* (125 Mb) (Kaul et al., 2000), and potato (844 Mb) (Diambra, 2011). Thus, it can be seen that *GRAS* gene families are widely distributed in plants, and the number of family members is independent of genome size (Xu et al., 2016), which may be due to the degree of genetic expansion between species. Collinearity analysis showed that nine *SmGRAS* gene pairs had collinearity relationships, and all of them were segmental duplication events, indicating that the amplification of eggplant *GRAS* protein was mainly dependent on segmental duplication. Meanwhile, these collinear genes were also grouped into the same group in evolutionary tree analysis, such as *SmGRAS21*, *SmGRAS33*, *SmGRAS44*,and *SmGRAS47*.

Phylogenetic tree, gene structure, and conserved motif analysis of *SmGRASs* showed that gene family members in the same group often had similar gene structures and conserved domains, but it was not excluded that some members had specificity. For example, in the HAM subgroup, only *SmGRAS36* contains an intron, which may be due to the evolution of genes within the group. In addition, the

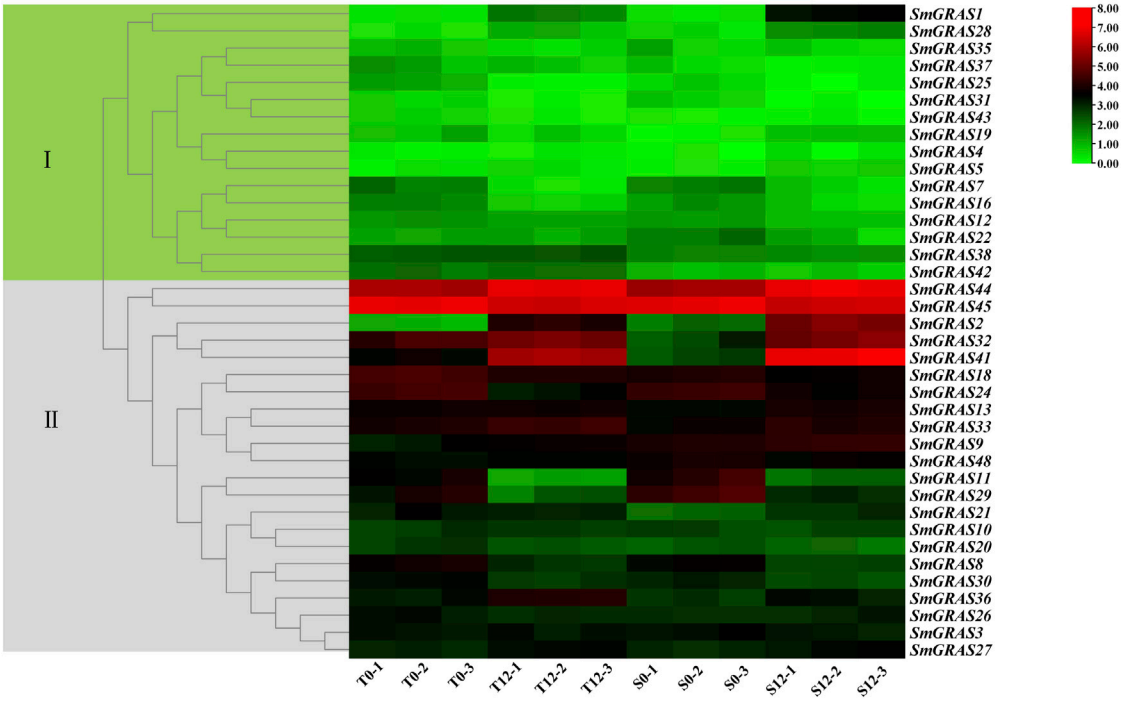


FIGURE 7
Expression changes of *SmGRASs* under cold stress. S0 and T0 represent the sensitive and tolerant varieties before low-temperature treatment, respectively. S12 and T12 represent the sensitive and tolerant varieties after low-temperature treatment, respectively. This experiment was repeated three times. The genes on the right of the image correspond to the two color-coded boxes on the left.

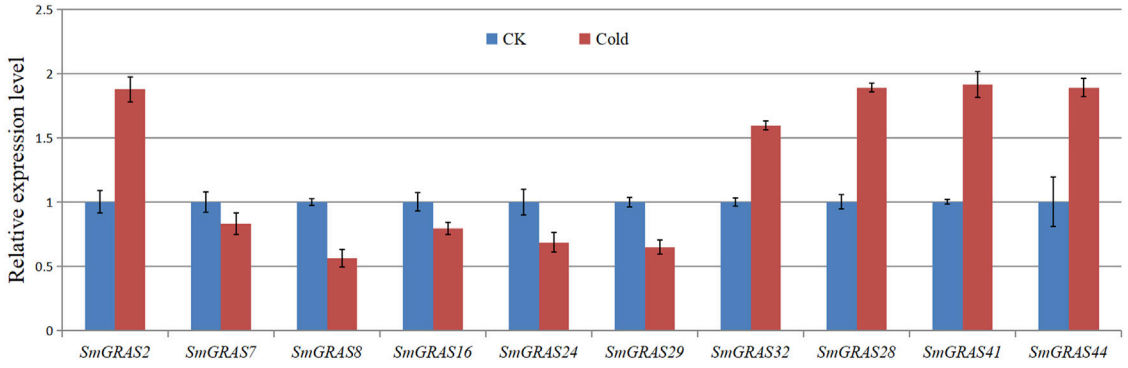


FIGURE 8
Relative expression levels of eight *SmGRASs* analyzed using qPCR in response to cold after 12 h treatment compared with that of control (CK). The bars of expression with and without hypothermia treatment are shown in red and blue, respectively. *Y*-axis represents relative expression values, and *X*-axis represents different genes. Error bars are standard deviations of three biological replicates.

exon–intron analysis showed that 60.42% of *SmGRAS* genes were intron-less (Figure 1B), with proportions of 82,2%, 77,4%, 67.6%, 55%, and 54.7% in *P. mume,* tomato, *Arabidopsis*, rice, *and Populus* (Liu and Widmer, 2014; Huang et al., 2015; Lu et al., 2015), respectively. The high percentage of intron-less genes in the *GRAS* gene family in plants implies the close evolutionary relationship of GRAS proteins.

The functional studies on *GRAS* gene families have shown that they play an important role in various abiotic stress responses (Fan et al., 2017) and are extensively involved in stress resistance, signal

transduction, and mycorrhizal formation in plants (Sun et al., 2012). This study analyzed the relative expression levels of *SmGRAS*s under low-temperature stress and found that the expression levels of 10 *SmGRASs* were significantly different after low-temperature stress compared with normal conditions. Among them, *SmGRAS2*, *SmGRAS32*, and *SmGRAS44* responded positively to low-temperature stress (Figure 8). Notably, *SmGRAS32* and *SmGRAS41* were segmentary replicate pairs (Figure 4), and their promoter regions both contained cis-acting elements related to cold stress (Figure 5), which was consistent with our qPCR results (Figure 8). The result suggested that segmentary replicate pairs may play similar functions. As we know, *SmGRAS32* and *SmGRAS44* belong to the PAT1 subfamily, indicating that members of this family can actively respond to low-temperature stress, which is consistent with the report of Yuan et al. (2016), and the genes included in this subfamily have potential applications in improving cold stress tolerance of plants. This study provides important information for further research on the function of GRAS transcription factors in eggplant under cold stress.

Conclusion

In this study, a total of 48 *SmGRASs* were identified by genome-wide analysis, and then a comprehensive analysis of the identified *SmGRASs* was conducted. The structural diversity of *SmGRASs* may reflect their functional diversity. The collinearity analysis of *SmGRASs* found nine pairs of genes in collinear relationships, and they belonged to segmental duplication. The analysis of expression patterns of *SmGRASs* showed that these genes were expressed distinctly in different tissues of eggplant, and five genes (*SmGRAS2/32/28/41/44*) were positively responsive to cold stress. Our results provide vital information for the further exploration of the functional aspects of the gene family.

Data availability statement

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

Author contributions

YW and TY conceived and designed the experiments. TY and CL performed the experiments. JW prepared the materials. TY and Hui Zhang analyzed the data. XX and YW helped to revise the manuscript. All authors read and approved the final manuscript.

Funding

This research was funded by the National Natural Science Foundation of China (grant no. 32071892), the National Natural Science Foundation of Fujian (grant no. 2021J01126), and the Science and Technology Innovation Special Foundation of Fujian Agriculture and Forestry University of China (no. CXZX2020109A).

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/fgene.2022.932731/full#supplementary-material

References

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

Bolle, C. (2004). The role of GRAS proteins in plant signal transduction and development. *Planta* 218, 683–692. doi:10.1007/s00425-004-1203-z

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

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

Di Laurenzio, L., Wysocka-Diller, J., Malamy, J. E., Pysh, L., Helariutta, Y., Freshour, G., et al. (1996). The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. *Cell* 86, 423–433. doi:10.1016/S0092-8674(00)80115-4

Diambra, L. A., Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., et al. (2011). Genome sequence and analysis of the tuber crop potato. *Nature* 475, 189–195. doi:10.1038/nature10158

Fan, S., Zhang, D., Gao, C., Zhao, M., Wu, H., Li, Y., et al. (2017). Identification, classification, and expression analysis of GRAS gene family in Malus domestica. *Front. Physiol.* 8, 253. doi:10.3389/fphys.2017.00253

Fode, B., Siemsen, T., Thurow, C., Weigel, R., and Gatz, C. (2008). The *Arabidopsis* GRAS protein SCL14 interacts with class II TGA transcription factors and is essential for the activation of stress-inducible promoters. *Plant Cell* 20, 3122–3135. doi:10.1105/tpc.108.058974

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

Goff, A., Ricke, D., Lan, T. H., Presting, G., Wang, R., Dunn, M., et al. (2002). A draft sequence of the rice genome (*Qryza sativa* L. ssp. *japonica*). *Science* 296, 92–100. doi:10.1126/science.1068275

Greb, T., Clarenz, O., Schäfer, E., Müller, D., Herrero, R., Schmitz, G., et al. (2003). Molecular analysis of the LATERAL SUPPRESSOR gene in *Arabidopsis* reveals a conserved control mechanism for axillary meristem formation. *Genes Dev.* 17, 1175–1187. doi:10.1101/gad.260703

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

Heo, J. O., Chang, K. S., Kim, I. A., Lee, M.-H., Lee, S. A., Song, S.-K., et al. (2011). Funneling of gibberellin signaling by the GRAS transcription regulator scarecrow-like 3 in the A*rabidopsis* root. *Proc. Natl. Acad. Sci. U. S. A.* 108, 2166–2171. doi:10.1073/pnas.1012215108

Huang, W., Xian, Z., Kang, X., Tang, N., and Li, Z. (2015). Genome-wide identification, phylogeny and expression analysis of GRAS gene family in tomato. *BMC Plant Biol.* 15, 209–218. doi:10.1186/s12870-015-0590-6

Kamiya, N., Itoh, J. I., Morikami, A., Nagato, Y., and Matsuoka, M. (2003). The SCARECROW gene's role in asymmetric cell divisions in rice plants. *Plant J.* 36, 45–54. doi:10.1046/j.1365-313X.2003.01856.x

Kaul, S., Koo, H. L., Jenkins, J., Rizzo, M., Rooney, T., Tallon, L. J., et al. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi:10.1038/35048692

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

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

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

Liu, X., and Widmer, A. (2014). Genome-wide comparative analysis of the GRAS gene family in *Populus, Arabidopsis* and rice. *Plant Mol. Biol. Rep.* 32, 1129–1145. doi:10.1007/s11105-014-0721-5

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 (4), 402–408. doi:10.1006/meth.2001.1262

Lu, J., Wang, T., Xu, Z., Sun, L., and Zhang, Q. (2015). Genome-wide analysis of the GRAS gene family in Prunus mume. *Mol. Genet. Genomics* 290, 303–317. doi:10.1007/s00438-014-0918-1

Marchler, B., Bo, Y., Han, L., He, J., Lanczycki C, J., Lu S, N., et al. (2017). CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 45, D200–D203. doi:10.1093/nar/gkw1129

Peng, J., Carol, P., Richards, D. E., King, K. E., Cowling, R. J., Murphy, G. P., et al. (1997). The *Arabidopsis* GAI gene defines a signaling pathway that negatively regulates gibberellin responses. *Genes Dev.* 11, 3194–3205. doi:10.1101/gad.11.23.3194

Pysh, L. D., Wysocka-Diller, J. W., Camilleri, C., Bouchez, D., and Benfey, P. N. (1999). The GRAS gene family in *Arabidopsis*: Sequence characterization and basic expression analysis of the SCARECROW-LIKE genes. *Plant J.* 18, 111–119. doi:10.1046/j.1365-313X.1999.00431.x

Saini, D. K., and Kaushik, P. (2019). Visiting eggplant from a biotechnological perspective: A review. *Sci. Hortic.* 253, 327–340. doi:10.1016/j.scienta.2019.04.042

Silverstone, A. L., Ciampaglio, C. N., and Sun, T. p. (1998). The *Arabidopsis* RGA gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. *Plant Cell* 10, 155–169. doi:10.1105/tpc.10.2.155

Sun, X., Jones, W. T., and Rikkerink, E. H. (2012). GRAS proteins: The versatile roles of intrinsically disordered proteins in plant signalling. *Biochem. J.* 442, 1–12. doi:10.1042/BJ20111766

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25 (24), 4876–4882. doi:10.1093/nar/25.24.4876

Torres-Galea, P., Huang, L. F., Chua, N. H., and Bolle, C. (2006). The GRAS protein SCL13 is a positive regulator of phytochrome-dependent red light signaling, but can also modulate phytochrome a responses. *Mol. Genet. Genomics* 276, 13–30. doi:10.1007/s00438-006-0123-y

Wan, F., Pan, Y., Li, J., Chen, X., Pan, Y., Wang, Y., et al. (2014). Heterologous expression of *Arabidopsis* C-repeat binding factor 3 (AtCBF3) and cold-regulated 15A (AtCOR15A) enhanced chilling tolerance in transgenic eggplant (*Solanum melongena* L.). *Plant Cell Rep.* 33, 1951–1961. doi:10.1007/s00299-014-1670-z

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

Wang, Z., Wong, D., Wang, Y., Xu, G., Liang, Z., Liu, Y., et al. (2021). Gras-domain transcription factor PAT1 regulates jasmonic acid biosynthesis in grape cold stress response. *Plant Physiol.* 186, 1660–1678. doi:10.1093/plphys/kiab142

Wei, Q., Wang, J., Wang, W., Hu, T., and Bao, C. (2020). A high-quality chromosome-level genome assembly reveals genetics for important traits in eggplant. *Hortic. Res.* 7, 153. doi:10.1038/s41438-020-00391-0

Xu, W., Chen, Z., Ahmed, N., Han, B., Cui, Q., and Liu, A. (2016). Genome-wide identification, evolutionary analysis, and stress responses of the GRAS gene family in castor beans. *Int. J. Mol. Sci.* 17, 1004. doi:10.3390/ijms17071004

Yang, M., Yang, Q., Fu, T., and Zhou, Y. (2011). Overexpression of the Brassica napus BnLAS gene in *Arabidopsis* affects plant development and increases drought tolerance. *Plant Cell Rep.* 30, 373–388. doi:10.1007/s00299-010-0940-7

Yang, X., Liu, F., Zhang, Y., Wang, L., and Cheng, Y. F. (2017). Cold-responsive miRNAs and their target genes in the wild eggplant species *Solanum aculeatissimum*. *Bmc Genomics* 18 (1), 1000. doi:10.1186/s12864-017-4341-y

Yang, X., Yuan, J., Luo, W., Qin, M., Yang, J., Wu, W., et al. (2020). Genome-wide identification and expression analysis of the class III peroxidase gene family in potato (*Solanum tuberosum* L.). *Front. Genet.* 11, 593577. doi:10.3389/fgene.2020.593577

Yuan, Y., Fang, L., Karungo, S., Zhang, L., Gao, Y., Li, S., et al. (2016). Overexpression of VaPAT1, a GRAS transcription factor from Vitis amurensis, confers abiotic stress tolerance in *Arabidopsis*. *Plant Cell Rep.* 35, 655–666. doi:10.1007/s00299-015-1910-x

Zhou, L., He, Y. J., Li, J., Li, L. Z., Liu, Y., and Chen, H. Y. (2020). An eggplant *SmICE1a* gene encoding MYC-type ICE1-like transcription factor enhances freezing tolerance in transgenic *Arabidopsis thaliana*. *Plant Biol.* 22, 450–458. doi:10.1111/plb.13095

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 02 September 2022
doi: 10.3389/fgene.2022.858763

Identification and Functional Prediction of Circular RNAs Related to Growth Traits and Skeletal Muscle Development in Duroc pigs

Lixia Ma[1], *Wei Chen*[1], *Shiyin Li*[1], *Ming Qin*[2] *and Yongqing Zeng*[1]*

[1]*Shandong Provincial Key Laboratory of Animal Biotechnology and Disease Control and Prevention, College of Animal Science and Technology, Shandong Agricultural University, Tai'an City, China,* [2]*Institute of Animal Science and Veterinary Medicine, Yantai Academy of Agricultural Sciences, Yantai City, China*

Porcine skeletal muscle is a highly heterogeneous tissue type, and the Longissimus Dorsi muscle (LDM), as the most economical and physiologically metabolized skeletal muscle in pigs, has always been the focus of research and improvement in pig molecular breeding. Circular RNA, as an important new member of regulatory non-coding RNA after microRNA, has become a frontier hot spot in life science research. This study aims to explore candidate circRNAs related to growth, meat quality, and skeletal muscle development among Duroc pigs with different average daily gain (ADG). Eight pigs were selected and divided into two groups: H group (high-ADG) and L group (low-ADG), followed by RNA-Seq analysis to identify circRNAs. The results showed that backfat at 6-7 rib (BF) and Intramuscular fat (IMF) content in the H group was significantly lower than L group, but ribeye area (REA) in the H group was higher than in the L group. In RNA-seq, 296 Differentially expressed (DE) circRNAs (157 upregulated and 139 downregulated) were identified and exons flanking long introns are easier to circularize to produce circRNAs. Most of the DE circRNAs were enriched in Quantitative trait locus (QTL) regions related to meat quality and growth traits. In addition, a gene can produce one or more circRNA transcripts. It was also found that the source genes of DE circRNAs were enriched in MAPK, FoXO, mTOR, PI3K-Akt, and Wnt signaling pathways. The results showed that different ADG, carcass, and meat quality traits among half-sibling Duroc pigs with the same diet may be due to the DE circRNAs related to skeletal muscle growth and development.

Keywords: circRNAs, skeletal muscle, meat quality, growth traits, Duroc pigs

OPEN ACCESS

Edited by:
Aline Silva Mello Cesar, University of São Paulo, Brazil

Reviewed by:
Krzysztof Flisikowski, Technical University of Munich, Germany
Bárbara Silva-Vignato, University of São Paulo, Brazil

***Correspondence:**
Yongqing Zeng
yqzeng@sdau.edu.cn

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *20 January 2022*
Accepted: *24 June 2022*
Published: *02 September 2022*

Citation:
Ma L, Chen W, Li S, Qin M and Zeng Y (2022) Identification and Functional Prediction of Circular RNAs Related to Growth Traits and Skeletal Muscle Development in Duroc pigs. Front. Genet. 13:858763. doi: 10.3389/fgene.2022.858763

1 INTRODUCTION

Skeletal muscle is an important tissue that plays a key role in metabolism (Das et al., 2020), abnormal growth of skeletal muscle can cause many diseases, such as Duchenne muscular dystrophy (DMD) and Facioscapulohumeral dystrophy (FSHD) (Birnkrant et al., 2018; Campbell et al., 2018), and it is directly related to growth and meat quality of livestock and poultry (Xu et al., 2017). In recent years, more and more reports have shown that non-coding RNA can regulate skeletal muscle growth and development, including microRNAs (miRNAs), long non-coding RNAs (lncRNAs) and circular RNAs (circRNAs) (Legnini et al., 2017; Li Y. et al., 2018; Yin et al., 2020). In 1976, circRNAs were first

discovered in potato tuber (Sanger et al., 1976). Firstly, circRNAs have been regarded as abnormal products because of their low expression abundance, and the fact that it exists only in a few pathogenic bacteria and does not have biological functions (Sanger et al., 1976; Kos et al., 1986). With the rapid development of high-throughput sequencing technology and bioinformatics, more and more studies have found that circRNAs are widely spread in organisms, and their functions are gradually being discovered. Compared with lncRNAs and miRNAs, circRNAs have a closed loop structure without 5′ end caps and 3′ end poly(A) tails, resisting the degradation of RNaseRand, therefore, presenting higher conservation and stability (Hansen et al., 2013; Chen et al., 2015).

Regarding the regulation of circRNAs in muscle development is still in its infancy, current researches on circRNAs in skeletal muscle include identification of circRNAs, analysis of expression patterns, and researches on the regulation mechanism of circRNAs in skeletal muscle development. The current reports found that circRNAs mainly act as molecular "sponges" of miRNAs and participate in the regulation of muscle development. The *circLMO7* regulates myoblasts differentiation and survival by sponging *miR-378a-3p* (Wei et al., 2017). The circRNA *circHIPK3* produced by the third exon of the chicken HIPK3 gene, has the highest expression level compared to other circular RNAs generated from the *HIPK3* gene. It was also differentially expressed in different stages of skeletal development. The *circHIPK3* can act as a sponge of *miR-30a-3p*, and promote cell proliferation and differentiation, binding to *MEF2C*, while *miR-3p* could inhibit the proliferation of Chicken primary myoblasts (CPMs) and repress the differentiation of CPMs by decreasing the expression of *MEF2C* (Chen et al., 2019). In addition, the *circFGFR2* can promote myoblast proliferation by sponging *miR-29b-1-5p* and *miR-133a-5p* (Chen X. et al., 2018). The *circFUT10* can act as a molecular sponge of *miR-133a* and promote cell differentiation, apoptosis and inhibit cell proliferation by upregulating the expression of *miR-133a* target genes (Li H. et al., 2018).

At present, the researches on circRNAs in the muscles of different animal species mainly focus on the comparison between animals with different genetic backgrounds, and there are relatively few comparative studies between half-siblings. This study aims to explore candidate circRNAs related to skeletal muscle growth and development among Duroc pigs with different average daily gain (ADG).

2 MATERIALS AND METHODS

2.1 Ethics Statement

All animal care and treatment procedures were conducted in strict accordance with the Animal Ethics Committee of Shandong Agricultural University, China, and performed in accordance with the Committee's guidelines and regulations (Approval No.: 2004006).

2.2 Animals

Duroc pigs come from a core breeding farm. Combined with the existing measurement data, we selected the top 30% (at least 200) of the excellent individuals from the 30–110 kg pigs, and continue the performance measurements until the animals reached 130 kg. According to the average daily gain, eight pigs were selected and divided into two groups: high-average daily gain group (774.89 g) and low-average daily gain group (658.77 g). The Longissimus Dorsi muscle (LDM) tissues were sampled and snap-frozen in liquid nitrogen for extraction of total RNA.

2.3 RNA Extraction, Strand-Specific Library Construction and Sequencing

Total RNA was extracted from LDM tissues with Trizol reagent kit (Invitrogen, Carlsbad, CA, United States) according to the instructions. RNA quality was detected on an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, United States), and checked with RNase free agarose gel electrophoresis. Then total RNA was treated with RNase R to degrade the linear RNAs, and purified with the RNeasy MinElute Cleanup kit (Qiagen, Venlo, Netherlands). Strand-specific library was constructed with VAHTS Total RNAseq (H/M/R) Library Prep Kit (Vazyme, Nanjing, China) for Illumina. In a word, ribosome RNAs were removed to retain circRNAs.

The enriched circRNAs were fragmented into short fragments with fragmentation buffer and transcribed into cDNA with random primers. Second-strand were synthesized with DNA polymerase I, RNase H, dUTP and buffer. And then, the cDNA fragments were purified with VAHTSTM DNA Clean Beads (Vazyme, Nanjing, China), end repaired, poly(A) added, and ligated to Illumina sequencing adapters. Then Uracil-N-Glycosylase (UNG) was used to digest the second-strand cDNA. The digested products were purified with VAHTSTM DNA Clean Beads, PCR amplified, and sequenced by Illumina Novaseq60000 by Gene *Denovo* Biotechnology Co. (Guangzhou, China).

2.4 Quality Control for Raw Reads (Raw Datas) and Mapping

Raw reads contained adapters or low-quality reads. Thus, raw reads were further filtered by fastp (Chen S. et al., 2018) (version 0.18.0, parameter settings: -a AGATCGGAAGAGC -q 20 -u 50 -n 15 -l 50 -w 1) to get high quality clean reads. The quality control standards were as follows: 1) removing adapters reads; 2) removing reads containing more than 10% of unknown nucleotides (N); 3) removing low quality reads containing more than 50% of low quality (Q-value≤20) bases. Bowtie2 (Langmead and Salzberg., 2012) (version 2.2.8, parameter settings: -local -p 4 --mm) software was used for mapping reads to ribosome RNA (rRNA) database. The rRNA removed reads of each sample were then mapped to reference genome by TopHat2 (Kim et al., 2015) (version 2.1.1, parameter settings: -rna- strandness RF -p 4 -q -t --dta --new-summary --mm), respectively.

2.5 Identification of circRNAs and circRNAs Statistics

The 20mers from both ends of the unmapped reads were extracted and aligned to the reference genome to find unique anchor positions within splice site. Anchor reads that aligned in the reversed orientation (head-to-tail) indicated circRNA splicing and then were subjected to find_circ (Memczak et al., 2013) (version 1, parameter settings: -r <sample.txt> -G <genome.fa> -p novel_ -s <stats.file>) to identify circRNAs. The anchor alignments were then extended such that the complete read aligns and the breakpoints were flanked by GU/AG splice sites. A candidate circRNA was called if it was supported by at least two unique back spliced reads at least in one sample.

The identified circRNAs were analyzed including type, chromosome distribution and length distribution by find_circ (version 1) and the annotation information of the reference genome.

2.6 Analysis of differentially expressed circRNAs

Differentially expressed (DE) circRNAs across group were identified by edgeR package (version 3.12.1) (http://www.r-project.org/) (Robinson et al., 2010). Identified DE circRNAs with an absolute fold change ≥1.5 and a p value < 0.05 were considered significant DE circRNAs.

2.7 Functional Enrichment Analysis of Source Gene

Source gene is the origin gene of a circRNA, and the functional analysis of source genes were studied to know about the main functions of circRNAs source genes.

2.7.1 GO Enrichment Analysis

Gene Ontology (GO) is an international standardized gene functional classification system, which includes three parts: cellular component, biological process and molecular function. On the one hand, the GO function analysis includes the GO function classification annotation of genes; on the other hand, it includes GO functional significant enrichment analysis of genes. Firstly, all source genes were mapped to GO terms in the Gene Ontology database (http://www.geneontology.org/), gene numbers were calculated for every term, significantly enriched GO terms in source genes comparing to the genome background were defined by hypergeometric test. The calculating formula of p-value is as follows:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

N: the number of all genes with GO annotation; n: the number of source genes in N; M: the number of all genes that are annotated to the certain GO terms; m: the number of source genes in M. The calculated p-value were gone through FDR Correction, taking FDR ≤ 0.05 as a threshold. GO terms meeting this condition were defined as significantly enriched GO terms in source genes.

2.7.2 KEGG enrichment analysis

The analysis of Pathway helps to further know about the biological functions of genes. KEGG is the main public database about Pathway (Kanehisa et al., 2008). Pathway enrichment analysis identified significantly enriched metabolic pathways or signal transduction pathways in source genes comparing with the whole genome background. The calculating formula is the same as that in GO analysis:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

N: the number of all genes that with KEGG annotation; n: the nuber of source genes in N; M: the number all genes annotated to specific pathways; m: the number of source genes in M. The calculated p-value was gone through FDR Correction, taking FDR ≤ 0.05 as a threshold. Pathways meeting this condition were defined as significantly enriched pathways in source genes.

2.8 Quantitative trait locus analysis

Basic Local Alignment Search Tool (blast) comparison was performed on the QTL locus (<2Mb) with high confidence related to pig growth and meat quality traits. If the transcript or QTL interval is more than 50% duplicated, the transcript is located on this QTL. Genome information can be downloaded from AnimalQTLdb (PigQTLdb: http://www.animalgenome.org-/QTLdb/pig.html) Database.

2.9 Target Relationship Prediction

For the circRNAs included in the circBase database, the StarBase (Li et al., 2014) database provides the targeting relationship between circRNAs and miRNAs. Therefore, the target relationship between existing circRNAs and miRNAs can be found through the StarBase database. There may be some unreported targeting relationships between all circRNAs and all miRNAs of pigs. Therefore, the analysis will also predict the target relationship between all circRNAs and miRNAs of pigs. RNAhybrid + svm_light, Miranda, and TargetScan were used to predict the target relationship (Qiao et al., 2016).

2.10 Quantitative real-time PCR analysis

The total RNA was reverse-transcribed into cDNA by PrimeScript RT reagent kit (TaKaRa, Dalian, China), and the analyzed by SYBR® Green Pro Taq HS Premix (Accurate Biotechnology (Hunan) Co., Ltd., Changsha, China). Primers were compounded from Sangon Biotech (Shanghai, China), and sequences of primers are shown in **Table 1**. β-actin was used as a housekeeping gene. The fold change in expression was the obtained by $2^{-\Delta\Delta CT}$ method, $\Delta\Delta CT = (CT_{\text{Target gene}} - CT_{\beta\text{-actin}})_{\text{H group}} - (CT_{\text{Target gene}} - CT_{\beta\text{-actin}})_{\text{L group}}$.

TABLE 1 | Primers sequences.

Gene	Sequence (5′-3′)
novel_circ_008472-F	CACCAGATGCCTACTCTGTTACTT
novel_circ_008472-R	GCCTTTGTTGCTCCTTCTATGATC
novel_circ_010066-F	GGATCAAACCCCACTGGACAT
novel_circ_010066-R	AGTACATTGTGCCTGGTAGATTCA
novel_circ_008433-F	AGACATGCACATCCAGATCACAGA
novel_circ_008433-R	CAGGAACACAACACCACGCTG
novel_circ_005672-F	GTGTCGGGAAGGTGAACTTGT
novel_circ_005672-R	TCTGAGGTTTCTTGTGCTCTTGG
novel_circ_012322-F	TGCCAGAAGGTCTTGCCATAG
novel_circ_012322-R	GGCTACATTCAGACACATTGCTT

TABLE 2 | Meat quality and carcass traits description.

Traits	H group	L group
Carcass weight/ kg	97.57 ± 1.33	96.45 ± 2.47
BF/ mm	30.66 ± 3.22^{b}	35.68 ± 4.23^{a}
REA/ cm^2	59.86 ± 3.61^{a}	50.42 ± 1.61^{b}
IMF content/%	4.30 ± 0.27 ^{b}	4.80 ± 0.08^{a}

Note: $^{a, b}$ Those with the different letters in each line showed significant difference (p < 0.05). H group represents High-average daily gain (ADG) group, L group represents L-average daily gain (ADG) group. BF, represents backfat at 6-7 rib, REA, represents ribeye area, and IMF, represents Intramuscular fat.

2.11 Meat Quality and Carcass Traits Description

Backfat at 6-7 rib (BF): hang the right carcass upside down and measure the subcutaneous fat thickness at the 6-7 rib of the carcass with a vernier caliper. unit: mm.

Ribeye area (REA): The cross-sectional area of the Longissimus Dorsi at the junction of the thoracic and lumbar vertebrae. The carcass was laid flat and the height and width of the cross-section area were measured by vernier calipers. The formula was as follows:

$$\text{REA}\ (\text{cm}^2) = \text{height}\ (\text{cm}) \times \text{width}\ (\text{cm}) \times 0.7$$

Detailed experimental methods for Intramuscular Fat (IMF) analysis are described in-depth by Holman et al. (Holman et al., 2019).

3 RESULTS

3.1 Meat Quality and Carcass Traits Description

As shown in **Table 2**, there was no difference in carcass weight concerning the two groups of AVG (H and L). For BF, the H group was significantly lower than the L group ($p < 0.05$), and for REA, the H group was higher than the L group ($p < 0.05$). In addition, IMF content in H group was significantly lower than L group.

3.2 Overview of Sequencing Data and Quality Assessment

The eight libraries in this study were filtered to remove the reads containing adapters (about 0.19%) and low-quality reads (about 1.13%). Finally, about 98.68% of the raw data reached the quality control standard as clean data (**Supplementary Table S1**). Through ribosome alignment, it was found that only 52.41% (49.75%–55.05%) of the data in the eight samples were available for subsequent transcriptome analysis (**Supplementary Table S2**). With the alignment in the reference Sus scrofa genome (Ensembl-release104), on average about 84.10% of the clean reads were mapped to the reference, about 77.60% of them were considered Unique Mapped reads, and only 6.5% were Multiple Mapped reads (**Supplementary Table S3**).

3.3 Identification of circRNAs

To identify circRNAs related to skeletal muscle development, we identified the expression of circRNAs by removing the ribosomal RNA (rRNA) RNA-seq. A total of 13974 circRNAs were identified (**Supplementary Table S10**), and 9066 circRNAs were co-expressed in the two groups (**Figure 1A**). The length of circRNAs was less than 1600bp and most of the circRNAs were from exon regions (**Figures 1B,C**), in addition, circRNAs were distributed on every chromosome (**Figure 1D**).

3.4 Analysis of Genomic Characteristics on circRNAs

The average number of circRNAs exon was 3.91, the average length of the open reading frame (ORF) was 23 amino acids (aa), and the length of the transcript was 1665.73nt (**Figures 2A–C**), which were lower than mRNA, but the expression levels of circRNAs were higher than mRNAs.

3.5 Analysis of Source Genes and Molecular Characteristics of circRNAs

To explore the source of circRNA transcription, the source genes of 13974 circRNAs transcripts were analyzed. As shown in **Figure 3A**, a total of 608 circRNAs did not detect their source genes, and the remaining 13366 circRNAs transcripts were generated from 4,466 genes. The majority of genes 1935) produce one circRNA transcript, and only 265 genes could produce ten or more circRNA transcripts.

To explore the molecular characteristics of circRNAs, the 5 kb flanking sequence of pig circRNAs upstream and downstream was analyzed. The results showed that most of the flanking sequences of circularized exons were introns. About 55.25% of 5′ sequences upstream of circRNAs contained introns, and 55.28% of 3′ sequences downstream of circRNAs contained introns. In addition, 29% of 5′ flanking sequences and 27.66% of the 3′ flanking sequences contained CDS (Coding Sequence), respectively, and the remaining flanking sequences were intergenic and UTR (Untranslated Region) (**Figure 3C**).

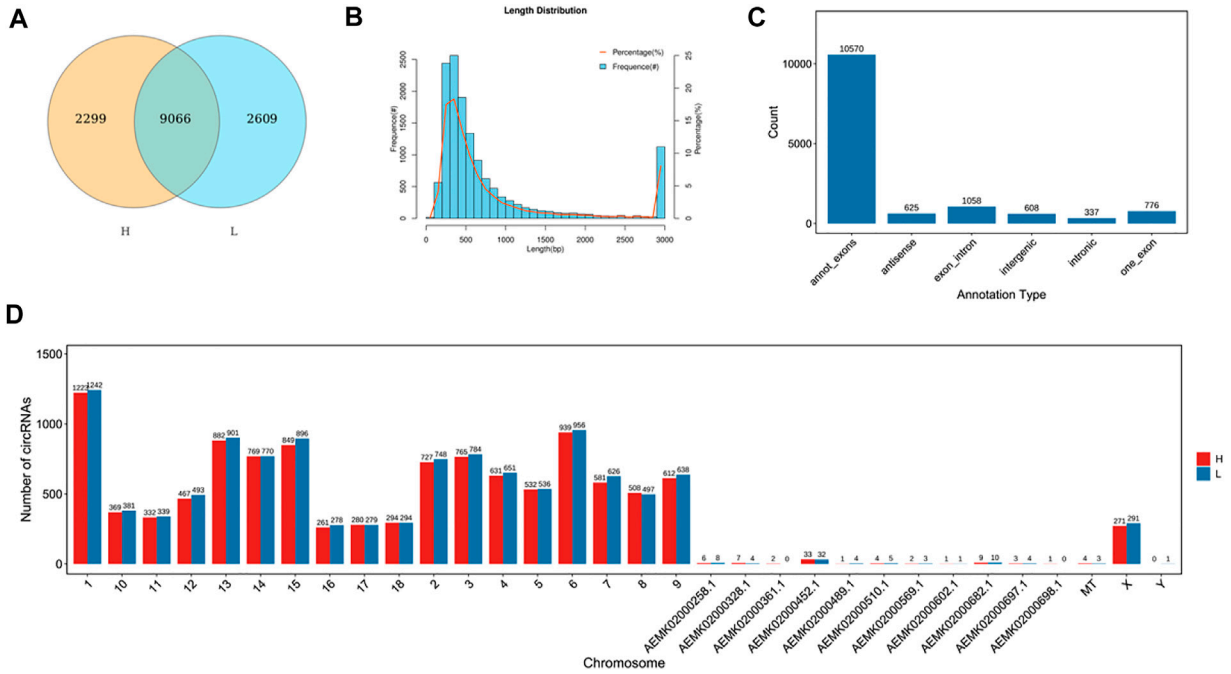


FIGURE 1 | Identification of circRNAs. **(A)** The number of identified circRNAs. Yellow represents H group, blue represents L group. The overlap represents the number of genes shared between the two groups, and non-overlap represents genes unique to each group. **(B)** The length of circRNAs. The abscissa represents the length of circRNAs, the ordinate represents frequency. **(C)** The location distribution of circRNAs. The abscissa represents the annotation type, the ordinate represents the number of circRNAs. **(D)** The chromosome distribution of identified lncRNAs. The abscissa represents the chromosome, the ordinate represents the number of circRNAs. Red represents H group, blue represents L group.

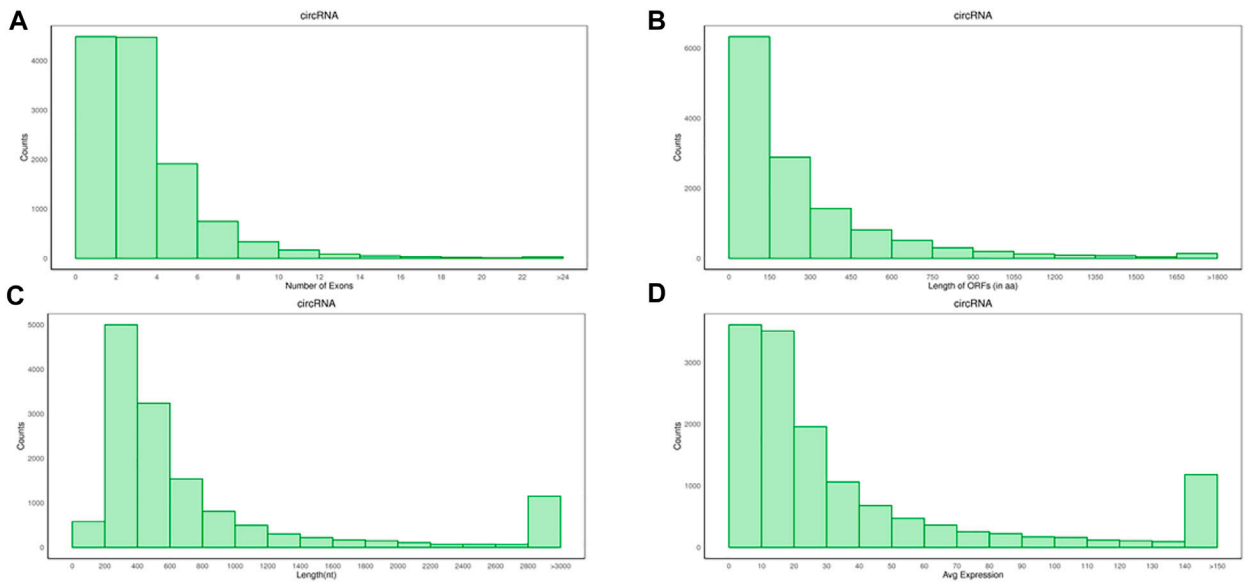


FIGURE 2 | Analysis of genomic characteristics on circRNAs. **(A)** Number of circRNAs exons. **(B)** Length of circRNAs ORF. **(C)** Transcript length of circRNAs. **(D)** The expression levels of circRNAs.

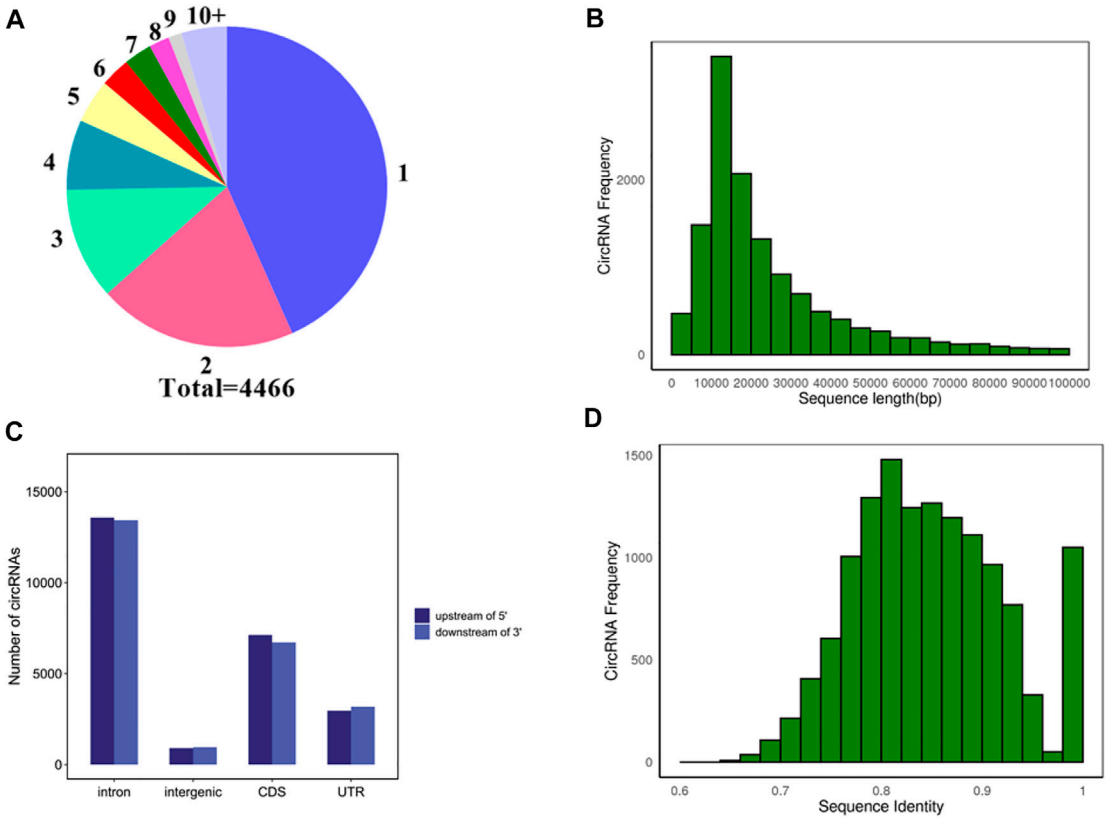


FIGURE 3 | Analysis of source genes and molecular characteristics of circRNAs. **(A)** Statistics of source genes. Different numbers represent different numbers of circRNA transcripts. 1, 2, 3. . .10+ represent one, two, three. . .and 10+ circRNA transcripts, respectively. Different colors represent the different number of source genes. For instance, blue represents the number of source genes that produce one circRNA transcript. Total 4,466 represents a total of 4,466 source genes. **(B)** Distribution of flanking intron sequences lengths of circRNAs. **(C)** Flanking sequences of a circRNA 5′ end upstream and 3′ end downstream. **(D)** Distribution of sequences identified on each intron pair that are flanking cricRNAs.

Statistics on the length of introns in the flanking sequences of circRNA found that most of the introns in the flanking sequence were longer than 5000 bp, indicating that the flanking sequence of pig circRNAs contains longer introns (**Figure 3B**). The basic Local Alignment Search Tool (BLAST) was used to align the introns of each intron pair that flanked circRNAs. Most of circRNAs shared reverse complementary matches (RCMs) of at least 200 bp (**Figure 3D**).

In addition, 296 DE circRNAs were identified, including 157 upregulated circRNAs and 139 downregulated circRNAs (**Figures 4A,B**, **Supplementary Table S4**). Among them, 28 circRNAs related to muscle development were identified (14 upregulated and 14 downregulated) (**Figure 4C**). Hierarchical clustering showed that the expressions of circRNAs were distinguishable between H group and L group, indicating a significant difference between H group and L group.

To verify RNA-seq results, five DE circRNAs were selected to perform by qRT-PCR. The results showed that qRT-PCR results were in agreement with those in RNA-seq (**Figure 4D**). The results indicated that circRNAs identified from RNA-seq were reliable.

3.6 GO and KEGG Analysis of Source Genes of DE circRNAs

To understand the functions of DE circRNAs, GO and KEGG analysis were performed. The results showed that the source genes of DE circRNAs were significantly enriched in regulation of DNA recombination, blood vessel development, response to oxidative stress, mitotic spindle organization, programmed cell death and other GO terms (**Figure 5A**, **Supplementary Table S5**). The source genes of DE circRNAs were significantly enriched in pathways including MAPK, FoXO, mTOR, PI3K-Akt and Wnt signaling pathways (**Figure 5B**, **Supplementary Table S6**).

The DE circRNAs related to muscle development were significantly enriched in regulation of endothelial cell proliferation, calcium ion transport, regulation of skeletal muscle contraction by calcium ion signaling, transition between fast and slow fiber, myofibril assembly, muscle cell differentiation and other GO terms (**Figure 6A**, **Supplementary Table S7**), and they were also enriched in Pathways in cancer, EGFR tyrosine kinase inhibitor resistance, PPAR signaling pathway, Insulin signaling pathway, cAMP signaling pathway, FoxO signaling pathway, PI3K-Akt signaling pathway, AMPK

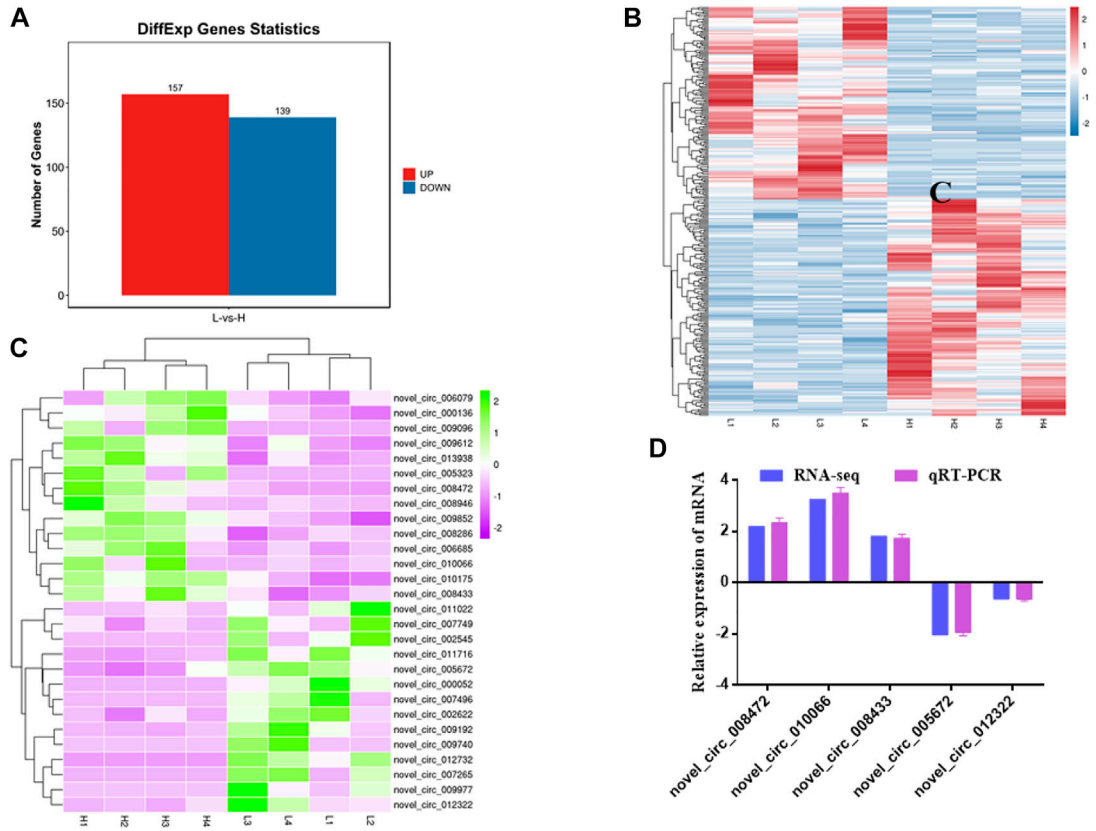


FIGURE 4 | Statistics and heat map analyses of DE circRNAs. **(A)** Statistics of DE circRNAs. Red represents up-regulation, blue represents down-regulation. **(B)** Heat map analysis of DE circRNAs. Each column represents a sample and each row represents a circRNA. Red and blue gradients indicate increase and decrease in gene expression abundance, respectively. **(C)** Heat map analysis of DE circRNAs related to skeletal muscle development. Each column represents a sample and each row represents a circRNA. Purple and green gradients indicate increase and decrease in gene expression abundance, respectively. **(D)** Validation of DE circRNAs. Blue represents RNA-seq result, purple represents qRT-PCR result.

signaling pathway, Jak-STAT signaling pathway and other pathways (**Figure 6B**, **Supplementary Table S8**).

3.7 Expression Analysis of Growth Traits and Meat Quality-Relevant QTLs

According to the QTL mapping analysis, 256 DE circRNAs source genes were enriched in QTLs related to growth and meat quality in pigs (**Table 3**, **Supplementary Table S9**). The results showed that DE circRNAs were closely related to the growth traits and meat quality. For example, circRNA ENSSSCG00000049158 was enriched in Drip loss QTL, Water holding capacity QTL, Average daily gain QTL and other QTLs. circRNA ENSSSCG00000004203 was enriched in Backfat at last rib QTL, Shoulder subcutaneous fat thickness QTL, Backfat at last rib QTL, Percentage type IIa fibers QTL, PH for Longissimus Dorsi QTL and Leaf fat weight QTL.

3.8 Analysis of Target Relationship Prediction

In this study, 252534 miRNA-circRNA pairs were identified between 453 miRNAs and 13836 circRNAs.

4 DISCUSSION

Average daily gain (ADG) is considered an important growth trait, it can directly affect the economic benefits of the pig producers. Growing fast and having a high lean meat rate is the best ideal state. In this study, the results showed that H group not only grew faster than L group, but also had less BF and IMF content than L group. Therefore, pigs in H group will be preferred by more pig producers. To explore the phenotype differences between the H and L groups, RNA-seq was used to identify circRNAs.

Due to the imperfect methods of circRNAs research and identification, the functional identification of circRNAs is still in its infancy. In recent years, the development of sequencing technology and bioinformatics have provided technical support for the study of circRNAs from the whole genome.

Genome-wide methods for circRNAs identification mainly include gene chip technology and RNA-seq technology. RNA-seq is widely used in circRNAs research in both model organisms and non-model organisms. At present, RNA-seq has been used in many species, such as mice (Jakobi et al., 2016), cattle (Wei et al., 2017), pigs (Liang et al., 2017) and sheep (Li et al., 2017), with high reliability and feasibility.

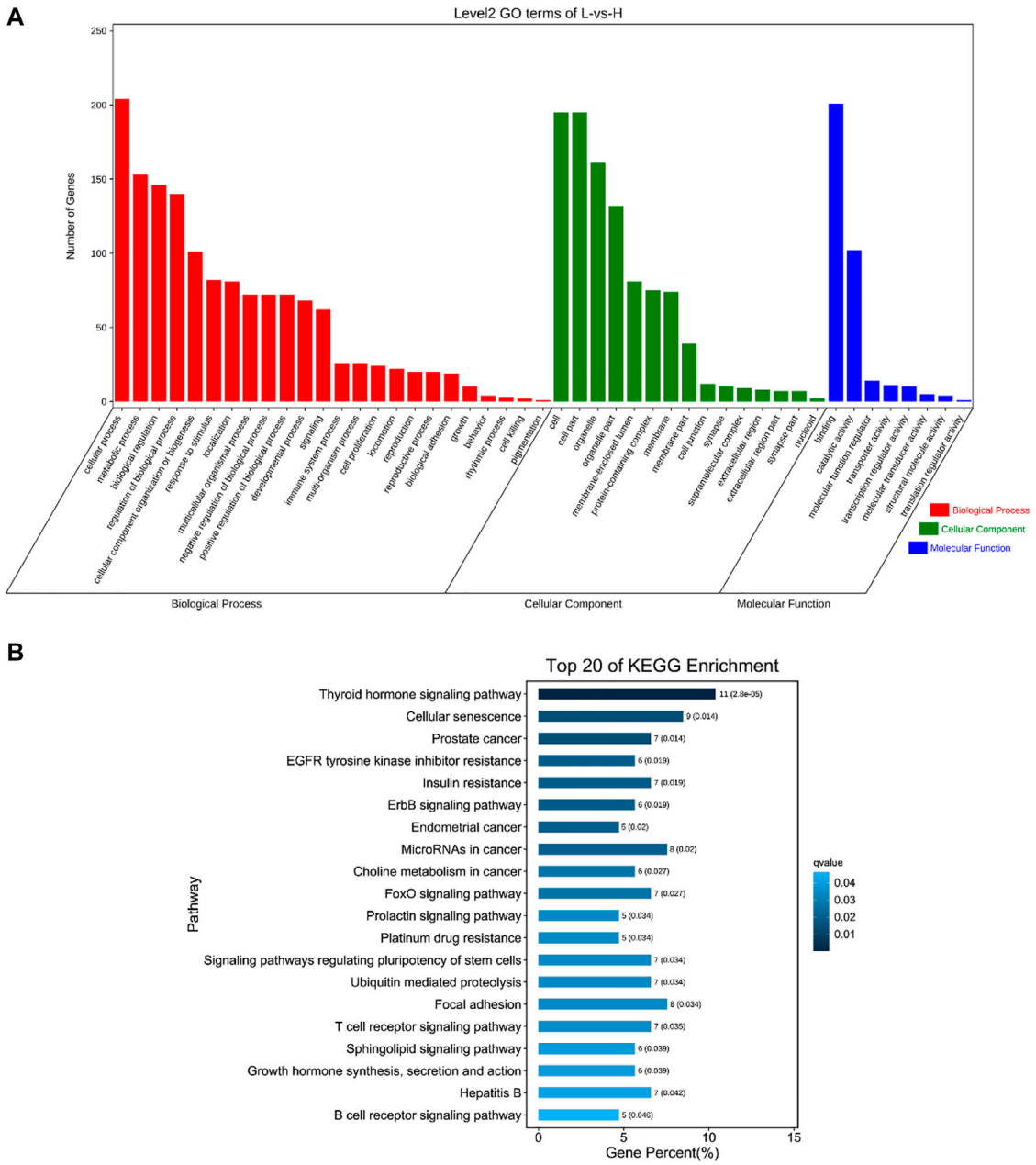


FIGURE 5 | GO and KEGG analysis of source genes of DE circRNAs. **(A)** GO enrichment. The abscissa is the secondary GO term, and the ordinate is the number of genes in this term. Different colors show different types of GO terms, red represents Biological Process, green represents Cellular Component, blue represents Molecular Function. **(B)** KEGG enrichment. Using the top 20 pathways with the smallest Q value to make a map, the ordinate is the pathway, the abscissa is the enrichment factor, the size of the column indicates the number of genes, the bluer the color, the smaller the Q value.

Firstly, to deeply understand the genomic characteristics of circRNAs, 13974 circRNA transcripts were analyzed. A total of 13366 circRNA transcripts were generated from 4,466 genes, and most genes can produce multiple circRNA transcripts. In previous research, 5,934 circRNAs were identified on different muscles in pigs, and 4,928 circRNAs corresponded to 2358 source genes, in addition, 85.7% of the source genes were transcribed to generate two or more circRNA transcripts (Liang et al., 2017). According to the source gene location, 13950 circRNAs were distributed on all chromosomes, and most of circRNAs contained 1-3 exons (Tao et al., 2017). The genome characteristics of identified circRNAs in this study were consistent with other researches, which further illustrated accuracy and reliability of the identified circRNAs, and also showed that circRNAs had similarities among various species.

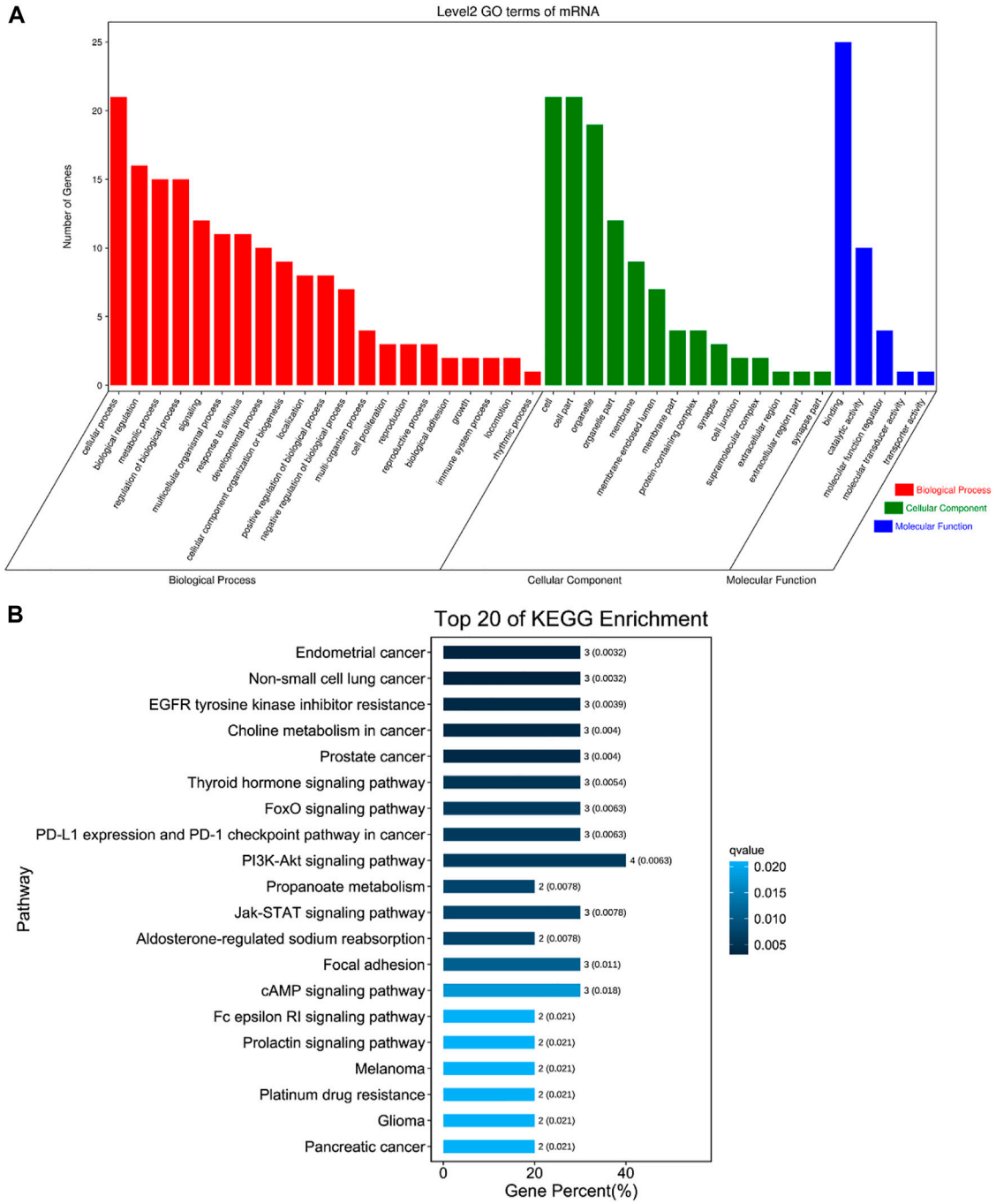


FIGURE 6 | The first 20 GO terms and KEGG. **(A)** The first 20 GO terms of DE circRNAs related to skeletal muscle. The abscissa is the secondary GO term, and the ordinate is the number of genes in this term. Different colors show different types of GO terms, red represents Biological Process, green represents Cellular Component, blue represents Molecular Function. **(B)** The first 20 KEGG of DE circRNAs related to skeletal muscle. Using the top 20 pathways with the smallest Q value to make a map, the ordinate is the pathway, the abscissa is the enrichment factor, the size of the column indicates the number of genes, the bluer the color, the smaller the Q value.

The conservation of circRNA sequences among species indicates that the formation mechanism of circRNAs may also have inter-species similarities, studies have shown that exons flanking long introns are easier to circularize to produce circRNAs, and the reverse complementary sequence of flanking introns promotes the reverse splicing of circRNAs (Ashwal-Fluss et al., 2014; Westholm et al., 2014). Our findings were consistent with literature reports.

At present, the functions of circRNAs are mainly annotated by GO and KEGG analysis of circRNA source genes. Through

TABLE 3 | Target gene of DE circRNAs distribution in chromosome and QTLs region.

Chromosome/Mitochondrion	DE gene number	DE gene number in QTL region	QTL region length(Mb)
1	17	16	272.538
2	17	17	152.419
3	14	13	129.972
4	12	11	126.574
5	14	13	98.508
6	20	20	146.584
7	15	15	126.113
8	19	19	143.374
9	12	12	138.044
10	7	7	67.1919
11	6	5	77.1184
12	10	10	58.0079
13	32	32	208.259
14	22	22	154.672
15	16	16	137.024
16	6	6	66.4773
17	10	9	52.1699
18	8	8	58.1156
X	6	5	83.7497
Y	0	0	0
MT	0	0	0

Note: Chromosome/Mitochondrion represents Chromosome number. DE, gene number represents differentially expressed genes in this Chromosome; DE, gene number in QTL, region represents differentially expressed genes in QTL, region.

enrichment analysis of DE circRNA (including related to muscle development) source genes, we found that the source genes were significantly enriched in histone methylation, programmed cell death, developmental growth, cytoskeletal protein binding and other GO terms. The significantly enriched pathways included Apoptosis, Regulation of actin cytoskeleton, FoxO signaling pathway, mTOR signaling pathway, PI3K–Akt signaling pathway, Wnt signaling pathway, Hedgehog signaling pathway and other pathways related to growth traits and skeletal muscle growth and development. Previous studies have found that the source genes of DE circRNAs were enriched in pathways related to skeletal muscle development, such as JAK-STAT, PI3K-Akt, Wnt and the transition between fast and slow fibers signaling pathways (Wang J. et al., 2019; Huang et al., 2021). The *Map3k20* gene is one of the upstream factors regulating the JNK/MAPK signaling pathway. The *circFgfr2* regulated Map3k20 expression by sponging *miR-133*, *circFgfr2* can regulate skeletal muscle development via the JNK/MAPK signaling pathway (Yan et al., 2021). Mechanistically, *circCCDC91* could absorb *miR-15a*, *miR-15b-5p*, and *miR-15c-5p* to regulate the expression of Insulin receptor substrate1 (*IRS1*), as well as activate insulin-1ike growth factor 1-phosphatidylinositol 3-kinase/AKT (IGF1-PI3K/AKT) signaling pathway. In addition, *circCCDC91* could rescue skeletal muscle atrophy by activating IGF1-PI3K/AKT pathway (Zhao et al., 2022). The *circPPP1R13B* promotes chicken SMSC proliferation and differentiation by targeting miR-9-5p and activating IGF/PI3K/AKT signaling pathway. The results showed that the DE circRNAs play an important role in meat quality and skeletal muscle development, and can be used as candidate functional circRNAs related to skeletal muscle.

circRNAs are found in almost all cells and tissues. At the same time, circRNAs also have a certain degree of sequence conservation (higher sequence conservation than lncRNAs), and have the miRNA response elements (MREs), therefore, circRNAs can be competing endogenous RNAs (ceRNAs) for regulating miRNAs (Song and Li., 2018). In this study, it was found that circRNAs have one or more binding sites for miRNAs. Previous research showed that the *ciRS-7* has multiple binding site for *miR-7*, thereby inhibiting *miR-7* activity and leading to upregulation of *miR-7* target gene expression (Hansen et al., 2013). In addition, many reports have demonstrated that circRNAs can inhibit the binding of miRNAs to target genes, thereby regulating cell growth and development. For instance, the *circZfp609* acts as a sponge of *miR-194-5p* to increase the expression of *BCLAF1* and inhibit myoblast differentiation (Wang Y. et al, 2019). The *circSVIL* can promote cell proliferation and differentiation by sponging *miR-203* and promoting the expression of *MEF2C* and *c-JUN* in chicken (Ouyang et al., 2018a). The *circRBFOX2* interacts with *miR-206* to increase the expression of *cyclin D2* and thereby promotes cell proliferation in chickens (Ouyang et al., 2018b).

In recent years, due to the extensive development of QTL research, a large number of QTLs have been discovered in pigs. Taking the QTL mapping study of domestic pigs in the AnimalQTLdb database as an example, more than 20,000 QTLs have been identified, involving more than 600 different traits. The QTLs mapping studies of other livestock and poultry, such as cattle, chickens, and horses, are also very common (Legarra et al., 2015). How to integrate this large amount of information, compare and locate, and finally determine the genetic basis of candidate genes and traits is a key issue in applying these localization results. Therefore, the research on QTL has become a new hot issue.

Some studies have shown that muscle fibers were related to meat quality and growth development (Lee et al., 2010). In QTL analysis, it was found that most DE circRNAs were related to meat quality and

growth traits. In other research, 54.01% of DE circRNAs were enriched in QTL regions related to meat quality and growth traits, such as the *circRNA227* was enriched in the Intramuscular fat content QTL and Loin muscle area QTL and *circRNA11553* was enriched in the Mean corpuscular hemoglobin content QTL (Shen et al., 2019). But compared with QTL, the number of circRNAs that control muscle traits in pigs is few. However, in the coming decades, the development of bioinformatics, genetics, and transgenic technology will greatly increase the number of circRNAs controlling muscle traits in pigs. This will facilitate the discovery of functional candidate circRNAs. The results of QTL analysis may provide some data support for the discovery of new functions of circRNAs.

circRNAs are involved in many physiological and pathological processes, including cancer, aging, and muscle development, but their specific functions are still unclear. Thus, our results could help to understand the genetic mechanisms in meat quality in livestock and poultry, and further studies are necessary for a better comprehension of circRNAs functions in skeletal muscle development.

5 CONCLUSION

In conclusion, 296 DE circRNAs were identified. Most DE circRNAs were enriched in QTLs related to meat quality and growth traits. In addition, the source genes of DE circRNAs were enriched in pathways related to cell growth and skeletal muscle development. Therefore, these DE circRNAs may play a key role in meat quality and growth traits. Researches on circRNAs in skeletal muscle development and disease are still in its infancy, and skeletal muscle plays a key role in body development. How to use these circRNAs and apply them to livestock breeding and disease treatment remains a challenge. Therefore, the understanding of the complex molecular mechanism of circRNAs requires further study.

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/, PRJNA812354.

ETHICS STATEMENT

The animal study was reviewed and approved by All animal care and treatment procedures were conducted in strict accordance with the Animal Ethics Committee of Shandong Agricultural University, China, and performed in accordance with the Committee's guidelines and regulations (Approval No.: 2004006). Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

WC and YZ designed the study. WC, LM, SL, and MQ performed the research. YZ contributed reagents and materials. LM analyzed the data and wrote the manuscript. WC and YZ provided substantial comments and revised the manuscript. All authors read and approved the final version of the manuscript.

FUNDINGS

This study was supported financially by National Key R&D Program of China (No. 2021YFD1301200), the Agricultural Animal Breeding Project of Shandong Province (No. 2020LZGC012), Shandong Province Pig Industry Technology System Project (No. SDAIT-08–02), Shandong Provincial Natural Science Foundation (No. ZR2019MC053).

ACKNOWLEDGMENTS

We are grateful to Yu Zhang, Jing Zhao, Yingbin Du and Meiqi Zhu for their assistance in sample collection and laboratory analyses.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.858763/full#supplementary-material

REFERENCES

Ashwal-Fluss, R., Meyer, M., Pamudurti, N. R., Ivanov, A., Bartok, O., Hanan, M., et al. (2014). circRNA biogenesis competes with pre-mRNA splicing. *Mol. Cell* 56 (1), 55–66. doi:10.1016/j.molcel.2014.08.019

Birnkrant, D. J., Bushby, K., Bann, C. M., Apkon, S. D., Blackwell, A., Brumbaugh, D., et al. (2018). Diagnosis and management of Duchenne muscular dystrophy, part 1: Diagnosis, and neuromuscular, rehabilitation, endocrine, and gastrointestinal and nutritional management. *Lancet. Neurol.* 17 (3), 251–267. doi:10.1016/S1474-4422(18)30024-3

Campbell, A. E., Belleville, A. E., Resnick, R., Shadle, S. C., and Tapscott, S. J. (2018). Facioscapulohumeral dystrophy: Activating an early embryonic transcriptional program in human skeletal muscle. *Hum. Mol. Genet.* 27 (R2), R153–R162. doi:10.1093/hmg/ddy162

Chen, I., Chen, C. Y., and Chuang, T. J. (2015). Biogenesis, identification, and function of exonic circular RNAs. *Wiley Interdiscip. Rev. RNA* 6 (5), 563–579. doi:10.1002/wrna.1294

Chen, X., Ouyang, H., Wang, Z., Chen, B., and Nie, Q. (2018a). A novel circular RNA generated by FGFR2 gene promotes myoblast proliferation and differentiation by sponging miR-133a-5p and miR-29b-1-5p. *Cells* 7 (11), 199. doi:10.3390/cells7110199

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

Chen, B., Yu, J., Guo, L., Byers, M. S., Wang, Z., Chen, X., et al. (2019). Circular RNA circHIPK3 promotes the proliferation and differentiation of chicken myoblast cells by sponging miR-30a-3p. *Cells* 8 (2), 177. doi:10.3390/cells8020177

Das, A., Das, A., and Das, D. (2020). Circular RNAs in myogenesis. *Biochim. Biophys. Acta. Gene Regul. Mech.* 1863 (4), 194372. doi:10.1016/j.bbagrm.2019.02.011

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., et al. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature* 495 (7441), 384–388. doi:10.1038/nature11993

Holman, B., Bailes, K. L., Meyer, R. G., and Hopkins, D. L. (2019). Effect of modified soxhlet (soxtec) and folch extraction method selection on the total lipid determination of aged beef. *J. Food Sci. Technol.* 56 (8), 3957–3961. doi:10.1007/s13197-019-03878-4

Huang, K., Chen, M., Zhong, D., Luo, X., Feng, T., Song, M., et al. (2021). Circular RNA profiling reveals an abundant circEch1 that promotes myogenesis and differentiation of bovine skeletal muscle. *J. Agric. Food Chem.* 69 (1), 592–601. doi:10.1021/acs.jafc.0c06400

Jakobi, T., Czaja-Hasse, L. F., Reinhardt, R., and Dieterich, C. (2016). Profiling and validation of the circular RNA repertoire in adult murine hearts. *Genomics Proteomics Bioinforma.* 14 (4), 216–223. doi:10.1016/j.gpb.2016.02.003

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36, D480–D484. doi:10.1093/nar/gkm882

Kim, D., Langmead, B., and Salzberg, S. L. (2015). Hisat: A fast spliced aligner with low memory requirements. *Nat. Methods* 12 (4), 357–360. doi:10.1038/nmeth.3317

Kos, A., Dijkema, R., Arnberg, A. C., van der Meide, P. H., and Schellekens, H. (1986). The hepatitis delta (delta) virus possesses a circular RNA. *Nature* 323 (6088), 558–560. doi:10.1038/323558a0

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9 (4), 357–359. doi:10.1038/nmeth.1923

Lee, S. H., Joo, S. T., and Ryu, Y. C. (2010). Skeletal muscle fiber type and myofibrillar proteins in relation to meat quality. *Meat Sci.* 86 (1), 166–170. doi:10.1016/j.meatsci.2010.04.040

Legarra, A., Croiseau, P., Sanchez, M. P., Teyssedre, S., Salle, G., Allais, S., et al. (2015). A comparison of methods for whole-genome QTL mapping using dense markers in four livestock species. *Genet. Sel. Evol.* 47 (1), 6. doi:10.1186/s12711-015-0087-7

Legnini, I., Di, Timoteo. G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., et al. (2017). Circ-ZNF609 is a circular RNA that can Be translated and functions in myogenesis. *Mol. Cell* 66 (1), 22–37. doi:10.1016/j.molcel.2017.02.017

Li, J. H., Liu, S., Zhou, H., Qu, L. H., and Yang, J. H. (2014). starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 42, D92–D97. doi:10.1093/nar/gkt1248

Li, C., Li, X., Ma, Q., Zhang, X., Cao, Y., Yao, Y., et al. (2017). Genome-wide analysis of circular RNAs in prenatal and postnatal pituitary glands of sheep. *Sci. Rep.* 7 (1), 16143. doi:10.1038/s41598-017-16344-y

Li, Y., Chen, X., Sun, H., and Wang, H. (2018a). Long non-coding RNAs in the regulation of skeletal myogenesis and muscle diseases. *Cancer Lett.* 417, 58–64. doi:10.1016/j.canlet.2017.12.015

Li, H., Yang, J., Wei, X., Song, C., Dong, D., Huang, Y., et al. (2018b). CircFUT10 reduces proliferation and facilitates differentiation of myoblasts by sponging miR-133a. *J. Cell. Physiol.* 233 (6), 4643–4651. doi:10.1002/jcp.26230

Liang, G., Yang, Y., Niu, G., Tang, Z., and Li, K. (2017). Genome-wide profiling of *Sus scrofa* circular RNAs across nine organs and three developmental stages. *DNA Res.* 24 (5), 523–535. doi:10.1093/dnares/dsx022

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495 (7441), 333–338. doi:10.1038/nature11928

Ouyang, H., Chen, X., Li, W., Li, Z., Nie, Q., and Zhang, X. (2018a). Circular RNA circSVIL promotes myoblast proliferation and differentiation by sponging miR-203 in chicken. *Front. Genet.* 9, 172. doi:10.3389/fgene.2018.00172

Ouyang, H., Chen, X., Wang, Z., Yu, J., Jia, X., Li, Z., et al. (2018b). Circular RNAs are abundant and dynamically expressed during embryonic muscle development in chickens. *DNA Res.* 25 (1), 71–86. doi:10.1093/dnares/dsx039

Qiao, Y., Mao, Y., Wang, J., Chen, R., Libing, Z., Su, Y. Q., et al. (2016). Analysis of liver and gill miRNAs of Larimichthys crocea against Cryptocryon irritans challenge. *Fish. Shellfish Immunol.* 59, 484–491. doi:10.1016/j.fsi.2016.10.027

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 (1), 139–140. doi:10.1093/bioinformatics/btp616

Sanger, H. L., Klotz, G., Riesner, D., Gross, H. J., and Kleinschmidt, A. K. (1976). Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc. Natl. Acad. Sci. U. S. A.* 73 (11), 3852–3856. doi:10.1073/pnas.73.11.3852

Shen, L., Gan, M., Tang, Q., Tang, G., Jiang, Y., Li, M., et al. (2019). Comprehensive analysis of lncRNAs and circRNAs reveals the metabolic specialization in oxidative and glycolytic skeletal muscles. *Int. J. Mol. Sci.* 20 (12), E2855. doi:10.3390/ijms20122855

Song, Y. Z., and Li, J. F. (2018). Circular RNA hsa_circ_0001564 regulates osteosarcoma proliferation and apoptosis by acting miRNA sponge. *Biochem. Biophys. Res. Commun.* 495 (3), 2369–2375. doi:10.1016/j.bbrc.2017.12.050

Tao, H., Xiong, Q., Zhang, F., Zhang, N., Liu, Y., Suo, X., et al. (2017). Circular RNA profiling reveals chi_circ_0008219 function as microRNA sponges in pre-ovulatory ovarian follicles of goats (*Capra hircus*). *Genomics* 110 (4), 257–266. doi:10.1016/j.ygeno.2017.10.005

Wang, J., Ren, Q., Hua, L., Chen, J., Zhang, J., Bai, H., et al. (2019a). Comprehensive analysis of differentially expressed mRNA, lncRNA and circRNA and their ceRNA networks in the Longissimus Dorsi muscle of two different pig breeds. *Int. J. Mol. Sci.* 20 (5), 1107. doi:10.3390/ijms20051107

Wang, Y., Li, M., Wang, Y., Liu, J., Zhang, M., Fang, X., et al. (2019b). A Zfp609 circular RNA regulates myoblast differentiation by sponging miR-194-5p. *Int. J. Biol. Macromol.* 121, 1308–1313. doi:10.1016/j.ijbiomac.2018.09.039

Wei, X., Li, H., Yang, J., Hao, D., Dong, D., Huang, Y., et al. (2017). Circular RNA profiling reveals an abundant circLMO7 that regulates myoblasts differentiation and survival by sponging miR-378a-3p. *Cell Death Dis.* 8 (10), e3153. doi:10.1038/cddis.2017.541

Westholm, J. O., Miura, P., Olson, S., Shenker, S., Joseph, B., Sanfilippo, P., et al. (2014). Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Rep.* 9 (5), 1966–1980. doi:10.1016/j.celrep.2014.10.062

Xu, M., Chen, X., Chen, D., Yu, B., and Huang, Z. (2017). FoxO1: A novel insight into its molecular mechanisms in the regulation of skeletal muscle differentiation and fiber type specification. *Oncotarget* 8 (6), 10662–10674. doi:10.18632/oncotarget.12891

Yan, J., Yang, Y., Fan, X., Liang, G., Wang, Z., Li, J., et al. (2021). Circrnaome profiling reveals circfgfr2 regulates myogenesis and muscle regeneration via a feedback loop. *J. Cachexia Sarcopenia Muscle* 13 (1), 696–712. doi:10.1002/jcsm.12859

Yin, H., He, H., Shen, X., Zhao, J., Cao, X., Han, S., et al. (2020). miR-9-5p inhibits skeletal muscle satellite cell proliferation and differentiation by targeting IGF2BP3 through the IGF2-PI3K/akt signaling pathway. *Int. J. Mol. Sci.* 21 (5), E1655. doi:10.3390/ijms21051655

Zhao, J., Zhao, X., Shen, X., Zhang, Y., Zhang, Y., Ye, L., et al. (2022). Circccdc91 regulates chicken skeletal muscle development by sponging mir-15 family via activating IGF1-PI3K/AKT signaling pathway. *Poult. Sci.* 101 (5), 101803. doi:10.1016/j.psj.2022.101803

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 Genetics

TYPE Original Research
PUBLISHED 30 September 2022
DOI 10.3389/fgene.2022.974619

OPEN ACCESS

EDITED BY
Shaojun Liu,
Hunan Normal University, China

REVIEWED BY
Biao Chen,
Jiangxi Agricultural University, China
Hong Wei Liang,
Yangtze River Fisheries Research Institute (CAFS), China

*CORRESPONDENCE
Yanyan Sun,
yanyansun2014@163.com
Jilan Chen,
chen.jilan@163.com

SPECIALTY SECTION
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

RECEIVED 21 June 2022
ACCEPTED 29 August 2022
PUBLISHED 30 September 2022

CITATION
Isa AM, Sun Y, Li Y, Wang Y, Ni A, Yuan J, Ma H, Shi L, Tesfay HH, Fan J, Wang P and Chen J (2022), MicroRNAs with non-additive expression in the ovary of hybrid hens target genes enriched in key reproductive pathways that may influence heterosis for egg laying traits. *Front. Genet.* 13:974619. doi: 10.3389/fgene.2022.974619

MicroRNAs with non-additive expression in the ovary of hybrid hens target genes enriched in key reproductive pathways that may influence heterosis for egg laying traits

Adamu Mani Isa[1,2], Yanyan Sun[1]*, Yunlei Li[1], Yuanmei Wang[1], Aixin Ni[1], Jingwei Yuan[1], Hui Ma[1], Lei Shi[1], Hailai Hagos Tesfay[1], Jing Fan[1], Panlin Wang[1] and Jilan Chen[1]*

[1]Key Laboratory of Animal (Poultry) Genetics Breeding and Reproduction, Ministry of Agricultural and Rural Affairs, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China, [2]Department of Animal Science, Usmanu Danfodiyo University, Sokoto, Nigeria

Heterosis has been extensively exploited in chicken breeding to improve laying traits in commercial hybrid stock. However, the molecular mechanisms underlying it remains elusive. This study characterizes the miRNAome in the pre-hierarchical follicles of purebred and hybrid laying hens, and investigate the functions of miRNAs with non-additive expression in the pre-hierarchical follicles as they modulate heterosis for egg number and clutch size. To achieve that aim, White Leghorn and Rhode Island Red chicken lines were reciprocally crossed to generate hybrids. The crossbreds demonstrated heterosis for egg number and clutch size, and pre-hierarchical follicles from 4 birds of each genotype were collected at 53 weeks of age. Mode of miRNA expression was characterized after miRNA sequencing. A total of 50 miRNAs including 30 novel ones, were found to exhibit non-additive expression. Dominance was the predominant mode of expression exhibited by majority of the miRNAs. Functional analysis of target genes of the known miRNAs with non-additive expression revealed Gene Ontology terms related to regulation of transcription, metabolic processes and gene expression. KEGG and REACTOME pathways including hedgehog, cellular senescence, wnt, TGF-β, progesterone-mediated oocyte maturation, oocyte meiosis, GnRH signaling, signal transduction and generic transcription, which can be linked to primordial follicle activation, growth and ovulation, were significantly enriched by target genes of miRNAs with non-additive expression. Majority of the genes enriched in these biological pathways were targeted by gga-miR-19a, gga-miR-19b, gga-miR-375, gga-miR-135a, and gga-miR-7 and 7b, thus, revealing their synergistic roles in enhancing processes that could influence heterosis for egg number and clutch size in hybrid hens.

KEYWORDS
crossbreeding, heterosis, egg (production), clutch size, miRNA, dominance

Introduction

Heterosis remains a central theme in the field of poultry breeding, and previous investigations established non-additive gene action as the primary cause. In chickens, heterosis is accomplished through crossbreeding genetically distinct lines and breeds such that the average performance of the crossbred population is superior to the mid performance of the purebred parental lines expected under additive gene assumptions. In addition to heterosis, breed complementarity is also exploited by crossbreeding. The success of crossbreeding schemes largely depends on crossing of genetically diverse lines (Amuzu-Aweh, 2020).

Egg laying efficiency typified by high laying rates, large number of eggs and larger clutches seem to be the traits that are enhanced in the hybrids laying birds, and were linked to well-orchestrated and organized follicular hierarchy in the hens (Johnson, 2012). The selection process involves recruitment from pre-hierarchical pool, follicles into the pre-ovulatory hierarchy. The changes that occur after follicle recruitment during the pre-ovulatory period were propound and differ even between pre-ovulatory follicles. The dynamics in gene expression profiles of pre-ovulatory follicles with its attendant variation between pullets of the same genotype poses great challenge to transcriptomics studies. Prior to selection event, many pre-hierarchical follicles either undergo atresia or remain steroidogenic incompetent in an effort to preserve the sanctity of small viable cohorts of pre-selected follicles associated with follicular reserve and subsequently, clutches of eggs (Johnson, 2015).

Previous studies have elucidated the mechanisms of maintaining the state of primordial follicles by local factors and intracellular pathways. This occurs via the action of multiple activators including growth differentiation factor 9 (GDF9), anti-mullerian hormone (AMH), zona pellucida 2 (ZP2), wingless-type MMTV integration site family member 4 (WNT 4) (Zhang et al., 2019), and recently, PPAR pathway (Yoon et al., 2020).

The discovery of the first microRNA (miRNA) gene member (lin4) in the early 90s (Lee et al., 1993) has ushered an ever-expanding field of miRNA research with new members being reported every day. MicroRNAs are a class of small non-coding RNAs (~22 nt) that regulate gene expression at a post-transcriptional level through complementary base pairing with the target mRNA, leading to degradation of mRNA and eventually repressing its translation (Riffo-Campos et al., 2016). Many investigations have shown that miRNAs are involved in the regulation of various pathways and exert influence on a wide variety of phenotypes. To date, there are 1,235 chicken miRNAs cataloged in the miRDB database (http://mirdb.org/statistics.html) and the number will continue to increase with the decline in the cost of high throughput sequencing, commensurate with Moore's law (Brock and Moore, 2006) and increasing curiosity of molecular scientist to address myriads of complex molecular phenomena. The array of functions miRNAs perform in the translation of coded information in the DNA to phenotype lies with its unique ability to bind to the canonical site of other RNA species called microRNA response elements (MRE), which allows it to interact with a wide variety of RNA species (including mRNAs, lncRNAs, and circRNAs). MiRNAs are therefore central molecules in the theory of cross-talking of ceRNAs proposed at the beginning of the last decade (Salmena et al., 2011). This class of RNAs represses the expression of mRNAs thereby lowering the expression of the genes encoding the larger species of ceRNAs.

In chickens, the roles of miRNAs in the expression of phenotypes including sperm motility in roosters (Liu et al., 2018) and egg number in laying hens (Zhang et al., 2017) have been documented in recent times. Despite the growing body of literature in the field of miRNA biogenesis, identification and target prediction on one hand and intensive search for molecular mechanisms of heterosis in domestic chickens (Mai et al., 2019; Zhuo et al., 2019) on the other hand, there is still a gap in literature on the mode of inheritance and possible influence of miRNAs on heterosis for egg number and clutch size in laying chickens. This study therefore, aims to identify miRNAs with non-additive mode of expression which may influence heterosis for egg number and clutch size in laying hens.

Results

Phenotypic data

Number of eggs laid by the White Leghorn (W) and Rhode Island Red (R) parental purebred lines were significantly less than ($p < 0.001$) the number laid by their reciprocal hybrids at 53 weeks of age (Figure 1). The synthesized mid-parent value (MPV) for the purebred was similar to the two parental lines. Heterosis for egg number was 14.1 and 18.5% in Rhode Island Red × White Leghorn (RW) and White Leghorn × Rhode Island Red (WR) respectively. Furthermore, heterosis for clutch size was 12.29 and 38.43% in RW and WR hybrids respectively.

Overview of miRNAome mapping statistics

MicroRNA sequences generated from the 16 libraries constructed from RNA extracted from pre-hierarchical white follicles in the two purebred (R and W) and their reciprocal hybrids (RW and WR) yield a total of 291, 129, 265 reads after

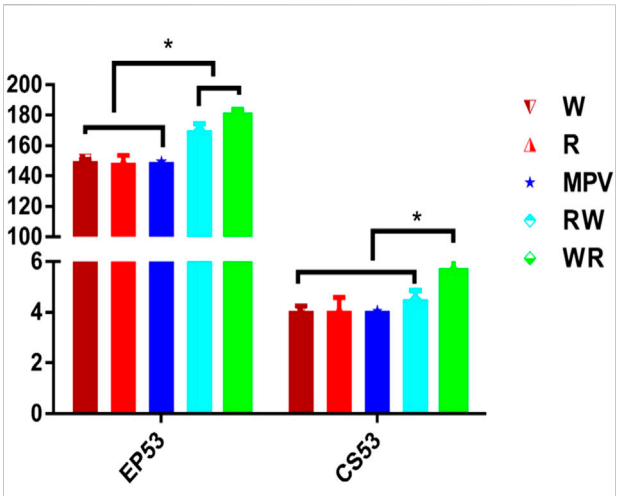


FIGURE 1
Egg number and clutch size at 53 weeks of age for parental lines and their reciprocal hybrids. Mean values with an asterisk (*) on top of them were significantly different (Tukey-Kramer HSD $p \leq 0.05$). W: Purebred White Leghorn, R: Purebred Rhode Island Red, RW: Rhode Island Red × White Leghorn cross, WR: White Leghorn × Rhode Island Red cross, MPV: Mid-parent value.

filtering for quality (Phred score >20), size selection (18–30 nt) and trimming of adapter sequences. Out of this, 272,842, 977 (93.66%) reads aligned to the chicken genome built GRCg6a of the Ensembl database and were preserved for further analysis. Specifically, 169, 706, 629 reads (62.30%) mapped with a perfect match while the remaining 103, 136, 348 reads (37.8%) mapped with a single nucleotide mismatch. For each library, more than 90% of the clean reads were successfully mapped to the chicken genome (Supplementary Figure S1).

Furthermore, mapped reads belonging to other species of non-coding RNAs including rRNA (8,103,438 (2.97%)), tRNA (4,251,546 (1.55%)), snRNA (176,885 (0.06%)), snoRNA (5,646,100 (2.07%)) and others (563,900, (0.21%)) collectively accounted for 6.44% of the total reads, and were discarded (Supplementary Table S1). Similarly, reads that mapped to the low diversity region of the genome including SINES, LINES, and LTR which collectively accounted for 0.83% of the total raw reads were not retained in the subsequent analyses.

A total of 899 and 289 known and predicted novel miRNAs were found to be expressed in the pre-hierarchical follicles. After filtering out the lowly expressed miRNAs with <10 read counts per million in all the libraries, a total of 595 known and 235 novel miRNAs were preserved for further downstream analyses (Supplementary Table S2). Principal component analysis (PCA) of the retained clean reads depicts clear separation of the four genetic groups (Figure 2). PC1 *vs.* PC2 plot assigned the genotypes into three clusters; two clusters for W and R purebred parental lines and another cluster for the two reciprocal hybrids (RW and WR). Furthermore, PC2 *vs.* PC3 plot separates the genotypes into four distinct clusters.

MiRNAs with abundant expression (~1 million raw reads/per library) in the ovary of laying hens at 53 weeks of age include gga-miR-143-3p, gga-miR-148a-3p, gga-miR-26a-2-5p, gga-miR-26a-5p, gga-miR-99a-5p, gga-miR-21-5p and gga-miR-10a-5p. The most abundant miRNA (gga-miR-143-3p) had raw reads between 9 and 13 million reads, accounting for 3.9, 3.8, 4.9 and 4.3% of all expressed miRNAs in R, W, RW and WR genotypes respectively. Twenty most abundantly expressed known miRNAs expressed in the pre-hierarchal follicles in the purebred and hybrids laying chickens were same and are presented in Table 1.

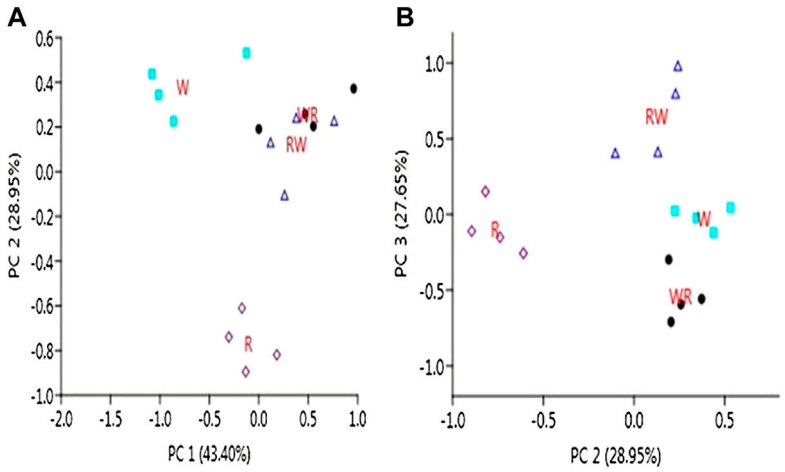


FIGURE 2
Separation of the pure lines and reciprocal hybrids by PCA **(A)**: PC1 vs. PC2 plot **(B)**: PC2 vs. PC3 plot.

Differentially expressed miRNAs and their mode of inheritance

Pair-wise comparisons of miRNA expression between the pre-hierarchical follicles in the ovaries of the reciprocal hybrids and their parental purebreds on one hand and between the hybrids and synthesized mid-parent average on the other hand are presented in Figure 3.

Comparing miRNA expression in the two parental purebred lines, 30 differentially expressed miRNAs (DEMiRs) were detected, including nine upregulated and 21 downregulated miRNAs. In RW vs. A, 21 DEMiRs were identified, 8 and 13 were upregulated and downregulated respectively. In RW vs R, a total of 19 DEMiRs were identified while 23 DEMiRs were detected in RW vs W. Further, WR vs. R yielded 65 DEMiRs consisting of eight upregulated and 57 downregulated miRNAs, and 139 miRNAs showed differential expression in WR vs. R. In WR vs. W, 127 DEMiRs were detected including 54 upregulated and 73 downregulated miRNAs.

Overall, all pairwise comparisons yielded 70 unique DEMiRs. Further classification of the DEMiRs based on their mode of inheritance pattern was achieved by overlapping the DEMiRs in Venn diagrams. Total of 20 DEMiRs exhibited additive mode of expression (Supplementary Figure S2), while the remaining unique DEMiRs exhibited non-additive mode of expression (Figure 4). Of these, 20 known miRNAs were grouped in to 14 miRNA family clusters (Table 2), and dominance mode of expression pattern was exhibited by 19 of them. The remaining 30 were suggested novel miRNAs (Supplementary Table S4).

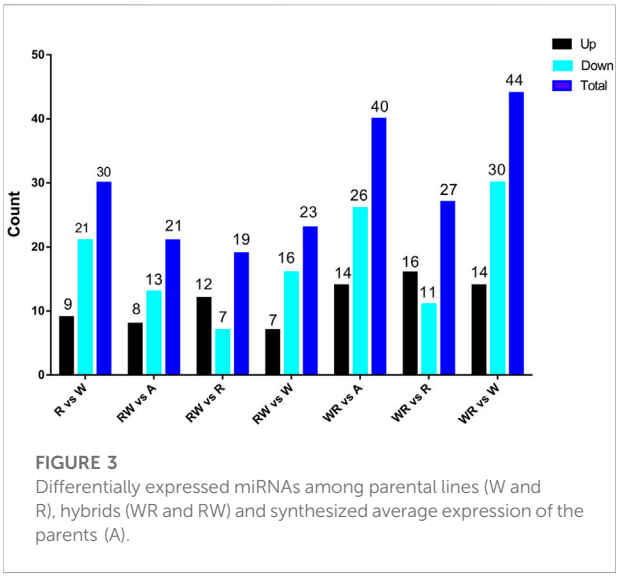


FIGURE 3
Differentially expressed miRNAs among parental lines (W and R), hybrids (WR and RW) and synthesized average expression of the parents (A).

TABLE 1 The 20 most abundantly expressed miRNAs in the ovary of purebred and hybrid chickens.

miRNA	R		RW		W		WR	
	Read counts	% of Total	Read counts	% of Total	Read counts	% of Total	Read counts	% of Total
gga-miR-143-3p	10688755	3.92	13462849	4.93	9219675	3.38	11713640	4.29
gga-miR-148a-3p	8918917	3.27	9317996	3.42	5912664	2.17	10132356	3.71
gga-miR-26a-2-5p	4442976	1.63	4900246	1.80	4018858	1.47	6210366	2.28
gga-miR-26a-5p	4442972	1.63	4900244	1.80	4018853	1.47	6210361	2.28
gga-miR-99a-5p	3900383	1.43	5325620	1.95	3461332	1.27	6754169	2.48
gga-miR-21-5p	4783473	1.75	3356219	1.23	2331764	0.85	3730281	1.37
gga-miR-10a-5p	2451540	0.90	2223960	0.82	2780188	1.02	2534866	0.93
gga-miR-146c-5p	1266042	0.46	1289721	0.47	1272549	0.47	1151559	0.42
gga-miR-145-5p	1130629	0.41	935879	0.34	1875808	0.69	1036686	0.38
gga-miR-199-3p	1107227	0.41	1352070	0.50	841458	0.31	1515098	0.56
gga-miR-126-3p	1111832	0.41	1235579	0.45	856975	0.31	1399615	0.51
gga-miR-100-5p	969498	0.36	1193374	0.44	801364	0.29	1205579	0.44
gga-miR-146b-5p	1278331	0.47	856633	0.31	1207839	0.44	821916	0.30
gga-let-7a-5p	999090	0.37	1048792	0.38	926779	0.34	1176787	0.43
gga-let-7j-5p	999090	0.37	1048792	0.38	926779	0.34	1176787	0.43
gga-let-7g-5p	918332	0.34	956672	0.35	919716	0.34	1035672	0.38
gga-miR-101-3p	774947	0.28	1272755	0.47	750151	0.27	976429	0.36
gga-let-7f-5p	940817	0.34	906388	0.33	918234	0.34	984032	0.36
gga-miR-125b-5p	637100	0.23	851730	0.31	605576	0.22	1125438	0.41
gga-let-7i	699908	0.26	649135	0.24	725435	0.27	666154	0.24

Overall, 14, 17, 13 and six miRNAs exhibited high-parent dominance, low parent dominance, over-dominance and under-dominance inheritance patterns respectively (Table 2; Supplementary Table S3).

Functional enrichment analysis of the target genes of the miRNA with a non-additive mode of expression

A total of 970 unique genes were identified as potential targets for 11 known miRNA with non-additive expression (Supplementary Table S4). Target genes with aggregate PTC score ≥50 were screened against the mRNA transcriptome of the pre-hierarchical follicles obtained by sequencing same sample from where miRNAome was sequenced. This yields a list of 970 unique mRNAs, which were uploaded into g. profiler for functional analysis including GO, KEGG and REAC. Overview of the result of gene enrichment analysis in GO, KEGG and REAC databases is depicted in Figure 5. The enriched GO terms consist of 918 biological processes (BP), 53 molecular functions (MF) and 80 cellular components (CC) (Supplementary Table S5). Specifically, GO annotation enrichment showed that target genes of miRNAs with non-additive expression were associated with regulation of cellular process (GO: BP term; FDR = 1.12×10^{-26}), nucleus (GO:CC term; FDR = 4.77×10^{-12}) and transcription regulatory activity (GO:MF term; FDR = 3.4×10^{-18}).

Further, 21 and 16 pathways respectively in KEGG and REACTOME databases were significantly enriched by target genes of miRNAs with non-additive expression. In KEGG database, the pathways include Hedgehog signaling, cellular senescence, focal adhesion, FoxO signaling Wnt signaling, MAPK signaling, mTOR signaling, insulin signaling, TGF-β signaling, oocyte meiosis, autophagy, progesterone-mediated oocyte maturation and GnRH signaling (Table 3). These pathways share common genes and were interconnected (Figure 6).

In the REACTOME database, enriched pathways include glucagon-like peptide (GLP1) regulates insulin secretion, post-transcriptional regulation by small RNA, generic transcription, gene expression, signal transduction, signal by WNT and regulation of insulin secretion (Table 4).

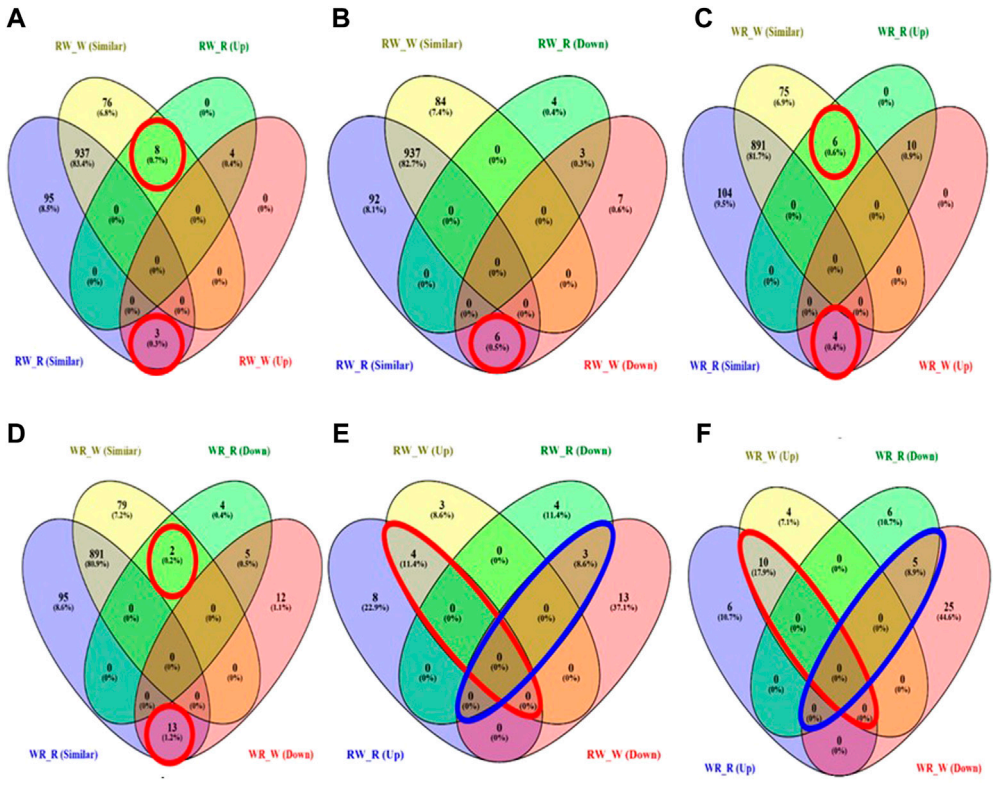


FIGURE 4
Venn diagram showing DEMiRs according to their pattern of inheritance **(A)** High parent dominance in RW **(B)** Low-parent dominance in RW **(C)** High parent dominance in WR **(D)** Low-parent dominance in WR **(E)** Overdominance (enclosed in red) and underdominance (enclosed in blue) expression of miRNAs in RW **(F)** Overdominance (enclosed in red) and underdominance (enclosed in blue) expression of miRNAs in WR.

10.3389/fgene.2022.974619

TABLE 2 Known miRNAs abundance and their non-additive modes of expression in the ovary of purebred White Leghorn, Rhode Island Red, and their reciprocal hybrids.

miRNA	Raw read counts				Mode of expression
	R	W	RW	WR	
gga-miR-10b-5p	56014	232040	49575	60899	LP Dominance both
gga-miR-122-5p	49067	8141	11986	6515	LP Dominance in WR
gga-miR-135a-5p	2515	4911	2724	1946	LP Dominance in WR
gga-miR-145-5p	1130629	1875808	935879	1036686	LP Dominance in RW
gga-miR-1684a-3p	6	437	280	316	HP Dominance in both
gga-miR-1684b-3p	869	21	571	409	HP Dominance in both
gga-miR-1720-5p	28	39	61	265	HP Dominance in WR
gga-miR-1747-5p	229	1	81	107	HP Dominance in both
gga-miR-19a-3p	28429	64340	28105	27508	LP Dominance in both
gga-miR-19b-3p	32190	72366	34010	32786	LP Dominance in WR
gga-miR-205a	1448	3705	2320	1009	LP Dominance in WR
gga-miR-217-5p	40	606	110	94	LP Dominance in RW
gga-miR-34b-3p	430	1731	262	154	LP Dominance in both
gga-miR-34b-5p	1892	5420	960	535	LP Dominance in WR
gga-miR-34c-3p	133	261	86	53	LP Dominance in WR
gga-miR-34c-5p	1889	5414	960	535	LP Dominance in WR
gga-miR-365-3p	744	1522	729	636	LP Dominance in both
gga-miR-375	589	610	4698	60891	Over-dominance in WR
gga-miR-7	61657	99612	63096	272947	HP Dominance in WR
gga-miR-7b	105796	149761	67781	72170	LP Dominance in both

LP, dominance: low parent dominance; HP, dominance: high parent dominance.

Regulatory network for miRNA with non-additive expression and their target genes

The co-expression network for miRNAs with non-additive expression and their target genes is presented in Figure 7. The network revealed that more than one gene can be targeted by one miRNA. MicroRNAs that exhibited low-parent dominance such as gga-miR-19-3p targeted many genes in the follicles. Contrarily, gga-miR-375 with overdominance in WR genotype targeted only CHRNA1 and FRRS1 genes.

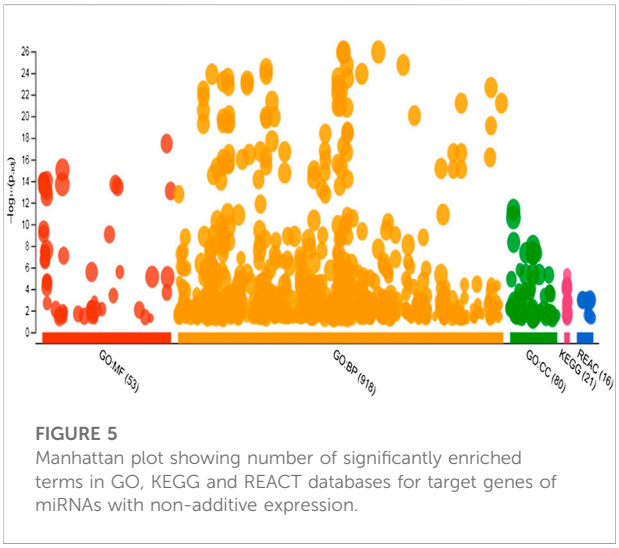


FIGURE 5
Manhattan plot showing number of significantly enriched terms in GO, KEGG and REACT databases for target genes of miRNAs with non-additive expression.

Real-time quantitative PCR validation

Expression analysis of the four miRNAs revealed that pattern of expression was consistent between the Illumina small RNA sequencing and real-time quantitative PCR (RT-qPCR) with a correlation co-efficient of 0.647 (Figure 8).

Discussion

Mode of miRNAs inheritance in reciprocal hybrids and their influence on heterosis for egg number and clutch size traits in laying chickens was not hitherto documented. By sequencing miRNAome of pre-hierarchical follicles of laying hens, we delineate the modes of miRNAs expression in parental purebred lines (R and W) and their reciprocal crossbreds (RW and WR). We identified 595 known miRNAs and

TABLE 3 Pathways enriched in the KEGG database for target genes of miRNAs with non-additive mode of expression.

KEGG ID	KEGG pathway	P- Adjusted
KEGG:04340	Hedgehog signaling pathway	4.83×10^{-6}
KEGG:04218	Cellular senescence	3.75×10^{-5}
KEGG:04068	FoxO signaling pathway	$3.75.59 \times 10^{-5}$
KEGG:04310	Wnt signaling pathway	4.23×10^{-5}
KEGG:04261	Adrenergic signaling in cardiomyocytes	4.23×10^{-5}
KEGG:04010	MAPK signaling pathway	8.49×10^{-5}
KEGG:04510	Focal adhesion	8.94×10^{-5}
KEGG:04150	mTOR signaling pathway	1.22×10^{-4}
KEGG:04910	Insulin signaling pathway	1.13×10^{-3}
KEGG:04114	Oocyte meiosis	1.33×10^{-3}
KEGG:04916	Melanogenesis	1.75×10^{-3}
KEGG:04012	ErbB signaling pathway	332×10^{-3}
KEGG:04140	Autophagy - animal	3.32×10^{-3}
KEGG:04810	Regulation of actin cytoskeleton	1.40×10^{-2}
KEGG:04371	Apelin signaling pathway	1.4136×10^{-3}
KEGG:04350	TGF-β signaling pathway	2.56×10^{23}
KEGG:04144	Endocytosis	2.74×10^{-2}
KEGG:04914	Progesterone-mediated oocyte maturation	2.74×10^{-2}
KEGG:04330	Notch signaling pathway	3.34×10^{-2}
KEGG:04912	GnRH signaling pathway	3.65×10^{-2}
KEGG:04270	Vascular smooth muscle contraction	4.82×10^{-2}

successfully predicted 325 novel miRNAs that were expressed in the ovary of the hybrid laying hens and their purebred parental lines. MicroRNAs with highest abundance were gga-miR-143-3p and gga-miR-148a-3p, which together accounted for approximately 8% of the total expressed miRNAs in the pre-hierarchical follicles of the hens. The top five most abundantly expressed miRNAs constitute approximately 10–15% of the expressed miRNAs in the purebred and the hybrids respectively. Interestingly, all of them were known miRNAs, suggesting that known miRNAs were most abundantly expressed while predicted novel miRNAs were lowly expressed, and hence, the inability to detect them in previous investigations. This is consistent with previous reports for most abundantly expressed miRNAs in the ovary of vertebrates (Kang et al., 2013; Wang et al., 2017; Wu et al., 2017; Wang and Ma, 2019). In particular, gga-miR-143-3p, gga-miR-145-5p, gga-miR-26, gga-miR-99, gga-miR-10a, gga-miR-21, gga-miR-148, gga-miR-199, gga-miR-126, gga-miR-125, gga-miR-101 and gga-let-7 family gene clusters were reported as the most abundantly expressed miRNAs in the ovaries of laying chickens (Kang et al., 2013; Wu et al., 2017). In addition, gga-miR-145-5p gene target mRNAs in the ovary of chickens were enriched in TGF-β signaling pathway by regulating the growth and development of primordial follicles (Kang et al., 2013). In mouse, gga-miR-143 homolog was also highly expressed in the ovary and regulates estradiol synthesis (Zhang et al., 2017) thereby controlling

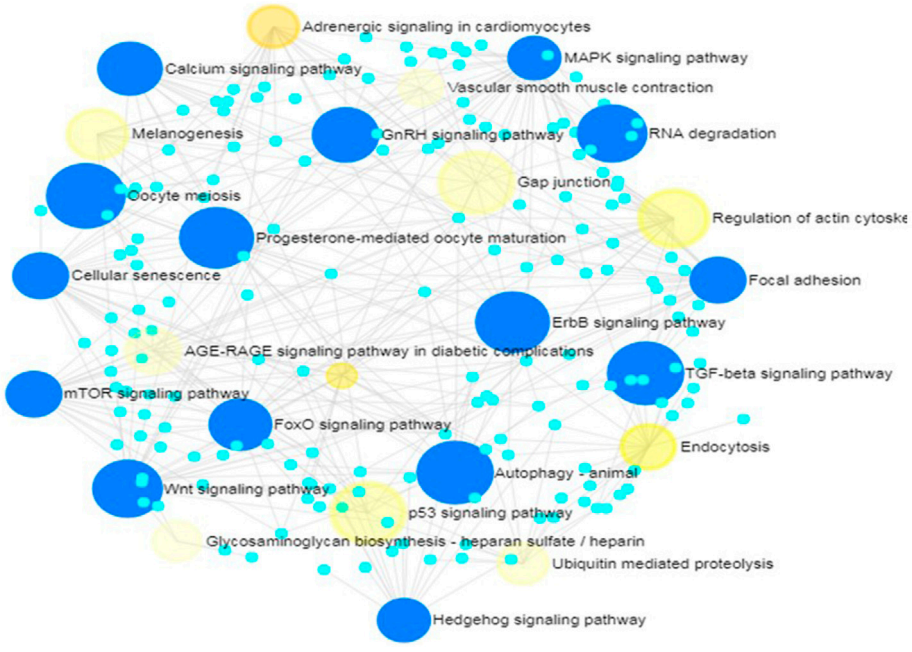


FIGURE 6
Network of enriched KEGG pathways for target genes of non-additively expressed miRNAs. Small cyan circles represent target genes of miRNAs with non-additive expression; Yellow circles represent KEGG pathways enriched by the target genes with p-values<0.05; Large blue circles represent KEGG pathways enriched by the target genes with corrected p-values<0.05.

TABLE 4 REACTOME terms enriched by target genes of miRNAs with non-additive modes of expression in crossbred laying hens.

TERM ID	REACTOME term	P- Adjusted
REAC:R-GGA-381676	Glucagon-like Peptide-1 (GLP1) regulates insulin secretion	0.001040923
REAC:R-GGA-163685	Integration of energy metabolism	0.001040923
REAC:R-GGA-426496	Post-transcriptional silencing by small RNAs	0.001040923
REAC:R-GGA-162582	Signal Transduction	0.001040923
REAC:R-GGA-212436	Generic Transcription Pathway	0.001040923
REAC:R-GGA-9006934	Signaling by Receptor Tyrosine Kinases	0.001040923
REAC:R-GGA-422356	Regulation of insulin secretion	0.001040923
REAC:R-GGA-112316	Neuronal System	0.001040923
REAC:R-GGA-74160	Gene expression (Transcription)	0.001040923
REAC:R-GGA-73857	RNA Polymerase II Transcription	0.001892943
REAC:R-GGA-388844	Receptor-type tyrosine-protein phosphatases	0.013307074
REAC:R-GGA-392517	Rap1 signalling	0.013307074
REAC:R-GGA-6794362	Protein-protein interactions at synapses	0.025576206
REAC:R-GGA-195721	Signaling by WNT	0.037665791
REAC:R-GGA-8878159	Transcriptional regulation by RUNX3	0.047377253
REAC:R-GGA-5627117	RHO GTPases Activate ROCKs	0.047377253

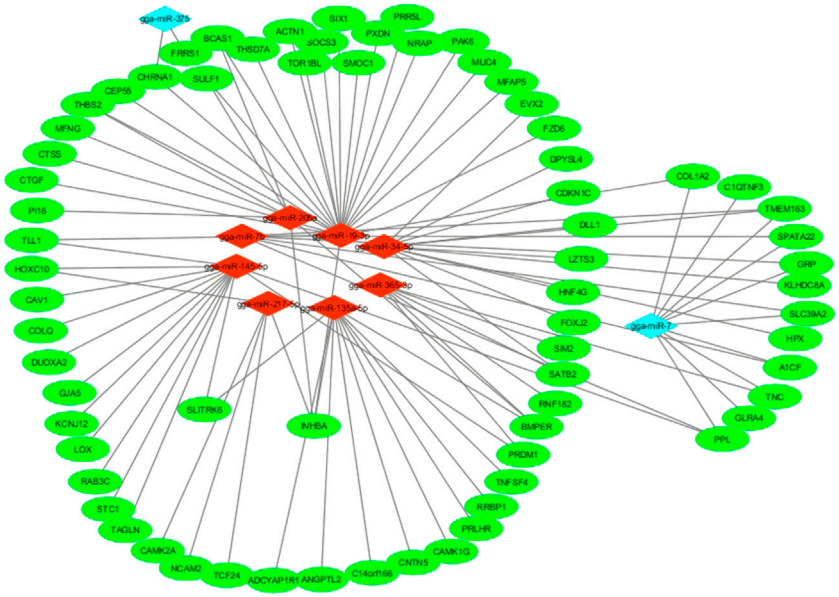


FIGURE 7
mRNA-miRNA regulatory network in the pre-hierarchical follicles of purebred and hybrid laying hens. Red diamonds represent miRNAs with low parent dominance expression; Cyan diamonds represent miRNAs with either high-parent dominance or over-dominance expression in the hybrids; Green ellipses represent target genes of miRNAs with non-additive expression.

ovulation. Loss of gga-miR-145 function in the ovary of mouse model was successfully linked to hyper-activation of primordial follicles, and deregulation of zona pellucida in actively growing follicles (Yang et al., 2013) which might lead to depletion of ovarian reserve. Further, gga-miR-21 in ovaries of chickens was construed to be involved in follicular growth and ovulation (Kang et al., 2013). Taken together, these suggest that miRNAs with abundant expression in the ovary play critical

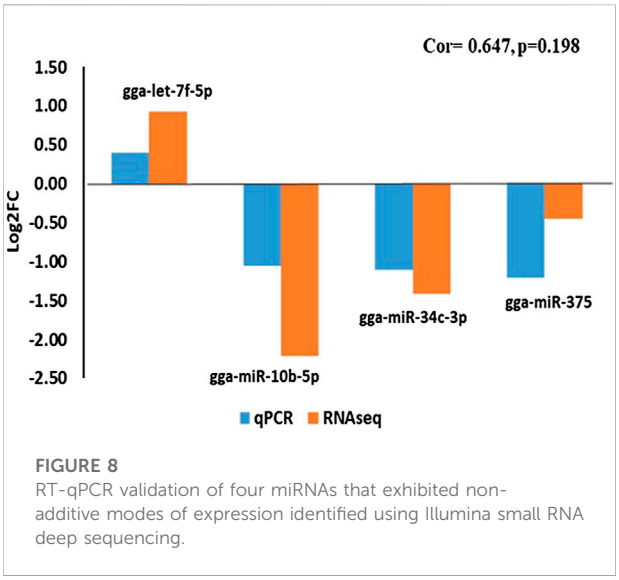


FIGURE 8
RT-qPCR validation of four miRNAs that exhibited non-additive modes of expression identified using Illumina small RNA deep sequencing.

roles in follicular development and ovulation, two processes that regulate egg reproduction in hens.

The main purpose of the present investigation was to identify miRNAs with non-additive expression in the pre-hierarchical ovarian follicles of laying hens. For the first time, our study identified 20 known and novel miRNAs with additive mode of expression, and 50 known and novel miRNAs that demonstrated non-additive expression. Of the 50 DEMiRs that demonstrated non-additive expression, 20 were known while 30 were not previously reported. MicroRNA families that showed non-additive mode of expression include gga-miR-1684, gga-miR-19 and gga-miR-7, which have two members each and gga-miR-34 family which has four members.

MicroRNAs in the gga-miR-1684 family exhibited high-parent dominant expression in both hybrids, while all the four members of the gga-miR-34 family showed low-parent mode of expression. Previous investigation of miRNA transcriptome in the hierarchical follicles of laying chickens revealed that gga-miR-34b and gga-miR-34c members were up-regulated in more efficient laying hens (Wu et al., 2017). In our study, dominance was the major mode of non-additive expression exhibited by known miRNAs (5 high-parent dominant and 14 low-parent dominant). Of note, gga-miR-375 was the only known miRNAs that demonstrated over-dominance mode of expression in the WR hybrid hens. Previously, *in vitro* over-expression study in granulosa cell (GC) line from pig ovary has established that miR-375 regulates E2 synthesis (Yu et al., 2017) and its mode of expression in the present study has confirmed its involvement not only in reproduction but also hybrid vigor for egg laying and clutch traits in hens. However, it is not clear how higher expression of gga-miR-375 in the pre-hierarchical follicles of WR hybrids that demonstrated high heterosis in egg and clutch traits may promote laying of more eggs. Yu et al. has reported co-expression between gga-miR-357 and corticotrophin releasing hormone in the GCs, which synergistically regulate E2 synthesis (Yu et al., 2017). It can be asserted that the higher expression may be exclusive only to the pre-hierarchical follicles. If this was the case, it can be a strategy for preserving follicle pool where only hierarchical follicles will show low expression thereby reducing disrupted follicular selection process typical in low efficient laying birds.

Evidence provided by the result in our study suggests the presence of synergistic effect between miRNAs that exhibited low and high parent dominance in dictating heterosis for egg number and clutch size. Results of the GO and REACTOME enrichment were populated with terms and pathways related to regulation of transcription and gene expression, and suggests the importance of these processes in modulating heterosis for egg number and clutch size in the crossbred laying hens. Further evidences inferred by KEGG enrichment analysis supporting the crucial role of target genes of miRNAs exhibiting non-additive mode of expression are centered in processes related to organization of follicular hierarchy and recruitment. Finely orchestrated follicular development is essential for efficient egg laying typical in commercial hybrid hens. This complex process begins by systematic activation of the quiescent, non-growing primordial follicles to primary oocyte. Follicles that did not follow this cellular fate undergo atresia, thus, reducing follicular reserve. Results of the current study suggests the involvement of cellular senescence in achieving follicle atresia, and genes in the pathway were targeted by miRNAs exhibiting low-parent dominance (gga-miR-19a-3p, gga-miR-1351-5p and gga-miR34c-3p). Hedgehog signaling, the most enriched pathway in the current study, was reported to regulate follicle development and female germ line stem cell proliferation (Jiang et al., 2019). Majority of the target genes enriched in this pathway were targeted by gga-miR-19a-3p and gga-miR-1351-5p, both of which exhibited low-parent dominance in the crossbred hens. Furthermore, pathways for signal transduction including FoxO, WNT, insulin, mTOR, and ErbB were reported to be crucial for follicle activation, growth and development (Kang et al., 2013; Sun et al., 2015; Lee et al., 2016; Tao et al., 2017; Zhao et al., 2018; Zhang et al., 2019; Zhou et al., 2020). Specifically, it was shown that elevated activity of mTORC1 in oocytes causes follicular depletion and premature ovarian failure (Adhikari et al., 2010). Our data corroborates these reports. Genes involved in insulin signaling pathway were reported to be targets for gga-miR-375 (Kang et al., 2013) similar to the observation in the current study.

TGF-β signaling and the genes that enriched the pathway (INHBB, INHBA, PITX2, Activin) support maintenance of ovarian reserve which diminishes owing to either follicle activation, recruitment, maturation and subsequent ovulation,

or through atresia (Pelosi et al., 2015). It is asserted here that oocyte maturation for follicles destined to ovulate were facilitated via key pathways including oocyte meiosis, GnRH-signaling and progesterone-mediated oocyte maturation. Many of these pathways were significantly enriched in the ovaries of actively laying geese, and laying chickens with high egg production (Luan et al., 2014; Zhang et al., 2019). Recent investigation reported that focal adhesion and gap junction pathways were enriched in the hypothalamus-pituitary-gonadal tissues, which were linked to high egg production in geese (Wu et al., 2020). Similarly, our data support the involvement of these pathways enriched by genes targeted by miRNAs with non-additive expression in the pre-hierarchical follicles of laying hens.

Based on mRNA-miRNA co-expression network, target genes for gga-miR-19-3p, gga-miR-34-5p, gga-miR-217-5p, gga-miR-135a-5p, gga-miR-205a, gga-miR-375, and gga-miR-7 may underlie heterosis for egg laying in hybrid hens. Specifically, gga-miR-34-5p and gga-miR-205a promote cell apoptosis and were significantly expressed in small yellow follicles of laying chickens and geese (Wu et al., 2020; Hou et al., 2021) while gga-miR-145a-5p promotes premature progesterone release in granulosa cells of pre-hierarchical follicles (McBride et al., 2012). These processes could induce atresia with consequent erosion of follicle reserve. Furthermore, the two candidate miRNAs with high-parent or overdominance expression (gga-miR-375 and gga-miR-7) were associated with regulation of cell proliferation (Yuan et al., 2015), inhibition of synthesis and secretion of gonadotropins (He et al., 2020) essential for finely regulated follicular hierarchy establishment. Taken together, the interplay of miRNAs with non-additive expression and their target genes could underlie heterosis in egg production and clutch traits observed in the hybrid chickens.

Conclusion

This study was designed to characterize miRNAs with non-additive expression in the follicles of purebred and crossbred hens, and investigate the functions of miRNAs in modulating heterosis for egg number and clutch size. To achieve that aim, mode of miRNA expression was characterized by miRNA sequencing. A total of 50 miRNAs including 30 novel, were found to exhibit non-additive expression. Dominance was the predominant mode of expression exhibited by majority of the miRNAs. Functional analysis of target genes of the known miRNAs with non-additive expression were significantly enriched in GO terms related to regulation of transcription, metabolic processes and gene expression. KEGG pathways enriched in target genes of non-additive genes include hedgehog, cellular senescence, WNT, TGF-beta, progesterone-mediated oocyte maturation, oocyte meiosis, GnRH signaling, signal transduction and transcription regulation, which could be linked to primordial follicle activation, growth and ovulation. mRNA-miRNA co-expression network constructed using mRNA and miRNA suggest gga-miR-19 family, gga-miR-375, gga-miR-205a-3p, gga-miR-375, and gga-miR-7 family are candidate miRNAs that play synergistic roles in maintenance of organized follicular growth and development which may influence heterosis for egg number and clutch size in crossbred hens.

Materials and methods

Experimental birds

Experimental procedure for all animal experiments was approved by the Animal Care and Use Committee of the Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing (IAS-CAAS). The procedure for the generation of the experimental birds involves crossbreeding of R and W purebred lines to produce offspring of the purebred lines; W, R and their reciprocal crossbred RW and WR. Briefly, semen from R and W purebred sires were artificially inseminated to purebred dams of their lines and the other lines in a reciprocal crossing design to obtain half-sib purebred and crossbred chicks. Four genotypes of chicks were hatched. Female chicks so generated were vaccinated, wing banded and raised. All hens were managed under standard housing and fed appropriate kind of diet specific to age and developmental stage of the birds. The birds were offered unrestricted access to feed and water throughout the period of brooding, rearing and laying.

Phenotypic data recording

Egg laying was recorded for individual hens once daily. Total egg number and clutch size up to 53 weeks of age were computed for individual hens and selected hens with phenotypic records corresponding to their population average. Least square means for egg number was compared between genetic groups using Tukey-Kramer HSD at $p < 0.05$, using the model below;

$$Y_{ij} = \mu + G_i + \xi_{ij}$$

Where $\mathbf{Y_{ij}}$ is the phenotype, $\boldsymbol{\mu}$ is the population average, $\mathbf{G_i}$ is the fixed effect of genetic group, and $\boldsymbol{\varepsilon_{ij}}$ is the random error.

Heterosis was calculated using the model below;

$$Heterosis\,(\%) = \left[\frac{F_1 - \left(\frac{P_1+P_2}{2}\right)}{\frac{P_1+P_1}{2}}\right] \times 100$$

Where F_1 is the performance value of the hybrid, P_1 and P_2 are the performance values of the two parental lines.

Tissue collection

Four birds each from the purebred parental and reciprocal crossbred populations (R, W, RW and WR) that have egg laying records corresponding to the average of their populations were exsanguinated by cervical dislocation. After slaughter, the pullets were dissected by ventral midline incision and the intact reproductive tract of the hens were collected, weighed and separated into ovary (and follicles) and oviduct. Pre-hierarchical follicles (4–8 mm in diameter) were sorted and collected from each bird, snap frozen in liquid nitrogen, taken to laboratory and stored at -80 °C until RNA extraction.

RNA purification and microRNA library preparation

Pre-hierarchical follicles (30–50 mg/sample) were used for total RNA extracted using Trizol reagent (Invitrogen). A total of 16 samples with four each belonging to the W, R, RW and WR were used. RNA concentration and integrity were determined using NanoDrop. ND-1000 (NanoDrop Technologies, Wilmington, DE, United States) and Agilent 2100 Bioanalyzer (Agilent Technologies, CA, United States) respectively. Only RNA sample with concentration >200 ng/μL, RIN >7 and 28S/18S rRNA ratio >1.7 was used for the small RNA library construction. An amount of 1 μg total RNA was used in the construction of each of the libraries (n = 16) using Illumina TruSeq.

Small RNA sequence data analysis

In silico processing of the generated sequence reads was carried out in accordance with established pipelines (Korpelainen et al., 2014; Do et al., 2018). Briefly, raw sequence reads (16 fasta files) were subjected to quality control using FastQC (He et al., 2020) where adapter primers and poor quality reads were removed. Trimming of 3′ and 5′ adapter sequences was achieved using in-house pipeline developed by Annaroad Gene Technology Co., Ltd. (Beijing, China). Only reads with Phred score >20 and sequence read between 18 and 30 nt were retained. Clean reads that passed the quality control criteria were parsed to build a reference genome index. Bowtie 1 (http://bowtie-bio.sourceforge.net/index.shtml) (Langmead et al., 2009) was used to align the built reference genome to GRCg6a (GCA_000002315.5) chicken genome downloaded from the Ensembl database.

Identification and discovery of known and novel miRNAs

The identification of known miRNAs was performed in miRBase (v22.1) (Andrews, 2017). Discovery of novel miRNAs was achieved after excluding known miRNAs. Reads that mapped to other small RNA species (rRNA, tRNA, snRNA, and snoRNA) Rfam (v.13.0) in the RNA family database (http://rfam.xfam.org/) (Kozomara et al., 2019) and those that mapped to the repeated regions of the genome were excluded. Identification of the repeated regions from the uniquely mapped reads was achieved using RepeatMasker v.4.0.9 (http://repeatmasker.org/cgi-bin/WEBRepeatMasker). MirDeep2 program that implements a probabilistic algorithm based on the miRNA biogenesis model was used to predict novel miRNAs in each library (Kalvari et al., 2018; Friedländer et al., 2012).

Differential miRNA expression

MiRNAs that met the criteria of being either known or novel were used for differential expression analysis. DEGseq (v1.18.0) package implemented in Bioconductor which follows the assumption of binomial distribution was used for differential gene expression analysis (http://www.bioconductor.org/packages/release/bioc/html/DEGseq.html). Only miRNAs that fulfill the Benjamini and Hochberg criteria for multiple testing correction (false discovery rate ($q < 0.05$) and $|\text{Log}_2\text{\_ratio}| \geq 1$ were identified as DEMiRs. Differential expression of miRNAs was achieved based on pairwise comparisons of R vs W, R vs RW, R vs WR, W vs RW, and W vs WR. Additionally, the reciprocal crosses were also compared with a synthetic group of mid-parent miRNA expression values, which were calculated by taking the means of normalized gene count from combinations of paternal lines (A = 1/2 (R + W)). Consequently, miRNAs with additive and various forms of non-additive modes of expression were identified. Delineation of DEMiRs into additive, high-parent dominant, low-parent dominant, over-dominant and under-dominant were in accordance with previous studies (Swanson-Wagner et al., 2006; Mai et al., 2019) with little modification. Briefly, additivity occurs when expression of the miRNA was significantly different between the two parental purebred lines (R vs. W Padj.< 0.05, $\text{Log}_2\text{FC} \geq 1$) and that the miRNA expression in the crossbred was similar to the synthesized mean of their parental purebred lines. The high-parent dominant mode was when expression of the miRNA in the crossbred was significantly higher than one parent but similar to the other parental line. The low-parent dominant mode was when expression in the hybrid was significantly lower than one parental line but similar to the other parental line. Over-dominant mode occurs when miRNA expression in the crossbred was significantly higher than either of the two parental purebred lines. Under-dominance was when gene expression in the hybrid was significantly lower than the two parental purebred lines. This classification was achieved by subjection the DEMiRs to systematic comparisons in Venny 2.0 (Oliveros, 2007).

Target genes of the DEMiRs prediction and their enrichment

Target genes for known DEMiRs that exhibited non-additive pattern of expression were predicted with TargeScan release 7.2 (http://www.targetscan.org/vert_72/) (Agarwal et al., 2015). Catalogue of such target genes were filtered against the mRNA transcriptome obtained from the pre-hierarchical follicle tissues of the experimental birds. Only mRNAs common in target genes and expressed genes were retained for gene enrichment analysis. The filtered target genes for the non-additively expressed miRNAs in the hybrid chickens were uploaded into g. Profiler (https://biit.cs.ut.ee/gprofiler/gost) (Raudvere et al., 2019) for gene set enrichment of analysis. Gene set enrichment was carried out in Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome (REAC) databases. The g. profiler algorithm adopts over-representation analysis approach that uses hypergeometric test to measure the significance of functional terms in the input gene list (Raudvere et al., 2019). Significant threshold used was Benjamini–Hochberg corrected FDR threshold at 0.05. Network mapping of interconnectedness of significantly enriched KEGG pathways was achieved using GeneAnalyst (Zhou et al., 2019).

Construction of mRNA-miRNA co-expression network

We constructed mRNA-miRNA regulatory network in order to identify candidate miRNAs with non-additive expression in the pre-hierarchical follicles of hybrid laying hens. Only mRNAs and miRNAs obtained from RNA sequencing of the samples were used. mRNA-miRNA pairs were delineated if their expression had pairwise correction coefficient of -0.9 or lower. The network was visualized using the cytoscape software.

Real-time quantitative PCR

The total RNA used for miRNA-sequencing was also used for Real-time quantitative PCR (RT-qPCR) to validate the expression of four non-additively expressed DEMiRs (gga-let-7f-5p, gga-miR-10b-5p, gga-miR-34c-3p and gga-miR-375) following previous report (Agarwal et al., 2015). miScript II RT kit (Qiagen) was used for the RT stage. Approximately 1 μL (1 μg) diluted RNA were added to the reaction mix containing 4 μL 5× miScript RT Buffer, 1 μL miScript Reverse Transcriptase Mix, and 15 μL RNase-free water to a final volume of 20 μL. RT reaction conditions consisted of incubation for 60 min at 37°C, incubating for 5 min at 95°C. Real-time PCR reactions were prepared using miScript SYBR Green PCR based on the manufacturer's protocol. Briefly, 1 μL of the cDNA was added to the reaction mix containing 10 μL 2 × QuantiTect SYBR Green PCR Master Mix, 2 μL 10× miScript universal Primer, 2 μL 10× miScript Primer Assay to a final volume of 20 μL. The qRT-PCR condition consisted of denaturing at 95 °C for 15 min, followed by 40 cycles consisting 95 °C for 15 s, 55 °C 30 s, 70 °C 30 s. Samples were run in three technical replicates. Relative abundance of the micRNA transcripts was calculated using the $2^{-\Delta\Delta CT}$ method. miRNAs expression result from qRT-PCR and their normalized expression from the TrueSeq were plotted in a histogram after calculating their correlation in each sample.

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: The transcriptome data are available in the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) at NCBI, with the BioProject ID: PRJNA859020 and SRA Accession Number: SAMN29766694-29766709.

Ethics statement

The animal study was reviewed and approved by the Animal Care and Use Committee of the Institute of Animal Science, Chinese Academy of Agricultural Sciences (IAS-CAAS, No. IAS2020–12).

Author contributions

Conceptualization, JC, AI, and YS; methodology, AI, YL, JF, PW, AN, HT, and LS; software, AI, YS, and YL; validation, AN and YM; formal analysis, AI, JY, and YS; investigation, AI, AN, PW, and YL; resources, JC, YS, and HM.; data curation, AI and YS; writing—original draft preparation, AI; writing—review and editing, YS, JC, and HM; visualization, AI; supervision, JC, YS, and HM; project administration, JC and YS; funding acquisition, JC. All authors have read and approved to the publication of the manuscript.

Funding

This research was funded by the National Natural Science Foundation (grant number 32172721), Chinese Agricultural Research System (grant number CARS-40), the Fundamental

Research Funds for Central Non-profit Scientific Institution (2021-YWF-ZYSQ-12), and the Agricultural Science and Technology Innovation Program (grant number ASTIP-IAS04).

Acknowledgments

The authors thank Yong Zhao and Chunjie Niu (Beijing Bainianliyuan Beijing-Yopu breeding company, Beijing, China) for their assistance in data recording.

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/fgene.2022.974619/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Mapping statistics for the 16 miRNA libraries in parental purebred lines and their reciprocal hybrids.

SUPPLEMENTARY FIGURE S2
Venn diagram depicting number of DEMiRs with additive mode of expression.

SUPPLEMENTARY TABLE S1
RNA-seq reads mapped to non-coding regions other than miRNAs expressed in the pre-hierarchical follicles of purebred chicken parental lines (R and W) and their reciprocal hybrids (RW and WR).

SUPPLEMENTARY TABLE S2
Expressed miRNAs in the pre-hierarchical follicles of purebred chicken parental lines (R and W) and their reciprocal hybrids (RW and WR).

SUPPLEMENTARY TABLE S3
Mode of expression of novel miRNAs in the pre-hierarchical follicles of purebred chicken parental lines (R and W) and their reciprocal hybrids (RW and WR).

SUPPLEMENTARY TABLE S4
Target genes for known miRNAs with non-additive mode of expression in the pre-hierarchical follicles of purebred chicken parental lines (R and W) and their reciprocal hybrids (RW and WR).

SUPPLEMENTARY TABLE S5
GO, KEGG and REAC terms enriched in target genes with non-additive mode of expression in the pre-hierarchical follicles of purebred chicken parental lines (R and W) and their reciprocal hybrids (RW and WR).

References

Adhikari, D., Zheng, W., Shen, Y., Gorre, N., Hämäläinen, T., Cooney, A. J., et al. (2010). Tsc/mTORC1 signaling in oocytes governs the quiescence and activation of primordial follicles. *Hum. Mol. Genet.* 19, 397–410. doi:10.1093/hmg/ddp483

Agarwal, V., Bell, G. W., Nam, J. W., and Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *eLife* 4, e05005. doi:10.7554/eLife.05005

Amuzu-Aweh, E. N. (2020). *Genomics of heterosis and egg production in White Leghorns. PhD thesis*. Uppsala: Swedish University of Agricultural Sciences.

Andrews, S. FastQC: A quality control tool for high throughput sequence data. Babr Bionf (2017).

Brock, D. C., and Moore, G. E. (2006). *Understanding moore's law: Four decades of innovation*. Philadelphia, United States: Chemical Heritage Foundation.

Do, D. N., Dudemaine, P. L., Fomenky, B. E., and Ibeagha-Awemu, E. M. (2018). Integration of miRNA and mRNA co-expression reveals potential regulatory roles of miRNAs in developmental and immunological processes in calf ileum during early growth. *Cells* 7, 134. doi:10.3390/cells7090134

Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40, 37–52. doi:10.1093/nar/gkr688

He, J., Xu, S., Ji, Z., Sun, Y., Cai, B., Zhang, S., et al. (2020). The role of miR-7 as a potential switch in the mouse hypothalamus-pituitary-ovary axis through regulation of gonadotropins. *Mol. Cell. Endocrinol.* 518, 110969. doi:10.1016/j.mce.2020.110969

Hou, L., Ji, W., Gu, T., Weng, K., Liu, D., Zhang, Y., et al. (2021). MiR-34c-5p promotes granulosa cells apoptosis by targeting Bcl2 in broody goose (*Anser cygnoides*). *Anim. Biotechnol.*, 1–9. doi:10.1080/10495398.2021.1886943

Jiang, Y., Zhu, D., Liu, W., Qin, Q., Fang, Z., and Pan, Z. (2019). Hedgehog pathway inhibition causes primary follicle atresia and decreases female germline stem cell proliferation capacity or stemness. *Stem Cell. Res. Ther.* 10, 198. doi:10.1186/s13287-019-1299-5

Johnson, A. L. (2015). Ovarian follicle selection and granulosa cell differentiation. *Poult. Sci.* 94, 781–785. doi:10.3382/ps/peu008

Johnson, P. A. (2012). Follicle selection in the avian ovary. *Repr. Dom. Anim.* 47, 283–287. doi:10.1111/j.1439-0531.2012.02087.x

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E. P., Rivas, E., Eddy, S. R., et al. (2018). Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* 46, D335–D342. doi:10.1093/nar/gkx1038

Kang, L., Cui, X., Zhang, Y., Yang, C., and Jiang, Y. (2013). Identification of miRNAs associated with sexual maturity in chicken ovary by Illumina small RNA deep sequencing. *BMC Genomics* 14, 352. doi:10.1186/1471-2164-14-352

Korpelainen, E., Tuimala, J., Somervuo, P., Huss, M., and Wong, G. (2014). *RNA-Seq data analysis: A practical approach*. New York, United States: CRC Press.

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 47, D155–D162. doi:10.1093/nar/gky1141

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25. doi:10.1186/gb-2009-10-3-r25

Lee, H. C., Lim, S., and Han, Y. (2016). Wnt/β-catenin signaling pathway activation is required for proliferation of chicken primordial germ cells *in vitro*. *Sci. Rep.* 6, 34510–34518. doi:10.1038/srep34510

Lee, R., Feinbaum, R. L., and Ambros, V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell.* 75, 843–854. doi:10.1016/0092-8674(93)90529-y

Liu, Y., Sun, Y., Li, Y., Bai, H., Xu, S., Xu, H., et al. (2018). Identification and differential expression of microRNAs in the testis of chicken with high and low sperm motility. *Theriogenology* 122, 94–101. doi:10.1016/j.theriogenology.2018.09.010

Luan, X., Liu, D., Cao, Z., Luo, L., Liu, M., Gao, M., et al. (2014). Transcriptome profiling identifies differentially expressed genes in Huoyan goose ovaries between the laying period and ceased period. *PLOS One* 9, e113211. doi:10.1371/journal.pone.0113211

Mai, C., Wen, C., Sun, C., Xu, Z., Chen, S., and Yang, N. J. G. (2019). Implications of gene inheritance patterns on the heterosis of abdominal fat deposition in chickens. *Genes.* 10, 824. doi:10.3390/genes10100824

McBride, D., Carre, W., Sontakke, S., Hogg, C. O., Law, A., Donadeu, F. X., et al. (2012). Identification of miRNAs associated with the follicular-lutealtransition in the ruminant ovary. *Reproduction* 144, 221–233. doi:10.1530/REP-12-0025

Oliveros, J. C. (2007). VENNY. An interactive tool for comparing lists with Venn Diagrams. https://bioinfogp.cnb.csic.es/tools/venny/index.html.

Pelosi, E., Simonsick, E., Forabosco, A., Garcia-Ortiz, J. E., and Schlessinger, D. (2015). Dynamics of the ovarian reserve and impact of genetic and epidemiological factors on age of menopause. *Biol. Reprod.* 92, 130–139. doi:10.1095/biolreprod.114.127381

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, W191–W198. doi:10.1093/nar/gkz369

Riffo-Campos, Á. L., Riquelme, I., and Brebi-Mieville, P. (2016). Tools for sequence-based miRNA target prediction: What to choose? *Int. J. Mol. Sci.* 17, 1987. doi:10.3390/ijms17121987

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA hypothesis: The rosetta stone of a hidden RNA language? *Cell.* 146, 353–358. doi:10.1016/j.cell.2011.07.014

Sun, X., Su, Y., He, Y., Zhang, J., Liu, W., Zhang, H., et al. (2015). New strategy for *in vitro* activation of primordial follicles with mTOR and PI3K stimulators. *Cell. cycle* 14, 721–731. doi:10.1080/15384101.2014.995496

Swanson-Wagner, R. A., Jia, Y., DeCook, R., Borsuk, L. A., Nettleton, D., and Schnable, P. S. (2006). All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. *Proc. Natl. Acad. Sci. U. S. A.* 103, 6805–6810. doi:10.1073/pnas.0510430103

Tao, Z., Song, W., Zhu, C., Xu, W., Liu, H., Zhang, S., et al. (2017). Comparative transcriptomic analysis of high and low egg-producing duck ovaries. *Poult. Sci.* 96, 4378–4388. doi:10.3382/ps/pex229

Wang, C., and Ma, W. (2019). Hypothalamic and pituitary transcriptome profiling using RNA-sequencing in high-yielding and low-yielding laying hens. *Sci. Rep.* 9, 10285. doi:10.1038/s41598-019-46807-3

Wang, P., Li, X., Cao, L., Huang, S., Li, H., Zhang, Y., et al. (2017). MicroRNA-148a overexpression improves the early development of porcine somatic cell nuclear transfer embryos. *PLoS One*, e0180535. doi:10.1371/journal.pone.0180535

Wu, N., Gaur, U., Zhu, Q., Chen, B., Xu, Z., Zhao, X., et al. (2017). Expressed micro RNA associated with high rate of egg production in chicken ovarian follicles. *Anim. Genet.* 48, 205–216. doi:10.1111/age.12516

Wu, Y., Zhao, X., Chen, L., Wang, J., Duan, Y., Li, H., et al. (2020). Transcriptomic analyses of the hypothalamic-pituitary-gonadal Axis identify candidate genes related to egg production in xinjiang yili geese. *Animals*. 10, 90. doi:10.3390/ani10010090

Yang, S., Wang, S., Luo, A., Ding, T., Lai, Z., Shen, W., et al. (2013). Expression patterns and regulatory functions of microRNAs during the initiation of primordial follicle development in the neonatal mouse ovary. *Biol. Reprod.* 89 (126), 126–111. doi:10.1095/biolreprod.113.107730

Yoon, S. Y., Kim, R., Jang, H., Shin, D. H., Lee, J. I., Seol, D., et al. (2020). Peroxisome Proliferator-activated receptor gamma modulator promotes neonatal mouse primordial follicle activation *in vitro*. *Int. J. Mol. Sci.* 21, 3120. doi:10.3390/ijms21093120

Yu, C., Li, M., Wang, Y., Liu, Y., Yan, C., Pan, J., et al. (2017). Exposure to the Chinese famine in early life and hypertension prevalence risk in adults. *J. Hypertens.* 153, 63–68. doi:10.1097/HJH.0000000000001122

Yuan, B., Sun, G. J., Zhang, G. L., Wu, J., Xu, C., Dai, L. S., et al. (2015). Identification of target genes for adenohypophysis-prefer miR-7 and miR-375 in cattle. *Genet. Mol. Res.* 14 (3), 9753–9763. doi:10.4238/2015.August.19.8

Zhang, L., Zhang, X-J., Lu, Y., Li, L., Sheng, C., and Cui, S. (2017). MiRNA-143 mediates the proliferative signaling pathway of FSH and regulates estradiol production. *J. Endocrinol.* 234, 1–14. doi:10.1530/JOE-16-0488

Zhang, T., Chen, L., Han, K., Zhang, X., Zhang, G., Dai, G., et al. (2019). Transcriptome analysis of ovary in relatively greater and lesser egg producing Jinghai Yellow Chicken. *Anim. Reprod. Sci.* 208, 106114. doi:10.1016/j.anireprosci.2019.106114

Zhao, Y., Zhang, Y., Li, J., Zheng, N., Xu, X., Yang, J., et al. (2018). MAPK3/1 participates in the activation of primordial follicles through mTORC1-KITL signaling. *J. Cell. Physiol.* 233, 226–237. doi:10.1002/jcp.25868

Zhou, G., Soufan, O., Ewald, J., Hancock, R. E., Basu, N., and Xia, J. (2019). NetworkAnalyst 3.0: A visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* 47, W234–W241. doi:10.1093/nar/gkz240

Zhou, S., Ma, Y., Zhao, D., Mi, Y., and Zhang, C. (2020). Transcriptome profiling analysis of underlying regulation of growing follicle development in the chicken. *Poult. Sci.* 99, 2861–2872. doi:10.1016/j.psj.2019.12.067

Zhuo, Z., Lamont, S. J., and Abasht, B. (2019). RNA-seq analyses identify additivity as the predominant gene expression pattern in F1 chicken embryonic brain and liver. *Genes.* 10, 27. doi:10.3390/genes10010027

frontiers | Frontiers in Genetics

TYPE Data Report
PUBLISHED 05 October 2022
DOI 10.3389/fgene.2022.1009127

OPEN ACCESS

EDITED BY
Turgay Unver,
FicusBio, Turkey

REVIEWED BY
Emre Ilhan,
Erzurum Technical University, Turkey
Hassan Ghazal,
National Center for Scientific and Technical Research (CNRST), Morocco

*CORRESPONDENCE
Dipankar Chakraborti,
dcgntcs@caluniv.ac.in

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

RECEIVED 01 August 2022
ACCEPTED 20 September 2022
PUBLISHED 05 October 2022

CITATION
Ghosh S, Purohit A, Hazra A, Mukherjee A, Bhar A, Gupta S, Chaudhuri RK and Chakraborti D (2022), Differential transcript expression profiles of susceptible and resistant pigeonpea cultivars at an early time point during Fusarium udum infection. *Front. Genet.* 13:1009127. doi: 10.3389/fgene.2022.1009127

Differential transcript expression profiles of susceptible and resistant pigeonpea cultivars at an early time point during *Fusarium udum* infection

Sanatan Ghosh[1], Arnab Purohit[2], Anjan Hazra[1], Aloleca Mukherjee[1], Anirban Bhar[3], Sumanti Gupta[4], Rituparna Kundu Chaudhuri[5] and Dipankar Chakraborti[1]*

[1]Department of Genetics, University of Calcutta, Kolkata, West Bengal, India, [2]EVA.4 Unit, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Prague, Czechia, [3]Post Graduate Department of Botany, Ramakrishna Mission Vivekananda Centenary College (Autonomous), Kolkata, West Bengal, India, [4]Department of Botany, Rabindra Mahavidyalaya, Hooghly, West Bengal, India, [5]Department of Botany, Barasat Govt. College, Kolkata, West Bengal, India

KEYWORDS

comparative transcriptomics, *Fusarium* wilt, next generation sequencing, fungal invasion, gene ontology

Introduction

Pigeonpea [*Cajanus* cajan (L.) Millspaugh] is ranked seventh among the legume crops, in terms of production, and is grown in arid and semiarid tropical regions of Asia, Africa, the Caribbean region, Latin America, and Australia. It is rich in vegetable protein (20%–22%), and its global productivity is nearly 5.012 million tonnes (FAO 2020). Pigeonpea is mostly grown as a field crop and as a backyard crop in more than 80 countries all over the world (Sameer Kumar et al., 2017). It is cultivated on 5.62 million hectares of land across the world, and India contributes 64% (2.85 million tons) of global production (Saxena et al., 2017). Pigeonpea is the second-most significant crop legume in India, mostly consumed as "dal." Seeds are an important source of protein for humans, whereas stems and leaves are used as fuel and animal feed. Vascular wilt caused by *Fusarium udum* (Butler) is the most damaging disease in pigeonpea and results in an annual loss of approximately 470,000 tons of grain in India (Saxena et al., 2017). *F. udum* is a soilborne, mitosporic, and necrotrophic fungus without known sexual stages in its lifecycle (Agrios 2008). *F. udum* produces three types of asexual spores, namely, thick-walled chlamydospores, 2–6 celled macroconidia, and 1–2 celled microconidia. The most frequently and abundantly produced spores are microconidia which are also found inside the infected host's vascular system. Macroconidia are primarily found on the surface of infected and dead host plants in sporodochia-like groupings. The old mycelium of the pathogen produces chlamydospores which can survive in the soil for a very long period (Purohit et al., 2017).

In pigeonpea, *F. udum*-mediated vascular wilt occurs at the early or late flowering, podding, or even seedling stages (Choudhury 2010). Xylem vessel clogging is an important phenomenon that leads to wilt in infected pigeonpea. Infected xylem

vessels of roots and stems become clogged with spores and mycelia of the pathogen, as well as polysaccharides produced by it. Additionally, xylem parenchyma cells of the infected root are induced by the pathogen to divide excessively. This situation combined with weaker and thinner xylem vessel walls causes a reduced diameter or complete collapse (Agrios 2008). *F. udum* also secretes toxins that are eventually carried to the leaves through xylem vessels, reducing chlorophyll synthesis and disrupting leaf cell membrane permeability and thereby leading to transpirational water loss. At the later stage of infection, the host plant shows wilting, yellowish leaf color, interveinal necrosis, and eventually death. Based on the growth stage of the infected plant and the severity of the wilt, yield loss can be up to 100% (Reddy et al., 1990).

Management of *F. udum* wilt was achieved through chemical treatment of seeds, crop rotation, and the development of biocontrol agents. Although wilt-resistant cultivars have been established through breeding programs dealing with pigeonpea, their usage was restricted due to the presence of variability among the pathogens and the existence of location-specific pathogenic races. As a result, the breakdown of resistance was evident in resistant cultivars (Dhar et al., 2012). Under this situation, resistance (R) gene pyramiding could be a promising strategy to develop resistant genotypes. However, all the R genes were not mapped, and the lack of information leads to very complex time-consuming marker-assisted breeding programs. Previous research works were also focused on the characterization of *F. udum* isolates from India, taken from various geographic regions, using molecular variability, cultural characteristics, and pathogenesis in pigeonpea. Thirteen *F. udum* isolates from India were identified through cultural and morphological methods and molecular fingerprinting techniques (Dhar et al., 2012; Purohit et al., 2017). Four variants and seven phylogenetic groups were established from the RAPD and AFLP data, respectively. The pathogenesis of nine infective isolates was studied, and the timing of the invasion of the pathogen, clogging of vascular bundles, drooping of leaves, and complete wilting were demonstrated in pigeonpea by the present group. Invasion of all nine isolates of *F. udum* in healthy pigeonpea root was identified at 24–36 hours post-inoculation (HPI), determined through anatomical, morphological, and biochemical studies of the infected root (Purohit et al., 2017). At 36 HPI (i.e., invasion stage), cDNA-AFLP-mediated comparative transcriptomics study on *F. udum* inoculated root of wilt-susceptible (ICP 2376) and wilt-resistant pigeonpea cultivars (ICP 8863) along with their mock inoculated controls were performed by Purohit et al. (2021). Among all differentially expressed transcript-derived fragments (TDFs), many were identified to be involved in disease resistance or tolerance mechanisms. The identified important defense responsive pathways were pathogen-triggered immunity, effector-triggered immunity, reactive oxygen species-mediated signaling, salicylic acid- and jasmonic acid-mediated defense responses, cell wall remodeling, vascular development and pattering, and abscisic acid-mediated responses. These pathways were found to be activated during pathogen attacks and played crucial roles in defense responses (Purohit et al., 2021).

Next-generation sequencing (NGS)-mediated transcriptome profiling is a convenient method to recognize the mechanism of host defense during *Fusarium* wilt. It is very useful to identify differentially regulated genes and novel signaling pathways associated with wilt-resistance mechanisms. RNA-seq-based NGS studies have been conducted during *Fusarium* wilt in the model system, *Arabidopsis thaliana*, and various legumes, such as common bean (*Phaseolus vulgaris*), pea, chickpea, soybean, and pigeonpea. Soren et al. (2021) reported RNA-seq data through comparative analysis in Bahar (susceptible) and KPL-44 (resistant) pigeonpea cultivars at late infection stages such as 72 HPI and 96 HPI in response to *F. udum* attack. They considered the untreated susceptible and resistant plants at 0 h as controls for both infection time points.

On the basis of our previous report, it was found that at the early stage of *F. udum* infection, maximum disease-responsive pathways were altered in pigeonpea (Purohit et al., 2021). Accordingly, the present study was designed to perform NGS-based comparative transcriptomic profiling of *F. udum*-induced transcripts at the early (36 HPI) infection stage in pigeonpea through an RNA-seq technique to accomplish a thorough understanding of the differentially expressed genes. This will help to obtain comprehensive knowledge of the tolerant or resistant mechanisms, underlying metabolic pathways influenced by *F. udum*, and putative regulatory genes which are involved in complex spatiotemporal regulation for wilt resistance.

Materials and methods

Cultivars of pigeonpea

Seeds of wilt-susceptible ICP 2376 and wilt-resistant ICP 8863 cultivars were collected from International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, Telangana, India. After surface sterilization of seeds, susceptible and resistant seedlings were grown in soilrite at 22–25°C, 35%–40% humidity, and a 16-h photoperiod (Purohit et al., 2021).

F. udum isolate

F. udum isolate MTCC 2204 was collected from the Microbial Type Culture Collection and Gene Bank of Institute of Microbial Technology, Chandigarh, India. A single colony of MTCC 2204 was inoculated in potato dextrose broth (PDB) and

potato dextrose agar (PDA) media and permitted to incubate at 25°C. After 8–10 days of growth, sporulation was checked.

Inoculation of resistant and susceptible cultivars with the MTCC 2204 pathogenic isolate

Seeds of the ICP 2376 and ICP 8863 cultivars were surface sterilized using 0.05% mercuric chloride ($HgCl_2$) and germinated in a sterile cotton bed. After germination, seedlings were carefully established in soilrite-filled pots. All the plants were maintained at 35%–40% humidity, and 22–25°C with a 16-h photoperiod in a plant growth chamber. After 14–15 days, inoculation of seedlings was performed with the *F. udum* MTCC 2204 isolate as described previously (Purohit et al., 2017).

M1 isolate was cultured in PDB, and conidia were harvested from two-week-old suspension culture. The conidial concentration was adjusted to 1×10^6 ml^{-1} in PDB. Two hundred grams of sand:chickpea meal (9:1) was mixed thoroughly with 50 ml of the MTCC 2204 spore suspension. This mixture was incubated in dark for 14 days at 25°C. Then, this sand:chickpea meal infested with *F. udum* was mixed thoroughly with a sand:soilrite (1:1) mix. The mix was filled in pots, and both susceptible and resistant seedlings were transplanted. A sand:soilrite (1:1) mix without fungal inoculum was used for transplanting control plants from both cultivars. Fifteen plants, each of ICP 2376 and ICP 8863, were inoculated with MTCC 2204 isolate and another 15 plants, each of ICP 2376 and ICP 8863, were used as negative controls. All of the seedlings were maintained under the previously mentioned growth conditions. Each experiment was repeated three times.

RNA isolation and double-stranded cDNA preparation

Roots of noninoculated susceptible (NIS), noninoculated resistant (NIR), MTCC 2204 inoculated susceptible (IS), and inoculated resistant (IR) plants were obtained at 36 hours post-inoculation (HPI). The roots of three plants (biological replications) of each noninoculated control and pathogen-inoculated treatment were pooled together for RNA isolation. Root tissues (500 μg) were washed thoroughly using distilled water and frozen immediately in liquid nitrogen. Frozen root tissues were crushed into a fine powder using mortar–pestle. Total RNA isolation was performed using TRIzol reagent (Ambion, Thermo Fisher Scientific, Massachusetts, United States) as mentioned in the protocol provided by the manufacturer. RNA isolated from three experimental replications was pooled for each control and treatment. Qualitative and quantitative assessments of the isolated and pooled total RNA were performed by measuring the absorbance using a nanodrop spectrophotometer followed by agarose gel electrophoresis. From the total RNA, mRNA purification was done using a Qiagen Oligotex mRNA Minikit (Qiagen, Hilden, Germany). Subsequently, a SMARTer PCR-cDNA synthesis kit (Clontech Laboratories, Inc., Dalian, China) was used to prepare double-stranded cDNA, from each of the purified poly-A mRNA samples (500 ng) following the instructions provided by the manufacturer.

Next-generation sequencing

The Illumina NextSeq500 platform was used for transcriptome sequencing. Double-stranded cDNA libraries were incorporated into an Illumina chamber for the generation of clusters. After obtaining the Qubit concentration for cDNA libraries, 2 × 150 bp paired-end sequencing was performed through the sequencing by synthesis method. Paired-end sequencing allowed the fragmented templates to be sequenced in both forward and reverse directions which were used for transcriptome sequencing. After sequencing, FASTQ files were generated for each sample.

Sequence quality analysis

Next-generation sequencing analysis was performed on cDNA samples collected at 36 hours post-inoculation with MTCC 2204. The quality control for raw paired-end sequence reads was examined using the FastQC version 0.11.3 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) program (Supplementary Material S1). The pair-end reads with <30 PHRED quality score were shortlisted, and >30 paired-end reads along with all the unpaired reads were eliminated. Following this step, processing of raw Illumina reads was performed using Trimmomatic software (version 0.32)-based analysis (Bolger et al., 2014) for removal of universal adapters and trimming of low-quality bases incorporated inside and also at the 3′ end region of the sequences.

De novo transcriptome assembly

De novo assembly of the Illumina processed datasets was performed using Trinityrnaseq_r20140717 software, and further filtration was performed to develop the Cluster of Genes (COGs) assembly. Trinityrnaseq_r20140717 (Grabherr et al., 2011) was used for de novo assembly after correction of all possible errors of the processed sequenced reads. Scaffolding of assembled contigs was carried out using the SSPACE software program (https://github.com/nsoranzo/sspace_basic) (Boetzer et al., 2011). Gap closer was performed to remove the polyN inserted during

TABLE 1 Statistical summary of the transcriptome data output with their NCBI-SRA accession numbers.

Sample identity	Genotype	Sample type	Read type	Total read count	PHRED quality score	GC content (%)	Reads mapped in pairs	Accession
NIR	ICP 8863	noninoculated resistant	150 × 2 Paired end	28,395,501	phred33	44	28,060,243 (98.82%)	SRR16990781
IR	ICP 8863	Inoculated resistant	150 × 2 Paired end	35,740,467	phred33	49	35,303,808 (98.78%)	SRR16990783
NIS	ICP 2376	noninoculated susceptible	150 × 2 Paired end	28,311,348	phred33	47	27,895,856 (98.53%)	SRR16990780
IS	ICP2376	Inoculated susceptible	150 × 2 Paired end	26,252,636	phred33	50	26,004,461 (99.05%)	SRR16990782

NIR, noninoculated resistant; IR, inoculated resistant; NIS, noninoculated susceptible; IS, inoculated susceptible; NCBI, national center for biotechnology information.

TABLE 2 Analysis of de novo assembly transcriptome data.

Transcript statistics	Transcriptome de novo assembly statistics	COG's statistics	Transcripts >500bp statistics
Sample name	Pooled samples	Pooled samples	Pooled samples
Tool used	Trinity 25	Trinity 25	Trinity 25
Hash length	25	25	25
Transcripts generated	3,22,688	2,22,874	1,15,970
Maximum transcript length	24,582	24,582	24,582
Minimum transcript length	300	300	300
Average transcript length	916.8 ± 875.9	853.3 ± 837.1	1,291.5 ± 971.3
Median transcript length	656	406	1,659.5
Total transcript length	29,58,30,572	19,01,78,549	14,97,79,299
Total number of non-ATGC characteristics	0	0	0
Transcripts ≥300 b	322,688	222,874	115,970
Transcripts ≥500 b	182,013	115,970	115,970
Transcripts >1 Kb	90,111	53,932	53,932
Transcripts >10 Kb	85	57	57

scaffolding using the GapCloser software (https://github.com/BGI-Qingdao/TGS-GapCloser) (Xu et al., 2019).

Transcript annotation

Candidate coding regions with the generated transcript sequences were identified using the TRANSDECODER tool, and processed transcripts were annotated using NCBI BLAST N 2.303 with the DNA and protein sequences of *C. cajan*, *Glycine max*, *Phaseolus vulgaris*, *Vigna radiata*, *Lotus Japonicas*, *Medicago truncatula*, and *Arabidopsis thaliana*. The fungal sequence contaminations present in the transcriptome data were removed. Then, the purified sequences were used as the reference to map the reads of all four samples to generate the gene matrix with the normalized fragments per kilobase of transcript per million (FPKM) values. Based on the matrix and the provided conditions, the sequences were further filtered and grouped into the corresponding categories. Gene Ontological classification (GO) was performed with the aid of the Blast2GO software program (Gotz et al., 2008), and pathway annotation of all the transcripts was done using KEGG pathway analysis software.

Identification of differentially expressed genes

The DESeq package (version 1.8.1) (http://www-huber.embl.de/users/anders/DESeq/) was used to perform differential gene expression analysis and a total number of up-, down- and neutrally regulated transcripts were identified between noninoculated vs. inoculated resistant cultivar (NIR vs IR) and noninoculated vs. inoculated susceptible cultivar (NIS vs IS). By comparing the base mean expression values of the inoculated samples with the matching control samples, fold

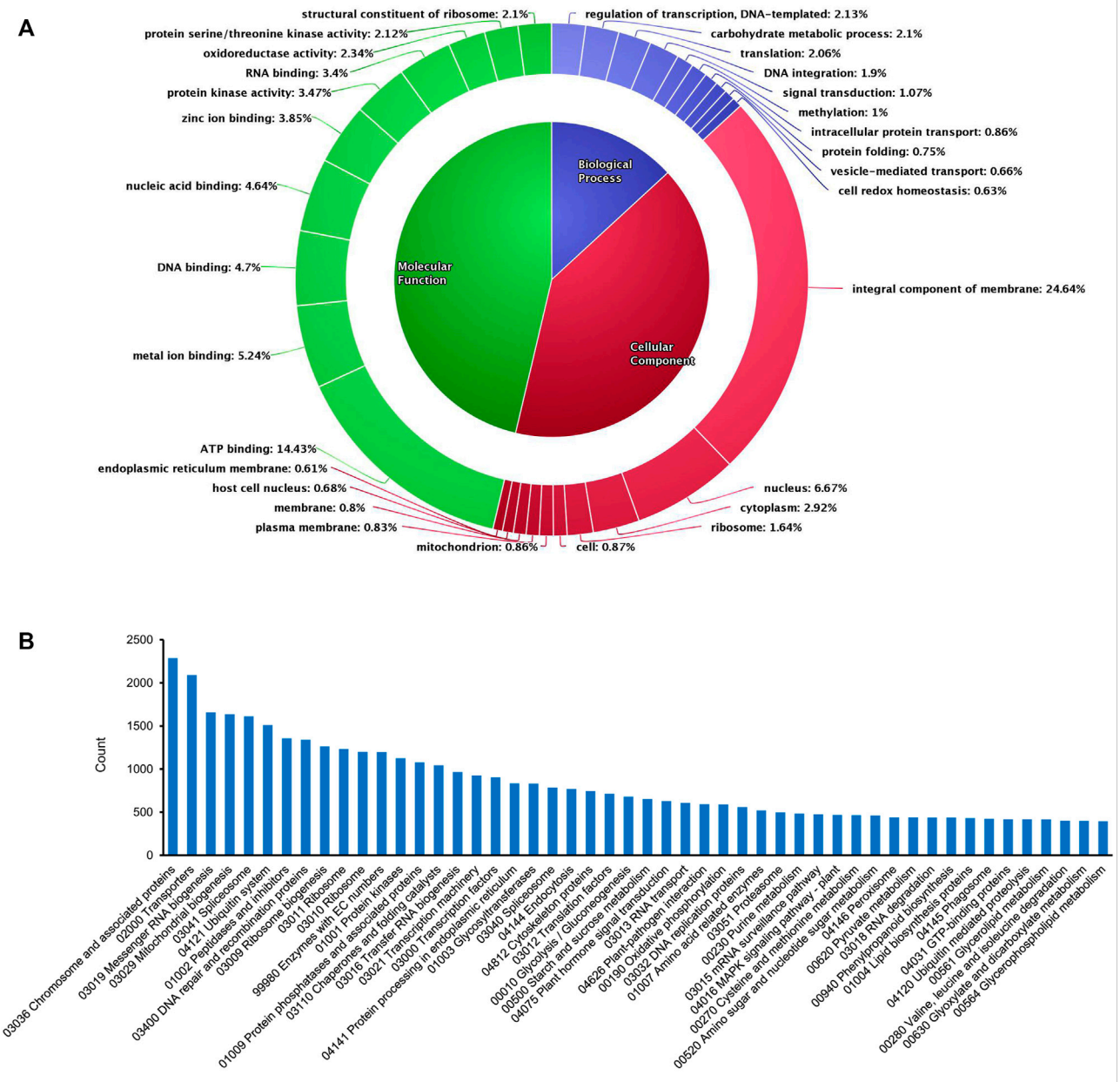


FIGURE 1
Overview of GO and KEGG enrichment analysis of the de novo assembled transcripts in the present study. **(A)** Gene Ontology chart showing the distribution of various transcripts. **(B)** Top 50 pathways observed in KEGG pathway analysis.

change was obtained. The number of common and uniquely occurred transcripts across all conditions in both the cultivar was represented using Venny 2.1.0 software (https://bioinfogp.cnb.csic.es/tools/venny/index.html). Fold-change distributions of the assembled transcripts were represented using volcano plots in R (version 4.2.1) along with the corresponding log-transformed *p*-values. Statistically significant outcomes from all the differentially expressed transcripts were filtered using a 0.05 *p*-value limit.

Results and discussion

NGS raw read data

The total numbers of raw reads were 28.39 million (m), 28.31, 35.74, and 26.25 m for noninoculated resistant (NIR), noninoculated susceptible (NIS), inoculated resistant (IR), and inoculated susceptible (IS), respectively. Following Trimmomatic filtering, 27.2, 26.95, 4.35, and 25.36 m clean reads were obtained

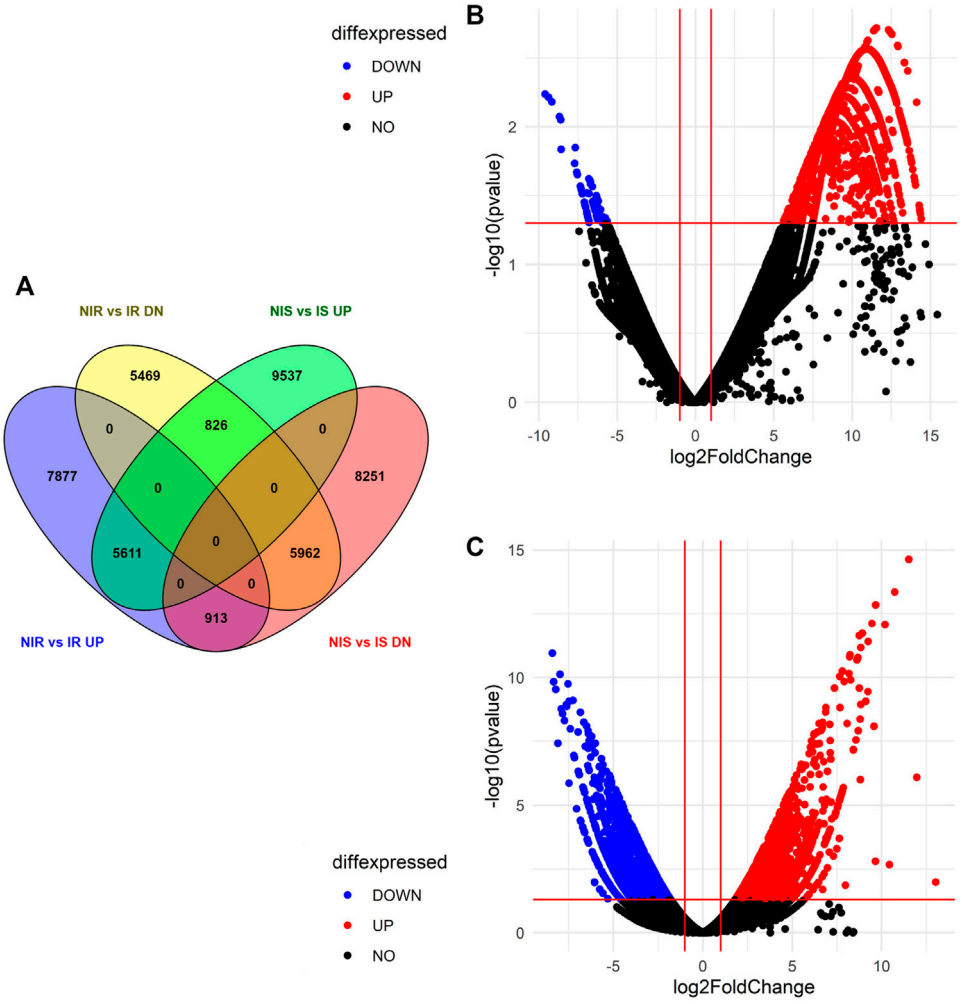


FIGURE 2
Differential expressed genes identified in both the susceptible and resistant cultivars. **(A)** Venny 2.1.0 software analysis of differential expression of genes in NIR vs IR and NIS vs IS. **(B)** Volcano plot representing DEG data in NIS vs IS. **(C)** Volcano plot representing DEG data in NIR vs IR.

from NIR, NIS, IR, and IS, respectively. These datasets were used for further downstream analysis. All the RNA-seq raw reads have been deposited in the NCBI Sequence-Read Archive (SRA) database under the bio project PRJNA782089. Table 1 shows the transcriptome data summary with corresponding NCBI-SRA accession numbers (Supplementary Materials S1, S2).

De novo transcriptome assembly of processed data

De novo transcriptome assembly of processed reads into transcripts generated 322,688 in total number with a maximum transcript length of 24,582 base pairs (bp) and a minimum length of 300 bp. The total transcript length was found to be 295,830,572 bp with 0% non-ATGC characteristics. After filtration, COG statistics showed 222,874 transcripts generated after clustering with 57 very large (Transcripts >10 Kb), 53,932 large (Transcripts >1 Kb), 115,970 medium (Transcripts≥500), and 222,874 small (Transcripts≥300) transcripts. The total identified transcript length after filtration was 190,178,549 bp without any non-ATGC characteristics in the matrix (Table 2).

Transcript annotation and Gene Ontological analysis

De novo assembly enabled us to acquire maximum transcripts from pigeonpea as annotations were done using the database of the *C. cajan* genome. However, in the case of transcripts with no possible match against *C. cajan*, *Glycine max* genome was

considered, as it was the closest relative to pigeonpea, followed by *Phaseolus vulgaris*, *Vigna radiata*, *Lotus Japonicas*, *Medicago truncatula*, and *Arabidopsis thaliana*. After the annotation program, out of 98,466 total transcripts, 70,981 transcripts were annotated and the remaining 27,485 were unannotated. Clustering of total annotated transcripts was performed using various database searches (NCBI, UNIPROT, and pFAM) based on their gene ontology, which revealed varying distributions of transcripts across biological processes (13.6%), molecular functions (46.29%), and cellular components (40.52%) (Figure 1A). Among the biological processes, transcripts involved in the regulation of transcription (2.13%), carbohydrate metabolic processes (2.1%), translation (2.06%), DNA integration (1.9%), signal transduction (1.07%), methylation (1%), intracellular protein transport (0.86%), protein folding (0.75%), vesicle-mediated transport (0.66%), and cell redox homeostasis (0.63%) were observed. In the case of cellular components, transcripts of integral components of membrane and signaling (24.64%) were mostly observed. In addition, transcripts with cellular locations in the nucleus (6.67%), cytoplasm (2.92%), ribosome (1.64%), mitochondrion (0.86%), and plasma membrane (0.83%) were expressed. Transcripts associated with molecular functions were found in nearly half of the total number. This group consisted of transcripts of ATP binding (14.43%), metal ion binding (5.24%), DNA binding (4.7%), nucleic acid binding (4.64%), zinc ion binding (3.85%), protein kinase related (3.47%), RNA binding (3.4%), oxidoreductase (2.34%), serine/threonine kinase (2.12%), and structural constituent of ribosome (2.1%). KEGG pathway analysis of the transcriptome data showed that 64,768 annotated transcripts were involved in 204 different pathways under six major groups: brite hierarchies, cellular processes, environmental information processing, genetic information processing, metabolism, and organismal systems. Figure 1B shows the top 50 pathways observed in the KEGG enrichment analysis.

Analysis of differentially expressed transcripts

Differential expression of gene study between noninoculated vs. inoculated resistant cultivar (NIR vs IR) revealed that a total of 75,799 transcripts were expressed in both samples, out of which 14,401 upregulated, 12,257 downregulated, and 49,141 neutrally regulated transcripts were present. A total of 5,153 and 6,538 transcripts were found only in NIR and IR, respectively. Similarly, the differential expression of genes between noninoculated vs. inoculated susceptible cultivar (NIS vs IS) depicted a total of 67,634 expressed transcripts with 15,974 up-, 15,126 down-, and 36534 neutrally regulated transcripts. A total of 10,022 and 19,979 transcripts are only expressed in NIS and IS, respectively. Venny's analysis of differentially regulated genes showed that 5,611 (12.6%) upregulated and 5,962 (13.4%) downregulated transcripts were common in both NIR vs IR and NIS vs IS. A total of 7,877 (17.7%) and 5,469 (12.3%) transcripts were exclusively up- and downregulated in NIR vs IR, respectively. Similarly, a total of 9,537 (21.5%) and 8,251 (18.6%) transcripts were exclusively up- and downregulated in NIS vs IS, respectively. Interestingly, 826 (1.9%) transcripts were found to be upregulated in NIS vs IS but downregulated in NIR vs IR, and 913 (2.1%) transcripts were found to be upregulated in NIR vs IR but downregulated in NIS vs IS (Figure 2A). After visualizing the p (0.05) values and the distribution of fold change of the DEGs in the NIS vs IS and NIR vs IR groups, they were presented by volcano plot (Figures 2B,C).

Conclusion

From the gene ontological as well as the differential gene expression data of both control and inoculated wilt-susceptible and -resistant transcriptomes, stress- and defense-related genes such as peroxidase73, RPM1, RIN4, RPP8, RPS2, MLO-like protein, wound-induced protein WIN, WRKY70, WRKY33, PR1, TMV-resistant protein, WAT1, MYB46, CESA, and many more such genes can be considered proposed leads for future molecular-based research of pigeonpea. These identified important putative genes along with their associated pathways, activated during the attack of *F. udum* can be investigated through time kinetics studies to understand signaling molecules for host–pathogen interactions and varied metabolic pathways that have crucial roles in inherent resistance mechanisms. Further research can be extended to validate the expression of these genes, as well as the identified disease susceptibility or resistance pathways through functional genomics approaches. This will provide improved knowledge of the resistance mechanisms in pigeonpea during *F. udum* wilt which can be utilized for genomics-assisted breeding programs, genome editing, and biotechnological improvement.

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 NCBI Sequence-Read Archive (SRA) database under the bio-project PRJNA782089.

Author contributions

SaG and DC conceived the idea of the work and designed the experiments. SaG and AP performed all the experiments and collected data. SaG, AH, AM, AB, SuG, RC, and DC were responsible for the data analysis and presentation. SaG and

DC drafted, edited, and finalized the manuscript. All authors read and approved the final manuscript.

Funding

The authors acknowledge the financial support of the Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India (DST No: SR/FT/LS-83/2010), and infrastructure support of the University of Calcutta, India.

Acknowledgments

SG and AH thank the Department of Science and Technology, Govt. of India INSPIRE fellowship (DST/INSPIRE Fellowship/2018/IF180158) and SERB (National PostDoctoral Fellowship PDF/2021/003920), respectively, for providing fellowships.

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/fgene.2022.1009127/full#supplementary-material

References

Agrios, G. N. (2008). *Plant pathology*. Academic Press, 522–534.

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27 (4), 578–579. doi:10.1093/bioinformatics/btq683

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114–2120. doi:10.1093/bioinformatics/btu170

Choudhary, A. K. (2010). A wilt resistant line 'IPA 204'of long-duration pigeonpea (Cajanus cajan). *Indian J. Agric. Sci.* 80, 907–909.

Dhar, V., Datta, S., Chaudhary, R. G., Upadhyay, J. P., Saifulla, M., Mishra, S., et al. (2012). Pathogenic and molecular characterisations of pigeonpea wilt pathogen Fusarium udum. *Archives phytopathology plant Prot.* 45 (4), 423–436. doi:10.1080/03235408.2011.587974

FAO *Food and agricultural organization of the united nation, FAO statistical database*. http://faostat.fao.org 2020 (Accessed September 8, 2022).

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., et al. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36 (10), 3420–3435. doi:10.1093/nar/gkn176

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Trinity: Reconstructing a full-length transcriptome without a genome from RNA-seq data. *Nat. Biotechnol.* 29 (7), 644–652. doi:10.1038/nbt.1883

Purohit, A., Ganguly, S., Ghosh, G., Kundu Chaudhuri, R., Datta, S., and Chakraborti, D. (2017). Variability among isolates of Fusarium udum and the effect on progression of wilt in pigeonpea. *Eur. J. Plant Pathol.* 149 (1), 73–87. doi:10.1007/s10658-017-1167-z

Purohit, A., Ghosh, S., Ganguly, S., Negi, M. S., Tripathi, S. B., Chaudhuri, R. K., et al. (2021). Comparative transcriptomic profiling of susceptible and resistant cultivars of pigeonpea demonstrates early molecular responses during Fusarium udum infection. *Sci. Rep.* 11 (1), 22319. doi:10.1038/s41598-021-01587-7

Reddy, M. V., Nene, Y. L., Kannaiyan, J., Raju, T. N., Saka, V. W., Daudi, A. T., et al. (1990). Pigeonpea lines resistant to wilt in Kenya and Malawi. *Int. Pigeonpea Newsl.* (12), 25–26.

Sameer Kumar, C. V., Satheesh Naik, S. J., Mohan, N., Saxena, R. K., and Varshney, R. K. (2017). "Botanical description of pigeonpea [Cajanus cajan (L.) Millsp.]," in *The pigeonpea genome* (Cham: Springer), 17–29.

Saxena, R. K., Singh, V. K., Kale, S. M., Tathineni, R., Parupalli, S., Kumar, V., et al. (2017). Construction of genotyping-by-sequencing based high-density genetic maps and QTL mapping for Fusarium wilt resistance in pigeonpea. *Sci. Rep.* 7 (1), 1911–11. doi:10.1038/s41598-017-01537-2

Soren, K. R., Tripathi, S., Pareek, S., Hembram, M., Gangwar, P., Abrol, S., et al. (2021). Comparative time series RNA-seq analysis of pigeonpea root tissues in response to Fusarium udum infection. *Front. Fungal Biol.* 2, 19. doi:10.3389/ffunb.2021.664953

Xu, G. C., Xu, T. J., Zhu, R., Zhang, Y., Li, S. Q., Wang, H. W., et al. (2019). LR_Gapcloser: A tiling path-based gap closer that uses long reads to complete genome assembly. *Gigascience* 8 (1), giy157. doi:10.1093/gigascience/giy157

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 24 October 2022
DOI 10.3389/fgene.2022.961474

OPEN ACCESS

EDITED BY
Suxu Tan,
Qingdao University, China

REVIEWED BY
Stafford Vigors,
University College Dublin, Ireland
Merete Fredholm,
University of Copenhagen, Denmark

*CORRESPONDENCE
K. H. Mellits,
ken.mellits@nottingham.ac.uk

SPECIALTY SECTION
This article was submitted to Livestock Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 04 June 2022
ACCEPTED 24 August 2022
PUBLISHED 24 October 2022

CITATION
Le Bon M, Tötemeyer S, Emes RD and Mellits KH (2022), Gut transcriptome reveals differential gene expression and enriched pathways linked to immune activation in response to weaning in pigs.
Front. Genet. 13:961474.
doi: 10.3389/fgene.2022.961474

Gut transcriptome reveals differential gene expression and enriched pathways linked to immune activation in response to weaning in pigs

M. Le Bon[1,2], S. Tötemeyer[3], R. D. Emes[3,4] and K. H. Mellits[1]*

[1]School of Biosciences, Division of Microbiology, Brewing and Biotechnology, University of Nottingham, Loughborough, United Kingdom, [2]School of Animal, Rural and Environmental Sciences, Nottingham Trent University, Loughborough, United Kingdom, [3]School of Veterinary Medicine and Science, University of Nottingham, Loughborough, United Kingdom, [4]Advanced Data Analysis Centre, University of Nottingham, University of Nottingham, Nottingham, United Kingdom

Weaning represents one of the most critical periods in pig production associated with increase in disease risk, reduction in performance and economic loss. Physiological changes faced by piglets during the weaning period have been well characterised, however little is currently known about the underlying molecular pathways involved in these processes. As pig meat remains one of the most consumed sources of protein worldwide, understanding how these changes are mediated is critical to improve pig production and consequently sustainable food production globally. In this study, we evaluated the effect of weaning on transcriptomic changes in the colon of healthy piglets over time using an RNA-sequencing approach. The findings revealed a complex and coordinated response to weaning with the majority of genes found to be rapidly differentially expressed within 1 day post weaning. Multiple genes and pathways affected by weaning in the colon were associated with immune regulation, cell signalling and bacterial defence. NOD-like receptors, Toll-like receptor and JAK-STAT signalling pathways were amongst the pathways significantly enriched. Immune activation was evidenced by the enrichment of pathways involved in interferon response, cytokines interactions, oxidoreductase activities and response to microbial invasion. Biosynthesis of amino acids, in particular arginine, was also amongst the most enriched KEGG pathways in weaned pigs, reinforcing the critical role of arginine in gut homeostasis under stress conditions. Overall, transcriptomic and physiological results suggest that pigs going through the weaning transition undergo a transient period of inflammatory state with a temporary breakdown of barrier functions in the gut. These findings could provide valuable tools to monitor host response post weaning, and may be of particular relevance for the investigation and development of intervention strategies aimed to reduce antibiotic use and improve pig health and performance.

KEYWORDS

pig, weaning, RNA-sequencing, transcriptomic, gut, immune response

Introduction

Livestock production is expected to produce more food than ever before. As the expanding world population is getting wealthier, the demand for safe and secure animal protein is increasing (Henchion et al., 2017). The challenge is to meet this demand in ways that are environmentally, socially and economically sustainable. Together with poultry, pork is one of the fastest growing livestock sectors and also one of the most consumed meats world-wide (FAO, 2019). Pig production is also widely recognised as one of the most efficient in terms of carbon footprint and climate change potential compared to other animal protein source (Macleod et al., 2013). Ensuring livestock animals can fulfil their full genetic potential is essential for sustainable food production.

As a result of abrupt dietary, social, and environmental changes, weaning is recognised as the most critical period in modern pig production associated with an increase in disease risk, reduction in performance and welfare leading to significant economic loss (Gresse et al., 2017; Nowland et al., 2019). At weaning, the pigs gastrointestinal tract (GIT) undergoes rapid changes in size, protein turnover rates, microbiome composition, and detrimental alterations in digestive and barrier functions (Pluske et al., 2018). Although the physiological changes faced by piglets over weaning have been well characterised, little is known about the underlying genes and pathways involved in these processes. Understanding how these changes are regulated or mediated is critical to improve pig production and consequently sustainable food production globally.

Furthermore, due to similarities in anatomy and physiology, the pig is widely recognised as a translational animal model to study human gastrointestinal diseases and to understand biological pathways related to mucosal function, development and nutritional regulation (Zhang et al., 2013; Roura et al., 2016; Sciascia et al., 2016). Previous studies have highlighted the importance of improving the knowledge on molecular mechanisms responsible for phenotypic differences especially at an early age with the dual purpose of improving production and providing adequate models for human studies (Ayuso et al., 2015).

Recent advances in sequencing technologies now provides novel opportunities to comprehensively explore the complex gut ecosystem of humans and animals. RNA-sequencing (RNA-Seq) is a powerful high-throughput approach to profile gene expression that, in contrast with microarray-based technologies, allows for the characterisation and quantification of both known and unknown transcripts (Mach et al., 2014). Fundamental understanding of the host response to stress is paramount to develop tools or best practices to improve pig health, productivity and welfare. To date, RNA-Seq has been used to study production traits of livestock animals but transcriptomic studies in pigs using this technology is relatively scarce and have mainly focused on disease response or regulation mechanism of fat deposition and muscle development to evaluate growth and meat quality between pig genotypes (Piórkowska et al., 2018; Xu et al., 2019).

Much attention and focus has been given to the gut microbiome and its taxonomic and metabolic changes through pig development and weaning (Frese et al., 2015; Bian et al., 2016; Guevarra et al., 2018a) but we are still lacking understanding about the host gene expression change in response to weaning. The current study aims to investigate the transcriptomic changes in the pig gut through weaning over time.

Material and methods

Animals and experimental design

All animals were treated in accordance with the University of Nottingham ethical guidelines and codes of practice applying to care and management of animals. Twelve litters (Landrace x Large White) over three batches (4 litters per batch) were used in the study and housed at the School of Biosciences, Sutton Bonington Campus, University of Nottingham, United Kingdom. For the phenotypic data, all 3 batches (12 litters) were used. For the transcriptomic data, only one batch (the first batch) was used (4 litters).

For prevention of iron deficiency and coccidiosis, all piglets received a 1 ml IM injection of Gleptosil (Alstoe Ltd., York, United Kingdom) 24 h after birth, and 0.7 ml of Baycox (Bayer, Newbury, United Kingdom) orally 3 days after birth. At 21 days of age, 6 weight-matched piglets per litter were randomly allocated by random selection of coded balls to treatment (baseline average weight at day 19: 6.77 ± 0.189 kg). Piglets allocated to the weaned treatment were separated from their dam, moved and mixed with non-littermates in pens of 4 individuals and received ad lib commercial diet (wheat, whey powder and soya based) containing: 21.25% protein, 7.50% fat, 2.00% fibre, 5.00% ash, 1.70% lysine, 13.80% moisture). Weaned piglets did not receive creep feed supplementation before weaning. Piglets allocated to the unweaned treatment as control remained with their dam and littermate up to 35 days of age with access to creep feed from day 25 days of age (same commercial diet as above). No antibiotic or anthelmintic treatment were used during the trial. At day 1, 4, and 14 post weaning: one weaned and one unweaned piglet from each litter were weighed euthanised by intraperitoneal injection of Dolethal (1 ml.kg-1 body weight; 20% w/v Pentobarbitone Sodium, Vétoquinol, Buckingham, United Kingdom). At slaughter, body lesion was scored for each pig on a 3-point scale with 1 for no lesion, 2 for moderate scratches on back, flank, head, ear and tail, and 3 for intense, deep or bleeding scratches on back, flank, head, ear and tail. Graphic visualisation of the trial design can be found in Supplementary Figure S1.

Intestinal measurements

All sample processing and analysis was blinded using randomly generated numerical codes. Tissue samples from the 0.5 section of the colon were immediately collected post-slaughter, rinsed in sterile buffered saline solution and preserved in RNA later (Ambion, CA United States) at 4°C for 24–48 h to allow tissue penetration then stored at −80°C.

Tissue samples of the 0.5 small intestine (as proportions) along from the gastric pylorus to the ileocecal valve were fixed in Bouin's solution, embedded in paraffin and cut in 5 µm transverse sections. Histological section were stained with H&E for histometric measurement of villus length and crypt depth, with Periodic Acid Schiff stain for goblet cell counts (Matsuo et al., 1997) and with Toluidine Blue for quantification of mast cells (Moeser et al., 2007).

Secretory IgA (sIgA) was measured in ileal flushes using the methods previously described by Lessard et al. (2009). At slaughter, a 20 cm segment of ileum taken upstream from the cecum was flushed using 5 ml of sterile PBS, and centrifuged for 10 min at 500 g. The supernatant was collected and stored at −80°C. Secretory IgA was measured in duplicate using a sandwich Porcine IgA ELISA Quantitation Kit (Bethyl Laboratories, TX, United States) according to the manufacturer protocol.

RNA-sequencing

Total RNA was extracted using NucleoSpin® RNA II kit (Macherey-Nagel, Düren · Germany), briefly 10–30 mg of tissue was placed in a 2 ml round bottomed tube (Fisher Scientific, United Kingdom) with 350 µl of guanidine thiocyate buffer, 3.5 µl of 0.1 mM β-mercaptoethanol and a 5 mm stainless steel bead for homogenisation for 2 × 2 min at 3,000 rpm using a Retsch MM200 homogeniser (Retsch, Haan, Germany). The extraction process was then followed according to the manufacturer instruction until the last step where final elution volume was 40 µl of RNAse-free water instead of 60 µl. All samples were quantified and tested for RNA integrity using the Nanodrop 1,000 spectrophotometer (NanoDrop Technologies, DE, United States) and Agilent 2,100 Bioanalyzer (Agilent Technologies, CA, United States). Mean (±SE) of RNA concentration was 306.5 ± 29 ng/ml and RNA integrity number was 9.21 ± 0.13.

RNA samples were sent to ARK Genomics (Edinburgh, United Kingdom) for library preparation and sequencing of the whole transcript expression. Illumina TruSeq mRNA library prep kit was used, and paired end sequencing was performed on an Illumina HiSeq2000 sequence analyser (Illumina, Inc. San Diego, United States). Animal details and metadata are provided in Supplementary Table S3 and Supplementary Table S4. Mean number of raw reads per sample was 101.4 M (min = 66.1M, max = 133.6 M), equivalent to 5,139 Mbp per sample on average. Reads in fastq format are available at the European Nucleotide Archive (ENA) with the accession number PRJEB54752 https://www.ebi.ac.uk/ena/browser/view/PRJEB54752.

Raw fastq reads were processed with Cutadapt (Version 1.12) (Martin, 2011) to remove adapters and low-quality bases (3′ quality cutoff of 10, -q 10), paired reads with a minimum length of 20 bases were retained prior to quantitation (−m 20). Sequence alignment and read quantification was performed using the pseudo-alignment-based tool Kallisto v0.43.0 (Bray et al., 2016). Differential expression of transcripts between weaned and unweaned animals was determined using Sleuth v0.28.1 (Pimentel et al., 2017). Identification of genes enriched in specific functions were determined with reference to the Kyoto Encyclopedia of Genes and Genomes (KEGG) and gene ontology (GO), using "NIPA" software available on https://github.com/ADAC-UoN/NIPA to conduct hypergeometric tests for enrichment with Benjamini-Hochberg multiple testing correction.

Statistical analysis

For phenotypic data, statistical analysis was performed in IBM SPSS v24 to determine the effect of weaning treatment at different time point treatment on pig weight, intestinal and blood measurement using linear mixed model analysis. In the mixed model, time point*treatment was used as the fixed effect and trial batch and litter as random effects.

Gene expression cluster analysis was performed and revealed that sex was the only factor that showed a grouping effect. Therefore sex was included in the model as a confounder for the gene expression analysis. Sequence alignment and read quantification was performed using the pseudo-alignment-based tool Kallisto v0.43. (Bray et al., 2016). Differential expression was determined using the Wald test in Sleuth v0.28.1 (Pimentel et al., 2017) with sex as a confounder in the model. Transcripts with a false detection rate corrected p-value < 0.1 and a log2 fold change (log2FC) greater than 1 and less than minus 1 were considered to be differentially expressed. False detection rate correction was performed using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995).

Results

Phenotypic data

All piglets were found in good health during the trial and displayed no clinical signs of disease or scour. As expected, weaning caused a number of physiological changes which

TABLE 1 Phenotypic measurements between weaned and unweaned pigs at 1, 4, and 14 days post weaning. Data shown are means and the pooled SEM for weaned and unweaned pigs for each time points.

Time point (days post weaning)	1		4		14		p values			
Treatment	**unwean**	**wean**	**unwean**	**wean**	**unwean**	**wean**	**SEM**	**time**	**weaning**	**time*weaning**
n	12	12	12	12	12	12				
Pig performance										
Pig weights (kg)	7.16	6.60	8.31	7.20	12.68	12.41	0.37	<0.001	0.002	0.238
Body lesion scores	1.3	2.4	1.7	2.0	1.6	1.4	0.2	0.007	<0.001	<0.001
Blood and plasma measurement										
RBC (10^6/µl)	5.83	5.92	6.04	6.15	6.78	6.67	0.12	<0.001	0.726	0.432
HCT (%)	35.56	35.78	34.32	37.16	35.14	39.89	1.06	0.035	<0.001	0.019
HGB (g/dl)	11.43	11.55	11.19	12.00	11.16	12.61	0.37	0.309	<0.001	0.045
PLT (10^3/µl)	552.6	551.9	579.4	491.0	595.5	436.0	28.8	0.356	<0.001	0.009
WBC (10^3/µl)	9.41	9.40	10.08	12.33	13.40	13.30	0.84	<0.001	0.202	0.152
Granulocytes (10^3/µl)	2.650	3.854	3.504	4.958	4.096	4.933	0.38	0.003	<0.001	0.722
Total plasma protein (g/l)	45.700	48.688	46.937	45.420	45.663	43.875	0.74	0.005	0.860	0.002
Plasma globulin (g/l)	16.254	17.418	16.528	15.757	14.693	13.262	0.47	<0.001	0.368	0.018
Plasma cholesterol (mmol/l)	3.925	3.687	3.910	2.441	3.681	2.432	0.13	<0.001	<0.001	<0.001
Plasma triglyceride (mmol/l)	0.848	0.872	0.971	0.470	0.939	0.511	0.06	0.001	<0.001	<0.001
Plasma cortisol (ng/ml)	15.886	57.744	16.125	30.884	10.999	17.767	4.18	<0.001	<0.001	<0.001
Intestinal measurement										
Villus length (µm)	0.372	0.293	0.373	0.341	0.403	0.418	0.018	<0.001	0.023	0.026
Crypt depth (µm)	0.151	0.145	0.155	0.171	0.175	0.192	0.005	<0.001	0.018	0.013
Villus width (µm)	0.100	0.098	0.115	0.114	0.128	0.135	0.004	<0.001	0.572	0.217
V/C ratio	2.508	2.106	2.500	2.054	2.368	2.276	0.137	0.944	0.005	0.355
Mast cells (/mm^2 submucosa)	100.0	109.8	96.4	89.6	182.3	161.2	13.6	<0.001	0.582	0.517
Mast cells (/mm^2 mucosa)	138.2	167.3	188.3	106.2	222.3	204.7	20.0	<0.001	0.089	0.005
Goblet cells in villus (/100µm^2)	1931.1	2087.6	763.8	993.0	863.5	908.5	152.8	<0.001	0.260	0.831
Goblet cells in crypt (/100µm^2)	4,204.1	3,959.8	1,699.4	1,684.0	1815.0	1770.1	601.4	<0.001	0.643	0.899
Goblet cells total (/100µm^2)	6,136.7	6,047.4	2,463.2	2,677.0	2,678.5	2,678.5	451.4	<0.001	0.898	0.925
Ileal flush sIgA (µg/ml)	596.3	22.3	169.6	10.8	72.5	12.7	88.9	0.007	<0.001	0.011

RBC, red blood cells; HCT, haematocrit; HGB, Haemoglobin; PLT, platelet; WBC, white blood cells; V/C, Villus/Crypt; sIgA, Secretory Immunoglobulin A

were found to be time-dependant. Significant differences were observed in pig weight, blood and plasma measurements, body lesion scores and intestinal measurement (Table 1). These results also indicate the significant effect of time on almost all variables measured between day 1 and day 14 post weaning. This highlights that the pigs are undergoing rapid period of development at this age and the importance of designing studies using age-matched controls when evaluating the effect of weaning in pigs as opposed to using pre-weaning values as controls.

Plasma analysis revealed that cortisol levels remained stable over time in the unweaned group but increased almost 4-fold at day 1 and 2-fold at day 4 post weaning in the weaned pigs suggesting an activation of the HPA axis under weaning stress. Interestingly, while most parameters showed a rapid spike followed by a progressive return to the unweaned level by day 14, cortisol level still remained significantly higher at 14 days post weaning but to a lesser magnitude.

Haematology and biochemistry profiles are used as indicators of health status to evaluate the metabolic, nutritional and energy state of the pig. In blood, circulating granulocyte levels, haemoglobin and haematocrit were significantly increased in weaned pigs, while platelet counts decreased. In plasma, cholesterol, triglycerides and globulin levels decreased at day 4 and day 14 in the weaned group compared to unweaned pigs.

Measurement of intestinal architecture were also affected by weaning, with significant reduction in jejunal villus height and

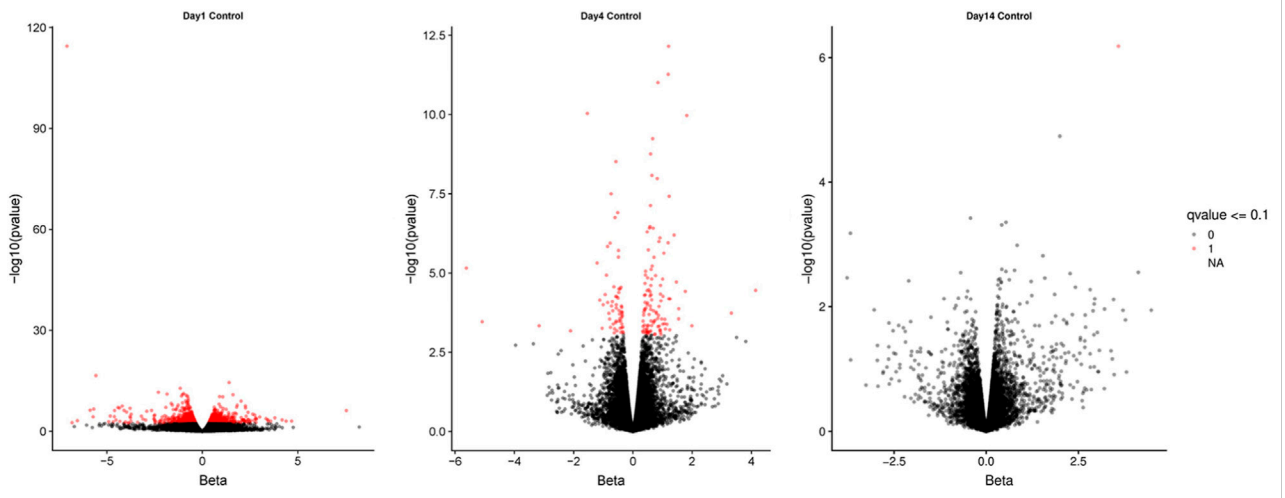


FIGURE 1
Volcano plots of differentially expressed transcripts between weaned and unweaned pigs at 1, 4, and 14 days post weaning. Red points indicates genes significant at Qval ≤0.1.

villus/crypt ratio. However, villus width remained unaffected by weaning and crypt depth was increased. Goblet cell numbers in crypt and in villus remained unaffected by weaning. Mucosal mast cells decreased at day 4 with an overall statistical trend for weaned pigs to show reduced mast cell count compared to unweaned controls.

Ileal sIgA was also greatly reduced at all time points post weaning, suggesting a decrease in immune protection from maternal milk. sIgA is considered one of the first line of defence and a critical factor for piglet growth and survival. Removal of the piglets from the sow at weaning causes dramatic drop in gut IgA levels increasing the immune vulnerability of piglets post weaning.

Transcriptomic analysis

RNAseq (HiSeq) was used to identify differentially expressed transcripts (DET) between weaned and unweaned pigs in colonic tissue. An average of 52.3 million trimmed paired reads were obtained for each sample. Reads mapped as pairs (83.0%–85.3%) to the porcine genome ref sequence (sus scrofa 10.2). The Wald test was used on TPM to establish the total number of DET between weaned and unweaned pigs for each time point and identified a total of 239 transcripts at q value ≤0.1 and FC ≥ 2. The volcano plots visually represent significant DET for each time point and show an even distribution between up- and down-regulated genes (Figure 1).

The majority of transcripts 171 (71.6%) were found to be differentially expressed after 1 days post weaning. After 4 days and 14 days post weaning only 67 (28.0%) and 1 (0.4%) genes were differentially expressed between weaned and unweaned pigs, respectively (Figure 2). Among DET with the cut off values of Qval ≤0.1 & FC ≥ 2, two transcripts were identified in common between day 1 and day 4 post weaning: ENSSSCT00000014128 and ENSSSCT00000017308 which are of unknown function (Figure 3). They were no DET common at all three time point or shared between day 4 and 14 or day 1 and 14 post weaning. Transcription according to time point and weaning treatment are shown in PCA plots and revealed distinct clusters between weaned and unweaned pigs across time points (Figure 4).

The full list of differentially expressed transcripts at Qval ≤0.1 & FC ≥ 2 is provided in Supplementary Table S1. DETs with FC > 2 across all time points were subject to Gene Ontology (GO) and KEGG pathway analysis using NIPA https://github.com/richarddemes/NIPA to identify significantly enriched pathways between weaned and unweaned pigs at Qval ≤0.05 and minimum number of gene in term of two (Supplementary Table S2). Top 10 significantly enriched GO terms and KEGG pathways are shown in Figure 5.

Discussion

In this study, the typical characteristics previously reported in other studies were observed validating our weaning model. Piglet weights were negatively affected by weaning at day 1 and day 4 but recovered at day 14. Post weaning growth check is a commonly reported problem in pigs with detrimental impact

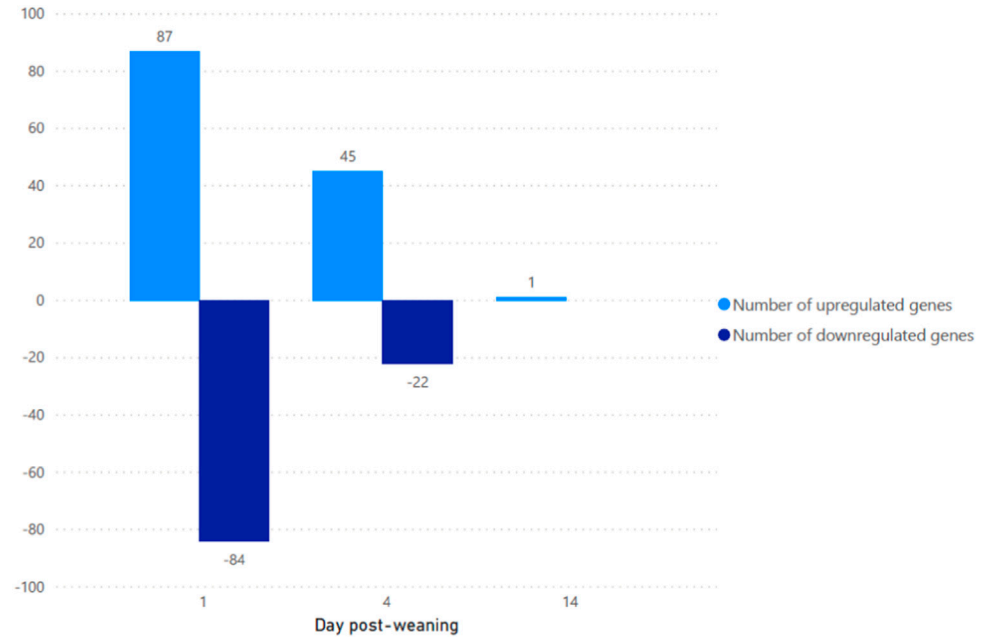


FIGURE 2
Number of up and downregulated transcripts between weaned and unweaned pigs at 1, 4, and 14 days post weaning (Qval ≤0.1 & FC ≥ 2).

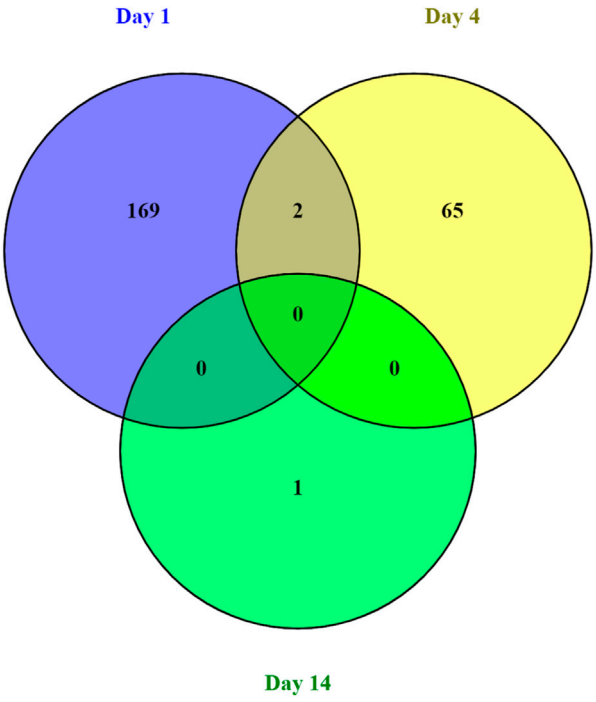


FIGURE 3
Venn diagram of the number of differentially expressed transcripts between weaned and unweaned pigs for each time point (Qval ≤ 0.1 & FC ≥ 2).

on life-time growth performances representing a large economic loss to the industry (Collins et al., 2017). Evidence of stress was also observed in this study with elevated plasma cortisol and increased lesion scores in the weaned pigs.

Measurements of intestinal structures such as villus and crypt have been commonly used to evaluate gut health as a measure of absorptive capacities for nutrients (Bontempo et al., 2006; Le Bon et al., 2010). As reported by others, the current study shows that villus morphometry is impaired by weaning. Villus atrophy during the first few days post weaning has been commonly reported and is often associated with loss in barrier function, decreased enzyme activities and decreased performance as observed in this study (Lallès, 2008).

This study has revealed a large number of genes up and down regulated as consequences of weaning in the pig colon highlighting the profound physiological impact of weaning on healthy pigs. The large majority of differentially expressed transcripts were found to peak on day 1 post weaning and progressively returned to unweaned level by day 14, emphasising the abrupt response of pigs to weaning (Lallès et al., 2007). There was little overlap in transcripts between time points suggesting a coordinated regulation of the host in response to weaning. These findings could inform selection of appropriate time points for further trials in weaned pigs.

Due to the large number of DET identified in this study, this discussion will mainly focus is on pathways significantly affected by weaning rather that individual genes. Metabolic, immune and barrier function activities have previously been

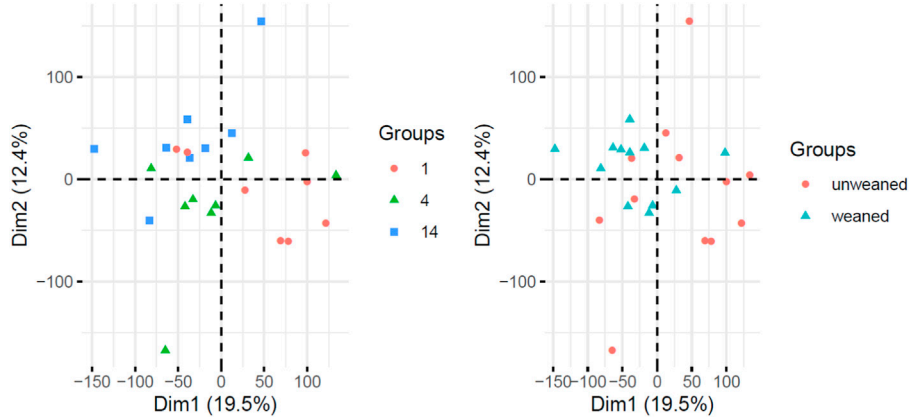


FIGURE 4
Principal Component Analysis (PCA) of gene expression data from pigs labelled as time point (left panel) and weaning status (right panel).

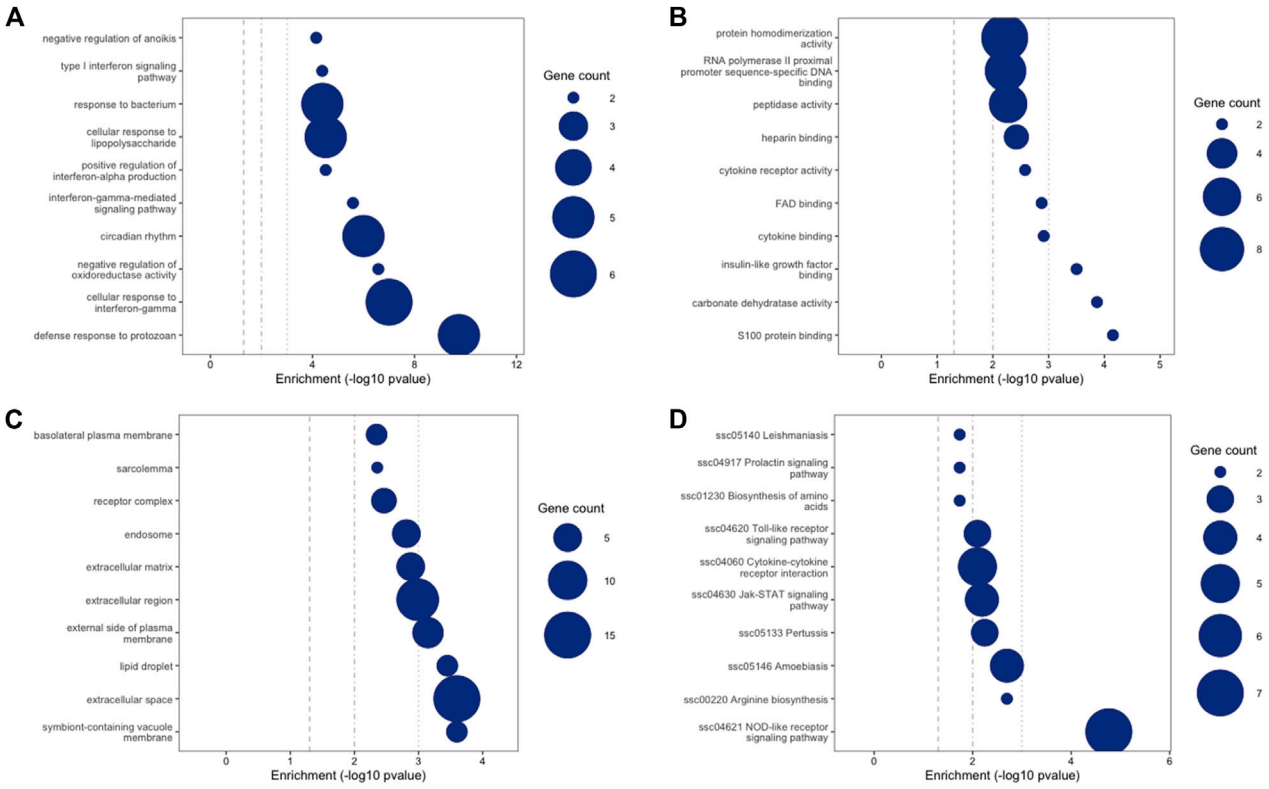


FIGURE 5
Gene ontology (GO) and KEGG pathway enrichment analysis showing top 10 enriched terms between weaned and unweaned pigs (all time points). Size of circles represents number of gene in each term. **(A)** GO Biological Process, **(B)** GO Molecular Function, **(C)** GO Cellular Compartment, **(D)** KEGG.

implicated with weaning response in both the small and large intestine but the underlying mechanism for these changes have not been previously identified. In the current study, we found several pathways related to microbial response and immune functions activated post weaning; specifically, NOD-like receptors, Toll-like receptor and JAK-STAT

signalling pathways are shown to be significantly activated. These signalling pathways indicate a cellular response to Microbial Associated Molecular Patterns (MAMPs) that drives the activation innate immune response. As such, we also observe positive regulation of Interferon alpha production and Interferon gamma mediated pathways.

In agreement with these results, Wang et al. (2008) also reported that weaning at 21 days of age showed an increase in expression of genes associated with oxidative stress and immune activation but decreased expression of genes related to nutrient utilisation and cell proliferation in the jejunum of pigs using microarray analysis. A previous study also reported that pigs weaned at a later age (28 days) showed increase in pro-inflammatory cytokines expression in the small intestine and colon during the first 2 days post weaning before returning to pre-weaned level after 8 days post weaning (Pié et al., 2004). The interferon response following weaning stress in pigs has previously been characterised and showed that weaning causes the release of IFN-α and the transient shut-off of the corresponding gene transcriptions in PBMC (Razzuoli et al., 2011).

In addition, Moeser et al. (2017) demonstrated that colonic and jejunal transepithelial resistance was increased resulting in impaired permeability of the gut barrier as a result of weaning, which may facilitate infiltration of luminal component such as bacteria or bacterial products. Here we see at least two GO biological terms that would agree with this hypothesis: “cellular responses to LPS” and “response to bacterium” where genes such as NOS2, CD274, IRF3, CXCL11, and ACOD1, are upregulated.

The immune regulation observed through our transcriptomic results also reflects the significant increase in granulocyte levels found in blood, suggesting both a local and systemic activation of immunity in response to weaning. The combined effect of immune and stress activation can lead to an energy cost whereby growing animals divert their energy resource towards these responses instead of growth which can explain reduced performance observed in weaned pigs. In growing pigs, transcriptomic multi-tissue analysis revealed that activated immune response, protein metabolism, defence against pathogens and oxidative stress were the main biological pathways associated with feed efficiencies (Gondret et al., 2017). The authors suggested that dietary intervention with anti-inflammatory or antioxidant properties could be evaluated to improve efficiencies in growing pigs. As feed accounts for more than 60% of the cost of food production, improving feed efficiency is a major target to improve profitability of the pig industry (Jing et al., 2015).

The colon also harbours the richest and most diverse microbial population of the gut. In recent years, a large number of studies have documented the development of the gut microbiota of the pig over time (Frese et al., 2015; Mach et al., 2015; Guevarra et al., 2018b; Han et al., 2018; Ke et al., 2019; Wang et al., 2019). Weaning is typically associated with a decrease in gut microbial diversity and major shifts in bacterial taxa composition (Guevarra et al., 2019). In the current study, we observe activation of immune signalling pathways and pro-inflammatory cytokines in the early days post weaning, and we also observe regulation of pathways involved in oxido-reducatase activity. Amongst the activated pathways related to anti-microbial defence, genes involved in oxidative burst such as NOS2 (Nitric oxide synthesase 2) and NOX1 (NAPDH oxidase 1) are upregulated which are involved in the production of nitic oxide and superoxide radicals. These nitrogen rich compounds are rapidly converted into nitrate (NO3-) in the lumen providing favourable condition for the growth and proliferation of gut bacteria that carry nitrate-reductase genes such as *Enterobacteriaceae* (Winter et al., 2013). In addition, inflammatory conditions provide increased level of luminal oxygen due to elevated blood flow and haemoglobin, this favours aerobic respiration of *Enterobacteriaceae* while inhibiting the growth of obligate anaerobes such as *Bacterioides* and *Clostridia* (Zeng et al., 2017). *Enterobacteriaceae* have a detrimental effect on pig health and growth and is one of the leading cause of diarrhoea in pig production (Rhouma et al., 2017). *E. coli* is highly prevalent post weaning and can lead to mortality and zoonosis (Luppi, 2017). As a result, antibiotic usage is commonly used around weaning and have led to increasing reports in colistin-resistant *E. coli* in pigs (Rhouma et al., 2017). Tackling post weaning inflammatory response could represent a step towards reduction in antimicrobial use in pig production.

At weaning, pigs are abruptly transitioned from sow’s milk to a complex plant-based diet with distinct nutritional profile and as such transcriptomic modulation of pathways involved in nutritional metabolism would be expected. In the current study, biosynthesis of amino acids, in particular arginine biosynthesis was the second most significant KEGG pathway enriched in weaned pigs *via* the up regulation of argininosuccinate synthase 1 gene (ASS1). Young mammals, including piglets, have a particularly high requirement of arginine for growth and metabolic function (Flynn and Wu, 1997). L-Arginine is also the biological precursor of nitric oxide (NO), and alteration of arginine uptake and metabolism has been found to be associated with inflammatory bowel diseases, (Stuehr, 2004; Luiking et al., 2012; Coburn et al., 2016). A number of studies have reported that arginine administration significantly attenuate intestinal inflammation associated with down-regulation of the JAK-STAT signalling pathway and increased growth performance and survival in pigs (Wu et al., 2010; Zheng et al., 2018; Che et al., 2019: Luiking et al., 2012; Zhang et al., 2019). Reports of reduced arginine availability in conditions of acute and chronic stress, often associated with increase in NOS2 activity, aligns with the results observed in the current study. The specific mechanisms of regulation and

interaction between cortisol, NOS, immune response, and arginine metabolism remain unknown, but could provide further evidence to suggest that arginine requirement should be carefully evaluated when designing diets to support pigs during the weaning transition.

Finally, although the current study has identified a large number of transcripts and pathways regulated at the mRNA level, post transcriptional and post translational mechanisms could also regulate the host response to weaning and should be investigated in future studies as a complementary approach to transcriptomic methods.

Conclusion

Weaning is a multifactorial event that results in complex interactions between gut, brain and metabolism. Understanding these responses and the molecular mechanisms that underpins these changes is critical to improve sustainable pig production. This study has identified multiple genes and pathways differentially regulated by weaning. These results revealed that pigs going through the weaning transition undergo a transient period of inflammatory state with temporary breakdown of barrier functions in the gut. The condition of the inflamed gut have been previously shown to provide favourable growth advantage for the expansion of *Enterobacteriaceae*, a leading cause of enteric disease in pigs. Under the experimental and controlled conditions of this trial, differential gene expression returned to unweaned control levels by day 14 post weaning. However, the translation of the study results to commercial production setting remains to be explored.

Indicators of weaning stress and response have previously been used including histology, systemic markers of immunity and characterisation of the microbiota composition. Here, we have identified a number of target genes and pathways that could also be used as biomarkers of intestinal inflammation to complement these measures. Together, these could provide valuable tools to monitor host response post weaning, especially in context of intervention strategies aimed to reduce antibiotic use and improve pig health and performance. Finally, as weaning in pigs have been used as a model for stress-related bowel dysfunction in humans, it would be of interest to investigate if the similar transcriptomic changes are involved in these disorders.

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: European Nucleotide Archive (ENA) with the accession number PRJEB54752.

Ethics statement

All trials and procedures were reviewed and approved by the University of Nottingham ethics committee, the Animal Welfare and Ethical Review Body (AWERB).

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

We thank the Nottingham Trent University and Lallemand for funding.

Acknowledgments

We acknowledge the team at Bio-Support Unit at the University of Nottingham for help in running the trial. We thank Prof. Ian Connerton for critical reading and review of the manuscript.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.961474/full#supplementary-material

References

Ayuso, M., Fernández, A., Núñez, Y., Benitez, R., Isabel, B., Barragán, C., et al. (2015). Comparative analysis of muscle transcriptome between pig genotypes identifies genes and regulatory mechanisms associated to growth, Fatness and metabolism. *PLoS One* 10, e0145162. doi:10.1371/journal.pone.0145162

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* 57, 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Bian, G., Ma, S., Zhu, Z., Su, Y., Zoetendal, E. G., Mackie, R., et al. (2016). Age, introduction of solid feed and weaning are more important determinants of gut bacterial succession in piglets than breed and nursing mother as revealed by a reciprocal cross-fostering model. *Environ. Microbiol.* 18, 1566–1577. doi:10.1111/1462-2920.13272

Bontempo, V., Di Giancamillo, A., Savoini, G., Dell'Orto, V., and Domeneghini, C. (2006). Live yeast dietary supplementation acts upon intestinal morpho-functional aspects and growth in weanling piglets. *Anim. Feed Sci. Technol.* 129, 224–236. doi:10.1016/j.anifeedsci.2005.12.015

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527. doi:10.1038/nbt.3519

Che, D., Adams, S., Zhao, B., Qin, G., and Jiang, H. (2019). Effects of dietary L-arginine supplementation from conception to post- weaning in piglets. *Curr. Protein Pept. Sci.* 20, 736–749. doi:10.2174/1389203720666190125104959

Coburn, L. A., Horst, S. N., Allaman, M. M., Brown, C. T., Williams, C. S., Hodges, M. E., et al. (2016). L-arginine availability and metabolism is altered in ulcerative colitis. *Inflamm. Bowel Dis.* 22, 1847–1858. doi:10.1097/MIB.0000000000000790

Collins, C. L., Pluske, J. R., Morrison, R. S., McDonald, T. N., Smits, R. J., Henman, D. J., et al. (2017). Post-weaning and whole-of-life performance of pigs is determined by live weight at weaning and the complexity of the diet fed after weaning. *Anim. Nutr. (Zhongguo xu mu shou yi xue hui)* 3, 372–379. doi:10.1016/j.aninu.2017.01.001

FAO (2019). *Meat market review*. Rome, Italy: Food Agric. Organ. United Nations.

Flynn, N. E., and Wu, G. (1997). Glucocorticoids play an important role in mediating the enhanced metabolism of arginine and glutamine in enterocytes of postweaning pigs. *J. Nutr.* 127, 732–737. doi:10.1093/jn/127.5.732

Frese, S. A., Parker, K., Calvert, C. C., and Mills, D. A. (2015). Diet shapes the gut microbiome of pigs during nursing and weaning. *Microbiome* 3, 28. doi:10.1186/s40168-015-0091-8

Gondret, F., Vincent, A., Houée-Bigot, M., Siegel, A., Lagarrigue, S., Causeur, D., et al. (2017). A transcriptome multi-tissue analysis identifies biological pathways and genes associated with variations in feed efficiency of growing pigs. *BMC Genomics* 18, 244. doi:10.1186/s12864-017-3639-0

Gresse, R., Chaucheyras-Durand, F., Fleury, M. A., Van de Wiele, T., Forano, E., and Blanquet-Diot, S. (2017). Gut microbiota dysbiosis in postweaning piglets: Understanding the keys to health. *Trends Microbiol.* 25, 851–873. doi:10.1016/j.tim.2017.05.004

Guevarra, R. B., Hong, S. H., Cho, J. H., Kim, B. R., Shin, J., Lee, J. H., et al. (2018a). The dynamics of the piglet gut microbiome during the weaning transition in association with health and nutrition. *J. Anim. Sci. Biotechnol.* 9, 54–59. doi:10.1186/s40104-018-0269-6

Guevarra, R. B., Hong, S. H., Cho, J. H., Kim, B. R., Shin, J., Lee, J. H., et al. (2018b). The dynamics of the piglet gut microbiome during the weaning transition in association with health and nutrition. *J. Anim. Sci. Biotechnol.* 9, 54. doi:10.1186/s40104-018-0269-6

Guevarra, R. B., Lee, J. H., Lee, S. H., Seok, M. J., Kim, D. W., Kang, B. N., et al. (2019). Piglet gut microbial shifts early in life: Causes and effects. *J. Anim. Sci. Biotechnol.* 10, 1. doi:10.1186/s40104-018-0308-3

Han, G. G., Lee, J. Y., Jin, G. D., Park, J., Choi, Y. H., Kang, S. K., et al. (2018). Tracing of the fecal microbiota of commercial pigs at five growth stages from birth to shipment. *Sci. Rep.* 8, 6012. doi:10.1038/s41598-018-24508-7

Henchion, M., Hayes, M., Mullen, A., Fenelon, M., and Tiwari, B. (2017). Future protein supply and demand: Strategies and factors influencing a sustainable equilibrium. *Foods* 6, E53. doi:10.3390/foods6070053

Jing, L., Hou, Y., Wu, H., Miao, Y., Li, X., Cao, J., et al. (2015). Transcriptome analysis of mRNA and miRNA in skeletal muscle indicates an important network for differential Residual Feed Intake in pigs. *Sci. Rep.* 5, 11953. doi:10.1038/srep11953

Ke, S., Fang, S., He, M., Huang, X., Yang, H., Yang, B., et al. (2019). Age-based dynamic changes of phylogenetic composition and interaction networks of health pig gut microbiome feeding in a uniformed condition. *BMC Vet. Res.* 15, 172–185. doi:10.1186/s12917-019-1918-5

Lallès, J.-P. (2008). Nutrition and gut health of the young pig around weaning : What news. *Arch. Zootech.* 11, 5–15.

Lallès, J. P., Bosi, P., Smidt, H., and Stokes, C. R. (2007). Nutritional management of gut health in pigs around weaning. *Proc. Nutr. Soc.* 66, 260–268. doi:10.1017/S0029665107005484

Le Bon, M., Davies, H. E., Glynn, C., Thompson, C., Madden, M., Wiseman, J., et al. (2010). Influence of probiotics on gut health in the weaned pig. *Livest. Sci.* 133, 179–181. doi:10.1016/j.livsci.2010.06.058

Lessard, M., Dupuis, M., Gagnon, N., Nadeau, É., Matte, J. J., Goulet, J., et al. (2009). Administration of Pediococcus acidilactici or *Saccharomyces cerevisiae* boulardii modulates development of porcine mucosal immunity and reduces intestinal bacterial translocation after *Escherichia coli* challenge. *J. Anim. Sci.* 87, 922–934. doi:10.2527/jas.2008-0919

Luiking, Y. C., Ten Have, G. A. M., Wolfe, R. R., and Deutz, N. E. P. (2012). Arginine de novo and nitric oxide production in disease states. *Am. J. Physiol. Endocrinol. Metab.* 303, E1177–E1189. doi:10.1152/ajpendo.00284.2012

Luppi, A. (2017). Swine enteric colibacillosis: Diagnosis, therapy and antimicrobial resistance. *Porc. Health Manag.* 3, 16. doi:10.1186/s40813-017-0063-4

Mach, N., Berri, M., Esquerré, D., Chevaleyre, C., Lemonnier, G., Billon, Y., et al. (2014). Extensive expression differences along porcine small intestine evidenced by transcriptome sequencing. *PLoS One* 9, e88515. doi:10.1371/journal.pone.0088515

Mach, N., Berri, M., Estellé, J., Levenez, F., Lemonnier, G., Denis, C., et al. (2015). Early-life establishment of the swine gut microbiome and impact on host phenotypes. *Environ. Microbiol. Rep.* 7, 554–569. doi:10.1111/1758-2229.12285

Macleod, M., Gerber, P., Mottet, A., Tempio, G., Falcucci, A., Opio, C., et al. (2013). *Greenhouse gas emissions from pig and chicken supply chains - a global life cycle assessment*. Rome, Italy: Food and Agriculture Organization of the United Nations FAO.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10, doi:10.14806/ej.17.1.200

Matsuo, K., Ota, H., Akamatsu, T., Sugiyama, A., and Katsuyama, T. (1997). Histochemistry of the surface mucous gel layer of the human colon. *Gut* 40, 782–789. doi:10.1136/gut.40.6.782

Moeser, A. J., Pohl, C. S., and Rajput, M. (2017). Weaning stress and gastrointestinal barrier development: Implications for lifelong gut health in pigs. *Anim. Nutr.* 3, 313–321. doi:10.1016/j.aninu.2017.06.003

Moeser, A. J., Ryan, K. A., Nighot, P. K., and Blikslager, A. T. (2007). Gastrointestinal dysfunction induced by early weaning is attenuated by delayed weaning and mast cell blockade in pigs. *Am. J. Physiol. Gastrointest. Liver Physiol.* 293, G413–G421. doi:10.1152/ajpgi.00304.2006

Nowland, T. L., Plush, K. J., Barton, M., and Kirkwood, R. N. (2019). Development and function of the intestinal microbiome and potential implications for pig production. *Animals.* 9, E76. doi:10.3390/ani9030076

Pié, S., Lallès, J. P., Blazy, F., Laffitte, J., Sève, B., and Oswald, I. P. (2004). Weaning is associated with an upregulation of expression of inflammatory cytokines in the intestine of piglets. *J. Nutr.* 134, 641–647. doi:10.1093/jn/134.3.641

Pimentel, H., Bray, N. L., Puente, S., Melsted, P., and Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. Methods* 14, 687–690. doi:10.1038/nmeth.4324

Piórkowska, K., Żukowski, K., Ropka-Molik, K., Tyra, M., and Gurgul, A. (2018). A comprehensive transcriptome analysis of skeletal muscles in two Polish pig breeds differing in fat and meat quality traits. *Genet. Mol. Biol.* 41, 125–136. doi:10.1590/1678-4685-gmb-2016-0101

Pluske, J. R., Turpin, D. L., and Kim, J. C. (2018). Gastrointestinal tract (gut) health in the young pig. *Anim. Nutr.* 4, 187–196. doi:10.1016/j.aninu.2017.12.004

Razzuoli, E., Villa, R., Sossi, E., and Amadori, M. (2011). Characterization of the interferon-α response of pigs to the weaning stress. *J. Interferon Cytokine Res.* 31, 237–247. doi:10.1089/jir.2010.0041

Rhouma, M., Fairbrother, J. M., Beaudry, F., and Letellier, A. (2017). Post weaning diarrhea in pigs: Risk factors and non-colistin-based control strategies. *Acta Vet. Scand.* 59, 31. doi:10.1186/s13028-017-0299-7

Roura, E., Koopmans, S. J., Lallès, J. P., Le Huerou-Luron, I., De Jager, N., Schuurman, T., et al. (2016). Critical review evaluating the pig as a model for human nutritional physiology. *Nutr. Res. Rev.* 29, 60–90. doi:10.1017/S0954422416000020

Sciascia, Q., Daş, G., and Metges, C. C. (2016). Review: The pig as a model for humans: Effects of nutritional factors on intestinal function and health. *J. Animal Sci.* 94, 441–452. doi:10.2527/jas2015-9788

Stuehr, D. J. (2004). Enzymes of the L-arginine to nitric oxide pathway. *J. Nutr.* 134, 2748S–2751S. doi:10.1093/jn/134.10.2748S

Wang, J., Chen, L., Li, P., Li, X., Zhou, H., Wang, F., et al. (2008). Gene expression is altered in piglet small intestine by weaning and dietary glutamine supplementation. *J. Nutr.* 138, 1025–1032. doi:10.1093/jn/138.6.1025

Wang, W., Hu, H., Zijlstra, R. T., Zheng, J., and Gänzle, M. G. (2019). Metagenomic reconstructions of gut microbial metabolism in weanling pigs. *Microbiome* 7, 48. doi:10.1186/s40168-019-0662-1

Wu, X., Ruan, Z., Gao, Y., Yin, Y., Zhou, X., Wang, L., et al. (2010). Dietary supplementation with L-arginine or N-carbamylglutamate enhances intestinal growth and heat shock protein-70 expression in weanling pigs fed a corn-and soybean meal-based diet. *Amino Acids* 39, 831–839. doi:10.1007/s00726-010-0538-y

Xu, X., Mishra, B., Qin, N., Sun, X., Zhang, S., Yang, J., et al. (2019). Differential transcriptome analysis of early postnatal developing longissimus dorsi muscle from two pig breeds characterized in divergent myofiber traits and fatness. *Anim. Biotechnol.* 30, 63–74. doi:10.1080/10495398.2018.1437045

Zeng, M. Y., Inohara, N., and Nuñez, G. (2017). Mechanisms of inflammation-driven bacterial dysbiosis in the gut. *Mucosal Immunol.* 10, 18–26. doi:10.1038/mi.2016.75

Zhang, B., Gan, L., Shahid, M. S., Lv, Z., Fan, H., Liu, D., et al. (2019). *In vivo* and *in vitro* protective effect of arginine against intestinal inflammatory response induced by *Clostridium perfringens* in broiler chickens. *J. Anim. Sci. Biotechnol.* 10, 73. doi:10.1186/s40104-019-0371-4

Zhang, Q., Widmer, G., and Tzipori, S. (2013). A pig model of the human gastrointestinal tract. *Gut Microbes* 4, 193–200. doi:10.4161/gmic.23867

Zheng, P., Song, Y., Tian, Y., Zhang, H., Yu, B., He, J., et al. (2018). Dietary arginine supplementation affects intestinal function by enhancing antioxidant capacity of a nitric oxide-independent pathway in low-birth-weight piglets. *J. Nutr.* 148, 1751–1759. doi:10.1093/jn/nxy198

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 28 November 2022
DOI 10.3389/fgene.2022.1060529

OPEN ACCESS

EDITED BY
Turgay Unver,
FicusBio, Turkey

REVIEWED BY
Jianzhao Li,
Ludong University, China
Yong-Fang Li,
Henan Normal University, China

*CORRESPONDENCE
Xining Gao,
syaugxn@syau.edu.cn

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

RECEIVED 03 October 2022
ACCEPTED 14 November 2022
PUBLISHED 28 November 2022

CITATION
Li M, Li H, Sun A, Wang L, Ren C, Liu J and Gao X (2022), Transcriptome analysis reveals key drought-stress-responsive genes in soybean.
Front. Genet. 13:1060529.
doi: 10.3389/fgene.2022.1060529

Transcriptome analysis reveals key drought-stress-responsive genes in soybean

Mingqian Li[1], Hainan Li[1], Anni Sun[1], Liwei Wang[1], Chuanyou Ren[1], Jiang Liu[1] and Xining Gao[1,2]*

[1]College of Agronomy, Shenyang Agricultural University, Shenyang, China, [2]Liaoning Key Laboratory of Agrometeorological Disasters, Shenyang, China

Drought is the most common environmental stress and has had dramatic impacts on soybean (*Glycine* max L.) growth and yield worldwide. Therefore, to investigate the response mechanism underlying soybean resistance to drought stress, the drought-sensitive cultivar "Liaodou 15" was exposed to 7 (mild drought stress, LD), 17 (moderate drought stress, MD) and 27 (severe drought stress, SD) days of drought stress at the flowering stage followed by rehydration until harvest. A total of 2214, 3684 and 2985 differentially expressed genes (DEGs) in LD/CK1, MD/CK2, and SD/CK3, respectively, were identified by RNA-seq. Weighted gene co-expression network analysis (WGCNA) revealed the drought-response TFs such as WRKY (*Glyma.15G021900, Glyma.15G006800*), MYB (*Glyma.15G190100, Glyma.15G237900*), and bZIP (*Glyma.15G114800*), which may be regulated soybean drought resistance. Second, *Glyma.08G176300 (NCED1), Glyma.03G222600 (SDR), Glyma.02G048400 (F3H), Glyma.14G221200 (CAD), Glyma.14G205200 (C4H), Glyma.19G105100 (CHS), Glyma.07G266200 (VTC)* and *Glyma.15G251500 (GST),* which are involved in ABA and flavonoid biosynthesis and ascorbic acid and glutathione metabolism, were identified, suggesting that these metabolic pathways play key roles in the soybean response to drought. Finally, the soybean yield after rehydration was reduced by 50% under severe drought stress. Collectively, our study deepens the understanding of soybean drought resistance mechanisms and provides a theoretical basis for the soybean drought resistance molecular breeding and effectively adjusts water-saving irrigation for soybean under field production.

KEYWORDS

drought stress, soybean, transcriptome, WGCNA, metabolic pathway, yeild

1 Introduction

Soybean (*Glycine* max L.), as one of the most important oil crops with significant economic value, has been cultivated worldwide. In addition to its macronutrients and minerals, soybean has many positive effects on human health due to its contents of oil, protein and isoflavones (Sakai and Kogiso, 2008; Choudhary and Tran, 2011; He and

Chen, 2013). However, drought is the most common environmental stress encountered by plants under the current situation of climate change, which has had dramatic impacts on plant growth and crop yield (Barnab´as et al., 2008; Qin et al., 2011; Lesk et al., 2016), such as up to a 40% reduction in soybean yield (Stacey et al., 2004; Fahad et al., 2017). Consequently, the molecular mechanisms and molecular breeding of drought tolerance in soybean remain to be explained (Lawlor, 2013).

Drought stress has negative effects at the physiological, developmental, and molecular levels in plants, including photosynthesis inhibition, reactive oxygen species (ROS) generation, and cellular tissue and membrane damage (Xu et al., 2010; Golldack et al., 2014; Zhu, 2016; Anjum et al., 2017). Thus far, a number of studies have suggested that plants use multiple physiological and molecular strategies in response to drought stress. For instance, they rapidly accumulate osmotic regulators (proline and soluble sugar), a process crucial for plant drought resistance (Dien et al., 2019; La et al., 2019). Furthermore, the increase in superoxide dismutase (SOD) and peroxidase (POD) activities enhances vitamin C production and glutathione metabolism, playing a key role in avoiding drought damage (Sun et al., 2020; Zhang et al., 2020). In addition, plants can activate drought stress defense though the altered expression of related genes, such as *4CL5* and *F5H1* involved in flavonoid pathway and lignin biosynthesis, which results in increased lignin content and subsequent drought tolerance (Xu et al., 2020; Sun et al., 2022).

Plant hormones responding to abiotic and biotic stresses play significant roles in plant growth and development. It has been well documented that drought stress can cause the biosynthesis and signal transduction of various plant hormones, especially abscisic acid (ABA) (Khan et al., 2015; Vishwakarma et al., 2017). Significantly, the NCED gene encodes 9-cis-epoxycarotenoid dioxygenase, which can increase abscisic acid content and induce drought stress related genes, stomatal closure and other physiological processes in plants (Chimungu et al., 2014; Huang et al., 2019). The ABA responsive elements binding factor (ABF) has been reported to regulate the expression of drought-responsive genes and enhance enzyme activity to maintain plant resistance to drought stress (Kerr et al., 2018).

Various omics analyses, such as genomics, transcriptomics, proteomics, and metabolomics, have gained insight into plant responses to abiotic stresses (Basim et al., 2021). Due to technological advancements and reduced cost, RNA sequencing, has become one of the most effective methods for evaluating the interaction between plants and abiotic stresses (Zhang and Song, 2017; Xuan et al., 2022). Based on transcriptomics analysis, many transcription factors (TFs) have been detected in plants associated with drought stress, including WRKY, MYB and DREB (Lindemose et al., 2013; Wang et al., 2016), AP2/ERF (Xie et al., 2019), bZIP (Kang et al., 2019), and NAC (Thirumalaikumar et al., 2018). In addition, the mitogen-activated protein kinase (MAPK) signaling pathway and the Ca^{2+} signaling pathway have been found to be enriched under drought stress based on KEGG analysis (Zhao et al., 2020; Liu et al., 2022).

Soybean is known to be highly sensitive to water deficit; hence, an adequate water supply is crucial for its growth and development to achieve high primary production (Buezo et al., 2019; Yu et al., 2020). However, there can be a certain degree of drought stress in soybean, which has no significant effects on soybean yield due to plant growth compensation after rehydration (Hao et al., 2010; Liu et al., 2012; Xue et al., 2013). Previous studies have shown that the flowering stage of soybean is the most sensitive period for drought stress (Meckel et al., 1984). Therefore, to investigate the response mechanism of soybean under varying degrees of drought, the drought-sensitive cultivar "Liaodou 15" was exposed to different levels of drought stress at the flowering stage by gradually decreasing the amount of irrigation at different levels and durations. Subsequently, physiological evaluation and transcriptomic analyses in soybean leaves under drought stress were performed. Then, WGCNA based on the transcriptome data and physiological indices was performed to investigate the function of TFs and identify genes of several key pathways in soybean response under drought stress. Meanwhile, we analyzed the crop yield of soybean after rehydration in the harvesting period. Collectively, the aim of this work was to elucidate the mechanism underlying the response of soybean to drought stress, provide a theoretical basis for the molecular breeding of drought resistance, and effectively adjust water-saving irrigation for soybean production under field conditions.

2 Materials and methods

2.1 Plant materials and equipment

The soybean, the drought-sensitive cultivar "Liaodou 15", was grown in the scientific observation and experimental station of crop cultivation in Northeast China, Ministry of Agriculture and Rural Affairs, P. R. China, located at Shenyang Agricultural University (123.53°E, 41.73°N, Shenyang, China).

The area has a brown soil type, and the soil capacity at the time was 30%. The basic soil fertility data are shown in Supplementary Table S1. The experimental equipment included a sliding plastic film rain shelter with a reinforced steel frame, which was used on rainy days, and a soil moisture and temperature sensor buried approximately 30 cm deep in the soil (SMTS-II-485, China). This study applied a drip system to control the amount of water released during each treatment to ensure uniform irrigation. Except for the water control, the other cultivation measures were the same as in standard procedures during the entire growth period.

2.2 Experimental design

There were three drought stress treatments in the experiment and three repetitions for the control. This study adopted a randomized block design. Each drought treatment had three plot replicates, for a total number of 12. The size of each plot was 2 × 3.6 m. A total of 108 soybean plants were grown in each plot. The relative soil water content was measured by a soil moisture and temperature sensor (SMTS-II-485, China). Before flowering, all plots should maintain the same soil moisture conditions and sufficient water content.

For drought treatments, soybean plants at the early flowering stage were continuously subjected to drought stress. 1) Water was continuously withheld from the first group for 7 days. As a result, the soybean leaves appeared curled (mild, the soil moisture content was 24.3%, LD). 2) Water was continuously withheld from the second group for 17 days, wilting and curling (moderate, the soil moisture content was 20.6%, MD). 3) Water was continuously withheld from the third group for 27 days, severe wilting and curling (severe, the soil moisture content was 16.9%, SD). 4) The control leaves of each treatment with the same developmental stage were grown in soil with 30% relative water content (CK1, CK2, and CK3), which remained green, fully expanded and healthy. After the drought stress period was completed, rehydration was performed with the control level on the same day, and this was maintained until harvest. Meanwhile, the top two to three leaf samples with each drought stress treatment and corresponding controls were collected on the day of completing the drought experiment, and the samples were quickly frozen in liquid nitrogen, then stored at −80°C until measurement. The leaves of each soybean sample type (n = 30) were used for the evaluation of physiological indices and transcriptome analysis. Soybean seeds were used for yield measurement. All experiments were performed with at least three biological replicates.

2.3 Determination of physiological indices

The collected soybean leaves under drought stress were examined for seven physiological indices. The contents of soluble sugar (cat no. YX-W-B602), soluble protein (cat no. YX-W-C202), chlorophyll (YX-W-A304), proline (ca no. YX-W-A605) and malondialdehyde (MDA, cat no. YX-W-A401), as well as the SOD (cat no. YX-W-A500 -WST-8) and CAT (cat no. YX-W-A501) activities were determined according to the instructions of the physiological index assay kit provided by Sinobestbio Biotechnology Co., Ltd. (Shanghai, China). Each sample was used for three technical replications.

2.4 Soybean RNA-seq analysis

Leaves of soybean were collected at 7, 17, and 27 d after drought stress with three biological replicates. Transcriptome sequencing was performed by Genedenovo Biotechnology Co., Ltd (Guangzhou, China). Briefly, total RNA was extracted and checked for purity and integrity using a NanoDrop 2000 (Thermo Scientific, Waltham, MA, United States) and the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 System (Agilent Technologies, Santa Clara, CA, United States), respectively. The qualified RNA was prepared for the construction of cDNA libraries and sequenced using an Illumina sequencing platform. The clean reads were mapped to the soybean genome (https://phytozome-next.jgi.doe.gov/info/Gmax_Wm82_a4_v1) by HISAT2 tools (Langmead et al., 2009; Kim et al., 2015). The raw sequencing data generated from this study were archived in NCBI SRA (http://www.ncbi.nlm.nih.gov/sra) with the BioProject accession number PRJNA852689. A power analysis for sequencing depth was calculated by RNASeqPower (https://doi.org/doi:10.18129/B9.bioc.RNASeqPower), which uses RNA-seq data analysis to examine transcription patterns. Differential expression analysis of soybean plants under drought stress was performed using DESeq2 (Love et al., 2014). DEGs with $|\log2FC| > 1$ and p-values ≤ 0.05 were retained and considered significantly upregulated or downregulated groups, respectively. The expressed gene function annotations were conducted by the Gene Ontology (GO) database. All molecular pathways were explored by the Kyoto Encyclopedia of Genes and Genomes (KEGG).

2.5 Weighted gene co-expression network analysis

The R package "weighted gene co-expression network analysis" (WGCNA) was used to identify the network of genes from the transcriptome data and seven physiological indices. The DEGs were divided into different modules marked with different colors based on similar expression patterns. Combined with the changes in seven physiological indices, the DEGs associated with drought resistance were analyzed in the modules using the significant correlation coefficient, which was performed by KEGG and GO enrichment analysis.

2.6 Validation of RNA–seq data by qRT–PCR analysis

Total RNA from soybean leaves was extracted by an ultrapure RNA kit (Cat#CW0581, CWbio. Co. Ltd., Beijing, China) and treated with DNase I (DNA free, Takara, Dalian, China). The concentration and quality of total RNA was checked by a NanoDrop device (Thermo, Fisher Scientific). First-strand cDNA synthesis was carried out from 1 μg of treated total RNA using PrimeScript™ RT Master Mix (Takara, Dalian, China) in a total volume of 20 μl.

Relative gene expression was quantified by qRT–PCR using SuperReal PreMix Plus (SYBR Green) (Takara, Dalian, China) according to the manufacturer's protocol on an ABI PRISM 7500 sequence detection system (Applied Biosystems, Thermo Fisher Scientific, United States). The gene-specific primers for real-time quantitative PCR (qRT–PCR) were designed using the NCBI online tool Primer-blast (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and are listed in Supplementary Table S2. The amplification program was as follows: one cycle of 30 s at 95 °C, followed by 45 cycles of 5 s at 95°C, 15 s at 60°C and 72°C for 15 s. The soybean gene *GmUKN1 (Glyma12g02310)* served as the internal reference gene and the $2^{-\Delta\Delta Ct}$ method was used to analyze relative changes of gene expression (Livak and Schmittgen, 2001; Lu et al., 2016). Three biological and three technical replicates were included in the qRT–PCR analysis.

2.7 Soybean yield measurement

After soybean ripening, thirty soybean plants were randomly selected from each plot and harvested in the laboratory. The seed weight and soybean yield in each plot were investigated.

2.8 Statistical analysis

All measurements were repeated at least three times. The data are presented as the mean values ± standard deviations (SD) and subjected to analysis of variance (ANOVA). Multiple comparisons were conducted by Duncan's multiple range test at $p < 0.05$ using the SPSS statistics program 18.0. The differences between the physiological index and gene expression in response to drought stress were considered statistically significant at p-values < 0.05. The charts were drawn by Excel 2010 (Microsoft, Redmond, WA, United States) and TBtools (Chen et al., 2020).

3 Results

3.1 Physiological responses of soybean under drought stress

To gain insights into the mechanism of the drought stress response in soybean at the flowering stage, we triggered a controlled water deficit in the experimental application to simulate different drought levels. Plants in each treatment group to a different level of soil moisture content are shown in Supplementary Figure S1, which was measured by a soil moisture and temperature sensor (SMTS-II-485, China). When exposed to drought stress for 7 d, the soybean leaves showed slightly curled edges, corresponding to mild drought stress (24.3% soil moisture content, LD). After 17 d of water deprivation, moderate drought stress was attained (20.6% soil moisture content, MD), and some soybean plants turned curved and partially yellowed surfaces with signs of water loss, while control plants remained healthy. Moreover, following 27 days without watering, almost all soybean plants appeared severe wilting and curling, indicating severe drought stress-induced damage (16.9% soil moisture content, SD), whereas the leaves of the control remained green and fully expanded (Figure 1A). Accompanied by changes in leaf phenotype, the chlorophyll content steadily decreased with decreasing soil moisture content from 30% to 16.9% and was markedly lower than that under control conditions (Figure 1E). The MDA content peaked in the SD treatment at 138.4 μmol/g, MDA content in SD treatment did differ significantly among the rest of the treatments, and each drought treatment was significantly higher than that of the control (Figure 1D).

Meanwhile, for the antioxidant enzyme activity in soybean leaves, the SOD activity tended to increase first and then decrease as drought time increased from 7 d to 27 d in the soybean (Figure 1B). The SOD activity of soybean in the MD treatment reached a maximum of 791.59 U/g min at 17 days, whereas significant differences were noted among the other treatments. The CAT activity was not significantly different under the LD condition compared to the control, and CAT activity was significantly increased (Figure 1C). Furthermore, for osmotic adjustment in soybean leaves, the proline content increased following the LD, MD and SD treatments. Soybean reached a maximum proline content in the SD treatment of 193.29 μg/g (Figure 1F). Consistent with this, the contents of soluble sugars and soluble protein were not significantly different under LD compared with the control, whereas soluble sugars content of 25.64 mg/g, and 32.33 mg/g and soluble protein content with 52.84 mg/g, 66.06 mg/g in soybean for the MD and SD treatments were noted compared with their control (Figures 1G,H).

In short, these results indicated that in soybeans exposed to varying degrees of drought stress due to prolonged water control, osmotic regulatory substances were increased and antioxidant enzyme activity was enhanced to maintain plant growth and development.

3.2 RNA-seq data revealed differentially expressed genes in soybean after drought treatments

To elucidate the mechanism of the response of differentially expressed genes in drought stress in soybean, RNA-seq was performed in soybean leaves subjected to different levels of drought stress. By comparing reads to the soybean genome, the genomic alignment of each sample was obtained, and the alignment rate was approximately 93% (Supplementary Table S3). Based on three biological replicates under control and

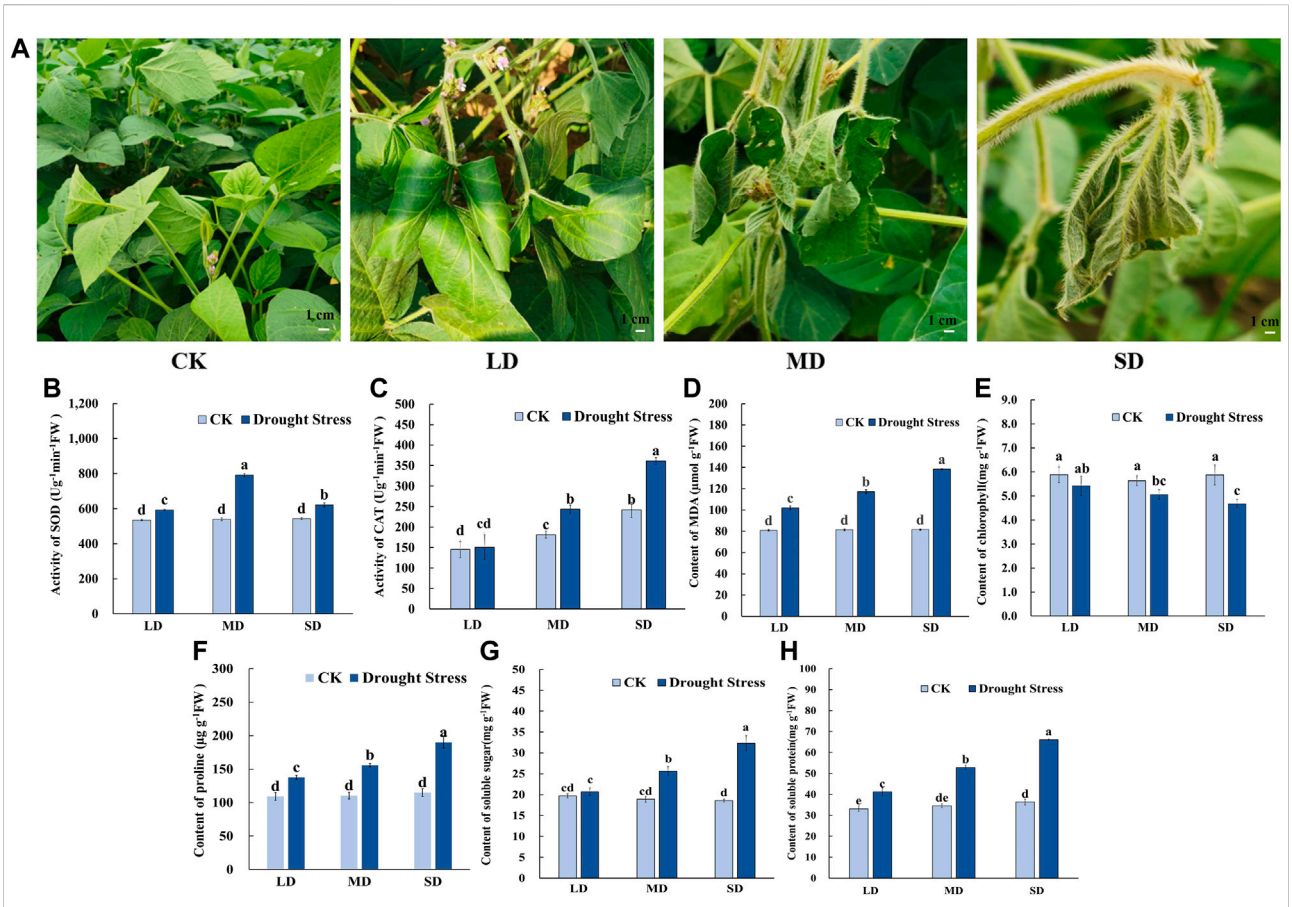


FIGURE 1
Phenotypic characteristics and physiological properties of soybean under drought stress. **(A)** Phenotypic characteristics of soybean under drought stress. **(B)** The activity of SOD. **(C)** The activity of CAT. **(D)** The content of MDA. **(E)** The content of chlorophyll. **(F)** The content of proline **(G)** The content of soluble sugars. **(H)** The content of soluble protein. Three biological replicates were performed and the data represent the means ± SDs. Different letters represent statistically significant differences ($p < 0.05$).

different drought conditions, we performed a principal component analysis (PCA). The first two principal components PC1 and PC2 accounted for 79.5% and 10.4%, respectively (Supplementary Figure S2A). These results indicated that there were different gene expression patterns between the soybeans under different levels of drought stress (Supplementary Figure S2B). In addition, the power analysis for sequencing depth was performed to identify DEGs (Supplementary Table S4). The statistical power of this experimental design, calculated in three drought treatments (CK1 vs. LD, CK2 vs. MD, and CK3 vs. SD), was 0.7655, 0.7654 and 0.7656, respectively (Supplementary Table S4).

Subsequently, the analysis of differentially expressed genes was performed in soybean leaves exposed to three levels of drought stress and compared with the control, which could identify the significantly changed genes regulated by drought stress (Figure 2). There were 1211 upregulated and 1003 downregulated genes in CK1 vs. LD, and 1865 upregulated and 1819 downregulated DEGs in CK2 vs. MD. A total of 2985 DEGs were identified in leaves exposed to SD, including 1238 upregulated and 1747 downregulated genes (Figure 2A). Furthermore, the Venn diagrams showed that 235 DEGs were common in soybeans under three drought treatments (CK1 vs. LD, CK2 vs. MD, and CK3 vs. SD) (Figure 2B). The gene expression of 235 common DEGs in soybean under three levels of drought stress is shown in Supplementary Table S5.

3.3 Functional enrichment analysis of DEGs by GO and KEGG

In order to analyze the differentially expressed genes in soybean in response to drought stress, gene ontology (GO) and KEGG analyses were performed, which highlighted the major functions of genes induced by drought stress. The GO enrichment analysis was

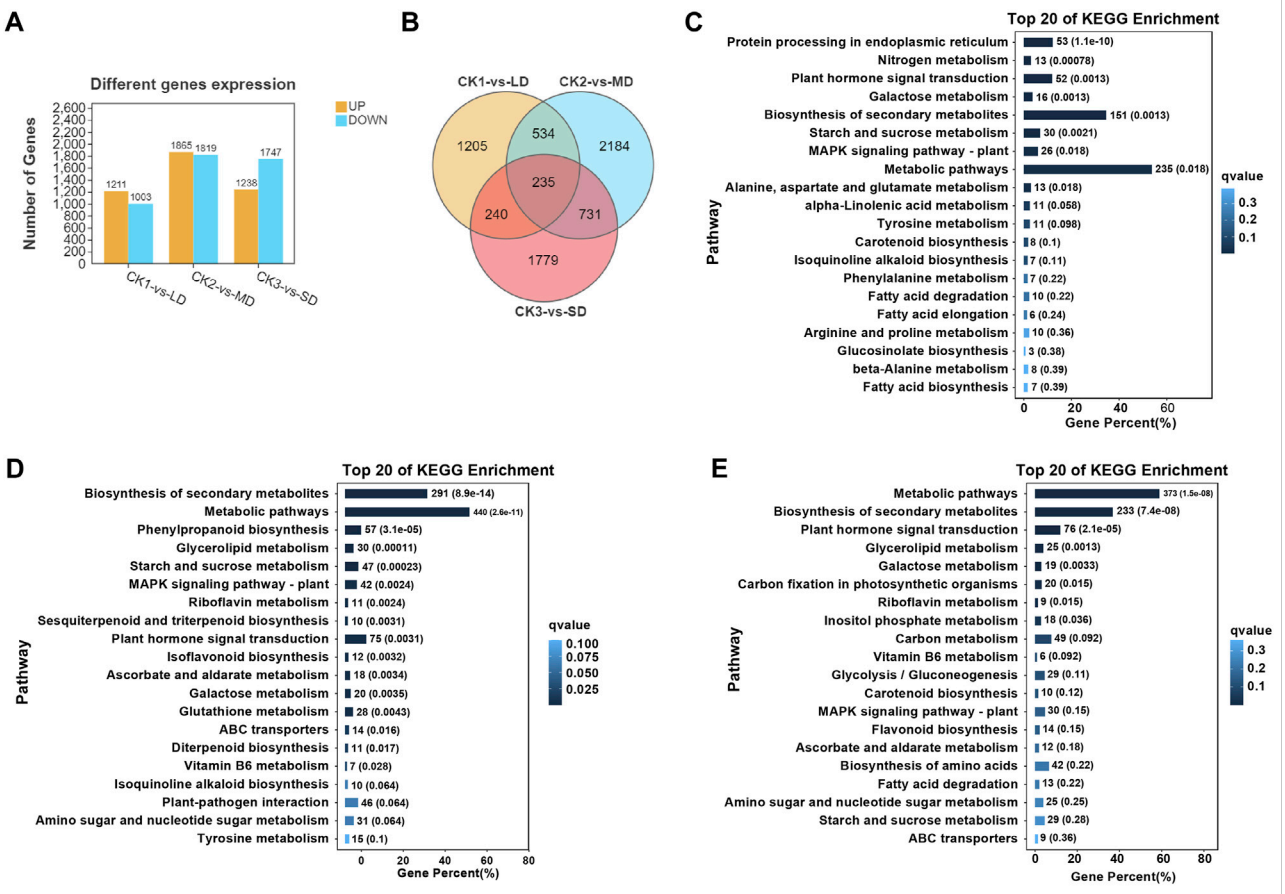


FIGURE 2
Transcriptional analysis of soybean under various drought stresses. **(A)** The numbers of DEGs. **(B)** Venn diagrams show the overlapping DEGs between the three drought stress. **(C)** The top20 enriched of KEGG pathways in soybean under LD/CK1. **(D)** The top20 enriched of KEGG pathways in soybean under MD/CK2. **(E)** The top20 enriched of KEGG pathways in soybean under SD/CK3.

conducted with division into biological process (BP), molecular function (MF) or cellular component (CC) terms. The majority of drought-responsive genes in soybean were significantly enriched regarding BP terms, including 'cellular process' and "metabolic process", as well as MF categories, including 'binding' and "catalytic activity". Several CC terms were mainly characterized as 'cell', "cell part", "organelle", and "membrane" (Supplementary Figures S3–5). The number of differentially expressed genes (up- or downregulated) in each pathway under mild drought was less than 800, while those under MD and SD were all higher than 1000. These results indicated that DEGs were related to metabolic pathways or other regulatory networks that respond to drought stress.

On the basis of the obtained transcriptome data, we specified the top 20 KEGG pathways under drought stress. The KEGG analysis showed that three pathways were significantly enriched (false discovery rate ≤5%) under varying degrees of drought stress, including starch and sucrose metabolism, plant hormone signal transduction, and MAPK signaling pathways (Figures 2C–E). Interestingly, significant enrichment of arginine and proline metabolism was detected only under mild drought stress (Figure 2C). MD and SD stress stimuli trigger the ascorbate and aldarate metabolism pathways. However, glutathione metabolism and flavonoid biosynthesis pathways were differentially expressed between MD and SD stress (Figures 2D,E). This observation indicates that the enrichment of antioxidant and secondary metabolite pathways are considered typical of the response of soybean to different drought stress treatments.

3.4 Co-expression network analysis of DEGs in soybean leaves

We further performed weighted correlation network analysis based on antioxidant enzyme activity and osmotic regulation substance content to identify gene clusters or modules that are associated with the drought stress response in soybean leaves. All genes were clustered into

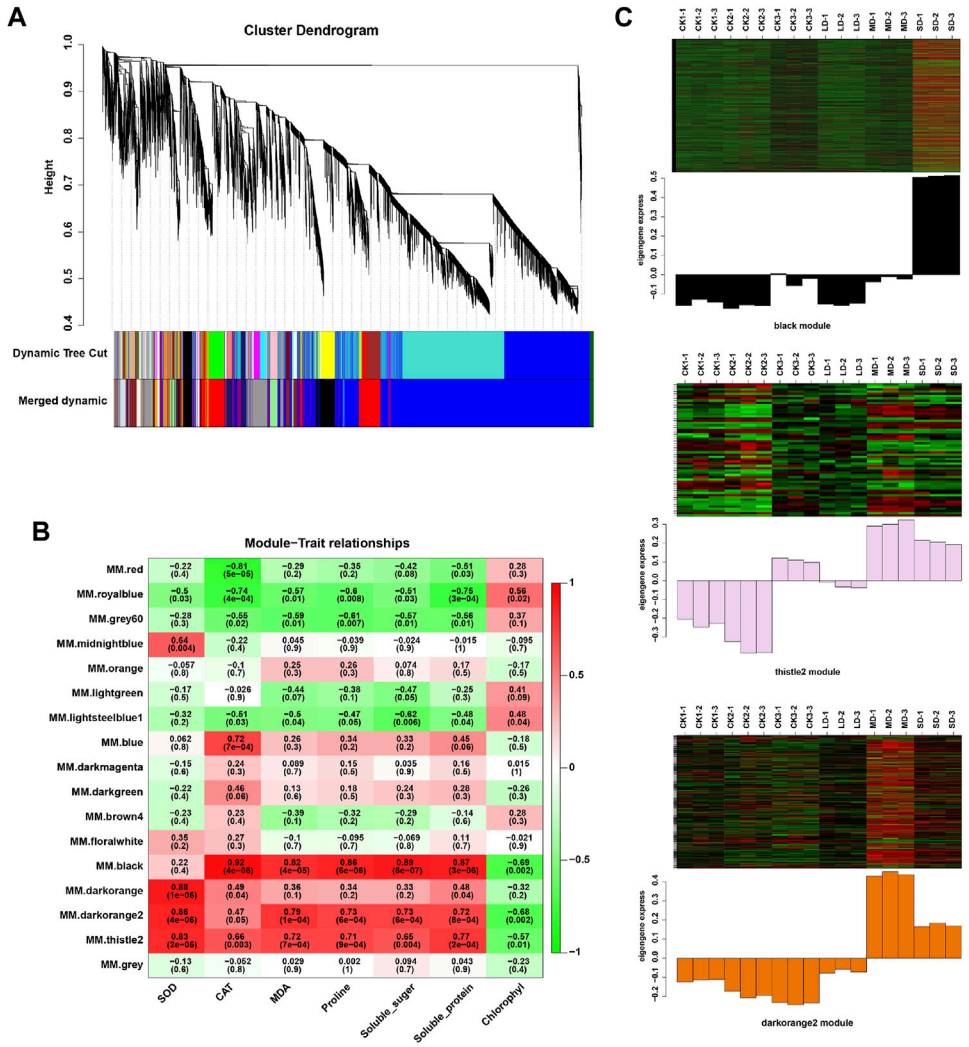


FIGURE 3
WGCNA of DEGs in soybean under various drought stresses. **(A)** Seventeen co-expression modules shown by a hierarchical cluster tree. **(B)** The correlation analysis between modules and physiological traits. **(C)** Heatmaps indicate the expression patterns of genes in the black, drakorange2, and thistle2 modules. The samples were CK1, CK2, CK3, LD, MD, and SD with three biological replicates.

27 modules (Figure 3A), of which the black modules were significantly correlated (80%) with six physiological indicators (Figure 3B). The orange and tiles modules were also significantly correlated (Figure 3B). The gene expression trends for the three modules are shown in Figure 3C. These results indicate that the network of genes and these indicators is complex, and the functioning pathways are in these modules under drought stress. The KEGG enrichment analysis of three modules showed that the MAPK signaling pathway was enriched (black modules) for most drought-stressed samples, and ascorbate and aldarate metabolism and biosynthesis of secondary metabolites were also enriched (orange modules) (Supplementary Figure S6). In addition, the expression of 135 DEGs, 15 DEGs and 5 DEGs was associated with the TFs in the respective three modules (Supplementary Figure S7). Among them, the MYB (6), bHLH (5), bZIP (3), NAC (10), WRKY (11) and AP2/ERF (9) families of TFs were predominantly differentially expressed, and a greater number of TFs were significantly upregulated under MD and SD stress, such as *WRKY* (*Glyma.15G021900, Glyma.15G006800*), *MYB* (*Glyma.15G190100, Glyma.15G237900*), and *bZIP* (*Glyma.15G114800*) (Figure 4). Taken together, the drought response in soybean leaves was positively regulated by multiple metabolic pathways and a number of transcription factors.

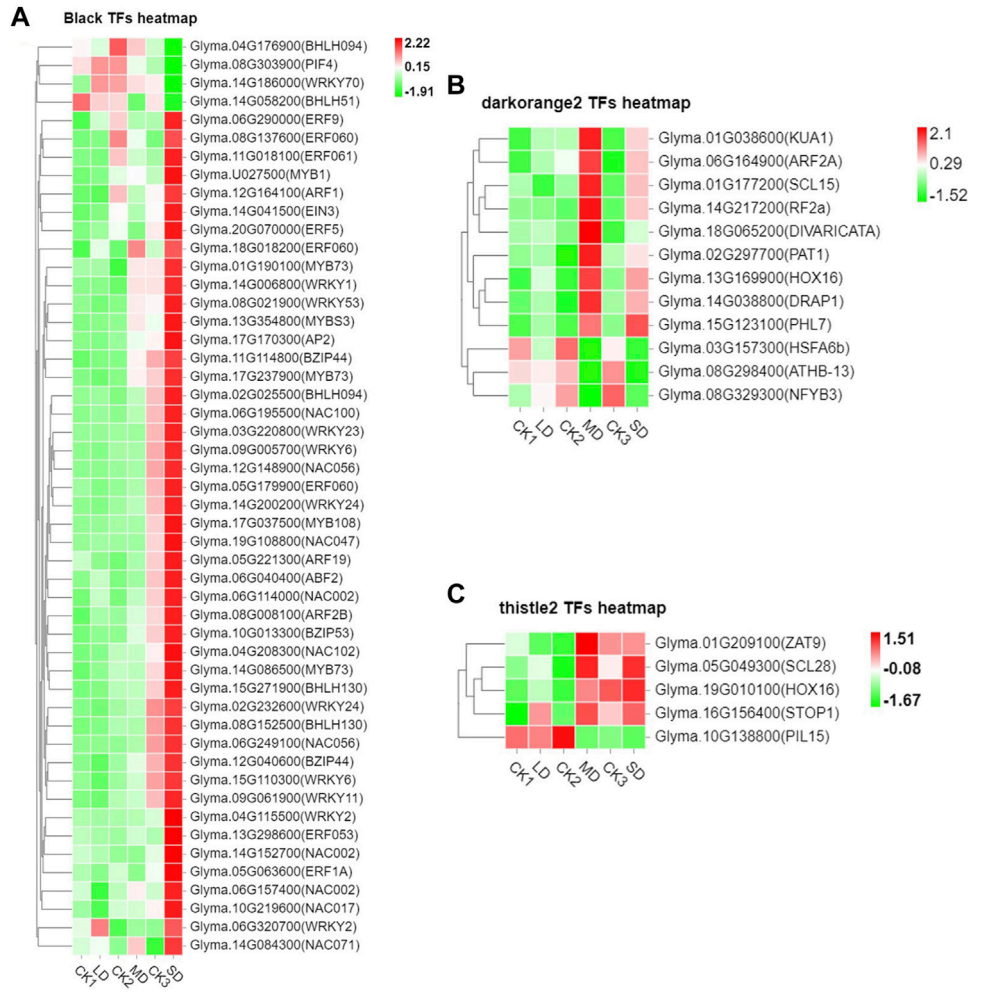


FIGURE 4
Heatmap of drought-induced TFs in three modules under drought stress. **(A)** Fifty drought-induced TFs in Black module. **(B)** Twelve drought-induced TFs in the Darkorange2 module. **(C)** Five drought-induced TFs in the Thistle2 module. The colors of the heatmap vary from green to red by normalizing the $\log_2$ (FPKM) of each gene.

3.5 DEGs involved in ABA biosynthesis and signaling pathways under drought stress

The drought-responsive DEGs were significantly enriched in the plant hormone signal transduction pathways. As an important plant hormone, ABA plays an essential role in drought stress. Therefore, we monitored the expression patterns of ABA biosynthesis and its signaling related genes in the transcriptome data. Multiple genes that synthesize ABA were upregulated. Two ZEPs were upregulated only in LD treatment. At the same time, *NCEDs* (*Glyma.15G250100 and Glyma.08G176300*), the key genes in ABA synthesis, and *SDR* (*Glyma.03G222600*) were upregulated in all groups. *AAO* was only upregulated in the MD treatment. In addition, the metabolic genes AOG and *CYP707A* (*Glyma.09G282900*) were downregulated (Figure 5). Five ABA-induced PP2Cs were significantly upregulated by drought stress in the soybean, of which *Glyma.14G162100, Glyma.19G069200,* and *Glyma.08G33800* were upregulated under LD, MD and SD stress. In addition, three *SnRK2* genes (*Glyma.17G148800, Glyma.09G066700, Glyma.02G176100*) were upregulated, and six such genes were downregulated under MD or SD stress. Finally, three ABF TFs (*Glyma.06G040400* in LD and MD, *Glyma.04G039300* in LD, *Glyma.12G184400* in SD) were upregulated under three drought stress levels and in turn activated ABA-responsive response genes (Figure 5). This finding suggested that ABA-related genes are induced by varying degrees of drought stress and thus play important roles in the soybean response.

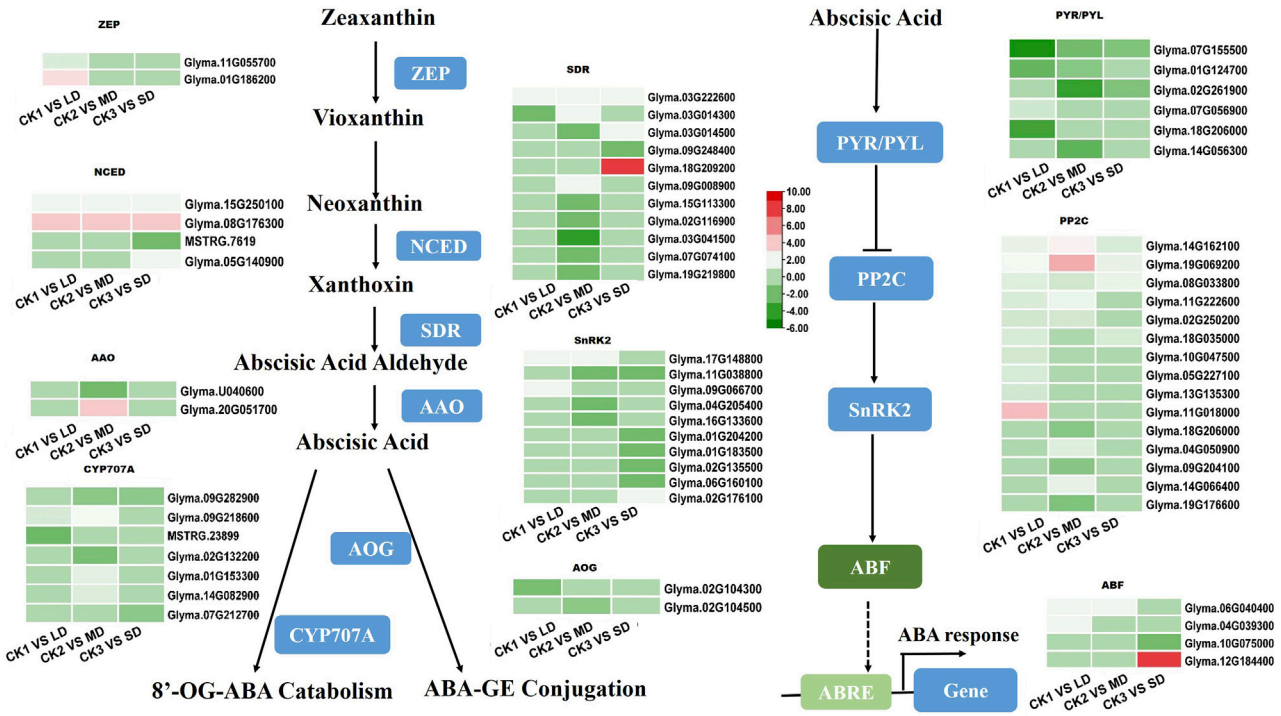


FIGURE 5
Heatmap of drought-responsive DEGs involved in ABA biosynthesis and signaling pathways under drought stress. Heatmap of drought-responsive DEGs involved in ABA biosynthesis and signaling pathways under drought stress. The colors of heatmap vary from green to red by normalizing the $\log_2$ (FPKM) of each gene.

3.6 DEGs involved in AsA and GSH biosynthesis pathways under drought stress

To identify the genes related to ascorbate biosynthesis and glutathione metabolism in soybean under drought stress, the DEGs related to ascorbate metabolism were analyzed (Figure 6). The six DEGs were found to be linked to ascorbate metabolism at all three drought stress levels. Some of these, DEGs, such as *GME* (*Glyma.10G162000*) and *VTC2-5* (*Glyma.02G292800*) were induced by SD stress and three *VTC4* genes (*Glyma.07G266200, Glyma.09G011100, Glyma.15G115500*) were upregulated under MD and SD stress (Figure 6A). Meanwhile, the DEGs were found to be related to glutathione metabolism under three levels of drought stress. Of these, six GST genes (glutathione S-transferase) and two *G6PDH* genes (*Glyma.18G284600, Glyma.17G096800*) were upregulated under SD stress, and some *GST* genes were downregulated under LD or MD stress. Furthermore, two *PGD* genes (*Glyma.19G038400, Glyma.05G214000*) and one *GPX* gene (*Glyma.01G219400*) were upregulated in the SD treatment (Figure 6B). These results showed that ascorbate and glutathione metabolism genes were activated by drought stress in soybean, especially by serious drought stress.

3.7 DEGs involved in flavonoid biosynthesis pathways under drought stress

Based on the KEGG and WGCNA analyses, we further identified the DEGs involved in phenylalanine and flavonoid biosynthesis pathways, that were significantly related to drought responses in soybean leaves. We found that the DEGs enriched in phenylalanine and flavonoid biosynthesis pathways were unique under MD and SD treatments. Mild stress had no effects on phenylalanine and flavonoid biosynthesis pathways and was not linked to differentially expressed genes (Figure 7). Two genes (*Glyma.10G209800, Glyma.20G180800*) encoding PAL proteins involved in the phenylalanine pathway showed markedly increased expression in soybean leaves under LD or SD conditions relative to control plants. It is worth noting that *C4H* genes (*Glyma.14G205200*) and two *4CL* genes (*Glyma.01G232400, Glyma.05G075100*) were upregulated by MD and SD treatments (Figure 7A). Next, we further analyzed the genes of the flavonoid pathway under drought stress. The genes *CHS* (*Glyma.19G105100*) and *F3H* (*Glyma.02G048400*) were significantly upregulated under MD and SD stress. Meanwhile, *DFR* (*Glyma.17G252200*) and two *ANS* genes

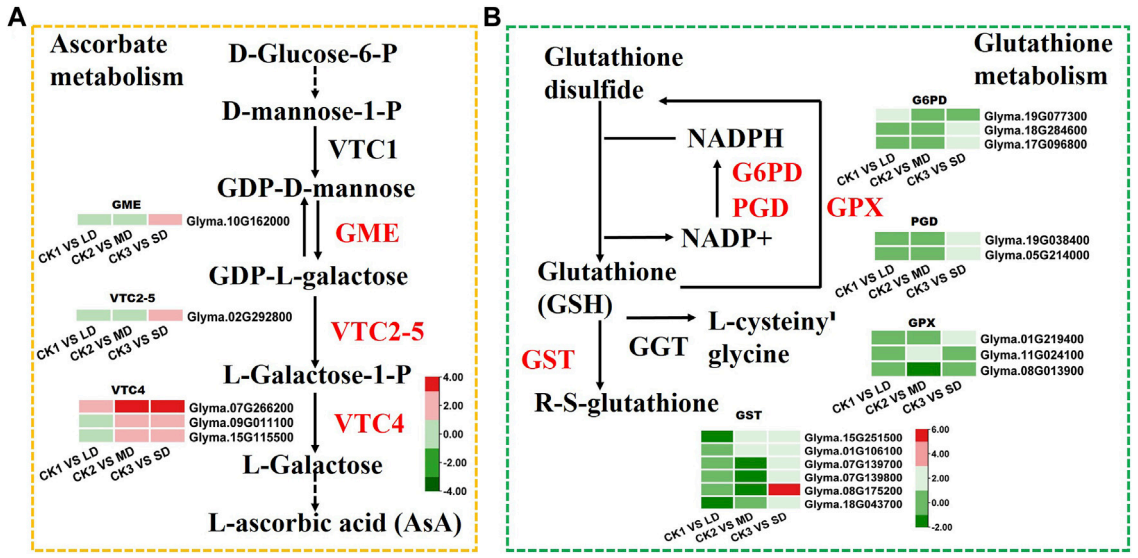


FIGURE 6
Heatmap of the DEGs involved in ascorbate and glutathione metabolism in response to drought stress. **(A)** The DEGs involved in ascorbate metabolism. **(B)** The DEGs involved in glutathione metabolism. The colors of heatmap vary from green to red by normalizing the $\log_2$ (FPKM) of each gene.

(*Glyma.11G027700, Glyma.01G214200*) were significantly upregulated under SD stress (up to 3-fold) (Figure 7B). These results suggest that the DEGs involved in the phenylalanine and flavonoid pathways are perhaps activated mainly by severe drought stress rather than mild drought stress.

3.8 The qRT-PCR analysis to validate RNA-seq data

To validate the gene expression data obtained from RNA-seq, we selected nine genes participated in ABA, AsA, GSH and flavonoid biosynthesis pathways under drought stress for qRT-PCR analysis. We found good agreement (r = 0.99–0.8) in with relative gene expression between RNA-Seq and qRT-PCR for all candidate genes except *Glyma.03G222600* under drought stress, which confirmed the reliability and accuracy of RNA-seq analyses in this study (Figure 8). When the soil moisture content was 20.3% (water withheld for 17 days), the upregulated genes (*Glyma.15G250100, Glyma.07G266200* and *Glyma.14G205200*) in the MD treatment increased by the largest proportion, to 3.7-, 6.5- and 7.6 fold, respectively. In turn, the up-regulated genes (*Glyma.15G251500, Glyma.02G048400, Glyma.19G105100 and Glyma.14G221200*) exhibited the largest increase (to 3.7-,6.5- and 7.6- fold) at SD treatment after 27 days without water, at which time the soil moisture content was 16.9%. Overall, when soybean exposed to MD and SD treatments, some genes involved in ABA, AsA, GSH and flavonoid biosynthesis pathways in soybean leaves were highly expressed, and as the water deficit treatment was extended, the degree of gene expression increased correspondingly.

3.9 Soybean yield after rehydration

To reveal the effect of varying degrees of drought stress on the yield of soybean after rehydration, soybean was harvested at the mature stage for analysis**.** Soybean yield significantly decreased with drought stress at different levels, whereas mild drought stress had no significant effect on soybean yield after rehydration in the harvested period ($p < 0.05$) (Figure 9). Moreover, soybean yield under MD and SD stress was approximately 0.75-and 0.5-fold lower than that of the controls, respectively ($p < 0.05$).

4 Discussion

As a symptom of a severe global climate disaster, drought leads to a decline in crop production and exacerbates direct economic losses each year (Molnár et al., 2021). In soybean, which is a global economic oilseed crop, drought stress is a major factor that can reduce yields by more than 40% (Specht et al., 1999). However, molecular insights into the drought resistance of soybean have been limited.

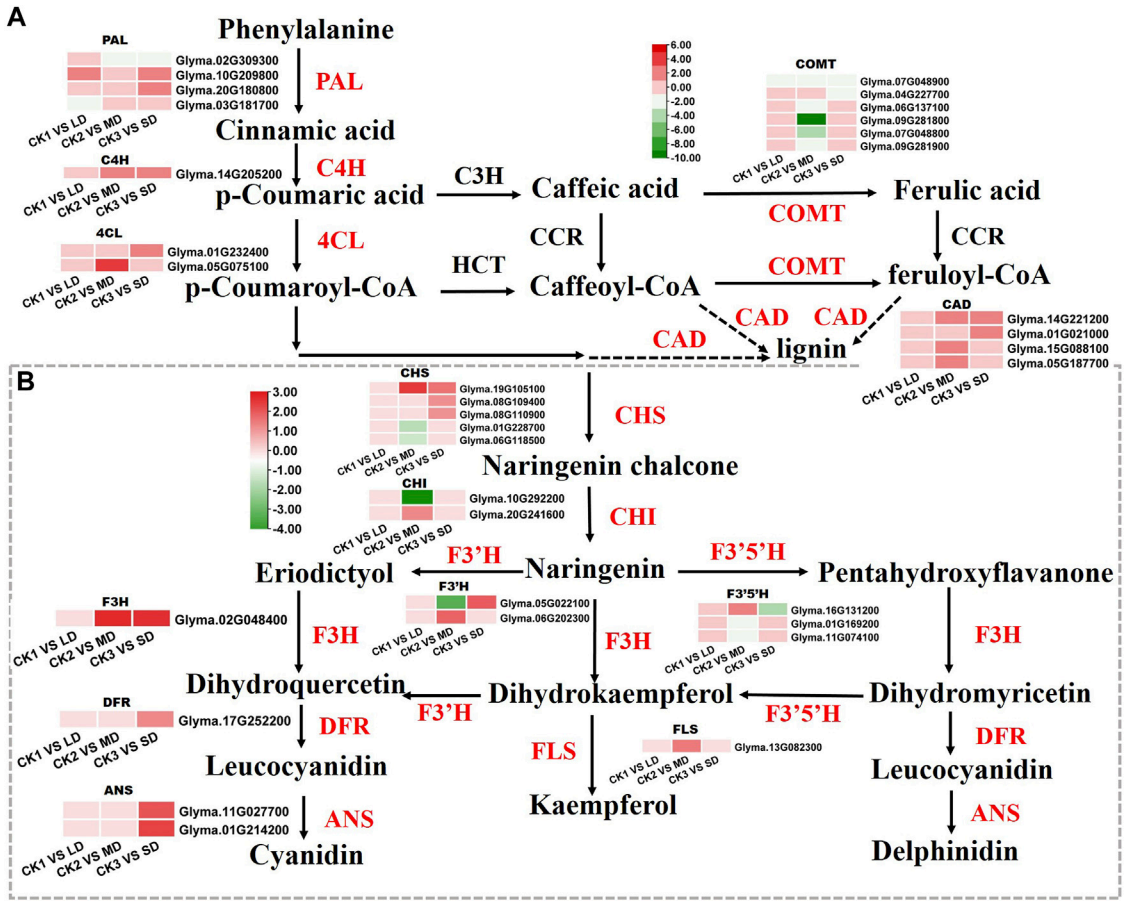


FIGURE 7
Heatmap of the DEGs involved in flavonoid biosynthesis in response to drought stress. **(A)** The DEGs involved in phenylalanine metabolism. **(B)** The DEGs involved in flavonoid biosynthesis. The colors of the heatmap vary from green to red by normalizing the $\log_2$ (FPKM) of each gene.

4.1 Physiological performance of soybean leaves subjected to different levels of drought stress

Plants have been shown to display a variety of changes to reduce drought damage through at stress response mechanisms in terms of physiology and biochemistry (Harauma et al., 2007; Song et al., 2009; Elansary and Salem, 2015; Martignago et al., 2020). In this study, mild drought stress had little effect on the growth of soybean, however, the leaves wilted, curled, and dehydrated dramatically under moderate and severe stress levels. The phenotype of plant leaves directly indicates the degree of drought stress (Li et al., 2020). When plants encounter drought stress, excess reactive oxygen species (ROS) will accumulate, and these will be scavenged by activating the antioxidant system to reduce oxidative damage, including POD, SOD and CAT activities (Miller et al., 2010). If the antioxidant system cannot remove excess reactive oxygen species in time, they will cause damage to membranes and protein function, thereby harming plant tissues and accelerating plant senescence (Mittler et al., 2004; Torres and Dangl., 2005). One of the associated damages to plants is the production of MDA and membrane lipid peroxidation (Gill and Tuteja, 2010). Our findings suggested that MDA content was increased, which was caused by ROS accumulation under drought stress (Figure 1B). The activities of SOD and CAT were successfully measured, and they were found to be upregulated to remove excess ROS and relieve membrane damage to maintain normal plants growth (Figure 1B) (Baldoni et al., 2016). Plants also adopt osmotic regulation strategies to confer drought tolerance and further reduce the damage caused by drought stress (Lotfi et al., 2010). Notably, soluble proteins, soluble sugars and proline increased approximately 2-fold under SD stress, which contributed to reducing the osmotic potential and membrane protection, as well as maintaining water uptake in plants in response to drought stress (Ashraf and Foolad, 2007; Seki et al., 2007).

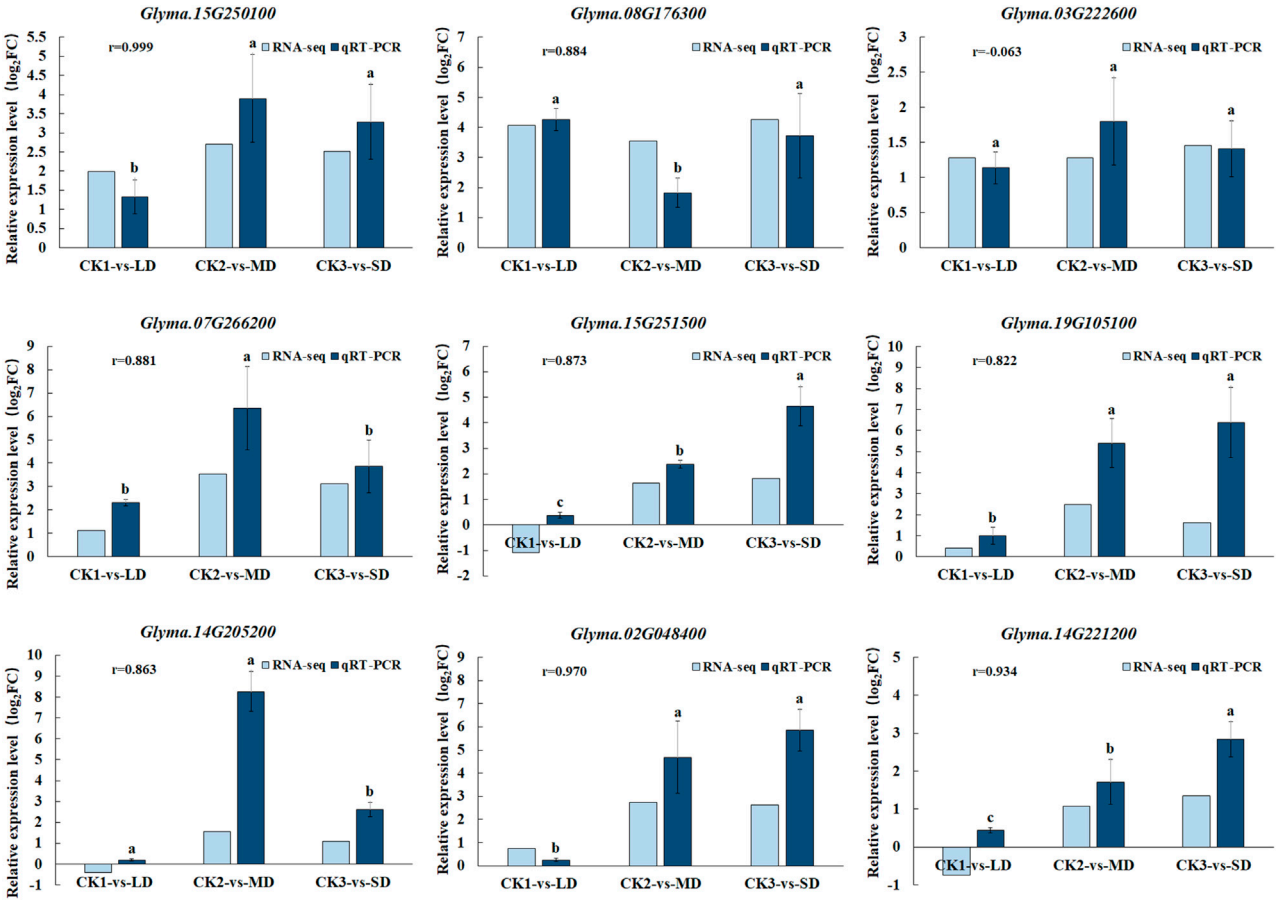


FIGURE 8
Comparison of gene expression patterns between RNA-seq and qRT-PCR under drought treatments. Data represent the means ± SD of three biological replicates, and each biological replicate contained three technical replicates ($p < 0.05$).

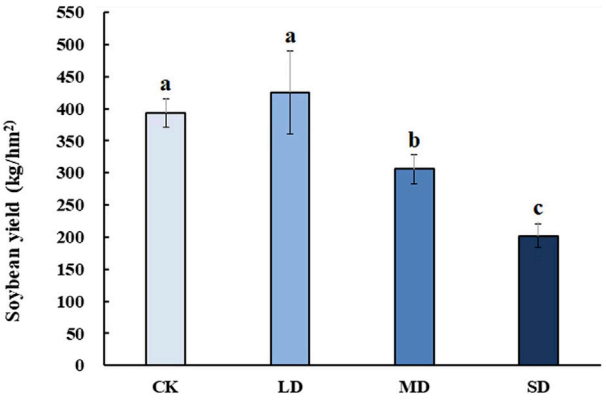


FIGURE 9
The effect of varying degrees of drought stress on soybean yield after rehydration. Three biological replicates were performed and the data represent the means ± SDs. Different letters represent statistically significant differences ($p < 0.05$).

4.2 The roles of ABA biosynthesis and signaling and TFs in drought stress responses in soybean

ABA plays a pivotal role in drought stress tolerance that results in closing the stomata as well as regulating the contents of proteins, soluble sugars and amino acids (Finkelstein, 2013; Maruyama et al., 2014; Gonz´alez-Villagra et al., 2019; Mathan et al., 2021). Using GO and KEGG enrichment analyses, 28 DEGs and 36 DEGs related to ABA biosynthesis and signaling pathways in soybean under different drought treatments were obtained (Figure 5). We found that the highest expression of the NCED gene occurred in soybean under three drought stress levels (Figure 5A), which demonstrated that the plant's 'commitment' increased ABA content (Qin and Zeevaart., 2002; Martínez-Andújar et al., 2011). Interestingly, it seems that the expression of *AOG* and *CYP707A* genes was significantly lower than that of the control after MD and SD treatments (Figure 5A).

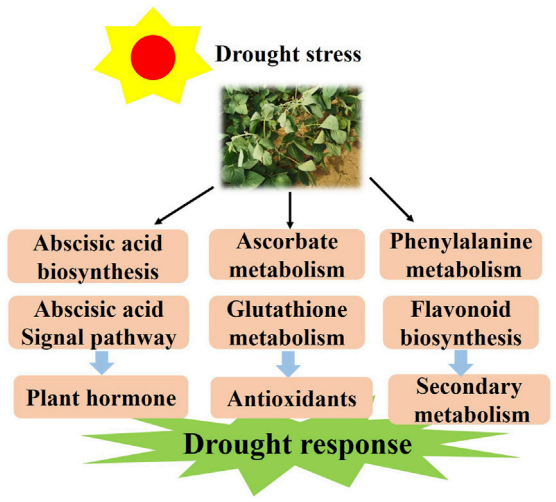


FIGURE 10
The model for the regulation of the drought response in soybean responds to drought conditions by regulating the plant hormone, antioxidant and flavonoid pathways.

This suggested that the ABA response to drought stress occurs through increased biosynthesis and catabolism (Ren et al., 2006). This phenomenon was also reflected at the transcription levels of ABA signaling pathways under drought stress. Drought treatments upregulated genes participating in the ABA signaling pathways, especially three *PP2C* genes (Figure 5B). Notably, we observed that the expression of three *ABF* genes in soybean treated with drought stress was higher than that in the controls. These results are compatible with the critical signaling role of ABA in plant defense.

Transcription regulation is a typical mechanism for plants encountering drought. Based on WGCNA, *MYB, WRKY, NAC*, and *bHLH* were upregulated under MD and SD stress (Figure 3). They function as activators or repressors to regulate target genes and then form a transcriptional regulatory network involved in abiotic stress response and tolerance. In addition, some plant hormone response factors were also enriched in the three modules, such as *ARF*, *EIN3* and *BES1*, which indicated that multiple hormones co-regulate plant responses to drought stress.

4.3 AsA and GSH metabolism involved in drought stress responses in soybean

When the plant is subjected to drought stress, the excess ROS will be scavenged by the antioxidant system, which involves the SOD, POD and CAT enzymes, as well as the nonenzymatic constituents (AsA, GSH, and flavonoids) that maintain the steady state of cell membranes (You et al., 2019). KEGG analysis detected the enrichment of AsA and GSH metabolism genes involved in the drought response (Figures 2C–F; Figure 6). Vitamin C, also known as L-ascorbic acid (AsA), protects cells from oxidative stress by maintaining the ROS balance (Gallie, 2013; Akram et al., 2017). The Smirnoff-Wheeler pathway (*VTC1, GME, VTC2, VTC4* and *VTC5*) was identified to participate in vitamin C biosynthesis (Wheeler et al., 1998; Conklin et al., 1999, 2000, 2006; Dowdle et al., 2007). The relevant pathway genes of AsA metabolism were all upregulated in both the MD and SD treatments, especially the VTC4 genes under SD stress. A recent study indicated that the molecular mechanism underlying AsA biosynthesis and ABA signaling pathways participates in plant drought tolerance (Zhang et al., 2020). The application of AsA can be used against drought stress in various plant species, such as wheat (Malik and Ashraf., 2012; Hussein et al., 2014), and maize (Darvishan et al., 2013). In short, these results provide evidence that AsA biosynthesis was significantly induced by drought stress in soybean, especially by severe drought stress.

An additional mechanism by which plants increase drought resistance is the accumulation of glutathione, a major antioxidant that conjugates with electrophilic compounds and facilitates peroxide reduction (Alscher, 1989; Anderson and Davis, 2004). Specifically, this involves upregulating G6PDH and GST enzymes involved in scavenging ROS and reducing secondary noxious products under drought stress (Wagner et al., 2002; Anderson and Davis, 2004; Landi et al., 2016). In the present study, two *G6PDH* and six *GST* genes were upregulated by approximately 2-fold under SD treatment, indicating that soybean can activate the *GSH* metabolism pathway to scavenge ROS under SD stress. Previous studies also reported that increasing *G6PDH* expression in the tomato and the overexpression of *GST* genes in transgenic *Arabidopsis* can enhance drought tolerance (Liu et al., 2013; Xu et al., 2015).

4.4 Flavonoid biosynthesis involved in drought stress responses in the soybean

Flavonoids are widely distributed in plants and exhibit antioxidant activities, which could increase the ROS scavenging ability to protect normal plant growth from drought stress (Falcone Ferreyra et al., 2012; Nakabayashi et al., 2014). Flavonoids are synthesized by the phenylalanine pathway in plants (Dixon et al., 2002). The C4H pathway is involved in the drought defense of cucumber (Bellés et al., 2008). In our study, severe drought stress significantly upregulated the transcript abundance of *C4H* genes and activated the transcription of two *4CL* genes (Figure 7A). We further detected that the transcription of four CAD genes (*Glyma.14G221200, Glyma.01G021000, Glyma.15G088100, Glyma.05G187700*) was markedly increased, which may be involved in lignin synthesis to confer drought tolerance to soybean (Xu et al., 2020).

Earlier research analyzed the key genes related to flavonoid biosynthesis by transcript profiling under drought stress (Kang

et al., 2011). Similarly, the DEGs of flavonoid biosynthesis pathway were activated by drought stress in *Populus euphratica* (Jiao et al., 2021). The KEGG analysis in our study established that the DEGs were enriched in flavonoid metabolism biosynthesis only under MD and SD stress (Figure 2). The key genes *CHS, F3H, DFR* and *ANS* showed prominent increases under SD stress, which may have elevated the contents of flavonoids and mobilized them in response to drought resistance (Figure 7). These results indicate that drought stress regulated the synthesis of flavonoids, which played a key role in enhancing drought tolerance in soybean.

4.5 A proposed model and the effect of drought stress on soybean yield

Plant growth compensation by rehydration is one of the strategies against short-term drought to alleviate its effect on plant yield (Acevedo et al., 1971), which was also performed in this study. Short periods of drought had no effect on the soybean yield, but prolonged periods of drought reduced the yields by 50%.

It is known that half of the drought response genes are activated by ABA (Seki et al., 2002). However, the molecular mechanism induced by drought stress in soybean remains elusive and needs to be further studied. Herein, we discuss the probable molecular mechanism of soybean response to drought stress. According to the results, we propose a model for the soybean response to varying levels of drought stress. First, drought stress activates the ABA biosynthesis and signaling and drought-related TFs in plants. Meanwhile, it promotes AsA and GSH metabolism by regulating the expression of the *VTC, GME* and *G6PDH* genes. With longer time exposure to drought stress, flavonoids may be a positive regulator and function to mediate drought resistance (Figure 10). The above results not only deepen the insights into the drought response in soybean but also provide a theoretical basis for the genetic improvement and water-efficient irrigation of this crop.

5 Conclusion

In this study, the response mechanism of soybean exposed to three levels of drought stress at the flowering stage was investigated. Physiological, transcriptomic and WGCNA analyses were carried out and the TFs and potential pathways of genes in soybean under drought stress were identified. The results suggested that the soybean plant is able to activate the genes of antioxidants, secondary metabolism and hormone signaling pathway, *Glyma.08G176300(NCED1), Glyma.03G222600(SDR), Glyma.02G048400(F3H), Glyma.14G221200(CAD), Glyma.14G205200(C4H) and Glyma.19G105100 (CHS), Glyma.07G266200 (VTC5-2) and Glyma.15G251500 (GST)*, response to drought stress by promoting MDA accumulation and the activities of SOD and CAT to partly cope with drought stress. Furthermore, the soybean yield after rehydration in the harvesting period was reduced by 50% under severe drought stress. Finally, we further deepen the understanding of molecular mechanism of soybean in response to drought stress, which provides a theoretical basis for the molecular breeding of drought resistance.

Data availability statement

The data presented in the study are deposited in the NCBI SRA repository, accession number PRJNA852689.

Author contributions

The experiments were conceived and designed by CR, JL, and XG. The experiments were performed by ML, HL, AS, and LW. ML analysed the data. ML and XG contributed to writing the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Key Research and Development Program of China (2019YFD1002204).

Acknowledgments

We thank all members of our laboratory for providing technical assistance.

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/fgene.2022.1060529/full#supplementary-material

References

Acevedo, E., Hsiao, T. C., and Henderson, D. W. (1971). Immediate and subsequent growth responses of maize leaves to changes in water status. *Plant Physiol.* 48 (5), 631–636. doi:10.1104/pp.48.5.631

Akram, N. A., Shafiq, F., and Ashraf, M. (2017). Ascorbic acid-A potential oxidant scavenger and its role in plant development and abiotic stress tolerance. *Front. Plant Sci.* 8, 613. doi:10.3389/fpls.2017.00613

Alscher, R. G. (1989). Biosynthesis and antioxidant function of glutathione in plants. *Physiol. Plant.* 77 (3), 457–464. doi:10.1111/j.1399-3054.1989.tb05667.x

Anderson, J. V., and Davis, D. G. (2004). Abiotic stress alters transcript profiles and activity of glutathione S-transferase, glutathione peroxidase, and glutathione reductase in Euphorbia esula. *Physiol. Plant.* 120 (3), 421–433. doi:10.1111/j.0031-9317.2004.00249.x

Anjum, S. A., Ashraf, U., Tanveer, M., Khan, I., Hussain, S., Shahzad, B., et al. (2017). Drought induced changes in growth, osmolyte accumulation and antioxidant metabolism of three maize hybrids. *Front. Plant Sci.* 8, 69. doi:10.3389/fpls.2017.00069

Ashraf, M., and Foolad, M. R. (2007). Roles of glycine betaine and proline in improving plant abiotic stress resistance. *Environ. Exp. Bot.* 59, 206–216. doi:10.1016/j.envexpbot.2005.12.006

Baldoni, E., Bagnaresi, P., Locatelli, F., Mattana, M., and Genga, A. (2016). Comparative leaf and root transcriptomic analysis of two rice japonica cultivars reveals major differences in the root early response to osmotic stress. *Rice* 9, 25. doi:10.1186/s12284-016-0098-1

Barnab´as, B., J¨ager, K., and Feh´er, A. (2008). The effect of drought and heat stress on reproductive processes in cereals. *Plant Cell Environ.* 31, 11–38. doi:10.1111/j.1365-3040.2007.01727.x

Basim, H., Basim, E., Tombuloglu, H., and Unver, T. (2021). Comparative transcriptome analysis of resistant and cultivated tomato lines in response to *Clavibacter michiganensis* subsp. *Genomics* 113 (4), 2455–2467. doi:10.1016/j.ygeno.2021.05.033

Bellés, J. M., López-Gresa, M. P., Fayos, J., Pallás, V., Rodrigo, I., and Conejero, V. (2008). Induction of cinnamate 4-hydroxylase and phenylpropanoids in virus-infected cucumber and melon plants. *Plant Sci.* 174 (5), 524–533. doi:10.1016/j.plantsci.2008.02.008

Buezo, J., Sanz-Saez, Á., Moran, J. F., Soba, D., Aranjuelo, I., and Esteban, R. (2019). Drought tolerance response of high-yielding soybean varieties to mild drought: Physiological and photochemical adjustments. *Physiol. Plant.* 166 (1), 88–104. doi:10.1111/ppl.12864

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

Chimungu, J. G., Brown, K. M., and Lynch, J. P. (2014). Reduced root cortical cell file number improves drought tolerance in maize. *Plant Physiol.* 166, 1943–1955. doi:10.1104/pp.114.249037

Choudhary, S., and Tran, L. (2011). Phytosterols: Perspectives in human nutrition and clinical therapy. *Curr. Med. Chem.* 18 (29), 4557–4567. doi:10.2174/092986711797287593

Conklin, P. L., Gatzek, S., Wheeler, G. L., Dowdle, J., Raymond, M. J., Rolinski, S., et al. (2006). *Arabidopsis thaliana* VTC4 encodes L-galactose-1-P phosphatase, a plant ascorbic acid biosynthetic enzyme. *J. Biol. Chem.* 281, 15662–15670. doi:10.1074/jbc.M601409200

Conklin, P. L., Norris, S. R., Wheeler, G. L., Williams, E. H., Smirnoff, N., and Last, R. L. (1999). Genetic evidence for the role of GDP-mannose in plant ascorbic acid (vitamin C) biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 96, 4198–4203. doi:10.1073/pnas.96.7.4198

Conklin, P. L., Saracco, S. A., Norris, S. R., and Last, R. L. (2000). Identification of ascorbic acid-deficient *Arabidopsis thaliana* mutants. *Genetics* 154, 847–856. doi:10.1093/genetics/154.2.847

Darvishan, M., Moghadam, H. R. T., and Nasri, M. (2013). Effect of foliar application of ascorbic acid (vitamin C) on yield and yield components of corn (Zea mays L.) as influenced by withholding of irrigation at different growth stages. *Res. Crops.* 14, 736–742.

Dien, D. C., Mochizuki, T., and Yamakawa, T. (2019). Effect of various drought stresses and subsequent recovery on proline, total soluble sugar and starch metabolisms in rice (Oryza sativa L.) varieties. *Plant Prod. Sci.* 22 (4), 530–545. doi:10.1080/1343943x.2019.1647787

Dixon, R. A., Achnine, L., Kota, P., Liu, C. J., Reddy, M. S., and Wang, L. (2002). The phenylpropanoid pathway and plant defence-a genomics perspective. *Mol. Plant Pathol.* 3, 371–390. doi:10.1046/j.1364-3703.2002.00131.x

Dowdle, J., Ishikawa, T., Gatzek, S., Rolinski, S., and Smirnoff, N. (2007). Two genes in *Arabidopsis thaliana* encoding GDP-L-galactose phosphorylase are required for ascorbate biosynthesis and seedling viability. *Plant J.* 52, 673–689. doi:10.1111/j.1365-313X.2007.03266.x

Elansary, H. O., and Salem, M. Z. M. (2015). Morphological and physiological responses and drought resistance enhancement of ornamental shrubs by trinexapac-ethyl application. *Sci. Hortic.* 189, 1–11. doi:10.1016/j.scienta.2015.03.033

Fahad, S., Bajwa, A. A., Nazir, U., Anjum, S. A., Farooq, A., Zohaib, A., et al. (2017). Crop production under drought and heat stress: Plant responses and management options. *Front. Plant Sci.* 8, 1147. doi:10.3389/fpls.2017.01147

Falcone Ferreyra, M. L., Rius, S. P., and Casati, P. (2012). Flavonoids: Biosynthesis, biological functions, and biotechnological applications. *Front. Plant Sci.* 3, 222. doi:10.3389/fpls.2012.00222

Finkelstein, R. (2013). Abscisic acid synthesis and response. *Arab. Book* 11, e0166. doi:10.1199/tab.0166

Gallie, D. R. (2013). The role of L-ascorbic acid recycling in responding to environmental stress and in promoting plant growth. *J. Exp. Bot.* 64, 433–443. doi:10.1093/jxb/ers330

Gill, S. S., and Tuteja, N. (2010). Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiol. biochem.* 48, 909–930. doi:10.1016/j.plaphy.2010.08.016

Golldack, D., Li, C., Mohan, H., and Probst, N. (2014). Tolerance to drought and salt stress in plants: Unraveling the signaling networks. *Front. Plant Sci.* 5, 151. doi:10.3389/fpls.2014.00151

Gonz´alez-Villagra, J., Cohen, J. D., and Reyes-Díaz, M. M. (2019). Abscisic acid is involved in phenolic compounds biosynthesis, mainly anthocyanins, in leaves of Aristotelia chilensis plants (Mol.) subjected to drought stress. *Physiol. Plant.* 165 (4), 855–866. doi:10.1111/ppl.12789

Hao, S., Guo, X., and Zhang, Z. (2010). Compensation effects of water stress and rewatering on the structure of rice canopy. *Trans. Chin. Soc. Agric. Mach.* 41 (3), 52–61.

Harauma, A., Murayama, T., Ikeyama, K., Sano, H., Arai, H., Takano, R., et al. (2007). Mulberry leaf powder prevents atherosclerosis in apolipoprotein E-deficient mice. *Biochem. Biophys. Res. Commun.* 358, 751–756. doi:10.1016/j.bbrc.2007.04.170

He, F. J., and Chen, J. Q. (2013). Consumption of soybean, soy foods, soy isoflavones and breast cancer incidence: Differences between Chinese women and women in western countries and possible mechanisms. *Food Sci. Hum. Wellness* 2 (3–4), 146–161. doi:10.1016/j.fshw.2013.08.002

Huang, Y., Jiao, Y., Xie, N., Guo, Y., Zhang, F., Xiang, Z., et al. (2019). OsNCED5, a 9-cis-epoxycarotenoid dioxygenase gene, regulates salt and water stress tolerance and leaf senescence in rice. *Plant Sci.* 287, 110188. doi:10.1016/j.plantsci.2019.110188

Hussein, N. M., Hussein, M. I., Gadel Hak, S. H., and Hammad, M. A. (2014). Effect of two plant extracts and four aromatic oils on tuta absoluta population and productivity of tomato cultivar gold stone. *Nat. Sci.* 12, 108–118.

Jiao, P., Wu, Z., Wang, X., Jiang, Z., Wang, Y., Liu, H., et al. (2021). Short-term transcriptomic responses of Populus euphratica roots and leaves to drought stress. *J. For. Res.* 32, 841–853. doi:10.1007/s11676-020-01123-9

Kang, C., Zhai, H., He, S., Zhao, N., and Liu, Q. (2019). A novel sweetpotato bZIP transcription factor gene, IbbZIP1, is involved in salt and drought tolerance in transgenic Arabidopsis. *Plant Cell Rep.* 38, 1373–1382. doi:10.1007/s00299-019-02441-x

Kang, Y., Han, Y., Torres-Jerez, I., Wang, M., Tang, Y., Monteros, M., et al. (2011). System responses to long-term drought and re-watering of two contrasting alfalfa varieties. *Plant J.* 68, 871–889. doi:10.1111/j.1365-313X.2011.04738.x

Kerr, T. C. C., Abdel-Mageed, H., Aleman, L., Lee, J., Payton, P., Cryer, D., et al. (2018). Ectopic expression oftwo AREB/ABF orthologs increases drought tolerance in cotton (Gossypium hirsutum). *Plant Cell Environ.* 41, 898–907. doi:10.1111/pce.12906

Khan, M. I. R., Fatma, M., Per, T. S., Anjum, N. A., and Khan, N. A. (2015). Salicylic acid-induced abiotic stress tolerance and underlying mechanisms in plants. *Front. Plant Sci.* 6, 462. doi:10.3389/fpls.2015.00462

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

Landi, S., Nurcato, R., De Lillo, A., Lentini, M., Grillo, S., and Esposito, S. (2016). Glucose-6phosphate dehydrogenase plays a central role in the response of tomato

(Solanum lycopersicum) plants to short and long-term drought. *Plant Physiol. biochem.* 105, 79–89. doi:10.1016/j.plaphy.2016.04.013

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, 25. doi:10.1186/gb-2009-10-3-r25

Lawlor, D. W. (2013). Genetic engineering to improve plant performance under drought: Physiological evaluation of achievements, limitations, and possibilities. *J. Exp. Bot.* 64 (1), 83–108. doi:10.1093/jxb/ers326

Lesk, C., Rowhani, P., and Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. *Nature* 529, 84–87. doi:10.1038/nature16467

Li, W., Fu, L. F., Geng, Z. W., Zhao, X. J., Liu, Q. H., and Jiang, X. Q. (2020). Physiological characteristic changes and full-length transcriptome of rose (rosa chinensis) roots and leaves in response to drought stress. *Plant Cell Physiol.* 61 (12), 2153–2166. doi:10.1093/pcp/pcaa137

Lindemose, S., O'Shea, C., Jensen, M. K., and Skriver, K. (2013). Structure, function and networks of transcription factors involved in abiotic stress responses. *Int. J. Mol. Sci.* 14, 5842–5878. doi:10.3390/ijms14035842

Liu, H., Zheng, W., Zheng, F., and Wang, L. (2012). Influence of rewatering on compensatory effect of maize seedling roots with diluted seawater irrigation. *Trans. Chin. Soc. Agric. Eng.* 28 (3), 101–106.

Liu, J., Wang, X., Hu, Y., Hu, W., and Bi, Y. (2013). Glucose-6-phosphate dehydrogenase plays a pivotal role in tolerance to drought stress in soybean roots. *Plant Cell Rep.* 32 (3), 415–429. doi:10.1007/s00299-012-1374-1

Liu, X. Q., Chen, A. P., Wang, Y. X., Jin, G. L., Zhang, Y. H., Gu, L. L., et al. (2022). Physiological and transcriptomic insights into adaptive responses of Seriphidium transiliense seedlings to drought stress. *Environ. Exp. Bot.* 194, 104736. doi:10.1016/j.envexpbot.2021.104736

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

Lotfi, Y. N., Vahdati, K., Hassani, D., Kholdebarin, B., and Amiri, R. (2010). Peroxidase, guaiacol peroxidase and ascorbate peroxidase activity accumulation in leaves and roots of walnut trees in response to drought stress. *Acta Hortic.* 861, 309–316. doi:10.17660/actahortic.2010.861.42

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 (12), 550. doi:10.1186/s13059-014-0550-8

Lu, X., Li, Q. T., Xiong, Q., Li, W., Bi, Y. D., Lai, Y. C., et al. (2016). The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication. *Plant J.* 86, 530–544. doi:10.1111/tpj.13181

Malik, S., and Ashraf, M. (2012). Exogenous application of ascorbic acid stimulates growth and photosynthesis of wheat (*Triticum aestivum* L.) under drought. *Soil Environ.* 31, 72–77.

Martignago, D., Medina, A. R., Escaez, D. B., Manzaneque, J. B. F., and Delgado, A. I. C. (2020). Drought resistance by engineering plant tissue-specific responses. *Front. Plant Sci.* 10, 1676. doi:10.3389/fpls.2019.01676

Martínez-Andújar, C., Ordiz, M. I., Huang, Z., Nonogaki, M., Beachy, R. N., and Nonogaki, H. (2011). Induction of 9-cis-epoxycarotenoid dioxygenase in *Arabidopsis thaliana* seeds enhances seed dormancy. *Proc. Natl. Acad. Sci. U. S. A.* 108, 17225–17229. doi:10.1073/pnas.1112151108

Maruyama, K., Urano, K., Yoshiwara, K., Morishita, Y., Sakurai, N., Suzuki, H., et al. (2014). Integrated analysis of the effects of cold and dehydration on rice metabolites, phytohormones, and gene transcripts. *Plant Physiol.* 164 (4), 1759–1771. doi:10.1104/pp.113.231720

Mathan, J., Singh, A., and Ranjan, A. (2021). Sucrose transport in response to drought and salt stress involves ABA-mediated induction of OsSWEET13 and OsSWEET15 in rice. *Physiol. Plant.* 171 (4), 620–637. doi:10.1111/ppl.13210

Meckel, L., Egli, D. B., Phillips, R. E., Radcliffe, D., and Leggett, J. E. (1984). Moisture stress and N redistribution in soybean [1]. *Agron. J.* 75, 1027–1031. doi:10.2134/agronj1983.00021962007500060036x

Miller, G., Suzuki, N., Ciftci-Yilmaz, S., and Mittler, R. (2010). Reactive oxygen species homeostasis and signalling during drought and salinity stresses. *Plant Cell Environ.* 33, 453–467. doi:10.1111/j.1365-3040.2009.02041.x

Mittler, R., Vanderauwera, S., Gollery, M., and Breusegem, F. V. (2004). Reactive oxygen gene network of plants. *Trends Plant Sci.* 9, 490–498. doi:10.1016/j.tplants.2004.08.009

Molnár, I., Cozma, L., D´enes, T.´E., Vass, I., Vass, I. Z., and Rakosy-Tican, E. (2021). Drought and saline stress tolerance induced in somatic hybrids of Solanum chacoense and potato cultivars by using mismatch repair deficiency. *Agriculture* 11 (8), 696. doi:10.3390/agriculture11080696

Nakabayashi, R., Yonekura-Sakakibara, K., Urano, K., Suzuki, M., Yamada, Y., Nishizawa, T., et al. (2014). Enhancement of oxidative and drought tolerance in Arabidopsis by overaccumulation of antioxidant flavonoids. *Plant J.* 77, 367–379. doi:10.1111/tpj.12388

Qin, F., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2011). Achievements and challenges in understanding plant abiotic stress responses and tolerance. *Plant Cell Physiol.* 52, 1569–1582. doi:10.1093/pcp/pcr106

Qin, X., and Zeevaart, J. A. D. (2002). Overexpression of a 9-cis-epoxycarotenoid dioxygenase gene in Nicotiana plumbaginifolia increases abscisic acid and phaseic acid levels and enhances drought tolerance. *Plant Physiol.* 128, 544–551. doi:10.1104/pp.010663

Ren, H., Gao, Z., Chen, L., Wei, K., Liu, J., Fan, Y., et al. (2006). Dynamic analysis ofABA accumulation in relation to the rate of ABA catabolism in maize tissues under water deficit. *J. Exp. Bot.* 58, 211–219. doi:10.1093/jxb/erl117

Sakai, T., and Kogiso, M. (2008). Soy isoflavones and immunity. *J. Med. Invest.* 55, 167–173. doi:10.2152/jmi.55.167

Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., et al. (2002). Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J.* 31, 279–292. doi:10.1046/j.1365-313x.2002.01359.x

Seki, M., Umezawa, T., Urano, K., and Shinozaki, K. (2007). Regulatory metabolic networks in drought stress responses. *Curr. Opin. Plant Biol.* 10, 296–302. doi:10.1016/j.pbi.2007.04.014

Song, W., Wang, H. J., Bucheli, P., Zhang, P. F., Wei, D. Z., and Lu, Y. H. (2009). Phytochemical profiles of different mulberry (Morus sp.) species from China. *J. Agric. Food Chem.* 57, 9133–9140. doi:10.1021/jf9022228

Specht, J. E., Hume, D. J., and Kumudini, S. V. (1999). Soybean yield potential - a genetic and physiological perspective. *Crop Sci.* 39, 1560–1570. doi:10.2135/cropsci1999.3961560x

Stacey, G., Vodkin, L., Parrott, W. A., and Shoemaker, R. C. (2004). National Science Foundation-sponsored workshop report. Draft plan for soybean genomics. *Plant Physiol.* 135 (1), 59–70. doi:10.1104/pp.103.037903

Sun, J. H., Qiu, C., Ding, Y. Q., Wang, Y., Sun, L. T., Fan, K., et al. (2020). Fulvic acid ameliorates drought stress-induced damage in tea plants by regulating the ascorbate metabolism and flavonoids biosynthesis. *BMC Genomics* 21 (1), 411. doi:10.1186/s12864-020-06815-4

Sun, T. T., Su, Z. H., Wang, R., Liu, R., Yang, T., Zuo, W. T., et al. (2022). Transcriptome and metabolome analysis reveals the molecular mechanisms of Tamarix taklamakanensis under progressive drought and rehydration treatments. *Environ. Exp. Bot.* 195, 104766. doi:10.1016/j.envexpbot.2021.104766

Thirumalaikumar, V. P., Devkar, V., Mehterov, N., Ali, S., Ozgur, R., Turkan, I., et al. (2018). NAC transcription factor JUNGBRUNNEN 1 enhances drought tolerance in tomato. *Plant Biotechnol. J.* 16, 354–366. doi:10.1111/pbi.12776

Torres, M. A., and Dangl, J. L. (2005). Functions of the respiratory burst oxidase in biotic interactions, abiotic stress and development. *Curr. Opin. Plant Biol.* 8, 397–403. doi:10.1016/j.pbi.2005.05.014

Vishwakarma, K., Upadhyay, N., Kumar, N., Yadav, G., Singh, J., Mishra, R. K., et al. (2017). Abscisic acid signaling and abiotic stress tolerance in plants: A review on current knowledge and future prospects. *Front. Plant Sci.* 8, 161. doi:10.3389/fpls.2017.00161

Wagner, U., Edwards, R., Dixon, D. P., and Mauch, F. (2002). Probing the diversity of the Arabidopsis glutathione S-Transferase gene family. *Plant Mol. Biol.* 49, 515–532. doi:10.1023/a:1015557300450

Wang, H., Wang, H., Shao, H., and Tang, X. (2016). Recent advances in utilizing transcription factors to improve plant abiotic stress tolerance by transgenic technology. *Front. Plant Sci.* 7, 67. doi:10.3389/fpls.2016.00067

Wheeler, G. L., Jones, M. A., and Smirnoff, N. (1998). The biosynthetic pathway of vitamin C in higher plants. *Nature* 393, 365–369. doi:10.1038/30728

Xie, Z., Nolan, T., Jiang, H., Tang, B., Zhang, M., Li, Z., et al. (2019). The AP2/ERF transcription factor TINY modulates brassinosteroid- regulated plant growth and drought responses in Arabidopsis. *Plant Cell* 31, 1788–1806. doi:10.1105/tpc.18.00918

Xu, J., Xiao, J. X., Yong, S. T., Ri, H. P., Yong, X., Wei, Z., et al. (2015). Transgenic Arabidopsis plants expressing tomato glutathione S-transferase showed enhanced resistance to salt and drought stress. *PLoS One* 10 (9), e0136960. doi:10.1371/journal.pone.0136960

Xu, W. Y., Tang, W. S., Wang, C. X., Ge, L. H., Sun, J. C., Qi, X., et al. (2020). SiMYB56 confers drought stress tolerance in transgenic rice by regulating lignin biosynthesis and ABA signaling pathway. *Front. Plant Sci.* 11, 785. doi:10.3389/fpls.2020.00785

Xu, Z., Zhou, G., and Shimizu, H. (2010). Plant responses to drought and rewatering. *Plant Signal. Behav.* 5, 649–654. doi:10.4161/psb.5.6.11398

Xuan, H. D., Huang, Y. Z., Zhou, L., Deng, S. S., Wang, C. C., Xu, J. Y., et al. (2022). Key soybean seedlings drought-responsive genes and pathways revealed by comparative transcriptome analyses of two cultivars. *Int. J. Mol. Sci.* 23, 2893. doi:10.3390/ijms23052893

Xue, H. Y., Zhang, Y. J., Liu, L. T., Sun, H. C., and Li, C. D. (2013). Responses of spectral reflectance, photosynthesis and chlorophyll fluorescence in cotton during drought stress and rewatering. *Sci. Agric. Sin.* 46 (1), 2386–2393.

You, J., Zhang, Y., Liu, A., Li, D., Wang, X., Dossa, K., et al. (2019). Transcriptomic and metabolomic profiling of drought tolerant and susceptible sesame genotypes in response to drought stress. *BMC Plant Biol.* 19, 267. doi:10.1186/s12870-019-1880-1

Yu, Y. H., Wang, P., Bai, Y. C., Wang, Y., Wan, H. N., Liu, C., et al. (2020). The soybean F-box protein GmFBX176 regulates ABA-mediated responses to drought and salt stress. *Environ. Exp. Bot.* 176, 104056. doi:10.1016/j.envexpbot.2020.104056

Zhang, H., and Song, B. (2017). RNA-seq data comparisons of wild soybean genotypes in response to soybean cyst nematode (*Heterodera glycines*). *Genom. Data* 14, 36–39. doi:10.1016/j.gdata.2017.08.001

Zhang, H., Xiang, Y. L., He, N., Liu, X. G., Liu, H. B., Fang, L. P., et al. (2020). Enhanced vitamin C production mediated by an ABA-induced PTP-like nucleotidase improves plant drought tolerance in Arabidopsis and maize. *Mol. Plant* 13, 760–776. doi:10.1016/j.molp.2020.02.005

Zhao, X., Bai, S., Li, L., Han, X., Li, J., Zhu, Y., et al. (2020). Comparative transcriptome analysis of two *Aegilops tauschii* with contrasting drought tolerance by RNA-Seq. *Int. J. Mol. Sci.* 21, 3595. doi:10.3390/ijms21103595

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

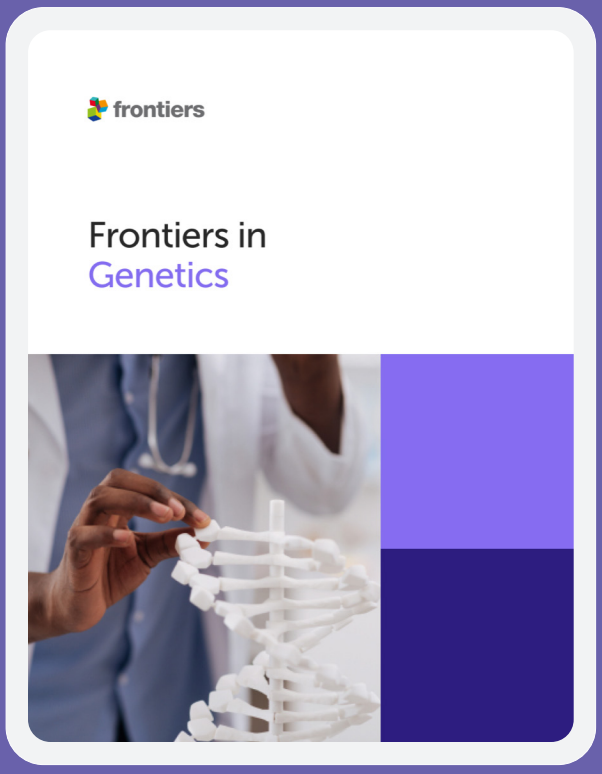
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

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