BMC Genomics

RESEARCH

Open Access

Genome-wide identification and expression profiling of MYB transcription factors in *Artemisia argyi*

Sainan Peng[1], Qianwen Zhang[1], Yun Yang[1], Yukun Li[1], Wunian Feng[1], Dan Zhao[1], Bisheng Huang[1,2], Dahui Liu[1,2*] and Yuhuan Miao[1,2*]

Abstract

Artemisia argyi, a significant medicinal plant in China, is known for its high content of essential oils, flavonoids, and other bioactive compounds. MYB transcription factors are the largest gene family in plants and are widely reported to play important roles in plant development, metabolism, defense, and stress resistance. However, the MYB family of *A. argyi* has not been systematically studied. The aim of this study was to comprehensively analyze the MYB gene family of *A. argyi* and explore its potential role in flavonoid biosynthesis. Here, the phylogeny, chromosome location, gene structure, *cis*-acting elements, expression patterns and Gene ontology (GO) annotation of MYB gene family members were investigated using bioinformatics methods based on the whole-genome and transcriptome data of *A. argyi*. In total, 227 *AYMYB* transcription factors were identified from *A. argyi* genome, including 22 *1R-MYB*, 165 *R2R3-MYB*, 16 *3R-MYB*, 5 *4R-MYB* and 19 atypical *MYB* members. These *AYMYBs* were unevenly distributed across the *A. argyi* genome. Subcellular localization prediction revealed that all the AYMYBs were localized in the nucleus. The protein motifs, conserved domains, and gene structures of *AYMYBs* were identified, and the results showed that *AYMYB*s from the same subfamily exhibited similar motifs and gene structures. *Cis*-acting elements and GO analysis suggested that *AYMYB*s may be involved in many biological processes related to plant development, metabolism, defense, and stress resistance. Moreover, quantitative real-time PCR (qRT-PCR) analysis showed that approximately 50 genes showed high expression levels in the leaves of *A. argyi* and *AYMYBs* showed specific expression patterns under MeJA treatment. Together, our research will offer useful information for future investigations into the functions of *MYB* genes in *A. argyi*, especially in regulating the process of flavonoid biosynthesis in leaves and in response to MeJA treatment.

Keywords *Artemisia argyi*, *AYMYB*, Bioinformatics tools, Expression pattern, MeJA

*Correspondence:
Dahui Liu
liudahui@hbucm.edu.cn
Yuhuan Miao
miaoyh@hbucm.edu.cn
[1]School of Pharmacy, Hubei University of Chinese Medicine, Wuhan 430065, China
[2]Hubei Shizhen Laboratory, Hubei University of Chinese Medicine, Wuhan 430065, China

Introduction

Transcription factors bind specific motifs on the target gene promoter to regulate gene expression. The MYB family is one of the largest families of transcription factors in plants and is characterized by a highly conserved DNA-binding domain, known as the MYB domain. This domain usually consists of 1 to 4 distinct tandem repeats. Each repeat sequence consists of 50–52 amino acids containing three conserved tryptophan residues arranged at 18–19 amino acid intervals and involved in forming hydrophobic nuclei in the spatial structure. Depending on the number of repeats of the MYB domain, MYB proteins can be classified into 1R-MYB (R1/2, R3-MYB), 2R-MYB (R2R3-MYB), 3R-MYB (R1R2R3-MYB) and 4R-MYB [1]. The functions of MYB proteins have been extensively studied, revealing that different types of MYBs exhibit distinct functions. Numerous studies have examined the roles of MYB transcription factors in plants [2]. 1R-MYB proteins are mainly involved in regulating cell morphology, root epidermal patterning, biological clock regulation, and other life processes [3–4]. *MaMYBR30*, a novel 1R-MYB, can positively regulate the resistance of mulberry to drought and play an important role in flower development and reproductive growth [5]. The largest class of plant MYB transcription factors is R2R3-MYB, which contains two MYB domain repeats. The main function of R2R3-MYB protein is to regulate plant development and metabolism. They can control phenylpropanoid biosynthesis, glucosinolate biosynthesis, plant hormone- and pathogen-mediated stress responses, and organ determination, including determining cell shape and the formation of root hairs and trichosomes [6–7]. For instance, *AaMYB108* positively regulates the artemisinin biosynthesis in *A. annua* [8]. Overexpression of *RcMYB1* significantly enhances anthocyanin accumulation in rose petals. Studies have found that *RcMYB1* could widely regulate the expression of genes related to anthocyanin biosynthesis, indicating its role in regulating anthocyanin synthesis in roses [9]. When *MdMYB3* was ectopically expressed in *Nicotiana tobacum.*, the transgenic lines showed pigmentation and higher anthocyanin and flavonol contents than the control lines. Furthermore, the peduncles of flowers and styles of pistils in transgenic lines are longer than those of wild-type plants [10]. 3R-MYB proteins can recognize the M phase-specific activator (MSA) motif contained by many plant cell cycle genes in their regulatory regions [11–12]. Overexpression of *AbMYB11* in *Arabidopsis* significantly increases the transcription of phenylpropanoid-related genes, increases the accumulation of total phenolics and flavonoids, and improves resistance to *P. tolaasii* [13]. 4R-MYB proteins are the smallest class with limited information available about their function.

Artemisia argyi Lévl. et Vant. (*A. argyi*), belonging to the *Asteraceae* family, is a medicinal plant that has been used in traditional Chinese medicine for thousands of years. The high medicinal and economic value of *A. argyi* is attributed to the presence of a large number of secondary metabolites, including volatile oils, flavonoids, terpenoids, phenolic acids, and other bioactive compounds [14]. Terpenoids (e.g., eucalyptol, thujone, borneol) and flavonoids (e.g., eupatilin, jaceosidin) are the main material foundation for the pharmacological effects [15]. However, the transcriptional regulatory mechanisms of the biosynthesis of the active compounds have not been reported. To date, many studies have identified the MYB family members in various plants [16–20]. However, a systematic analysis of the MYB transcription factor family in *A. argyi* has not been reported. Based on the whole-genome-wide study of *A. argyi*, we identified 227 members of the *MYB* family and analyzed these *MYBs* through bioinformatics analyses. We also predicted the functions of these *MYB* genes by analyzing the *cis*-acting elements, GO analysis, and expression profiles. We screened candidate *MYB* genes that may be involved in regulating the biosynthesis of flavonoids. This analysis will provide a theoretical basis for studying the regulatory mechanisms of MYB transcription factors in *A. argyi*.

Materials and methods

Plant materials

The *A. argyi* plant material we used was the 'Xiang Ai' cultivar. This cultivar was kept in the Medicinal Botanical Garden of Hubei University of Traditional Chinese Medicine in Wuhan, China.

Identification, characteristics and phylogenetic analysis of *AYMYBs* in *A. argyi*

Protein sequences of MYB transcription factors from *Arabidopsis thaliana* (AtMYB) were obtained from the TAIR database (https://www.arabidopsis.org/). Based on these protein sequences of the *Arabidopsis* MYB family, all possible MYB protein sequences from the *A. argyi* genome were identified using TBtools software (v2.096) with a BLASTP search under default parameters (e-value ≤ 1e-5) [21]. The reference genome could be obtained in the National Genomics Data Center under the accession number PRJNA804646. The conserved Myb_DNA-binding domain (PF00249) was obtained from the Pfam protein family database (http://pfam.sanger.ac.uk/) and this HMM file was then employed to identify the MYB proteins through the HMMER program (e-value ≤ 1e-5) with the aim of eliminating sequences lacking this conserved domain [22]. Cd-hit (Cluster Database at High Identity with Tolerance) was used to remove highly homologous redundant sequences [23], with the parameters configured to cluster sequences at a

90% identity threshold, producing a set of non-redundant sequences known as *AYMYBs*. Subsequently, the online tool ProtParam (https://wcb.expasy.org/protparam) was utilized to analyze the physicochemical properties of the predicted AYMYB proteins [24]. Plant-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/#) was used for subcellular localization prediction analysis. SMART (https://smart.embl.de/) was employed to analyze the repeat of MYB domains. We first aligned the protein sequences of 227 AYMYB proteins and 38 MYB proteins from other species by using the ClustalW program. Subsequently, the Neighbor-Joining (NJ) method is simple in algorithm and fast in computation, we constructed a phylogenetic tree using the neighbor-joining method [25] in MEGA 7.0 with 1000 bootstrap replicates [26]. Lastly, we enhanced the visualization of the phylogenetic tree using the ChiPlot website (https://www.chiplot.online/index.html#) [27].

Chromosomal localization analysis and syntenic analysis

Based on the previous genome annotation and structural files (GFF files and genome files) of *A. argyi*, we utilized TBtools software (version 2.096) to determine the chromosomal locations of all *MYB* genes in *A. argyi*. Concurrently, we employed the Gene Density Profile in TBtools to visualize gene density on chromosomes. In addition, we performed a syntenic analysis of *AYMYBs* using the MCScanX tool, obtained the relevant collinearity file, and visualized it using the Advanced Circos tool. The parameters of all tools were set to default values.

Intron-exon, conserved motifs and domain analysis

The Conserved Domain Database (https://www.ncbi.nlm.nih.gov/cdd/?term=) at NCBI was used to identify the conserved domains of AYMYB proteins. The Simple MEME Wrapper in TBtools was employed to find motifs (number 10, default parameters). The results were visualized using the Gene Structure View tool in TBtools.

Cis-regulatory features and GO annotation of *AYMYBs*

To analyze the *cis*-acting elements on the promoter regions of all identified putative *AYMYBs*, a 2000 bp (bp) region upstream from the transcription start site was extracted from the genome data and subjected to PlantCARE (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) for analysis. To further explore the biological processes involved with AYMYB proteins, the GO annotation of 227 *AYMYB* genes was analyzed using GO Enrichment at Majorbio (https://www.majorbio.com/).

Expression pattern analysis of *AYMYBs* in *A. argyi*

Samples treatment and collection

The young leaves from 90-day-old 'Xiang Ai' plants were used to construct the reference genome. Roots (R), rhizomes (Rh), stems (S), leaf buds (5 days, Leaf A), young leaves (15 days, Leaf B), mature leaves (30 days, Leaf C) and old leaves (45 days, Leaf D) of 'Xiang Ai' were collected. Additionally, young leaves treated with 300 μmol/L MeJA were collected at 0, 6, 12, 24 and 48 h post-treatment for RNA-seq analysis. All samples for RNA-seq and quantitative real-time PCR (qRT-PCR) analyses were stored at -80℃ in triplicate.

RNA extraction, cDNA library construction, and sequencing

In this experiment, total RNA was extracted from the tissue using TRIzol® Reagent according to the manufacturer's instructions. Then RNA quality was determined by 5300 Bioanalyser (Agilent) and quantified using the ND-2000 (NanoDrop Technologies). And then the RNA sample was used to construct sequencing library. RNA purification, reverse transcription, library construction and a sequencing were performed at Shanghai Majorbio Bio-pharm Biotechnology Co., Ltd. (Shanghai, China) according to the manufacturer's instructions. The RNA-seq transcriptome library was prepared following Illumina® Stranded mRNA Prep, Ligation (San Diego, CA) using 1 μg of total RNA. Shortly, messenger RNA was isolated according to the polyA selection method by oligo(dT) beads and then fragmented by fragmentation buffer first. Secondly, double-stranded cDNA was synthesized with random hexamer primers. Then the synthesized cDNA was subjected to end-repair, phosphorylation and adapter addition according to library construction protocol. Libraries were size selected for cDNA target fragments of 300–400 bp use magnetic beads followed by PCR amplified for 10–15 PCR cycles. After quantified by Qubit 4.0, the sequencing library was performed on NovaSeq X Plus platform (PE150) using NovaSeq Reagent Kit.

Expression pattern analysis of AYMYBs in A. argyi

The whole genome of *A. argyi* was used as the reference genome for the transcriptome sequencing data, and the Bowtie2 (Version 2.4.1) software was used for sequencing alignment. Then using the RSEM (Version 1.3.3) to obtain the number of Reads for each sample compared to each transcript, and the average expression level (Fragments Per Kilobase per Million bases FPKM) conversion was calculated and performed for each transcribed region [28]. Essentially, differential expression analysis was performed using the DESeq2 [29]. DEGs with $|\log2FC| \geqq 1$ and FDR<0.05(DESeq2) or FDR < 0.001(DEGseq) were considered to be significantly different expressed genes.

Real-time fluorescence quantitative PCR

To examine the tissue expression and MeJA induction expression profiles of the *AYMYB* genes, we analyzed their expression patterns using RNA-seq data obtained

from seven different plant tissues (Root, Rhizome, Stem, Leaf A, Leaf B, Leaf C, Leaf D) and leaves treated with 300 μmol/L MeJA. Total RNA was extracted from these samples using the RNA prep Pure kit (DP441, TIANGEN). First-strand cDNAs were synthesized from 2 μg of total RNA using the ABScript Neo RT Master Mix for qPCR with gDNA Remover (RK20433, ABclonal), followed by qRT-PCR analysis. In order to validate the transcriptome data, ten *AYMYB* genes were selected for qRT-PCR analysis. The relative expression levels of *AYMYB* genes were calculated using the $2^{-\Delta\Delta Ct}$ method [30]. *Actin* was used as a reference gene [31]. The standard qRT-PCR was performed using a SYBR Green Premix Pro *Taq* HS qPCR Kit (ACCURATE BIOTECHNOLOGY(HUNAN) CO.,LTD, ChangSha, China) with at least three replicates for each gene. Gene-specific primers used in qRT-PCR experiment were designed using the Primer Premier 5 tool (Version 5.00) (www.PremierBiosoft.com).

Results

Identification of the *AYMYB* family in *A. argyi*

We performed a BLASTP search against the *A. argyi* genome database using AtMYB protein sequences as a query. To ensure the accuracy of results, we also employed the MYB domain (PF00249) from the Pfam database as a query in an HMM search against the genome. After, redundant sequences were removed using the CD-HIT tool. A total of 227 protein sequences containing MYB or MYB-like repeats were obtained. Based on the number of MYB domain repeats, these transcription factors were categorized into four typical classes, namely *1R-MYB* (22 genes), *R2R3-MYB* (165 genes), *3R-MYB* (16 genes), *4R-MYB* (5 genes) and some atypical *MYB* genes (19 genes). The amino acid number of AYMYB proteins varied from 96 aa (AY029171-RA) to 1,057 aa (AY211561-RA), and the molecular weight of AYMYB proteins ranged from 11.01864 kDa to 120.1428 kDa. The pI values of 111 MYB members were less than seven, and the other 116 proteins were higher than seven. Only seven AYMYB proteins had an instability index below 40 (34.69–39.8), suggesting that most AYMYB proteins were unstable. The aliphatic index of AYMYB proteins ranged from 52.6 to 96.92, indicating its high aliphatic amino acid content and stability over a wide temperature range. The Grand Average of Hydropathicity (GRAVY) for all AYMYB proteins was below 0, showing that AYMYBs were hydrophilic proteins. All AYMYB proteins were predicted to be localized in the nucleus, which was consistent with the known localization characteristics of transcription factors (Table S1).

This study conducted an evolutionary analysis of *AYMYB* genes using the NJ method. To comprehensively analyze the *AYMYB*s, We selected 38 well-characterized *MYB* genes from different plant species as references (Table S2). The evolutionary analysis revealed that the 227 AYMYB proteins could be divided into ten clusters (C1-C10) (Fig. 1). Among the 227 AYMYBs, C1, C3 and C8 clusters accounted for the largest protein numbers with 34 AYMYBs each, followed by C2, C4, C5, C6, C7, C9 and C10 clusters, with 23, 27, 6, 28, 26, 4 and 11 members, respectively. Interestingly, C7 clusters had the highest number of atypical MYB proteins.

Chromosome location, gene duplication events and syntenic analysis of *AYMYBs*

To identify the distribution of *AYMYB*s on the *A. argyi* chromosomes, we mapped the location of *AYMYB*s using the GFF file of the *A. argyi* genome. The results indicated that *AYMYB*s were unevenly distributed across the 34 chromosomes in the *A. argyi* genome (Fig. 2A). Most of the *AYMYB*s were located on chromosomes 2, 4, 5, 6, 7 and 10 and a few genes were located on chromosomes 1, 3, 8 and 9. Notably, no genes were located on chromosome 1–1.

In order to clarify the evolutionary relationships of MYB family, genome-wide syntenic analysis of *AYMYBs* was performed in *A. argyi* genome. We employed the MCScanX and Advanced Circos to visualize the duplications within the *A. argyi* genome. The results showed that 668 syntenic gene pairs were identified in *A. argyi* (Fig. 2B). There were 15 gene pairs predicted to be derived from dispersed duplication, 9 gene pairs have proximal duplication origins, 7 gene pairs have tandem duplication origins and 196 gene pairs have whole-genome duplication (WGD) or segmental duplication origins (Table S3). These results suggest that WGD or segmental duplications may be the key driving force of MYB family expansion in *A. argyi.*

Conserved motifs, domain architectures, and gene structures of AYMYB proteins

In the 227 AYMYB proteins, ten conserved motifs with a length ranging from 11 to 41 amino acids (Table S4) were found using MEME software. Motif numbers, types, and orders were generally similar among AYMYB members in the same subfamily. For example, many R2R3-MYB subfamily members possessed motifs 3, 4, 7, 8, 9 and 10, which were located on the *N*-terminal of the protein. Atypical MYBs had the largest number of motifs, usually containing more than 10 motifs (Fig. 3A and B). All the AYMYB proteins contained the Myb_DNA-binding, PLN03091, REB1, and SANT domains (Fig. 3C). We also analyzed the gene structure of *AYMYBs* (Fig. 3D). The results showed that the number of exons in *AYMYB* genes ranged from 1 to 14, and the number of introns ranged from 0 to 13. Notably, *AYMYBs* in the same subfamily usually had similar gene structures. However, the number of gene exons in the same subfamily could vary

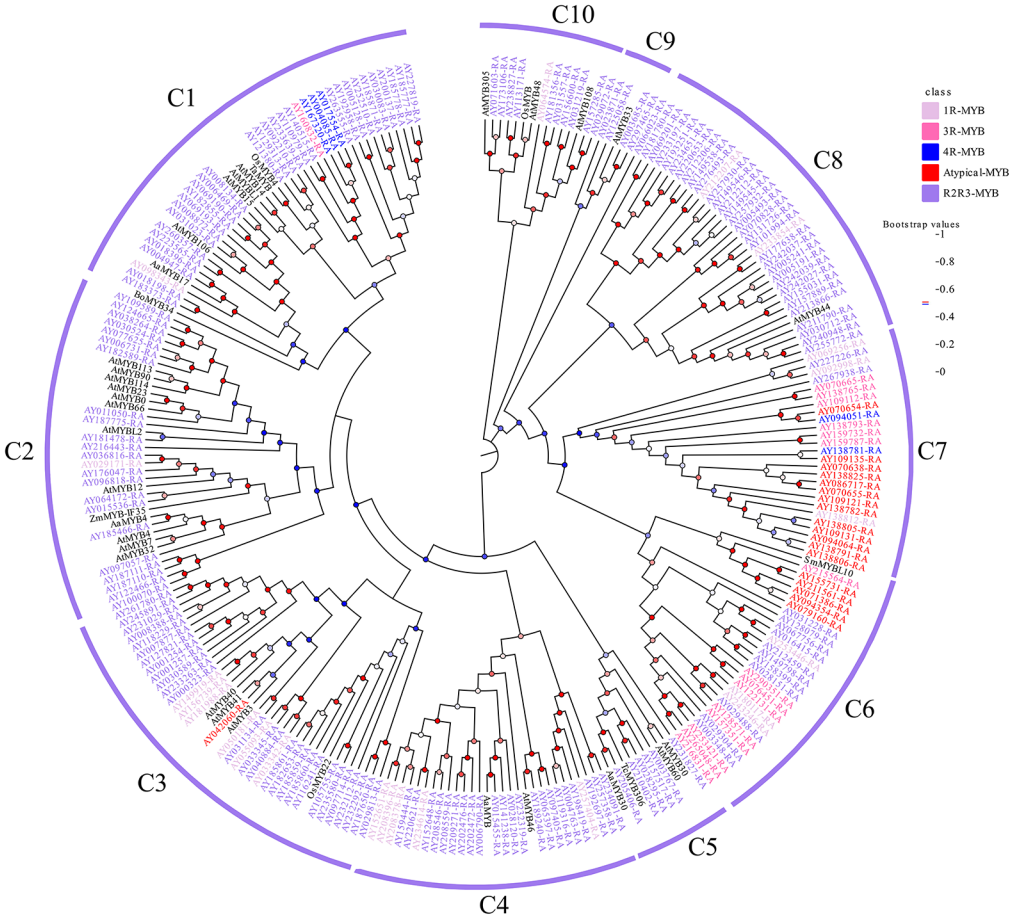

Fig. 1 Phylogenetic relationships of MYB proteins were analyzed between *A. argyi* and other species, such as *Arabidopsis thaliana*, *Artemisia annua*, *Oryza sativa*, *Zea mays*, *Triticum aestivum*, *Salvia miltiorrhiza*, *Brassica oleracea*, and *Tanacetum cinerariifolium*. The phylogenetic tree was constructed using MEGA 7.0, representing different classes in distinct colors

greatly. For example, *AY149798-RA*, which belonged to the R2R3-MYB subfamily, generally contained 12 exons. However, a few genes, such as *AY149538-RA* and *AY030712-RA*, contained only 2 or 1 exons, respectively. These results suggested that exon loss and acquisition may have occurred during the evolution of the *AYMYB*s in *A. argyi*.

Cis-acting element analysis of *AYMYBs* promoter regions

The promoter sequences of all 227 *AYMYB*s were extracted using TBtools, and *cis*-acting elements within these regions were predicted using the PlantCARE tool. A total of 24,654 *cis*-acting elements in the promoter region of *AYMYBs* were identified (Table S5 and Figure S1). TATA-box and CAAT-box were the most widely present elements in *AYMYB* promoters, accounting for 51.33% and 14%, respectively (Fig. 4A). We classified these *cis*-acting elements according to their functions and found that *AYMYBs* promoters mainly included the following elements: (1) essential promoter elements; (2) light response-related elements, including TCT-motif, G-Box, Box 4, GT1-motif, MRE, AE-box, ACE, and AT1-motif; (3) hormone response-related elements, containing CGTCA-motif, TCA-element, P-box, TGA-element, TATC-box, ABRE, AuxRR-core, and TGA-box; (4) biotic and abiotic stress-related elements, including MBS, ARE, LTR, and TC-rich repeats; (5) development and tissue specificity-related elements, including *cis*-acting regulatory element related to meristem expression, seed-specific regulation, and circadian control (Fig. 4B). *Cis*-acting elements play a vital role in the regulatory networks. Our findings indicated the potential regulatory functions of *AYMYBs*.

Gene ontology analysis of AYMYB proteins

To further determine the specific biological processes in which the *AYMYB*s are involved, we performed a Gene Ontology (GO) enrichment analysis. We utilized the Megorbio bioinformatics platform to functionally annotate the 227 *AYMYB* genes in *A. argyi*. Genes were classified based on biological processes, molecular functions, and cellular component categories. These proteins were

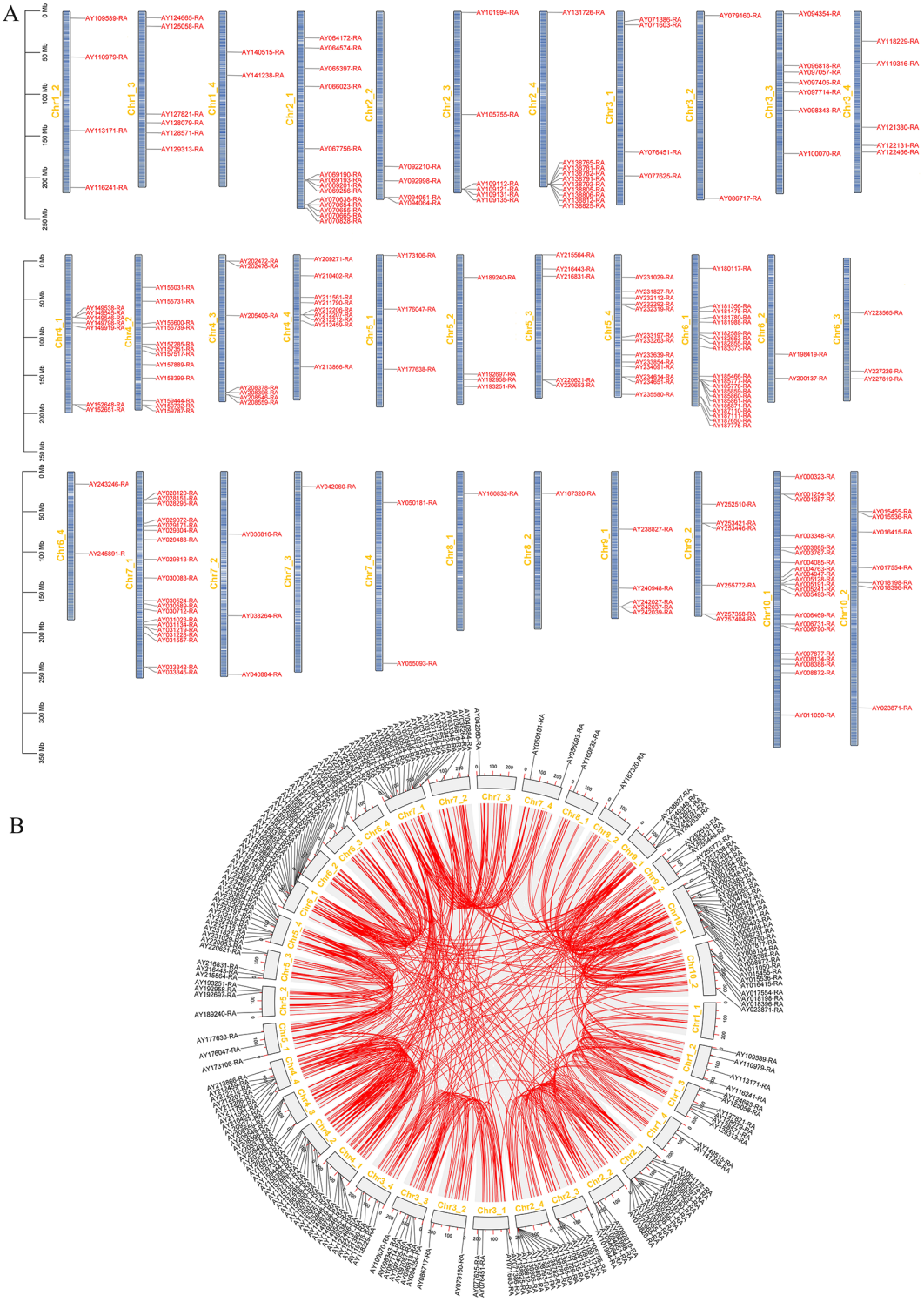

Fig. 2 Chromosomal distribution and collinearity analysis of *AYMYB*s. **A**. Chromosomal distribution of each *AYMYB* gene. **B**. Chromosomal relationships of *AYMYB*s. The gray blocks denoted the segments of *A. argyi* chromosomes. The gray lines in the background represent collinear blocks within the genome of *A. argyi*, while the red lines indicate syntenic *AYMYB* gene pairs. Here, 'Chr' denotes the chromosome

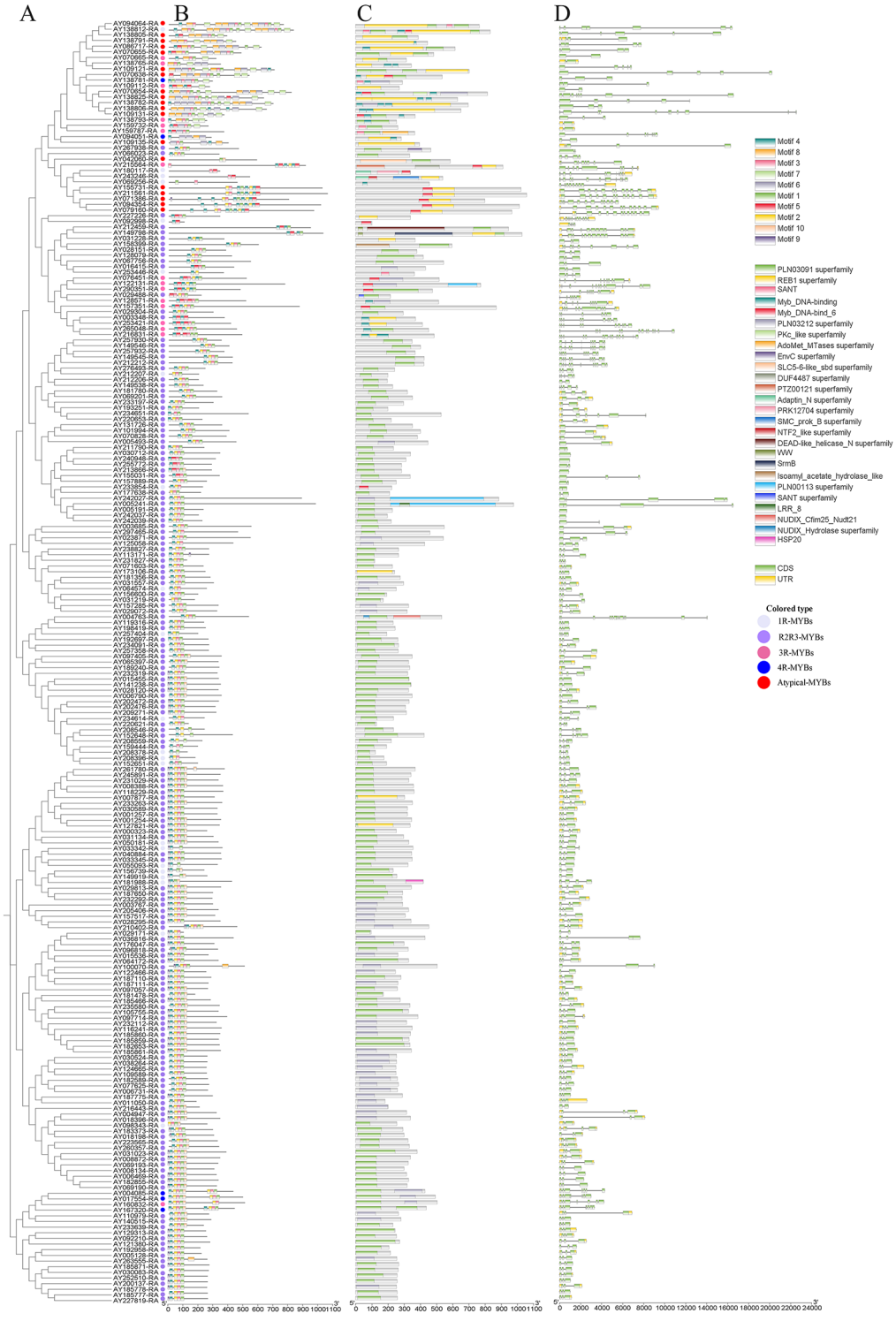

Fig. 3 Sequence analysis of *AYMYBs*. Phylogenetic relationships (**A**), conserved motifs (**B**), domains (**C**), and exon/intron organization (**D**) of AYMYB proteins or genes from *A. argyi*. UTR and CDS indicate non-translated regions and coding sequences, respectively

separated into 59 categories based on biological processes (Table S6). A total of 34 genes were identified to be involved in transcriptional regulation. Additionally, 27 genes were found to participate in the regulation of macromolecule biosynthetic processes as well as in biosynthetic process regulation. Furthermore, 32 genes were associated with transcription *cis*-regulatory region binding. Lastly, 13 genes were shown to be involved in the

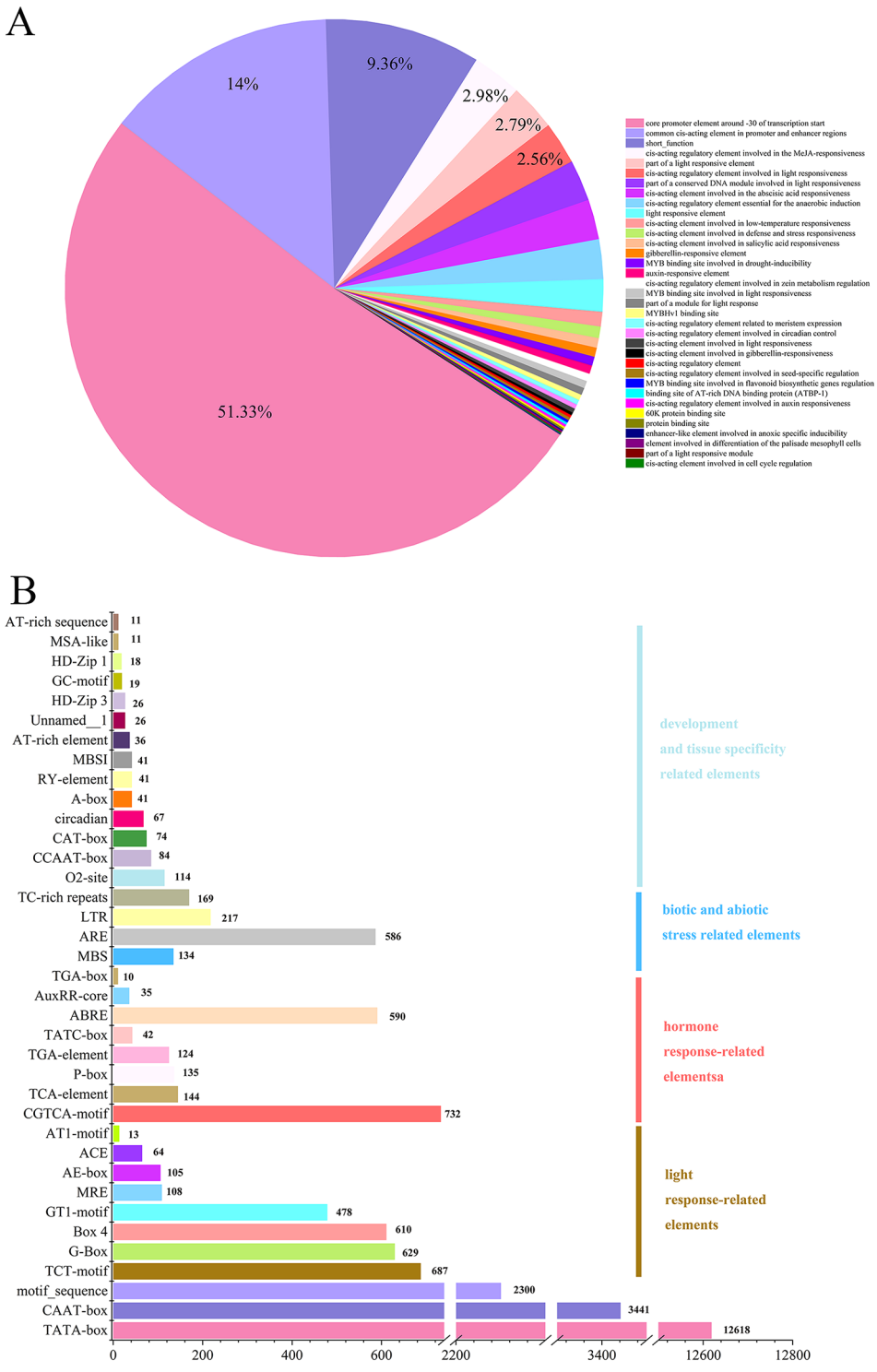

Fig. 4 *Cis*-acting element analysis of *AYMYBs* promoter regions. **A**. The proportion of *cis*-elements predicted in the promoters of *AYMYBs*. **B**. Numbers of *cis*-elements involved in light response, hormone response, biotic and abiotic stress, development, and tissue specificity

green leaf volatile biosynthetic process (Fig. 5). These results suggest that the *AYMYB* genes may play significant roles in transcriptional regulatory networks, providing evidence for the identification of candidate genes for further study.

Expression pattern analysis of *AYMYBs* in *A. argyi*

The expression patterns of *AYMYB* family members were characterized based on RNA-seq data. Out of 227 *AYMYB* genes analyzed, 142 were found to be expressed in seven distinct tissues (Root, Rhizome, Stem, Leaf A,

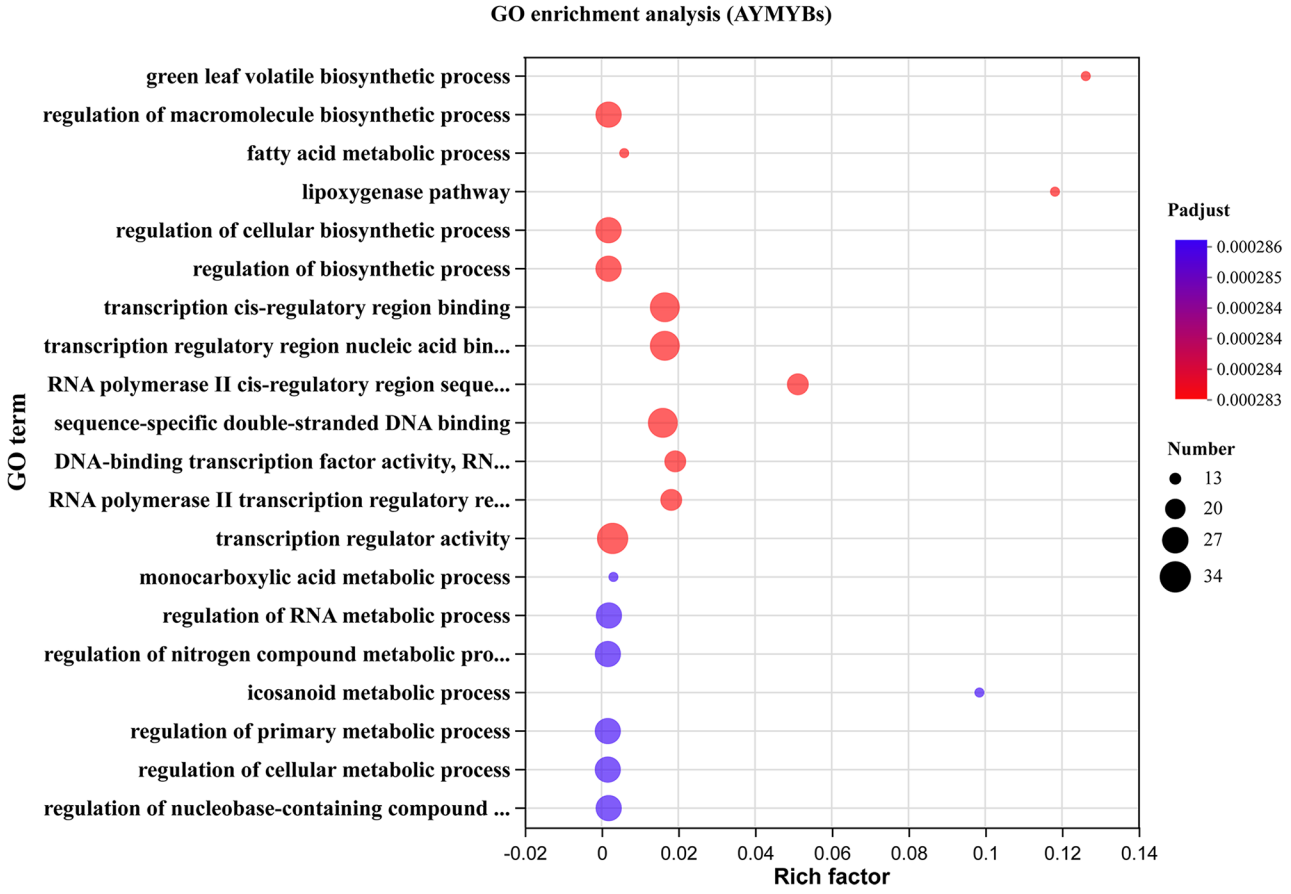

Fig. 5 GO analysis of *AYMYBs* in *A. argyi*

Leaf B, Leaf C, Leaf D). The majority of *AYMYB* genes exhibited specific expression patterns in the root, while approximately 50 genes showed high expression levels in the leaves of *A. argyi* (Fig. 6A). In addition, the expression of most *AYMYBs* was significantly increased at 6 and 12 h after MeJA treatment (Fig. 6B). We also found that *AYMYB*s belonging to the same branch in the evolutionary tree exhibited similar expression profiles. However, the expression patterns of *AYMYB*s did not significantly differ across categories (Figure S2).

To further validate the gene expression data obtained from transcriptome sequencing, we employed qRT-PCR to analyze the expression levels of ten selected *AYMYB*s in different tissues and MeJA treatments. We first examined the expression of ten selected *AYMYB*s in six different tissues (Root, Rhizome, Stem, Leaf A, Leaf B, Leaf C). A total of ten genes, including *AY121380-RA*, *AY003767-RA*, *AY015536-RA*, *AY030712-RA*, *AY187110-RA*, *AY205406-RA*, *AY211790-RA*, *AY267938-RA*, *AY240948-RA* and *AY141238-RA*, were selected for analysis. The primer information for qRT-PCR can be found in Table S7. The qRT-PCR results demonstrated that *AY030712-RA*, *AY211790-RA*, *AY240948-RA*, *AY121380-RA*, and *AY141238-RA* were mainly expressed in the roots, rhizomes, and stems. In contrast, *AY003767-RA*, *AY015536-RA*, and *AY205406-RA* exhibited high expression levels in the leaves. Meanwhile, the expression of *AY187110-RA* and *AY267938-RA* did not display significant tissue specificity (Fig. 7). We also analyzed the expression patterns of the above ten *AYMYB* genes following MeJA treatment. The results showed that all ten genes were up-regulated in response to MeJA treatment (Fig. 8). The expression patterns of the ten selected genes, as analyzed by qRT-PCR, were in accordance with the transcriptome data, suggesting the high reliability of the transcriptome data. Consequently, we identified several candidate genes based on the expression patterns, including *AY030712-RA* (*AYMYB1*), *AY267938-RA* (*AYMYB3*), *AY015455-RA* (*AYMYB5*), *AY255772-RA* (*AYMYB8*), *AY141238-RA* (*AYMYB13*), *AY187110-RA* (*AYMYBa*), *AY003767-RA* (*AYMYBe*), *AY015536-RA* (*AYMYBf*), *AY185466-RA* (*AYMYBh*), among others.

Discussion

The MYB family, one of the largest transcription factor families in plants, possesses a wide array of biological functions. *MYB* transcription factors play important roles in regulating primary and secondary metabolism,

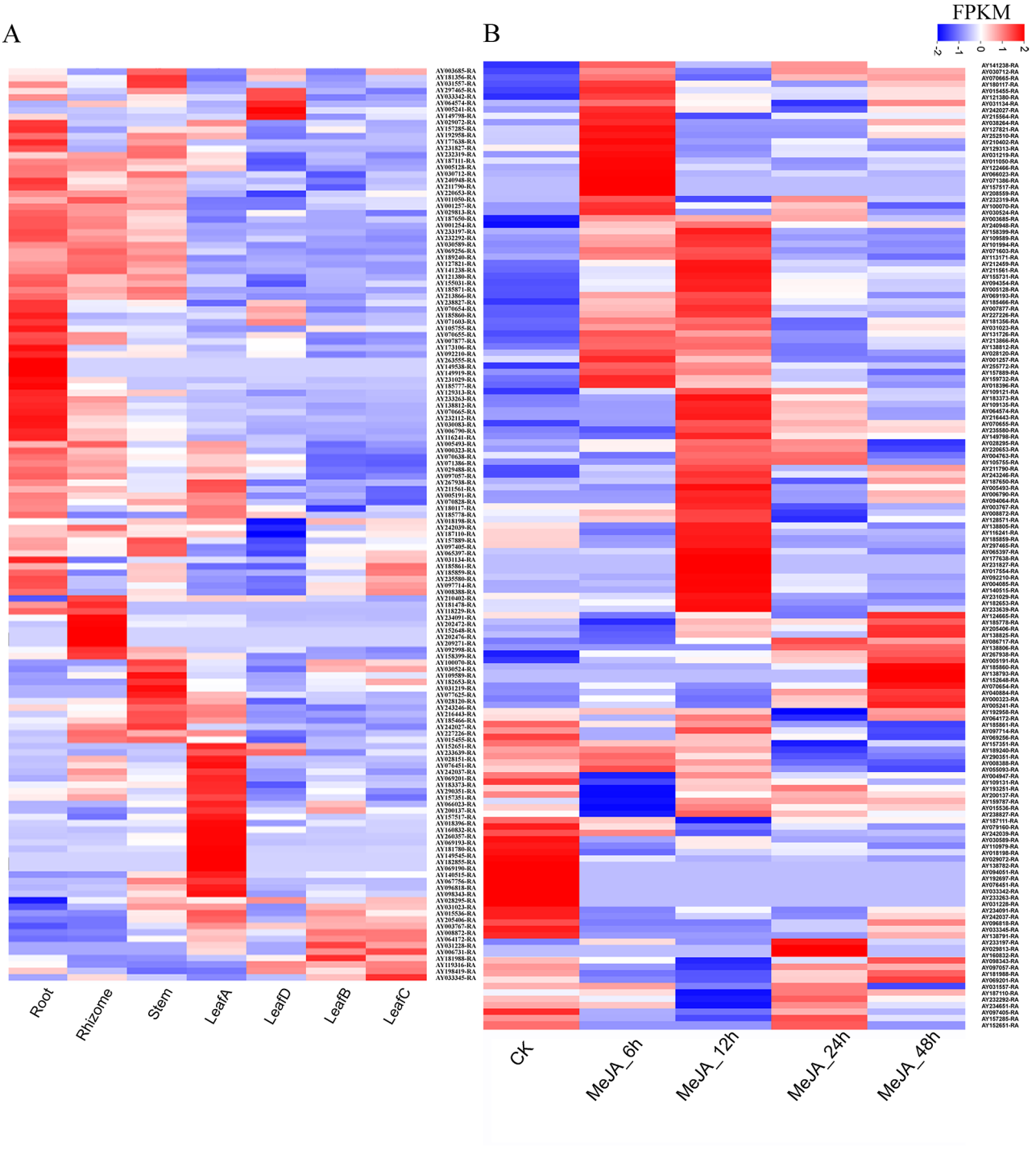

Fig. 6 Expression pattern analysis of *AYMYBs* in *A. argyi*. **A**. The expression profiles of *AYMYBs* in different tissues of *A. argyi*.**B**. The expression profiles of *AYMYB* genes under MeJA treatment. The color scale represents log10 expression values, with blue representing low expression and red indicating high expression. Leaf A represents 5 days of leaf buds; Leaf B represents 15 days of young leaves; Leaf C represents 30 days of mature leaves; Leaf D represents 45 days of old leaves

cell morphogenesis, plant growth and development, and responses to biotic and abiotic stresses [1]. Due to the diverse physiological roles of MYB transcription factors, this study performed basic bioinformatic analyses on the 227 identified *AYMYB* genes. We analyzed the physicochemical properties, conserved motifs, domains, gene structures, and key *cis*-acting elements of these genes. To date, genome-wide analyses of the MYB family have been conducted in several monocot and dicot plants. This article identifies 19 atypical MYB proteins

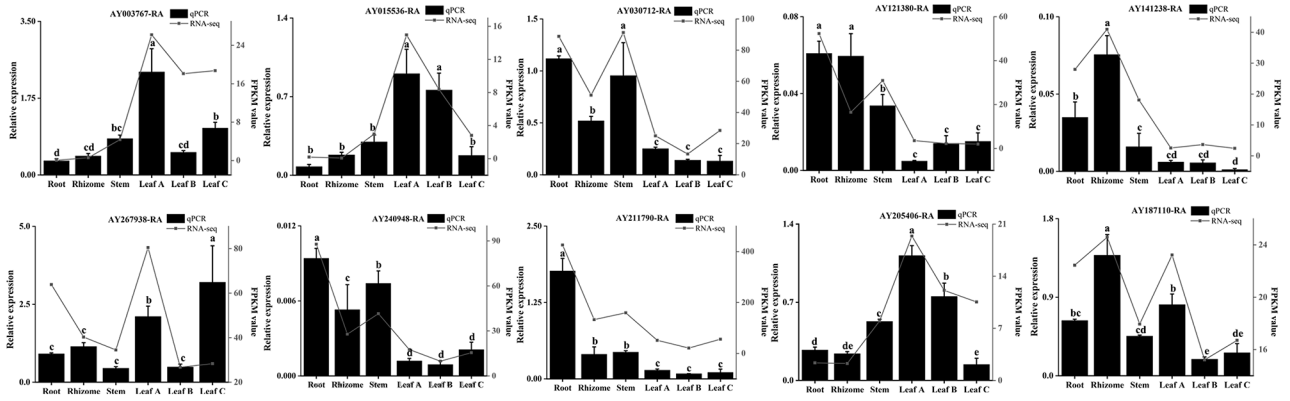

Fig. 7 qRT-PCR analysis of relative expression of *AYMYB*s in different *A. argyi* tissues. Leaf A represents 5 days of leaf buds; Leaf B represents 15 days of young leaves; Leaf C represents 30 days of mature leaves. Vertical bars indicate the mean ± SD calculated from three replicates. Different letters indicate significant differences among treatments in the same period using Duncan's test ($P < 0.05$)

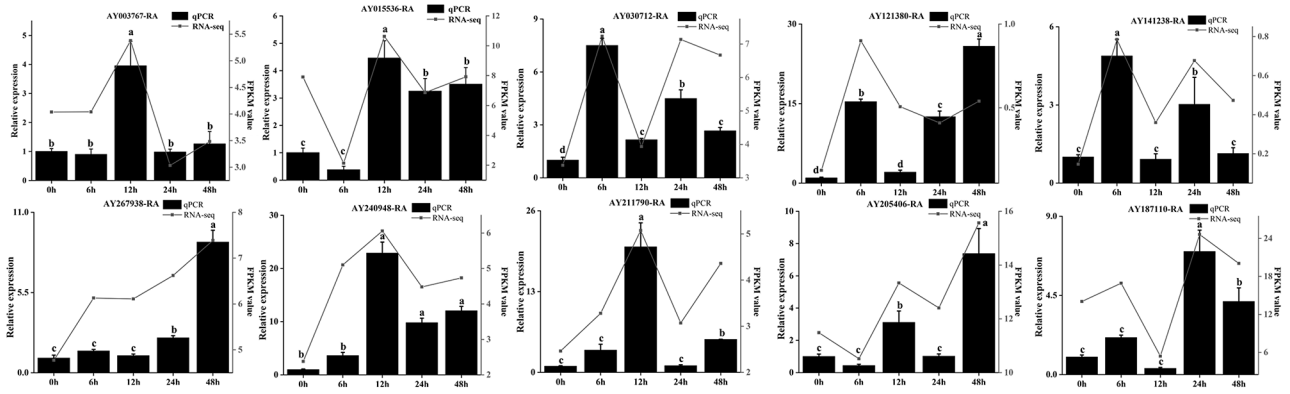

Fig. 8 qRT-PCR analysis of relative expression of *AYMYB*s under MeJA treatment of *A. argyi* at 0, 6, 12, 24 and 48 h. Vertical bars indicate the mean ± SD calculated from three replicates. Different letters indicate significant differences among treatments in the same period using Duncan's test ($P < 0.05$)

that contain more than four MYB domain repetitions. Additionally, five atypical *MYB* genes have been identified in *Chrysanthemum nankingense* [16], one atypical *MYB* gene in *Oryza sativa*, two atypical *MYB* genes in *Arabidopsis thaliana* [32], six atypical *MYB* genes in *Morella rubra* [33], and 14 atypical *MYB* genes in *Prunus persica* [34]. Although atypical *MYB* genes can be identified in these species, their numbers are lower than those identified in *A. argyi*. In addition, we found that the atypical MYB proteins had more conserved motifs and domains compared with other MYBs. Although there are very few reports about the function of atypical MYBs, this type of MYB will be worthy of further study in *A. argyi*.

Consistent with previous research, the R2R3-MYB subfamily is the largest subfamily of the MYB family. In our study, the number of identified *R2R3-MYB* genes was the largest, accounting for 72.68% of the total number of MYB families in *A. argyi*. Besides the large number of *R2R3-MYBs* in *A. argyi*, we also observed that the arrangement and composition of conserved domains within the R2R3-MYB subfamily members of the *A. argyi* genome are relatively conserved. This conservation is likely due to their widespread presence and essential functions in plants. In contrast, the 3R-MYB, 4R-MYB, and atypical-MYB subfamilies exhibit diverse compositions and arrangements of conserved domains, which may reflect their distinct roles in specific biological processes within plants [35]. In summary, The diversity of MYB transcription factors and the different combinations of their domains may be related to their various biological functions in *A. argyi*.

Gene duplication plays a vital role in the evolutionary process and the expansion of gene families, leading to changes in the number and function of duplicated genes, and is also essential for studying the inheritance and evolution of gene families. Tandem and segmental duplications are the primary mechanisms for the expansion of the gene family [36]. In this study, a total of 668 gene pairs of *AYMYBs* were identified by genome-wide syntenic analysis and most of these gene pairs were predicted to be derived from WGD or segmental duplications. Our previous study also showed that *A. argyi* has experienced two rounds of WGD events, and the most recent WGD event occurred 2.2 million years ago in *A.*

argyi [37]. This study suggested that WGD or segmental duplication events may be critical drivers for the expansion of the *MYB* family in *A. argyi*. Previous studies have demonstrated that these expansions and duplication events play a very important role in generating diversity, producing conserved subfamily members, and forming lineage-specific subfamilies with specific biological functions [38].

The flavonoids in *A. argyi* leaves are the main active components, which have anti-inflammatory, anti-tumor and other pharmacological effects. More than 50 kinds of flavonoids have been identified in *A. argyi* leaves. Among them, eupatilin, jaceosidin, apigenin, luteolin, quercetin, and naringin are representative components and have been proven to have biological activities. However, the transcriptional regulatory mechanisms of the biosynthesis of flavonoids in *A. argyi* have not been reported. In early studies, researchers found that MeJA could significantly enhance the content of terpenoids in *A. argyi* [39]. Additionally, in our study, we found a notable increase in the content of flavonoids in *A. argyi* following MeJA treatment. Moreover, it was found that exogenous MeJA treatment could activate the expression of *MYB* genes [40]. In this study, we also discovered that the expression of the majority of *AYMYB* genes was significantly up-regulated under MeJA treatment. Based on this, we hypothesized that *AYMYBs* with dominant expression in leaves and induced expression upon MeJA treatment might be involved in the synthesis of flavonoids.

Phylogenetic analysis showed that there are many *MYB* genes in *A. thaliana* involved in flavonoid biosynthesis in the C2 gene cluster of *A. argyi*. For example, *AtMYB12* has been shown to regulate the expression of key enzyme genes involved in flavonoid biosynthesis, leading to the increased accumulation of flavonoids by strongly activating the promoters of chalcone synthase (CHS), flavanone 3-hydroxylase (F3H), flavonol synthase (FLS) [41]. *AtMYB7* could repress several genes of the flavonoid pathway, DFR and UGT being early targets of this transcription factor which acts as a repressor of flavonol biosynthesis [42]. Therefore, it can be speculated that *AYMYB* genes in the C2 gene cluster may be involved in regulating flavonoid biosynthesis in *A. argyi*. Furthermore, among the 9 candidate genes, we found that *AY185466-RA* (*AYMYBh*) was closely related to *AaMYB4* in the evolutionary tree within the C2 gene cluster. The sequence similarity between the two genes is as high as 93.97%. *AaMYB4* was reported to play a role in mediating both artemisinin and flavonoid biosynthesis pathways by activating the expression of *AaADS* and *AaDBR2* in the artemisinin biosynthesis pathway, as well as *AaUFGT* in the flavonoid biosynthesis pathway [43]. The homologous comparison showed that the sequence similarity between *AY015536-RA (MYBf)* and *CmMYB3* gene was 93.92%. It has been reported that the CmMYB3 protein can bind to the promoters of the *CmCHI* and *CmFLS* genes, thereby enhancing their expression and increasing the flavonol content in *Chrysanthemums morifolium* [44].

In this study, we conducted a bioinformatics and expression analysis of the MYB family members in *A. argyi* and identified several key candidate genes. Furthermore, we plan to perform functional experiments on these important candidate genes to elucidate how *AYMYBs* respond to MeJA signals in regulating the accumulation of flavonoids and other compounds in *A. argyi*. Based on the findings of genome-wide analysis of the MYB family, this study will provide new insights into the role of *AYMYBs* in the biosynthesis of flavonoid of *A. argyi*, and provide a basis for improving the medicinal properties and germplasm innovation of excellent varieties in *A. argyi* in the future.

Conclusions

Here, we identified a total of 227 *AYMYB* genes in *A. argyi*, comprising 22 *1R-MYB*, 165 *R2R3-MYB*, 16 *3R-MYB*, 5 *4R-MYB*, and 19 atypical *AYMYB* members. The protein molecular characteristics, subcellular localization prediction, and chromosomal localization were accomplished using various bioinformatics tools. The protein motifs, conserved domains, and gene structures of *AYMYB* genes were analyzed, indicating that most *AYMYB* genes share similar motifs and gene structures. *Cis*-acting elements and GO analysis suggested that *AYMYB* genes participated in many biological processes associated with development, metabolism, defense, and stress resistance. Additionally, transcriptome data showed that *AYMYB* genes exhibited tissue-specific expression patterns and their expression was significantly up-regulated under MeJA treatment. The qRT-PCR examination confirmed the reliability of the transcriptome data. Our research offers useful information for subsequent functional analysis of the *AYMYB* genes in the process of flavonoid biosynthesis in leaves and under MeJA treatment.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11441-z.

Supplementary Material 1: Additional file 1: Figure S1. The C***is***-elements of 227 ***AYMYBs*** promoters. Promoter sequences (-2,000 bp) of ***AYMYBs*** were analyzed using PlantCARE. Different shapes and colors represented different elements. Figure S2. Different tissue expression profiles and phylogenetic tree of all identified MYB members in ***A. argyi*** (A). MeJA treatment expression profiles and phylogenetic tree of all identified MYB members in ***A. argyi*** (B).

Supplementary Material 2: Additional file 2: Table S1. The number of amino acids, molecular weight, isoelectric point, instability index, aliphatic index, GRAVY, subcellular localization and MYB-domain type of the members of the ***AYMYB*** gene. Table S2. NCBI entries of ***MYB*** gene from other organisms used in phylogenetic analysis. Table S3. The gene pair, the link region

and the gene type of ***AYMYBs*** in this study. Table S4. List of the motif sequences identified in this study. Table S5. List of the ***cis***-acting elements identified in this study. Table S6. The information about GO analysis for ***AYMYBs*** in this study. Table S7. List of the primers used for qRT-PCR in this study.

Acknowledgements
We are thankful to funding agencies for funding.

Author contributions
PSN completed the formal analysis and wrote the manuscript. LYK, YY, FWN, and ZD managed plant resources and collected samples. ZQW, HBS, LDH, and MYH contributed to the review and editing of the manuscript. All authors approved the manuscript for publication.

Funding
This work was supported by the National Natural Science Foundation of China (82304660); Hubei Provincial Natural Science Foundation and Traditional Chinese Medicine Innovation and Development Joint Foundation of China (2024AFD228). HuBei Provincial Natural Science Foundation Innovative Group Project (2023AFA032). China Postdoctoral Science Foundation (2024M753648).

Data availability
All materials and related data in this study are available upon request. The raw data of RNA sequencing from seven different plant tissues and leaves treated with 300 μmol/L MeJA have been deposited at the Sequence Read Archive (SRA) under the accessions PRJNA804653 and PRJNA1135560, respectively.

Declarations

Ethics approval and consent to participate
The cultivated *A. argyi* 'Xiang Ai' has been identified and preserved by the team of the corresponding author, Professor Dahui Liu. Therefore, we ensure that we have obtained permission from Professor Liu to collect and use *A. argyi* in this experiment. The collection of plant material complies with relevant institutional, national, and international guidelines and legislation.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 30 May 2024 / Accepted: 4 March 2025
Published online: 18 April 2025

References

1. Dubos C, Stracke R, Grotewold E, et al. MYB transcription factors in *Arabidopsis*. Trends Plant Sci. 2010;15:573–81.
2. Du H, Zhang L, Liu L, et al. Biochemical and molecular characterization of plant MYB transcription factor family. Biochem (Mosc). 2009;74:1–11.
3. Simon M, Lee MM, Lin Y, et al. Distinct and overlapping roles of single-repeat MYB genes in root epidermal patterning. Dev Biol. 2007;311:566–78.
4. Kuno N, Møller SG, Shinomura T, et al. The novel MYB protein EARLY-PHYTOCHROME-RESPONSIVE1 is a component of a slave circadian oscillator in Arabidopsis. Plant Cell. 2003;15:2476–88.
5. Liu L, Li S, Tang F, et al. MaMYBR30, a novel 1R-MYB, plays important roles in plant development and abiotic stress resistance. Plants (Basel). 2024;13:1794.
6. Feller A, Machemer K, Braun EL, et al. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. Plant J. 2011;66:94–116.
7. Liu J, Osbourn A, Ma P. MYB transcription factors as regulators of phenylpropanoid metabolism in plants. Mol Plant. 2015;8:689–708.
8. Liu H, Li L, Fu X, et al. AaMYB108 is the core factor integrating light and jasmonic acid signaling to regulate Artemisinin biosynthesis in *Artemisia annua*. New Phytol. 2023;237:2224–37.
9. He G, Zhang R, Jiang S, et al. The MYB transcription factor RcMYB1 plays a central role in Rose anthocyanin biosynthesis. Hortic Res. 2023;10:uhad080.
10. Vimolmangkang S, Han Y, Wei G, et al. An Apple MYB transcription factor, MdMYB3, is involved in regulation of anthocyanin biosynthesis and flower development. BMC Plant Biol. 2013;13:176.
11. Ito M, Araki S, Matsunaga S, et al. G2/M-phase-specific transcription during the plant cell cycle is mediated by c-Myb-like transcription factors. Plant Cell. 2001;13(8):1891–905.
12. Feng G, Burleigh JG, Braun EL, et al. Evolution of the 3R-MYB gene family in plants. Genome Biol Evol. 2017;9:1013–29.
13. Yuan S, Jiang H, Wang Y et al. A 3R-MYB transcription factor is involved in Methyl Jasmonate-Induced disease resistance in agaricus bisporus and has implications for disease resistance in *Arabidopsis*. J Adv Res Published Online September 2, 2024.
14. Song X, Wen X, He J, et al. Phytochemical components and biological activities of *Artemisia argyi*. J Funct Foods. 2019;52:648–62.
15. Cui Z, Li S, Chang J, et al. The pharmacophylogenetic relationships of two edible medicinal plants in the genus Artemisia. Front Plant Sci. 2022;13:949743.
16. Ai P, Xue J, Shi Z, et al. Genome-wide characterization and expression analysis of MYB transcription factors in *Chrysanthemum Nankingense*. BMC Plant Biol. 2023;23:140.
17. Salih H, Gong W, He S, et al. Genome-wide characterization and expression analysis of MYB transcription factors in *Gossypium hirsutum*. BMC Genet. 2016;17:129.
18. Du H, Yang SS, Liang Z, et al. Genome-wide analysis of the MYB transcription factor superfamily in soybean. BMC Plant Biol. 2012;12:106.
19. Li Z, Peng R, Tian Y, et al. Genome-Wide identification and analysis of the MYB transcription factor superfamily in *Solanum lycopersicum*. Plant Cell Physiol. 2016;57:1657–77.
20. Arce-Rodríguez ML, Martínez O, Ochoa-Alejo N. Genome-Wide identification and analysis of the MYB transcription factor gene family in Chili pepper (*Capsicum* spp). Int J Mol Sci. 2021;22:2229.
21. Chen C, Wu Y, Li J, et al. TBtools-II: A one for all, all for one bioinformatics platform for biological big-data mining. Mol Plant. 2023;16:1733–42.
22. Potter SC, Luciani A, Eddy SR, et al. HMMER web server: 2018 update. Nucleic Acids Res. 2018;46:W200–4.
23. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22:1658–9.
24. Wilkins MR, Gasteiger E, Bairoch A, et al. Protein identification and analysis tools in the expasy server. Methods Mol Biol. 1999;112:531–52.
25. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4:406–25.
26. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870–4.
27. Xie J, Chen Y, Cai G, et al. Tree visualization by one table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees. Nucleic Acids Res. 2023;51:W587–92.
28. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.
29. Love MI, Huber W, Anders S. Moderated Estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
30. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real–time quantitative Pcr and the $2^{-\Delta\Delta ct}$ method. Methods. 2001;25:402–8.
31. YI X, Wu L, Xiang L, et al. Screening of reference genes for quantitative real-time PCR in *Artemisia argyi*. China J Chin Materia Med. 2022;47:659–67.
32. Katiyar A, Smita S, Lenka SK, et al. Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis. BMC Genomics. 2012;13:544.
33. Cao Y, Jia H, Xing M, et al. Genome-Wide analysis of MYB gene family in Chinese bayberry (*Morella rubra*) and identification of members regulating flavonoid biosynthesis. Front Plant Sci. 2021;12:691384.
34. Zhang C, Ma R, Xu J, et al. Genome-wide identification and classification of MYB superfamily genes in Peach. PLoS ONE. 2018;13:e0199192.
35. Thakur S, Vasudev PG. MYB transcription factors and their role in medicinal plants. Mol Biol Rep. 2022;49:10995–1008.
36. Mao K, Yang J, Wang M, et al. Genome-wide analysis of the Apple CaCA superfamily reveals that MdCAX proteins are involved in the abiotic stress response as calcium transporters. BMC Plant Biol. 2021;21:81.
37. Miao Y, Luo D, Zhao T, et al. Genome sequencing reveals chromosome fusion and extensive expansion of genes related to secondary metabolism in *Artemisia argyi*. Plant Biotechnol J. 2022;20:1902–15.
38. Wu Y, Wen J, Xia Y, et al. Evolution and functional diversification of R2R3-MYB transcription factors in plants. Hortic Res. 2022;9:uhac058.

39. Yi X, Wang X, Wu L, et al. Integrated analysis of basic Helix loop Helix transcription factor family and targeted terpenoids reveals candidate AarbHLH genes involved in terpenoid biosynthesis in *Artemisia argyi*. Front Plant Sci. 2022;12:811166.
40. Wang J, Cui Y, Li S, et al. Transcriptome analysis of *Artemisia argyi* following Methyl jasmonate (MeJA) treatment and the mining of genes related to the stress resistance pathway. Front Genet. 2023;14:1279850.
41. Wang F, Kong W, Wong G, et al. AtMYB12 regulates flavonoids accumulation and abiotic stress tolerance in Transgenic *Arabidopsis thaliana*. Mol Genet Genomics. 2016;291:1545–59.
42. Fornalé S, Lopez E, Salazar-Henao JE, et al. AtMYB7, a new player in the regulation of UV-sunscreens in *Arabidopsis thaliana*. Plant Cell Physiol. 2014;55:507–16.
43. Li Y, Qin W, Fu X, et al. Transcriptomic analysis reveals the parallel transcriptional regulation of UV-B-induced Artemisinin and flavonoid accumulation in *Artemisia annua* L. Plant Physiol Biochem. 2021;163:189–200.
44. Yang F, Wang T, Guo Q, et al. The CmMYB3 transcription factors isolated from the Chrysanthemum morifolium regulate flavonol biosynthesis in *Arabidopsis thaliana*. Plant Cell Rep. 2023;42:791–803.

Publisher's note