



OPEN ACCESS

EDITED BY

Xuke Lu,
Institute of Cotton Research (CAAS), China

REVIEWED BY

Weixiong Long,
Jiangxi Academy of Agricultural Sciences
(CAAS), China
Rakibul Hasan,
Sylhet Agricultural University, Bangladesh

*CORRESPONDENCE

Shangqian Xie
✉ sqianxie@foxmail.com
Licao Cui
✉ cuilicao@jxau.edu.cn

[†]These authors have contributed
equally to this work

SPECIALTY SECTION

This article was submitted to
Plant Abiotic Stress,
a section of the journal
Frontiers in Plant Science

RECEIVED 11 January 2023

ACCEPTED 13 February 2023

PUBLISHED 02 March 2023

CITATION

Tang M, Liu L, Hu X, Zheng H, Wang Z,
Liu Y, Zhu Q, Cui L and Xie S (2023)
Genome-wide characterization of R2R3-
MYB gene family in *Santalum album* and
their expression analysis under cold stress.
Front. Plant Sci. 14:1142562.
doi: 10.3389/fpls.2023.1142562

COPYRIGHT

© 2023 Tang, Liu, Hu, Zheng, Wang, Liu,
Zhu, Cui and Xie. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License](#)
(CC BY). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Genome-wide characterization of R2R3-MYB gene family in *Santalum album* and their expression analysis under cold stress

Minqiang Tang^{1†}, Le Liu^{1†}, Xu Hu^{1†}, Haoyue Zheng¹,
Zukai Wang¹, Yi Liu¹, Qing Zhu¹, Licao Cui^{2*}
and Shangqian Xie^{1*}

¹Key Laboratory of Genetics and Germplasm Innovation of Tropical Special Forest Trees and
Ornamental Plants (Ministry of Education), School of Forestry, Hainan University, Haikou, China,

²College of Bioscience and Engineering, Jiangxi Agricultural University, Nanchang, China

Sandalwood (*Santalum album*) is a high-value multifunctional tree species that is rich in aromatic substances and is used in medicine and global cosmetics. Due to the scarcity of land resources in tropical and subtropical regions, land in temperate regions is a potential resource for the development of *S. album* plantations in order to meet the needs of *S. album* production and medicine. The R2R3-MYB transcription factor family is one of the largest in plants and plays an important role in the response to various abiotic stresses. However, the R2R3-MYB gene family of *S. album* has not been studied. In this study, 144 R2R3-MYB genes were successfully identified in the assembly genome sequence, and their characteristics and expression patterns were investigated under various durations of low temperature stress. According to the findings, 31 of the 114 R2R3-MYB genes showed significant differences in expression after cold treatment. Combining transcriptome and weighted gene co-expression network analysis (WGCNA) revealed three key candidate genes (*SaMYB098*, *SaMYB015*, and *SaMYB068*) to be significantly involved in the regulation of cold resistance in *S. album*. The structural characteristics, evolution, and expression pattern of the R2R3-MYB gene in *S. album* were systematically examined at the whole genome level for the first time in this study. It will provide important information for future research into the function of the R2R3-MYB genes and the mechanism of cold stress response in *S. album*.

KEYWORDS

R2R3-MYB gene family, cold stress, sandalwood, WGCNA, transcriptome analysis

1 Introduction

Santalum album L., an important plant belonging to the Class Magnoliopsida, is well known for its medicinal value and valuable carving wood. Unfortunately, owing to its high commercial value and applications in cosmetics, religion, and medicine, the population of *S. album* is gradually declining from overharvesting and illegal trading. This worrisome decline in its population underscores the need for effective *in situ* and *ex situ* conservation strategies. However, owing to the influence of temperature and climate, the natural distribution of sandalwood ranges from 30° north latitude to 40° south latitude, starting from Indonesia in the east, reaching Juan Fernandez Islands in the west, Hawaii Islands in the north, and New Zealand in the south (Kumar et al., 2012). Low temperature is one of the main abiotic factors that hinder the growth and geographical distribution of *S. album*. However, there are very few studies on the mechanisms of environmental adaptation of *S. album*, especially on its mechanisms of adaptation to low temperatures. Therefore, it is of great significance to explore the molecular genetic mechanisms of cold stress for the protection and utilization of *S. album*.

The effects of low-temperature stress on plants are mainly reflected in enzyme activity, membrane system, and cell dehydration that lead to disorders in cell metabolism and even cell death. The response of plants to cold stress at the molecular level is primarily mediated through changes in gene expression, protein levels, and metabolites (Li et al., 2018). The molecular mechanism of cold tolerance has been studied in the model plant, *Arabidopsis thaliana*, and other crop plants, such as maize, rice, wheat, tomato, and barley (Sanghera et al., 2011; Jeon and Kim, 2013). The MYB transcription factors (TFs), which are related to c-Myb, are involved in plant growth and development, metabolism, responses to biotic and abiotic stress, and other biological processes through interaction with the basic helix-loop-helix (bHLH) TFs (Stracke et al., 2001). The MYB TF MdMYB308L in apple was found to positively regulate cold tolerance and anthocyanin accumulation by interacting with MdbHLH33 and enhancing its binding to the promoters of MdCBF2 and MdDFR (An et al., 2020). In *A. thaliana*, the overexpression of the *Mallus baccata* MYB4 (MbMYB4) gene enhanced the tolerance of the transgenic plants to cold and drought stress and cold resistance-related traits, such as the proline and chlorophyll content, and the activity of peroxidase (POD) and catalase (CAT) increased significantly in the transgenic plants (Yao et al., 2022a). Many studies have shown that the MYB gene family is significantly associated with cold stress and cold resistance (Wu et al., 2021; Dar et al., 2022; Yao et al., 2022b). In *A. thaliana*, AtMYB15 negatively regulates the expression of the c-repeat binding factor (CBF) gene, leading to reduced cold resistance (Agarwal et al., 2006). Although considerable progress has been made in research on cold stress in *S. album*, the research has mainly focused on the physiological and phenotypic responses, and studies on the molecular mechanisms are limited (Zhang et al., 2017; Yan et al., 2019). Specifically, whole-genome identification of the MYB gene family and its expression pattern under cold stress has not yet been reported in *S. album*.

The MYB gene family plays an important role in plant evolution. Generally, a gene family consists of a group of genes from a common ancestor whose members share > 50% pairwise

amino acid sequence similarity and contain a common functional domain (Thornton and DeSalle, 2000). The MYB gene family constitutes one of the largest TF families in plants. The members of this family contain a unique conserved MYB domain, which is composed of 1–4 tandem and non-repetitive R motifs. Based on the number of MYB domains, the MYB genes can be classified into four categories: 1R-MYB/MYB-related (1R motif), R2R3-MYB (2R motifs), 3R-MYB (3R motifs), and 4R-MYB (4R motifs) (Dubos et al., 2010). R2R3-MYB genes are the most abundant subtype in plants; they contain two R structures (R2R3) at the N-terminal and generally contain a transcriptional activation domain at the C-terminal. They are involved in cell differentiation, hormone response, secondary metabolism, and environmental stress. Yao et al. performed genome-wide analysis of the R2R3-MYB gene family members in ginger (*Zingiber officinale* Roscoe) and identified 120 segmental duplications in R2R3-MYB using gene duplication analysis. Ten of the R2R3-MYB genes were significantly differentially expressed under abscisic acid (ABA) and low-temperature stress in ginger leaves (Yao et al., 2022). Zhang et al. identified 69 R2R3-MYB proteins containing a conserved functional domain in *Cyclocarya paliurus*, four of which were found to respond to salt stress by regulating plant hormone signals (Zhang et al., 2022). Yuan et al. (2021) identified 202 R2R3-MYB genes in the polyploid *Saccharum spontaneum* genome sequence. They used collinearity analysis and found that 70% of the genes had experienced duplication events, suggesting the contributors to the MYB gene family expansion. Four of these R2R3-MYB genes actively responded to drought treatment in stress expression analysis (Yuan et al., 2021). However, genome-wide identification of 2R3R-MYB genes and their responses to stress have not yet been reported in *S. album*.

Weighted gene co-expression network analysis (WGCNA) is a data mining method used to analyze the gene expression patterns in multiple samples, which can be used to analyzing the correlation between genes, identifying modules with high phenotypic correlation, and identifying hub genes in different modules (Langfelder et al., 2008). Sharma et al. performed meta-analysis and WGCNA on 390 samples from 29 studies in *A. thaliana* and identified 6,120 and 7,079 differentially expressed genes (DEGs) under drought and cold stress, respectively. They also found that 28% of the DEGs were common to both drought and cold stress, and most of them showed a similar expression pattern (Sharma et al., 2018). Using WGCNA at the transcriptome level, Zeng et al. provided a potential regulatory mechanism for cold stress and recovery of rice cultivars and identified key candidate genes involved in cold tolerance, which provided valuable information for cultivating rice strains with high cold tolerance in the future (Zeng et al., 2022). Li et al. used time-lag initiation of the two pathways and WGCNA in *Arabidopsis* to demonstrate that vernalization was independent of cold acclimation. WGCNA revealed three main networks involving response of ethylene and jasmonic acid, chromatin modification, and cold adaptation in response to prolonged cold exposure, which provided a comprehensive overview of the global changes mediated by cold stress and vernalization in *Arabidopsis* (Li et al., 2021). In the National Center for Biotechnology Information (NCBI) database, transcriptome datasets for 41 samples involving different tissues and

abiotic stress were found for *S. album*. However, because the quality of the reference genome was not optimal, these datasets have not been used effectively at present, especially for WGCNA analysis and identification of related functional genes. Re-mining these important data will help to identify the molecular genetic basis of cold tolerance in *S. album*.

In this study, we performed whole-genome identification and characterization of the R2R3-MYB gene family in *S. album* and revealed the classification and evolution of 75 plants. Furthermore, we performed transcriptome analysis and WGCNA of the RNA-seq data under cold stress at four time points and identified key R2R3-MYB genes responsive to cold stress. Our study provides a genome-wide overview, identifies the expression pattern of R2R3-MYB genes under cold treatment, and reveals the identity of key genes for improving cold resistance and *ex situ* protection of *S. album*.

2 Materials and methods

2.1 Data sources

We obtained a chromosome-level reference genome sequence of *S. album*. The reference genome size was 236.49 million bases (Mb) and included 28,665 coded proteins. For genes with multiple transcripts, the longest transcript was selected for subsequent analysis. The transcriptome and phenotypic data following cold treatments for 0, 12, 24, and 48h were obtained from a published article (Zhang et al., 2017). A total of 12 RNA-seq datasets were collected from the NCBI database under BioProject accession number PRJNA320980, which comprised three replicates of four time points of 0, 12, 24, and 48h at 4°C. The consistency of three replicability at the same time point was acceptable as shown in the previous research article (Zhang et al., 2017).

2.2 Identification of R2R3-MYB genes in *S. album*

Genome sequence and gene annotation information for *S. album* were acquired by our research group (unpublished data). The hidden Markov model (HMM) of the MYB domain was downloaded from the Pfam database (PF00249), and the *S. album* MYB genes were searched in the protein database via an HMM search using HMM files ($E < 1 \times 10^{-5}$) (Finn et al., 2011). Candidate MYB protein sequences, CDS, and conserved domain sequences were extracted using a Perl script (Supplementary File S1). The candidate protein sequences were aligned with Pfam (<http://pfam.xfam.org/>), SMART (<http://smart.embl.de/smart/batch.pl>), and CDD (<https://www.ncbi.nlm.nih.gov/cdd/>) databases to predict the conserved domains. There is a conserved DNA binding domain (DBD) in MYB proteins, made up of 1–4 imperfect amino acid sequence repeats “R.” There are approximately 52 amino acids in each “R,” forming every similar folding architecture with three well-defined α -helices. The helix-turn-helix hydrophobic core was formed by the second and third helices of each “R” with three regularly spaced tryptophans (W) or

other hydrophobic residues. MYB superfamily is divided into four subfamilies based on the number of “R” in MYB DBD: MYB-related subfamily gene with a single or partial “R”, R2R3-MYB subfamily gene with “R2” and “R3”, 3R-MYB subfamily gene with “R1”, “R2”, and “R3” as well as 4R-MYB subfamily gene with four “R1/R2”. A total of 154 MYB genes were ultimately identified in *S. album*, of which 31 were clustered into 1R-MYB type, 114 into R2R3-MYB type, two into 3R-MYB type, and the remaining seven genes could not be clustered in any group with certainty. The online tool ExPASy (<https://web.expasy.org/protparam/>) was used to analyze the amino acid number, isoelectric point (pI), and molecular weight of the R2R3-MYB proteins. Plant-mPLOC (<http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/>) predictor was used to predict the subcellular localization of the R2R3-MYB proteins.

2.3 Phylogenetic analysis, classification of R2R3-MYB genes, and gene duplication analyses

Multiple alignments of R2R3-MYB amino acids were conducted using ClustalW with the default parameters. A phylogenetic tree containing MYB proteins of *S. album* and 75 other plants was constructed using the maximum likelihood (ML) method in MEGA7.0 (Kumar et al., 2016). The parameters were as follows: Poisson model, pairwise deletion, and 1,000 bootstrap replications (Edgar, 2004). The R2R3-MYB family of genes from 75 plants was used as a reference for the classification of R2R3-MYB family members in *S. album*.

Conserved motifs in the R2R3-MYB proteins from *S. album* were analyzed using the online tool MEME (<http://meme-suite.org/>) with the following parameters: maximum number of motifs = 20 and optimum width = 6–100 residues (Bailey et al., 2009). Finally, the conserved motifs and domains were visualized using the TBtools software (Chen et al., 2020). The distribution of 114 R2R3-MYB genes identified in the reference genome annotation information of *S. album* was mapped to chromosomes and visualized using the software MG2C (Chao et al., 2021). MCScanX is used to analyze tandem and segmental gene duplications (Wang et al., 2012). All R2R3-MYB protein sequences of *S. album* were compared against themselves using BLASTP, with tabular output format (-m 8) and an e value of $< 1e^{-10}$. The BLASTP tabular file and a simplified *S. album* gene location file served as inputs for MCScanX to identify duplication types using default settings (Tang et al., 2014). Values of nonsynonymous (Ka) and synonymous (Ks) substitution rates were calculated using KaKs_Calculator software (Zhang et al., 2006).

2.4 Expression pattern of R2R3-MYB genes

All RNA sequence datasets from *S. album* leaves at the four cold treatment time points were filtered using Fastp with default parameters (Chen et al., 2018). The total filtered high-quality clean RNA-seq data were mapped to the reference genome of *S. album* (not officially published) using Hisat2 with default parameters (Kim et al., 2019). The mapped reads with mapping quality (MQ) ≤ 30 were

filtered using Samtools, and BAM files were sorted. StringTie (Pertea et al., 2015) was used to count unique and normalized mapped reads as fragments per kilobase per million (FPKM) mapped reads for each gene with the parameter of “-B -e -G.” Transcriptome gene expression data were visualized using ImageGP software (Chen et al., 2022). The R2R3-MYB gene expression data are listed in Supplementary Table S2.

2.5 WGCNA analysis and prediction of the three-dimensional structure of key genes

To identify the core gene modules and the hub genes within the modules related to cold stress, we performed a WGCNA (Langfelder and Horvath, 2008). We evaluated 10,483 genes after filtering total FPKM ≤ 100 at the four time points and three biological replicates to test their availability and used the R package termed “WGCNA” to construct a gene co-expression network. Candidate power values were set from 1 to 30. The threshold for scale-free topology model fit was set at 0.8. Subsequently, we constructed an adjacency matrix to describe the correlation strength between the modules and cold resistance-related traits, which were malondialdehyde (MDA), superoxide dismutase (SOD), peroxidase (POD) activity, soluble sugar content, intercellular CO₂ concentration, and respiration rate. These phenotypic data were collected from the previous research article (Zhang et al., 2017). Correlations between modules and cold stress-related traits were calculated using a function in “moduleTraitCor.” Finally, the genes in the modules with high correlation with cold stress-related traits were extracted and compared with the significant DEGs under different cold treatments. Genes present in both DEGs and modules were considered as the key genes associated with cold stress in *S. album*.

The spatial structure of genes determines their functions. The functions and three-dimensional (3D) structures of the key genes were predicted using the Swiss-Model website. From the database, sequences with coverage greater than 50% were selected as templates for structural prediction. The qualitative model energy analysis (QMEAN) values of > 0.6 for predicted structures was considered reliable.

3 Results

3.1 Identification and analysis of physicochemical properties of R2R3-MYB genes

Genome sequence and gene annotation information for *S. album* were obtained by our research group (unpublished). To determine the conserved sequence of the MYB protein in *S. album*, the conserved domains of the MYB genes in *S. album* were identified through alignment based on the classical HMM (PF00249). Two comparison searches were performed and online sites such as Simple Modular Architecture Research Tool (SMART), Pfam, and National Center for Biotechnology Information-

Conserved Domains Database (NCBI-CDD) were used to verify the structures of the conserved domains in the candidate genes. The remaining seven MYB genes could not be classified with certainty. A total of 114 R2R3-MYB genes were named based on their subfamilies as SaMYB001–SaMYB114 (Supplementary Table S1; Supplementary File S2). These genes were distributed across 10 chromosomes. The physicochemical properties of the MYB proteins were analyzed online using the ExPaSy tool. The results showed that the 114 MYB proteins contained 169–1452 amino acids, with a molecular weight of 18.36–159.44 kDa. SaMYB087 contained the lowest number of amino acids (169 aa) and consequently had the least molecular weight (18.36 kDa). The mean theoretical pI of all the R2R3-MYB genes was 7.03 (range: 4.71–9.58). The theoretical pI of 64 R2R3-MYB proteins was < 7 , indicating that these proteins were acidic, whereas that of the remaining 50 R2R3-MYB proteins was > 7 , indicating that these were alkaline. The predicted subcellular localization of all the proteins was nuclear, and only SaMYB049 was distributed in the chloroplast and nucleus (Supplementary Table S1).

3.2 Phylogenetic analysis and characterization of R2R3-MYB genes in *S. album*

The R2R3-MYB protein sequences of *S. album* and 75 other plants (Wu et al., 2022) were aligned using ClustalW program and a phylogenetic evolution tree was constructed. Toward subfamily classification of the R2R3-MYB proteins (Wu et al., 2022), the 114 MYB proteins in *S. album* were divided into 40 categories (Figure 1; Supplementary Table S1): the S21 category in watchet, with the largest number of eight R2R3-MYB genes in *S. album* and the S22 category in cyan, with the second largest number of seven R2R3-MYB genes in *S. album*. The number of SaR2R3-MYB genes in each subgroup was close to that of R2R3-MYB genes in *A. thaliana*. Interestingly, all genes in subfamily A were R2R3-MYB genes in *S. album*, and no genes from any other species were clustered with them, which indicates that Subfamily A may be unique to *S. album*. Genes of the same subfamily exhibit high homology and sequence similarity and may have similar gene functions. The S21 subfamily harboring eight R2R3-MYB genes in *S. album* is related to specialized metabolic processes, such as cell wall thickening, seed oil accumulation, and phenylpropanoid metabolism. The S22 subfamily, harboring seven R2R3-MYB genes in *S. album*, is related to abiotic stress. The genes in the S1 and S2 subfamilies were related to abiotic stress, such as salt tolerance, drought stress, and cold tolerance.

MEGA7 and WEBLOGO softwares were used to conduct multiple sequence alignments and to characterize the conserved regions in the 114 R2R3-MYB proteins in *S. album*. Figure 2 shows that the R2 structure contains three very conserved tryptophan (W) residues, and every two W residues are separated by 19 amino acids. The R3 structure contains two very conserved W residues: The first W is replaced by phenylalanine (F), isoleucine (I), and leucine (L) and the second and third W residues are separated by 18 amino acids. The online software MEME was used to analyze the

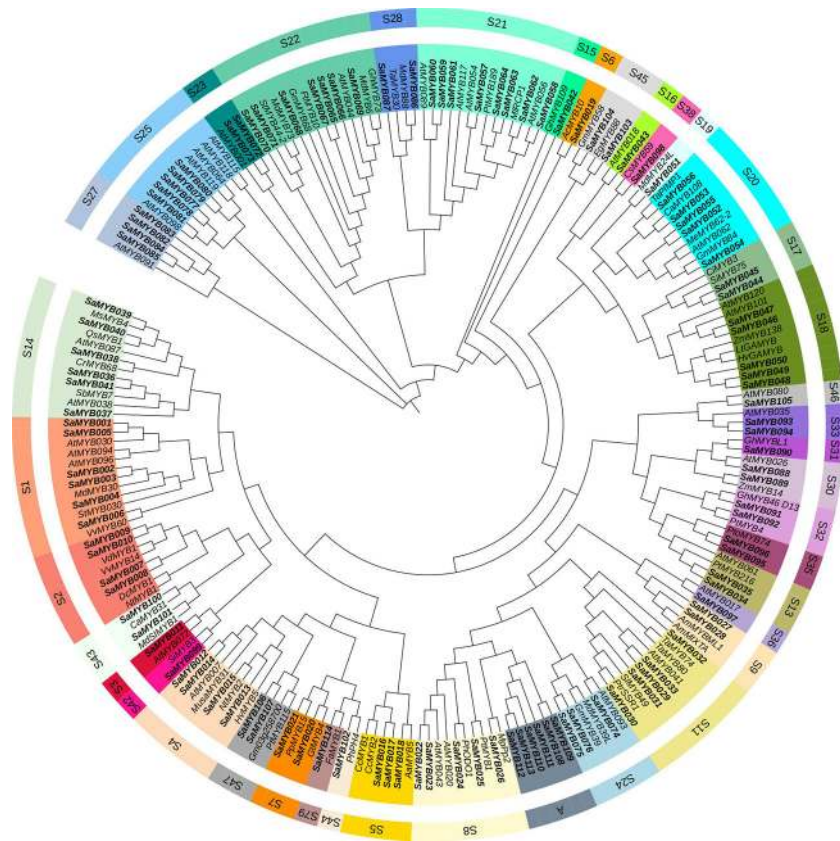


FIGURE 1
Result of phylogenetic analysis of R2R3-MYB proteins in *S. alba*. Same color font of genes represent that they were in common subfamilies.

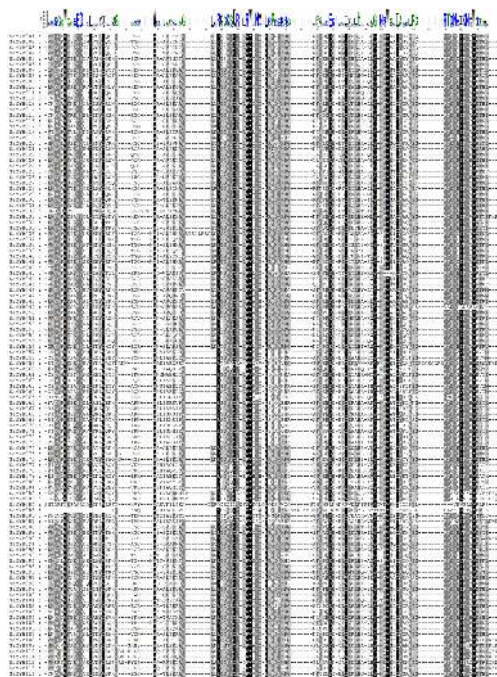


FIGURE 2
Results of conserved motifs and sequence alignment analyses of 114 R2R3-MYB proteins in *S. alba*.

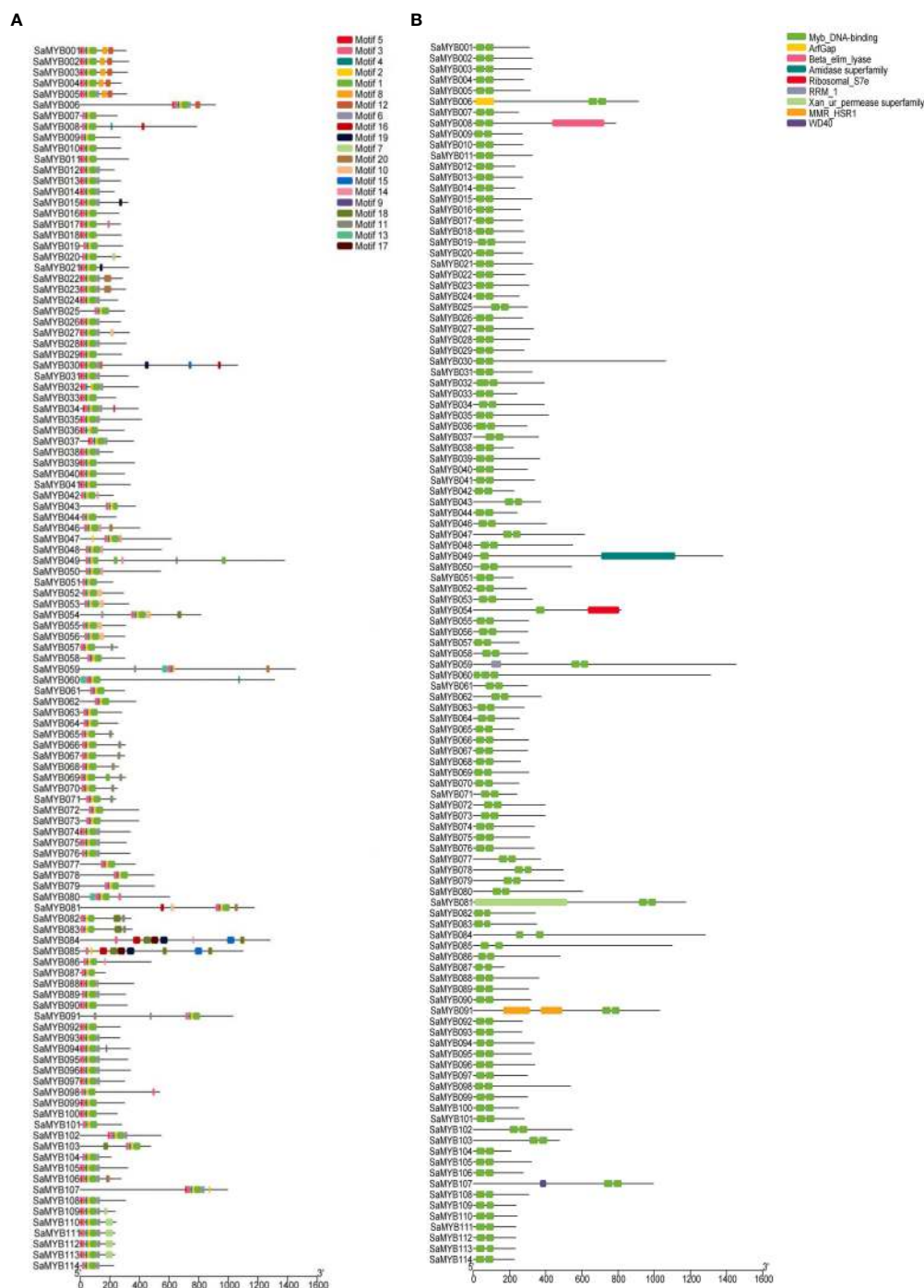


FIGURE 3

Conserved motif analyses of the 114 R2R3-MYB genes in *S. alba*. (A) The left panel shows the results of conserved motif analysis. A total of 20 predicted motifs are represented by different colored boxes. (B) The right panel shows the conserved domains in the 114 R2R3-MYB genes with different colors representing the different types of domains.

conserved domains in R2R3-MYB proteins (Figure 3). Analysis of the conserved motifs revealed that all R2R3-MYB proteins harbored motif 1 (green box), motif 2 (yellow box), and motif 3 (pink box), as shown in Figure 3A. All R2R3-MYB proteins in *S. alba* contained conserved MYB DNA-binding domains (Figure 3B). These conserved amino acids and structures indicate that the R2R3-MYB genes identified in our study are reliable.

3.3 Gene duplication and chromosomal location of R2R3-MYBs in *S. alba*

The location of the R2R3-MYB genes was extracted from the gene annotation general feature format (gff) file of *S. alba*. The results indicated that the 114 R2R3-MYB genes were unevenly distributed on ten chromosomes (Figure 4): six genes in

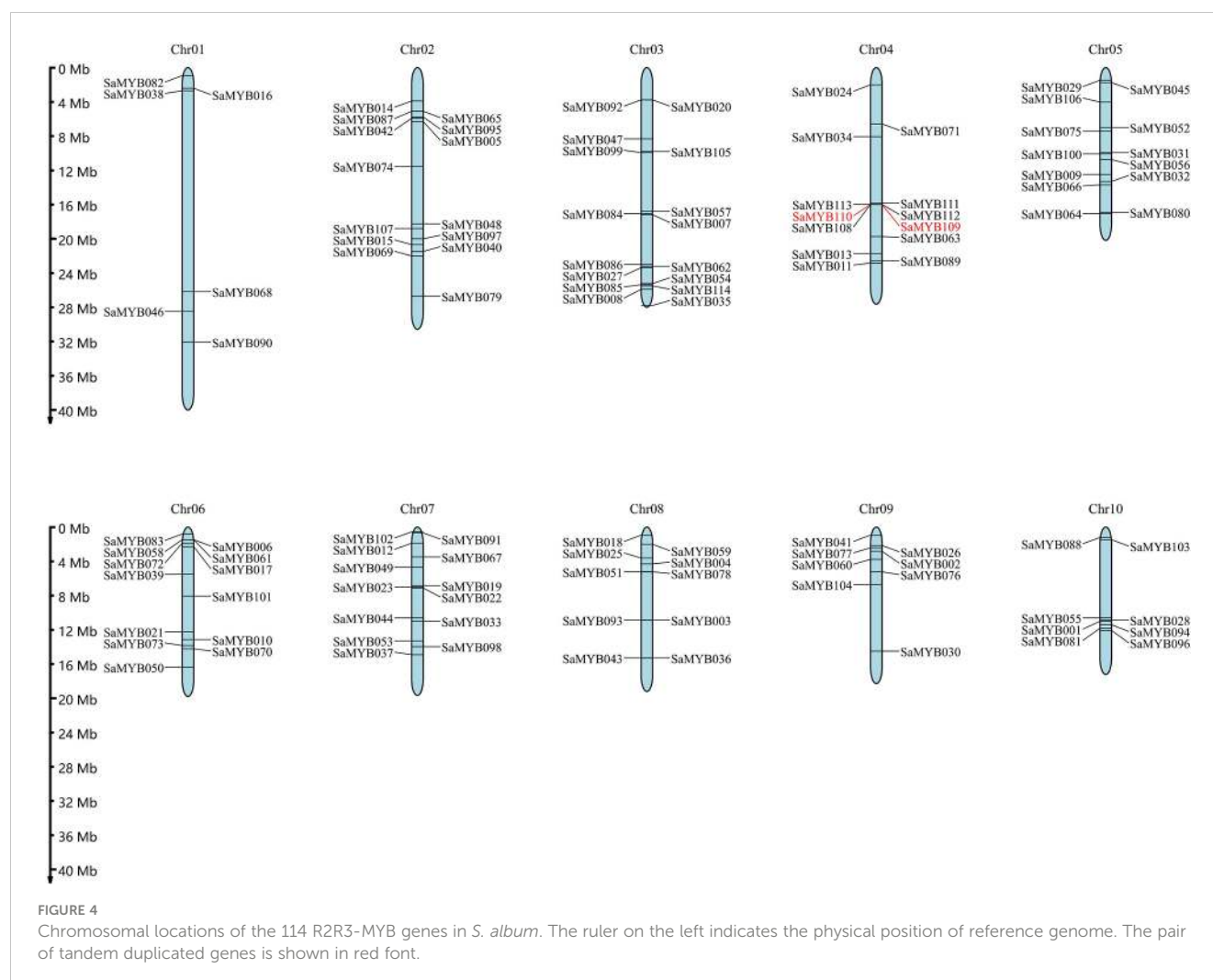


FIGURE 4

Chromosomal locations of the 114 R2R3-MYB genes in *S. album*. The ruler on the left indicates the physical position of reference genome. The pair of tandem duplicated genes is shown in red font.

chromosome 01 (5.26% of the total); 14 genes in chromosome 02 (12.30%); 16 genes in chromosome 03 (14.04%) with the most genes; 13 genes each in chromosome 04, chromosome 05, chromosome 06, and chromosome 07 (11.40%); 10 genes in chromosome 08 (8.77%); and eight genes each in chromosome 09 and chromosome 10 (7.02%).

On one of the aforementioned ten chromosomes was a pair of tandem duplicate genes, specifically distributed in chromosome 04 (SaMYB109 and SaMYB110). The two genes belong to Subfamily A, and the distance between them was 8116 bp. Their nonsynonymous rate (K_a) and synonymous rate (K_s) substitution rates were analyzed, the non-synonymous mutation rate was 0.39, and the synonymous mutation rate was 1.75. The K_a/K_s ratio was found to be 0.22 (Table 1), which was less than 1, indicating negative selection pressure. In addition to tandem duplication, MCScanX was used to analyze fragment duplication events in the R2R3-MYB gene family. The results showed that 33.33% (38/114) of the R2R3-MYB members showed segmental duplication. The K_a/K_s values of SaMYB103 and SaMYB017 were > 1 (1.17), implying that they evolved under positive selection pressure. The K_a/K_s value of the remaining gene pairs was < 1 , indicating that they evolved under the effect of purifying selection (Table 1). Thus, the analysis of

duplication events in the R2R3-MYB genes suggests that some genes were produced by tandem and segmental duplication, and these gene duplication events may be among the driving forces of gene evolution.

3.4 SaMYB gene expression pattern under cold stress

We collected transcriptome datasets corresponding to 6-month-old *S. album* seedling leaves subjected to 4°C treatment for 0 (0h), 12 (12h), 24 (24h), and 48 (48h) using three biological replicates from the NCBI database generated in a previous study (Zhang et al., 2017). After quality control assessments, a total of 7.62 Gb clean data was retained, Q30 of which accounted for over 91.25%. We used the Hisat2 alignment program to map the qualified RNA-sequencing data on our reference genome (unofficially published). The mapping rate ranged from 93.85 to 95.81%, with an average of 95.11%, which indicated that these RNA-sequence data were reliable to quantify the global abundance of R2R3-MYB gene expression following cold treatment. The square of the Pearson's correlation coefficient

TABLE 1 Ka, Ks and Ka/Ks of replication pairs of R2R3-MYB gene family in *S. album*.

Duplicated gene pairs	Non synonymous (Ka)	Synonymous (Ks)	Ka/Ks	Duplicated type
SaMYB110 & SaMYB109	0.39	1.75	0.22	tandem
SaMYB003 & SaMYB002	0.17	0.85	0.19	segmental
SaMYB004 & SaMYB003	0.17	0.84	0.20	segmental
SaMYB005 & SaMYB001	0.25	1.19	0.21	segmental
SaMYB009 & SaMYB010	0.20	0.70	0.28	segmental
SaMYB012 & SaMYB014	0.07	0.66	0.11	segmental
SaMYB016 & SaMYB017	0.28	1.16	0.24	segmental
SaMYB029 & SaMYB033	0.24	1.21	0.20	segmental
SaMYB036 & SaMYB041	0.21	0.93	0.22	segmental
SaMYB044 & SaMYB045	0.12	1.04	0.11	segmental
SaMYB059 & SaMYB060	0.21	0.50	0.43	segmental
SaMYB063 & SaMYB064	0.19	0.92	0.21	segmental
SaMYB065 & SaMYB067	0.35	0.74	0.48	segmental
SaMYB066 & SaMYB067	0.36	1.79	0.20	segmental
SaMYB068 & SaMYB069	0.15	0.68	0.23	segmental
SaMYB072 & SaMYB073	0.32	1.79	0.18	segmental
SaMYB077 & SaMYB078	0.26	0.94	0.28	segmental
SaMYB081 & SaMYB078	2.08	NA	NA	segmental
SaMYB082 & SaMYB083	0.07	0.57	0.13	segmental
SaMYB084 & SaMYB085	1.61	1.94	0.83	segmental
SaMYB095 & SaMYB096	0.18	1.04	0.18	segmental
SaMYB103 & SaMYB017	2.07	1.77	1.17	segmental

(r) was > 0.91 among the three biological replicates at each time point, indicating both operational stability and reliability of the experimental results. Of the 114 R2R3-MYB genes, 89 were found to be expressed at least once time. The expression patterns of the R2R3-MYB genes in *S. album* under different cold-stress conditions were visualized using heatmap analysis. The results showed that the expression levels of 31 R2R3-MYB genes were significantly up-regulated ($\log_2[\text{fold change}] > 1$) at least one time point of cold stress (Figure 5; Supplementary Table S2), whereas that of 13 R2R3-MYB genes were significantly down-regulated at least one time point of cold treatment. SaMYB078 and SaMYB100 were not expressed at 0h at 4°C. However, they were expressed at 12, 24, and 48h after cold treatment, which indicated that the two genes may be involved in cold-stress resistance. Five genes (SaMYB031, SaMYB068, SaMYB069, SaMYB004, and SaMYB015) were highly expressed following cold treatment at three different time points, whereas SaMYB114 and SaMYB058 were down-regulated at three time points of cold treatment (Figure 5; Supplementary Table S2).

3.5 Identification of key R2R3-MYB genes related to cold stress based on WGCNA

WGCNA can be used to identify modules of highly correlated genes, to summarize such clusters using the module eigengene or an intramodular hub gene, and to relate modules to one another and to external sample traits. Here, we used WGCNA to identify the hub R2R3-MYB genes associated with cold stress. After removing genes with a total FPKM ≤ 100 at four time points and three biological replicates, a total of 10,483 genes were reserved for WGCNA. Correlation coefficient cluster analysis of the expression levels of 12 samples revealed that the clustering among samples was acceptable and there were no outliers (Figure 6A). Twenty candidate power values from 1 to 30 were set for model fitting. The results showed that, when the power value was 20, there was an obvious inflection point where the topology model fit value was 0.8 and stable (Figure 6B). Therefore, we chose a power value of 20 for module construction and clustering. Based on correlation analysis and clustering according to the FPKM value of different genes, the

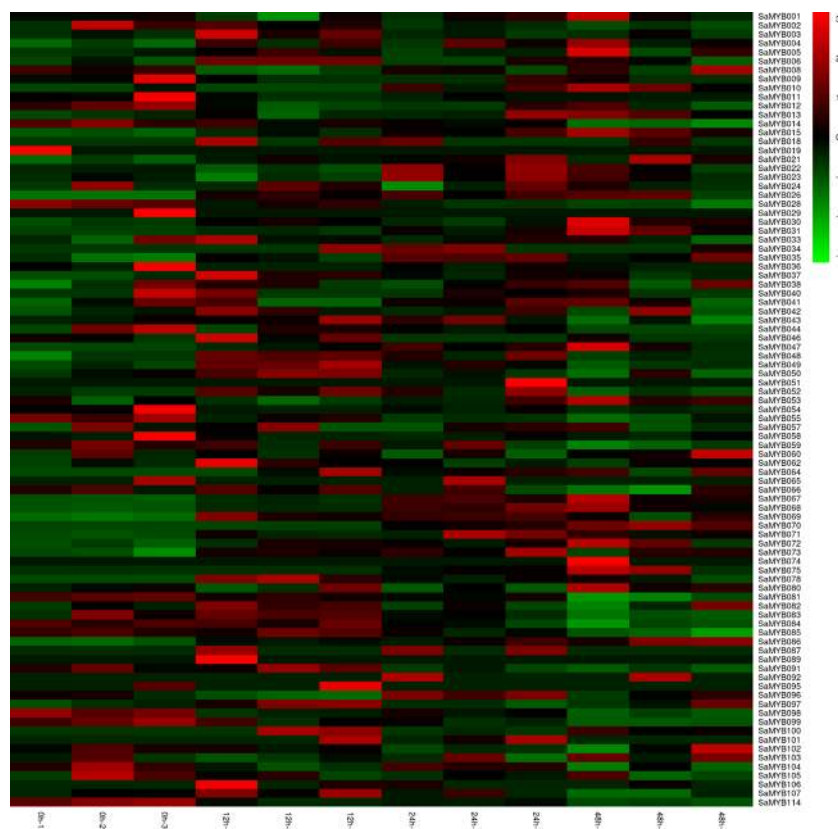


FIGURE 5

R2R3-MYB gene expression levels determined by RNA-seq at the four time points of cold (4°C) treatment. Expression profiles were normalized to log₁₀(FPKM). The color scale represents relative expression level from low (green) to high (red) values.

genes with high correlation were allocated to the same module. Different genes were divided into 17 modules according to their co-expression patterns, with different colors representing the different modules (Figure 6C; Supplementary Table S3). For correlation analysis between the trends in gene expression modules and cold stress-related traits, the Pearson correlation coefficient ($r > 0.5$) and $P < 0.05$ were set as thresholds. The MEred and MEgreen modules were found to significantly positively correlate with changes in malondialdehyde (MDA), superoxide dismutase (SOD), peroxidase (POD) activity, soluble sugar content, and intercellular CO₂ concentration in response to cold stress, whereas they negatively correlated with photosynthetic rate, conductance to H₂O, and respiration rate. MEblue was most significantly positively correlated with the photosynthetic rate, conductance to H₂O, and respiration rate; the correlation coefficients were up to 0.9. The MEblue module was most significantly negatively correlated with MDA, SOD, POD activity, soluble sugar content, and intercellular CO₂ concentration. This indicated that the genes in these significantly correlated modules were the core genes involved in cold-stress response.

Further analysis revealed that the MEblue module contained three (SaMYB014, SaMYB084, and SaMYB059), MEbrown module contained one (SaMYB098), and MERed module contained three R2R3-MYB genes (SaMYB015, SaMYB030, and SaMYB081). The METurquoise module contained five

R2R3-MYB genes (SaMYB068, SaMYB085, SaMYB083, SaMYB050, and SaMYB091). Combined with the expression levels of these genes at different durations of cold stress, the expression levels of three core genes were significantly different under cold stress treatment. The expression level of SaMYB098 was 85.49 at 0h cold treatment, but decreased to 26.55 at 12h cold treatment, and further decreased to 9.32 after 48h cold stress, indicating that SaMYB098 may be negatively regulated to cold stress. The expression of SaMYB015 in the MERed module was low at 0h (FPKM = 3.18); however, it increased progressively after cold stress treatment, and the expression levels at 12, 24, and 48h were 7.73, 12.87, and 18.18, respectively. The expression level of SaMYB068 in the METurquoise module was similar to that of SaMYB015 (Figure 7A). The 3D structure prediction analyses of the three important genes using the Swiss-Model online tool revealed that, although their structures were different at the 3D level (Figures 7B–D, QMEAN Z-scores > 0.65), they all harbored a classical R2R3-MYB binding domain. Therefore, it is suggested that these three genes may participate in cold stress in different ways. Overall, using a combination of gene family identification, transcriptome data, and WGCNA analysis, three core genes with differential expression significantly related to cold stress were identified, which can provide important information for the genetic improvement of *S. album* under cold stress.

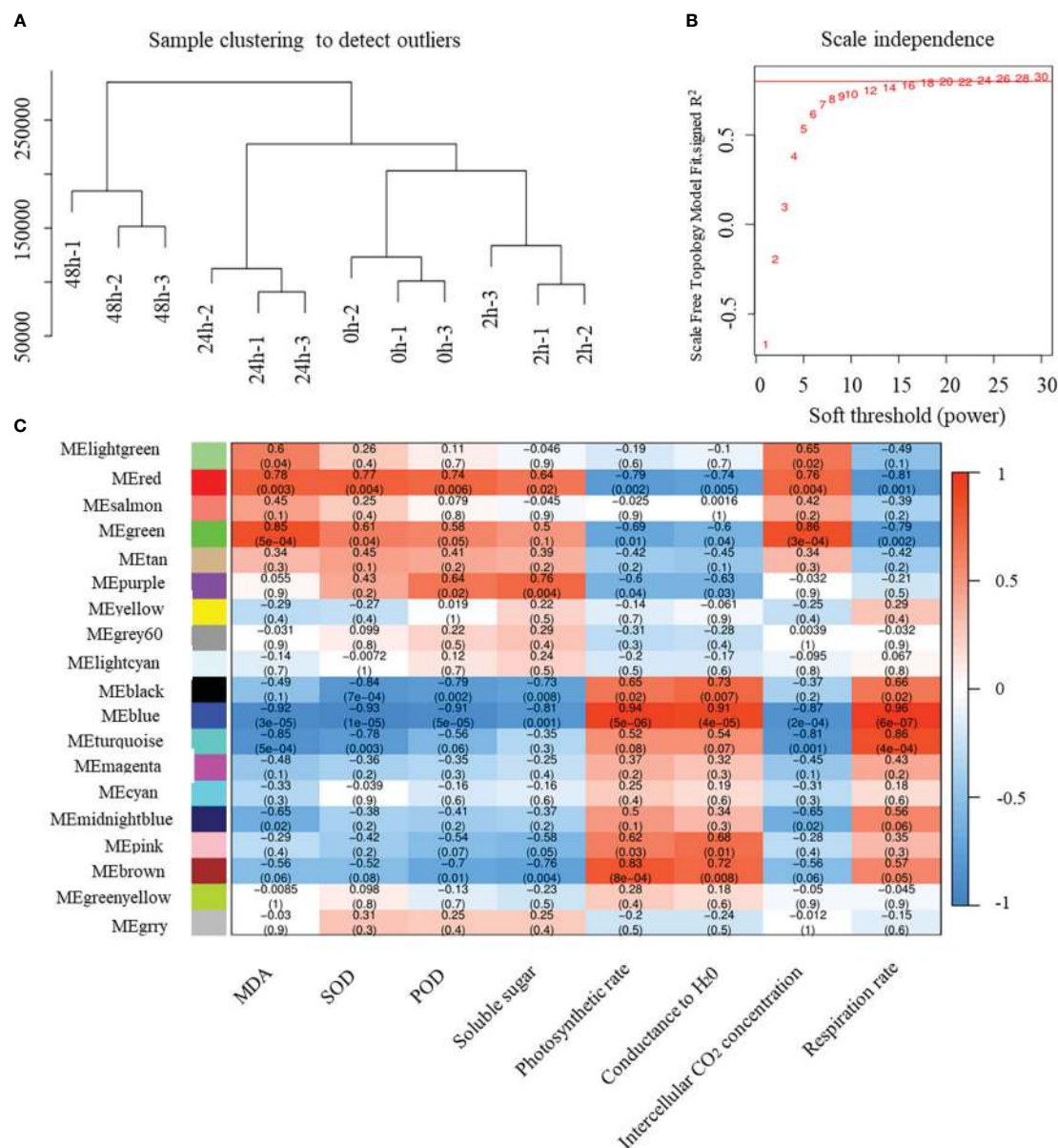


FIGURE 6

Results of WGCNA based on the gene expression level and phenotypic data. (A) Results of sample clustering; (B) results of scale independence, candidate power was from 1 to 30; (C) relationship of different modules and cold resistance-related traits. Numbers in the box represent the correlation coefficient, and the number in the brackets represent the corresponding *p*-value. The color scale represents correlation coefficient from 1 (red) to -1 (blue).

4 Discussion

Among the many transcription factor families in plants, the MYB family constitutes one of the largest transcription factor families. MYB proteins regulate plant growth and development, primary and secondary metabolism, and response to various abiotic stresses, such as drought, cold, and salt (Dubos et al., 2010; Chen et al., 2019; Millard et al., 2019). Genome-wide identification of the MYB gene family has been performed in various plants, such as *Arabidopsis* (Matus et al., 2008), *Cucumis sativus* (Li et al., 2012), maize (Dias et al., 2003), and wheat (Gao et al., 2016). The members of the plant MYB gene family contain 1–4 incomplete MYB repeat sequences and can be divided

into four subfamilies based on the number of repeat sequences: 1R-MYB type (containing one or two separate repeat sequences), R2R3-MYB type (containing two adjacent repeat sequences), and 3R-MYB type and 4R-MYB type (containing three and four adjacent repeat sequences, respectively). Among these, the R2R3-MYB gene family is most abundant in Plants (Jiang and Rao, 2020; Tombuloglu, 2020). However, the whole-genome MYB gene family has not yet been identified in *S. albus*. In this study, the MYB gene family in *S. albus* was identified at the whole-genome level. The physical and chemical properties, physical tree, preserved motifs, and gene duplications were systematically analyzed. Consequently, 154 MYB genes in *S. albus* were identified, 114 of which were classified as R2R3-MYB genes. The

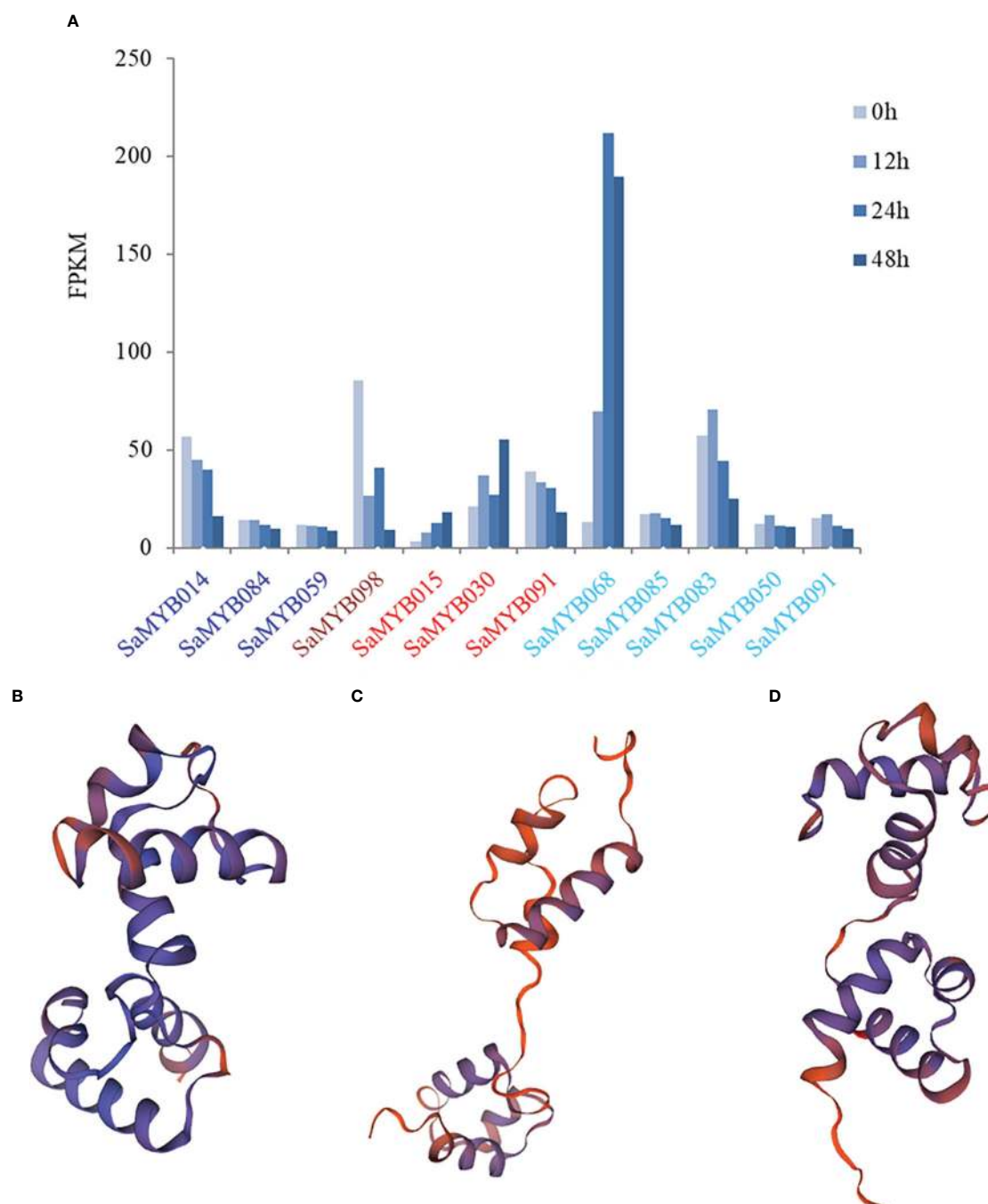


FIGURE 7

The expression patterns of module genes and the protein structure prediction results of the three important genes. (A) color of the text in the x-axis represents the corresponding module in same color in Figure 6. The y-axis represents the gene expression level; (B) the prediction result of protein spatial structure of *SaMYB098*; (C) the prediction result of protein spatial structure of *SaMYB01*; (D) the prediction result of protein spatial structure of *SaMYB068*.

R2R3-MYB gene family in plants contains two R structures at the N-terminal. The number of R2R3 genes in *S. album* was similar to that in other plants, such as 125 in *Arabidopsis* (Stracke et al., 2001), 119 in pea (Yang et al., 2022), 114 in moso bamboo (Hou et al., 2018), and 133 in white clover (Ma et al., 2022), indicating that the results for the R2R3 gene family in *S. album* were reliable. The functions of R2R3 MYBs can be divided into three main processes: development and cell differentiation, specialized metabolism, and stress response. Du et al.

classified the gene family into 73 subfamilies based on their highly conserved domain and motif composition (Du et al., 2015; Wu et al., 2022). Through multiple sequence alignment and evolutionary analysis, the 114 *S. album* MYB proteins were divided into 40 categories (Figure 1; Supplementary Table S1). Thirty-nine of the 40 subfamilies could be clustered with the corresponding subfamilies of other plants. Only Subfamily A contained six genes (*SaMYB108*, *SaMYB109*, *SaMYB110*, *SaMYB111*, *SaMYB112*, and *SaMYB113*),

indicating that these genes may participate in specific biological processes in *S. album*.

Gene duplication occurs through various mechanisms, such as tandem duplication, gene transformation, horizontal transfer and other translocations, hybridization, duplication, and doubling of the whole segment of the chromosome during the recombination process, resulting in effective doubling of the large-scale genome sequence (Sankoff, 2001). Regardless of the origin, duplication of genes in a genome can lead to increased expression of the gene. Differentiation of duplicated genes may be subject to sub-functionalization and produce neo-functionalization to enhance the ability of plants to adapt to the environment and respond to various stresses (Liu and Adams, 2007; Airolidi and Davies, 2012). In this study, a pair of tandem duplicated genes was identified in the R2R3-MYB gene family in *S. album*. About one-third of R2R3-MYB genes have replication events, indicating that these genes may participate in special biological processes in *S. album*. Nonsynonymous (Ka) and synonymous (Ks) substitution rates are important indicators for studying the pressures of gene selection, identifying deviations from neutrality, and estimating the occurrence time of replication events (Hurst, 2002; Echave et al., 2016). A comparative analysis of the positively and negatively selected genes was performed based on Ka and Ks, and artificial selection was found to be a major factor affecting the gene evolutionary rate, compactness, expression level, and genetic diversity in barley (Tao et al., 2022). Song et al. compared negatively and positively selected genes in *Arachis duranensis* and *Arachis ipaënsis* and found Ks to be a determining factor that affected the selection pressure (Song et al., 2018). In 22 pairs of duplicate genes in our study (excluding one pair [SaMYB110 and SaMYB109] with Ka/Ks value > 1), the Ka/Ks value of the duplicated genes was < 1 (range: 0.11–0.83), which indicates that most of the R2R3-MYB genes in *S. album* were subjected to purifying selection.

Although the non-MYB regions diverged in different plant species, the two conserved MYB structures are a signature feature of R2R3-MYB genes and include two main functional parts: a DNA-binding domain at the N-terminal and a regulatory region at the C-terminal (Wu et al., 2022). In this study, 114 R2R3-MYB genes harboring highly conserved R2R3 sequences were identified in *S. album* (Figures 2, 3), and the subcellular localization results showed that all of them localized to the nucleus excepting SaMYB049 distributed in the chloroplast and nucleus. These results are consistent with those reported for other plants (Stracke et al., 2001; Wilkins et al., 2009), indicating that the R2R3-MYB genes identified in this study are reliable. Although the R2R3-MYB genes have highly conserved sequences, they are functionally distinct. Some of them are essential for plant development and cell differentiation (Oshima et al., 2013; Liu et al., 2015), whereas some serve as key regulators of responses to diverse environmental stresses (abiotic and biotic stresses), such as drought, temperature, and salinity (Li et al., 2015). Some of them play an important role in specialized metabolic biosynthesis pathways (Stracke et al., 2001; Dubos et al., 2010). In this study, three key candidate genes that are significantly related to cold stress were identified through genome-wide identification of the R2R3-MYB gene family, transcriptome analysis, and WGCNA. Comparison of the 3D structure prediction revealed that they were significantly different in their spatial structure (except for the conserved R2R3-

MYB domain) and expression pattern, which indicates that they may respond to cold stress in different ways. These three key candidate genes may provide important information for genetic improvement and research on cold tolerance in *S. album*.

5 Conclusion

R2R3-MYB genes play important roles in the growth and development and abiotic stress of plants; however, there is no relevant report in *S. album* yet. In our study, the R2R3-MYB gene family of *S. album* was identified at the whole-genome pattern, and its characteristics were analyzed and its expression pattern under cold stress was studied. Three important candidate genes significantly related to cold resistance were found by combining transcriptome results and WGCNA results. These results fill in the gap of R2R3-MYB gene family of *S. album* and provide great significance for genetic improvement and molecular mechanism research to cold resistance in *S. album*.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. The RNA-seq data presented in the study are deposited in the NCBI repository, accession number PRJNA320980. Further inquiries can be directed to the corresponding authors.

Author contributions

SX and LC designed this study and revised the manuscript. MT carried out the study and wrote the manuscript. LL and HX performed the analyses of MYB gene family and transcriptome dataset. HZ and ZW revised the manuscript. YL and QZ helped to improve the language of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the National Natural Science Foundation of China (32201624 and 31760316), the Hainan Provincial Natural Science Foundation of China (321RC469), and Priming Scientific Research Foundation of Hainan University (KYQD(ZR)-21039).

Acknowledgments

We sincerely thankful to Dr. Guohua Ma (Chinese Academy of Sciences) published and uploaded the RNA-sequencing data to the public database. We would like to thank the reviewers for their constructive comments.

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1142562/full#supplementary-material>

References

- Agarwal, M., Hao, Y., Kapoor, A., Dong, C.-H., Fujii, H., Zheng, X., et al. (2006). A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance. *J. Biol. Chem.* 281 (49), 37636–37645. doi: 10.1074/jbc.M605895200
- Airolidi, C. A., and Davies, B. (2012). Gene duplication and the evolution of plant MADS-box transcription factors. *J. Genet. Genomics* 39 (4), 157–165. doi: 10.1016/j.jgg.2012.02.008
- An, J. P., Wang, X. F., Zhang, X. W., Xu, H. F., Bi, S. Q., You, C. X., et al. (2020). An apple MYB transcription factor regulates cold tolerance and anthocyanin accumulation and undergoes MIEL1-mediated degradation. *Plant Biotechnol. J.* 18 (2), 337–353. doi: 10.1111/pbi.13201
- 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
- Chao, J., Li, Z., Sun, Y., Aluko, O. O., Wu, X., Wang, Q., et al. (2021). MG2C: A user-friendly online tool for drawing genetic maps. *Mol. Horticulture* 1 (1), 1–4. doi: 10.1186/s43897-021-00020-x
- 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 (8), 1194–1202. doi: 10.1016/j.molp.2020.06.009
- Chen, C., Zhang, K. X., Khurshid, M., Li, J. B., He, M., Georgiev, M. I., et al. (2019). MYB transcription repressors regulate plant secondary metabolism. *Crit. Rev. Plant Sci.* 38 (3), 159–170. doi: 10.1080/07352689.2019.1632542
- Chen, S. F., Zhou, Y. Q., Chen, Y. R., and Gu, J. (2018). Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34 (17), 884–890. doi: 10.1093/bioinformatics/bty560
- Chen, T., Liu, Y.-X., and Huang, L. (2022). ImageGP: An easy-to-use data visualization web server for scientific researchers. *iMeta* 1 (1), e5. doi: 10.1002/imt.2.5
- Dar, N. A., Mir, M. A., Mir, J. I., Mansoor, S., Showkat, W., Parihar, T. J., et al. (2022). MYB-6 and LDOX-1 regulated accretion of anthocyanin response to cold stress in purple black carrot (*Daucus carota* L.). *Mol. Biol. Rep.* 49 (6), 5353–5364. doi: 10.1007/s11033-021-07077-3
- Dias, A. P., Braun, E. L., McMullen, M. D., and Grotewold, E. (2003). Recently duplicated maize R2R3 myb genes provide evidence for distinct mechanisms of evolutionary divergence after duplication. *Plant Physiol.* 131 (2), 610–620. doi: 10.1104/pp.012047
- Du, H., Liang, Z., Zhao, S., Nan, M. G., Tran, L. S. P., Lu, K., et al. (2015). The evolutionary history of R2R3-MYB proteins across 50 eukaryotes: New insights into subfamily classification and expansion. *Sci. Rep.* 5, 16. doi: 10.1038/srep11037
- Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in arabidopsis. *Trends Plant Sci.* 15 (10), 573–581. doi: 10.1016/j.tplants.2010.06.005
- Echave, J., Spielman, S. J., and Wilke, C. O. (2016). Causes of evolutionary rate variation among protein sites. *Nat. Rev. Genet.* 17 (2), 109–121. doi: 10.1038/nrg.2015.18
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32 (5), 1792–1797. doi: 10.1093/nar/gkh340
- Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367
- Gao, F., Zhao, H. X., Yao, H. P., Li, C. L., Chen, H., Wang, A. H., et al. (2016). Identification, isolation and expression analysis of eight stress-related R2R3-MYB genes in tartary buckwheat (*Fagopyrum tataricum*). *Plant Cell Rep.* 35 (6), 1385–1396. doi: 10.1007/s00299-016-1971-5
- Hou, D., Cheng, Z. C., Xie, L. H., Li, X. Y., Li, J., Mu, S. H., et al. (2018). The R2R3MYB gene family in *Phyllostachys edulis*: Genome-wide analysis and identification of stress or development-related R2R3MYBs. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00738
- Hurst, L. D. (2002). The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet. TIG* 18 (9), 486. doi: 10.1016/s0168-9525(02)02722-1
- Jeon, J., and Kim, J. (2013). Cold stress signaling networks in arabidopsis. *J. Plant Biol.* 56 (2), 69–76. doi: 10.1007/s12374-013-0903-y
- Jiang, C. K., and Rao, G. Y. (2020). Insights into the diversification and evolution of R2R3-MYB transcription factors in Plants. *Plant Physiol.* 183 (2), 637–655. doi: 10.1104/pp.19.01082
- Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37 (8), 907–90+. doi: 10.1038/s41587-019-0201-4
- Kumar, A. A., Joshi, G., and Ram, H. M. (2012). Sandalwood: history, uses, present status and the future. *Curr. Sci.* 103 (12), 1408–1416.
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33 (7), 1870–1874. doi: 10.1093/molbev/msw054
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9, 559. doi: 10.1186/1471-2105-9-559
- Li, F., Hu, Q., Chen, F. D., and Jiang, J. F. (2021). Transcriptome analysis reveals vernalization is independent of cold acclimation in arabidopsis. *BMC Genomics* 22 (1), 14. doi: 10.1186/s12864-021-07763-3
- Li, L. J., Lu, X. C., Ma, H. Y., and Lyu, D. G. (2018). Comparative proteomic analysis reveals the roots response to low root-zone temperature in *Malus baccata*. *J. Plant Res.* 131 (5), 865–878. doi: 10.1007/s10265-018-1045-6
- Li, C. N., Ng, C. K. Y., and Fan, L. M. (2015). MYB transcription factors, active players in abiotic stress signaling. *Environ. Exp. Bot.* 114, 80–91. doi: 10.1016/j.envexpbot.2014.06.014
- Li, Q., Zhang, C. J., Li, J., Wang, L., and Ren, Z. H. (2012). Genome-wide identification and characterization of R2R3MYB family in *Cucumis sativus*. *PLoS One* 7 (10), 18. doi: 10.1371/journal.pone.0047576
- Liu, Z., and Adams, K. L. (2007). Expression partitioning between genes duplicated by polyploidy under abiotic stress and during organ development. *Curr. Biol. CB* 17 (19), 1669–1674. doi: 10.1016/j.cub.2007.08.030
- Liu, J. Y., Osbourn, A., and Ma, P. D. (2015). MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Mol. Plant* 8 (5), 689–708. doi: 10.1016/j.molp.2015.03.012
- Ma, S. A., Yang, Z. F., Wu, F. F., Fan, J. W., Dong, X. T., Hu, R. C., et al. (2022). R2R3-MYB gene family: Genome-wide identification provides insight to improve the content of proanthocyanidins in *Trifolium repens*. *Gene* 829, 13. doi: 10.1016/j.gene.2022.146523
- Matus, J. T., Aquea, F., and Arce-Johnson, P. (2008). Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across vitis and arabidopsis genomes. *BMC Plant Biol.* 8, 83. doi: 10.1186/1471-2229-8-83
- Millard, P. S., Kragelund, B. B., and Burow, M. (2019). R2R3 MYB transcription factors - functions outside the DNA-binding domain. *Trends Plant Sci.* 24 (10), 934–946. doi: 10.1016/j.tplants.2019.07.003
- Oshima, Y., Shikata, M., Koyama, T., Ohtsubo, N., Mitsuda, N., and Ohme-Takagi, M. (2013). MIXTA-like transcription factors and wax inducer1/shine1 coordinately regulate cuticle development in *Arabidopsis* and *Torenia fournieri*. *Plant Cell* 25 (5), 1609–1624. doi: 10.1105/tpc.113.110783

- 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 (3), 290–294. doi: 10.1038/nbt.3122
- Sanghera, G. S., Wani, S. H., Hussain, W., and Singh, N. B. (2011). Engineering cold stress tolerance in crop plants. *Curr. Genomics* 12 (1), 30. doi: 10.2174/138920211794520178
- Sankoff, D. (2001). Gene and genome duplication. *Curr. Opin. Genet. Dev.* 11 (6), 681–684. doi: 10.1016/S0959-437X(00)00253-7
- Sharma, R., Singh, G., Bhattacharya, S., and Singh, A. (2018). Comparative transcriptome meta-analysis of *Arabidopsis thaliana* under drought and cold stress. *PLoS One* 13 (9), 18. doi: 10.1371/journal.pone.0203266
- Song, H., Sun, J., and Yang, G. F. (2018). Comparative analysis of selection mode reveals different evolutionary rate and expression pattern in *Arachis duranensis* and *Arachis ipaensis* duplicated genes. *Plant Mol. Biol.* 98 (4–5), 349–361. doi: 10.1007/s11103-018-0784-z
- Stracke, R., Werber, M., and Weissshaar, B. (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* 4 (5), 447–456. doi: 10.1016/S1369-5266(00)00199-0
- Tang, J., Wang, F., Hou, X. L., Wang, Z., and Huang, Z. N. (2014). Genome-wide fractionation and identification of WRKY transcription factors in Chinese cabbage (*Brassica rapa ssp. pekinensis*) reveals collinearity and their expression patterns under abiotic and biotic stresses. *Plant Mol. Biol. Rep.* 32 (4), 781–795. doi: 10.1007/s11105-013-0672-2
- Tao, W. J., Bian, J. X., Tang, M. Q., Zeng, Y., Luo, R. H., Ke, Q. L., et al. (2022). Genomic insights into positive selection during barley domestication. *BMC Plant Biol.* 22 (1), 19. doi: 10.1186/s12870-022-03655-0
- Thornton, J. W., and DeSalle, R. (2000). Gene family evolution and homology: genomics meets phylogenetics. *Annu. Rev. Genomics Hum. Genet.* 1, 41–73. doi: 10.1146/annurev.genom.1.1.41
- Tombuloglu, H. (2020). Genome-wide identification and expression analysis of R2R3, 3R- and 4R-MYB transcription factors during lignin biosynthesis in flax (*Linum usitatissimum*). *Genomics* 112 (1), 782–795. doi: 10.1016/j.ygeno.2019.05.017
- Wang, Y. P., Tang, H. B., DeBarry, J. D., Tan, X., Li, J. P., Wang, X. Y., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40 (7), 14. doi: 10.1093/nar/gkr1293
- Wilkins, O., Nahal, H., Foong, J., Provart, N. J., and Campbell, M. M. (2009). Expansion and diversification of the populus R2R3-MYB family of transcription factors. *Plant Physiol.* 149 (2), 981–993. doi: 10.1104/pp.108.132795
- Wu, Y., Wen, J., Xia, Y. P., Zhang, L. S., and Du, H. (2022). Evolution and functional diversification of R2R3-MYB transcription factors in plants. *Horticulture Res.* 9, 16. doi: 10.1093/hr/uhac058
- Wu, W. H., Zhu, S., Zhu, L. M., Wang, D. D., Liu, Y., Liu, S. Q., et al. (2021). Characterization of the *liriodendron chinense* MYB gene family and its role in abiotic stress response. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.641280
- Yan, H. F., Li, M. Z., Xiong, Y. P., Wu, J. M., da Silva, J. A. T., and Ma, G. H. (2019). Genome-wide characterization, expression profile analysis of WRKY family genes in *Santalum album* and functional identification of their role in abiotic stress. *Int. J. Mol. Sci.* 20 (22), 19. doi: 10.3390/ijms20225676
- Yang, Y. T., Yuan, Z., Ning, C. H., Zhao, B. L., Wang, R. R., Zheng, X. L., et al. (2022). The pea R2R3-MYB gene family and its role in anthocyanin biosynthesis in flowers. *Front. Genet.* 13. doi: 10.3389/fgene.2022.936051
- Yao, C. Y., Li, X. G., Li, Y. M., Yang, G. H., Liu, W. D., Shao, B. T., et al. (2022a). Overexpression of a *Malus baccata* MYB transcription factor gene MbMYB4 increases cold and drought tolerance in *Arabidopsis thaliana*. *Int. J. Mol. Sci.* 23 (3), 18. doi: 10.3390/ijms23031794
- Yao, C. Y., Li, W. H., Liang, X. Q., Ren, C. K., Liu, W. D., Yang, G. H., et al. (2022b). Molecular cloning and characterization of MbMYB108, a *Malus baccata* MYB transcription factor gene, with functions in tolerance to cold and drought stress in transgenic *Arabidopsis thaliana*. *Int. J. Mol. Sci.* 23 (9), 18. doi: 10.3390/ijms23094846
- Yao, X. Y., Meng, F., Wu, L. P., Guo, X. H., Sun, Z. P., Jiang, W. M., et al. (2022). Genome-wide identification of R2R3-MYB family genes and gene response to stress in ginger. *Plant Genome* 12, 1–12. doi: 10.1002/tpg2.20258
- Yuan, Y., Yang, X. P., Feng, M. F., Ding, H. Y., Khan, M. T., Zhang, J. S., et al. (2021). Genome-wide analysis of R2R3-MYB transcription factors family in the autopolyploid *Saccharum spontaneum*: an exploration of dominance expression and stress response. *BMC Genomics* 22 (1), 8. doi: 10.1186/s12864-021-07689-w
- Zeng, Z. C., Zhang, S. C., Li, W. Y., Chen, B. S., and Li, W. L. (2022). Gene-coexpression network analysis identifies specific modules and hub genes related to cold stress in rice. *BMC Genomics* 23 (1), 18. doi: 10.1186/s12864-022-08438-3
- Zhang, X. H., da Silva, J. A. T., Niu, M. Y., Li, M. Z., He, C. M., Zhao, J. H., et al. (2017). Physiological and transcriptomic analyses reveal a response mechanism to cold stress in *Santalum album* l. leaves. *Sci. Rep.* 7, 18. doi: 10.1038/srep42165
- Zhang, Z., Li, J., Zhao, X.-Q., Wang, J., Wong, G. K.-S., and Yu, J. (2006). KaKs_Calculator: calculating ka and ks through model selection and model averaging. *Genomics Proteomics Bioinf.* 4 (4), 259–263. doi: 10.1016/s1672-0229(07)60007-2
- Zhang, Z. J., Zhang, L., Liu, Y., Shang, X. L., and Fang, S. Z. (2022). Identification and expression analysis of R2R3-MYB family genes associated with salt tolerance in *Cyclocarya paliurus*. *Int. J. Mol. Sci.* 23 (7), 18. doi: 10.3390/ijms23073429