# Full-Length Transcriptomic Identification of R2R3-MYB Family Genes Related to Secondary Cell Wall Development in Cunninghamia Lanceolata (Chinese Fir)

**Hebi Zhuang**
  Zhejiang A & F University

**Sun-Li Chong**
  Zhejiang A & F University

**Priyanka Borah**
  Zhejiang A & F University

**Xiao Han**
  Zhejiang A & F University

**Erpei Lin**
  Zhejiang A & F University

**Zaikang Tong**
  Zhejiang A & F University

**Huahong Huang** ( ✉ huanghh@zafu.edu.cn )
  Zhejiang A & F University

---

**Research Article**

# Abstract

**Background:** R2R3-MYB is a class of transcription factor crucial in regulating secondary cell wall development during wood formation. The regulation of wood formation in gymnosperm has been understudied due to its large genome size. Using Single-Molecule Real-Time sequencing, we obtained full-length transcriptomic libraries from the developmental stem of *Cunninghamia lanceolata*, a perennial conifer known as Chinese fir. The R2R3-MYB of *C. lanceolata* (hereafter named as ClMYB) associated with secondary wall development were identified based on phylogenetic analysis, expression studies and functional study on transgenic line.

**Results:** The evolutionary relationship of 52 ClMYBs with those from *Arabidopsis thaliana*, *Eucalyptus grandis*, *Populus trichocarpa*, *Oryza sativa*, two gymnosperm species, *Pinus taeda*, and *Picea glauca* were established by neighbour-joining phylogenetic analysis. A large number of ClMYBs resided in the woody-expanded subgroups that predominated with the members from woody dicots. In contrast, the woody-preferential subgroup strictly carrying the members of woody dicots contained only one candidate. The results suggested that the woody-expanded subgroup emerges before the gymnosperm/angiosperm split, while most of the woody-preferential subgroups are likely lineage-specific to woody dicots. Nine candidates shared the same subgroups with the *A. thaliana* orthologs, with known function in regulating secondary wall development. Gene expression analysis inferred that *ClMYB1/2/3/4/5/26/27/49/51* might participate in secondary wall development, among which *ClMYB1/2/5/26/27/49* were significantly upregulated in the highly lignified compression wood region, reinforcing their regulatory role associated with secondary wall development. ClMYB1 was experimentally proven a transcriptional activator that localised in the nucleus. The overexpression of ClMYB1 in *Nicotiana benthamiana* resulted in an increased lignin deposition in the stems. The members of subgroup S4, ClMYB3/4/5 shared the ERF-associated amphiphilic repression motif with AtMYB4, which is known to repress the metabolism of phenylpropanoid derived compounds. They also carried a core motif specific to gymnosperm

m lineage, suggesting divergence of the regulatory process compared to the angiosperms.

**Conclusions:** This work will enrich the collection of full-length gymnosperm-specific R2R3-MYBs related to stem development and contribute to understanding their evolutionary relationship with angiosperm species.

# Background

The radial growth of woody stem in gymnosperm and angiosperm is contributed by bifacial development of cambium cells that forms xylem and phloem to the inner and outer part of vascular bundles, respectively [1]. Even though the xylems subsequently differentiate into vessels and fibres in angiosperms, gymnosperms form a primitive lineage having tracheids in woody tissues for water conductance and mechanical support. In a matured xylem cell, the secondary cell wall forms the thickest layer that determines wood properties. The secondary cell wall is formed by a structural network

consisting of cellulose as load-bearing support infiltrated by lignin and hemicellulose to reinforce wall support. To maintain sustainable growth, woody plants possess tight regulation on cambial activity and its wood formation incoherent with developmental phases and response to environmental stimuli [2]. Cell wall formation is spatiotemporally controlled by a hierarchical regulatory network. Transcription factors containing NAC (NAM/ATAF/CUC) [3–5] and MYB domains [6, 7] are essential for the transcriptional control of wood formation.

The MYB proteins constitute one of the largest transcription factor families in plants. All members have a highly conserved DNA-binding domain, i.e., the domain with one to three imperfect amino acid sequence repeats (R1, R2 and R3). They can be categorized according to the number of adjacent R proteins as R1, R2R3, 3R or 4R proteins with one, two, three or four R proteins, respectively [8]. The R2R3-MYB subfamily has the highest number of MYB members. Some R2R3-MYB members have crucial roles in regulating secondary cell wall formation [9, 10]. In *Arabidopsis thaliana*, the paralogous genes *AtMYB46* and *AtMYB83* are the second-level master switches that regulate secondary cell wall formation, and several other genes (*AtMYB4*, *AtMYB7*, *AtMYB20*, *AtMYB32*, *AtMYB42*, *AtMYB43*, *AtMYB52*, *AtMYB54*, *AtMYB58*, *AtMYB63*, *AtMYB69*, *AtMYB85*, and *AtMYB103*) may function as downstream transcription factors in this regulatory network [11]. Some R2R3-MYB members were involved in forming the secondary cell wall in tree species via the regulation of cellulose, hemicellulose, and lignin biosynthesis [12]. Like *AtMYB46* and *AtMYB83*, *PtrMYB3* and *PtrMYB20* from *Populus trichocarpa* are also second-level master switches involved in the regulation of secondary cell wall formation, which induce the expression of cellulose, xylan, and lignin biosynthesis genes [13]. In addition, *EgrMYB1*, *EgrMYB2*, and *EgrMYB88* of *Eucalyptus grandis* also involve in the regulation of secondary cell wall formation [14–16]. In conifers, *PtMYB1*, *PtMYB4*, and *PtMYB8* of *Pinus taeda* [17–19] have been suggested to regulate lignin biosynthesis in the differentiated xylem tissues.

Following land plants' territorialisation, R2R3-MYBs have undergone clade-asymmetric expansion and subsequent sub- or neofunctionalization correspond with species diversification [20]. The genome-wide duplication and tandem duplication events contribute to the lineage-specific expansion [21]. A genome-wide analysis of R2R3-MYBs in various woody dicots such as *Eucalyptus grandis, P. trichocarpa and Vitis vinifera* has identified specific lineage expansion as woody-preferential and woody-expanded subgroups. These are implicated for essential functions at regulating wood formation [12]. The formation of tracheid in gymnosperm is regulated by the VNS-MYB based transcriptional network [22], reflecting the conserved function of MYB regulators to some extent among vascular plants. The evolutionary relationship of R2R3-MYBs between gymnosperm and angiosperm species remains elusive.

*Cunninghamia lanceolata* (Chinese fir), an evergreen conifer belonging to the Cupressaceae family, is native to China and distributed mainly in 16 provinces in southern China. Due to its rapid growth, desirable wood quality and low disease and insect pest pressure, this species has been widely cultivated for over 3,000 years. Breeding efforts beginning in the 1960s, including provenance tests, cross-breeding and clonal selection aimed for rapid growth, have achieved remarkable successes. As a result, they currently account for 20–30% of China's total commercial timber production. Recently, exploration of the

molecular mechanism of wood formation in *C. lanceolata* was attempted by constructing a cDNA library and expressed sequence tag (EST) database of *C. lanceolata* var. Dugaushau [23] and a transcriptome database containing more than 80,000 sequences established using Illumina high-throughput sequencing [24]. Additionally, candidate genes related to lignin and cellulose biosynthesis were analysed in Wang et al. (2011) [25]. Despite that, identification and characterization of gene functions in *C. lanceolate* have been complicated by incomplete and redundant sequences obtained from Illumina short-read sequencing analysis. We used the PacBio long-read sequencing platform to obtain full-length transcriptomic libraries from the developmental stem of *C. lanceolata*. As such, the R2R3-MYBs of *C. lanceolata* (hereafter named as ClMYBs) were identified. Protein sequence characteristics, phylogenetic relationships, and expression patterns of ClMYB encoding genes were subsequently investigated. Based on their evolutionary relationship with the known genes from *Arabidopsis thaliana* and woody species, we identified nine ClMYBs associated with secondary cell wall development with some degree of diversification of tissue specificity compared to the angiosperm species. The involvement of *ClMYB1* in secondary wall development was identified and characterized by ectopic expression in tobacco. These results provide essential genetic resources for improving the wood quality of *C. lanceolata* at the molecular level.

## Materials And Methods

### Plant materials

The plant materials were harvested from two-year-old living plants belonging to *C. lanceolata* clone 2015-2, obtained from the Forest Germplasm Resource Conservation Nursery (E119.726, N30.267) located at the State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University. The nursery was maintained by our laboratory; hence no permission for the sample collection was required. To our knowledge there was no herbarium available for the deposition of the voucher specimen of this specific plant. Plant identification was carried out by the co-author, Professor Zaikang Tong. Plants were grown at 26-28$^{\circ}$C under a light intensity of 1,500 Lux and a photoperiod of 12 h daily. Eleven different plant organs and tissues were collected from mid-March until the end of June in 2017. Among these organs, female cones (FC) and male cones (MC) were harvested from upper branches at similar heights. The roots (RT) were cut from hydroponically grown clone 2015-2 plants. Leaves (L) consisted of the one-year-old mature coniferous needles. Five stem segments were collected from the top to bottom of an ongoing-year branch, with each segment being 2 cm long and marked sequentially as S1, S2, S3, S4, and S5. Phloroglucinol-HCl staining was performed on the sequential stem cross-sections to observe vascular development (Supplementary Figure S1). The bark (B) of a two-year-old stem segment was collected and the 1.0-mm-thick xylem (X) layer was scraped from the outer xylem. All samples were quickly frozen in liquid nitrogen and stored at -80°C. *Nicotiana benthamiana* was used for genetic transformation for subsequent analysis of the gene functions.

### Compression wood induction

An artificial bending treatment was performed to induce compression wood [26]. In early May of 2017, stem of *C. lanceolata* plants that displayed relatively uniform growth were bent at an angle of approximately 45° to the vertical direction of the ground. The trunk segments at the bending site were sampled after 10, 30 and 60 days of bending. After the bark was removed, the outer xylem was scraped from the compression wood (CW) and opposite wood (OW) and was quickly frozen in liquid nitrogen. Untreated, upright plants were used as controls (CK). Five biological replicates were tested. Transverse sections (thickness of 10 μm) were prepared with the use of stem segments harvested after 60 days of bending. To visualize the thickening of the cell wall, the sections were stained with safranin. In addition, the double-wall thickness of tracheid was measured using a modification of Franklin's method [27]. At least 90 tracheids from each wood type were measured using a light microscope, and the data were analysed for significant differences using SPSS 19.0 (IBM) with Student's t-test.

## Preparation of full-length transcriptomic library

The RNA from five stem segments were subjected to PacBio library construction and sequencing. The cDNAs were synthesized with Clontech SMARTer PCR cDNA Synthesis Kit according to the manufacturer's protocol and the PCR products were sequenced with the PacBio Sequel System. The long-read sequence was analysed by Single-Molecule Real-Time (SMRT) analysis software (v. 2.3.0) using the RS_IsoSeq.1 protocol (https://github.com/PacificBiosciences/). High-quality consensus isoforms were obtained based on the Iterative Clustering for Error Correction (ICE) algorithm and subsequent Quiver polishing. All these isoforms were further clustered and filtered using CD-Hit (v. 4.6) [28] to obtain unigenes.

## Identification of R2R3-MYB family members in *C. lanceolata*

The Hidden Markov Model (HMM) file for the MYB domain (PF00249) was downloaded from the Pfam database (http://pfam.sanger.ac.uk/). The putative MYB members obtained from the full-length transcriptome of *C. lanceolata* were identified by HMMER 3.0 [29]. A total of 116 putative sequences containing MYB domain were identified. The protein sequences were analysed via SMART (http://smart.embl-heidelberg.de/) and InterPro (http://www.ebi.ac.uk/interpro/) to examine the topological arrangement of MYB domains in the sequences. BLASTp searches in conjunction with the nonredundant (NR) library (https://blast.ncbi.nlm.nih.gov/Blast.cgi) were used to aid in the annotation of the sequences. As a result, 70 sequences harbouring R2R3-MYB domain were identified. ClustalX multiple sequence alignment tool (v. 2.1) [30] was employed to remove redundant sequences. High similarity sequences were filtered-out and as a result, 52 putative R2R3-MYB sequences remained in the dataset for further analysis.

## Protein sequence analysis and phylogenetic tree construction

The ExPASy database (https://expasy.org) was used to analyse the physicochemical properties of R2R3-MYBs in *C. lanceolata*. Multiple sequence alignments were performed using ClustalX, and sequence logos of the R2 and R3 domains were plotted via the WebLogo (http://weblogo.Berkeley.edu/logo.cgi). The

motifs of the R2R3-MYBs in *C. lanceolata* were analysed using the MEME program (http://meme-suite.org/tools/meme). The motif width was set to 6-50, and the number of motifs was 20, with the default values for the remaining parameters.

The evolutionary relationship of *C. lanceolata* (52) R2R3-MYBs with *E. grandis* (141), *P. trichocarpa* (187), *A. thaliana* (126), *Oryza sativa* (104), *P. taeda* (7), and *Picea glauca* (12) were studied via the construction of a Neighbour-Joining phylogenetic tree with 1000 bootstrap replicates. The sequences of the species other than *C. lanceolata* were downloaded from the Phytozome (http://www.phytozome.org/) and NCBI (http://www.ncbi.nlm.nih.gov/) databases. A total of 629 R2R3-MYBs sequences were aligned using ClustalW and the phylogenetic tree were constructed using MEGA7 [31]. The evolutionary distances were computed using the Jones-Taylor-Thornton matrix-based method and the rate variation among sites was modelled with a gamma distribution (shape parameter = 1).

## Expression profile analysis of *ClMYB*s based on RNA sequencing (RNA-Seq) and quantitative real-time PCR (qRT-PCR)

The RNA from various organs and tissues such as female cone (FC), male cone (MC), leaves (L), roots (RT), five stem segments (named as S1 to S5), bark (B) and xylem (X) were used to construct RNA-Seq libraries with three biological replicates were included. Total RNA from *C. lanceolata* samples was extracted according to the PureLink Plant RNA Reagent Kit manual (ThermoFisher Scientific, USA). Paired-end 150-bp sequencing was performed on the Illumina HiSeq Xten platform. The transcript abundance of the *ClMYB*s in the different organs and tissues was reported in terms of fragments per kilobase of exon model per million mapped reads (FPKM). After they were log-transformed, the expression data were subjected to hierarchical clustering, and an expression profile heatmap was constructed by TBtools software [32].

The qRT-PCR analysis was performed using a PrimeScript Reagent Kit in conjunction with a SYBR Premix Ex Taq Kit (Takara, Japan). The qRT-PCR system consisted of 5 µl of SYBR Premix Ex Taq (2×), 0.2 µl of forward primer (10 µM), 0.2 µl of reverse primer (10 µM), 0.8 µl of the first-strand cDNA, and enough double-distilled water (ddH$_2$O) to bring the total volume of 10 µl. The reaction was performed with the following the temperature program: 95°C for 2 min; 40 cycles of 95°C for 5 sec, and 60°C for 30 sec. The data obtained were analyzed by the $2^{-\Delta\Delta Ct}$ method, with the housekeeping gene *ClActin* used as an internal control. All calculations were performed via CFX96 Manager version 1.6 (Bio-Rad, USA) and Excel 2007 software. The qRT-PCR primer sequences are listed in Supplementary Table S1.

## Subcellular localization, transcriptional activation assay and ectopic expression analysis of *ClMYB1*

The open reading frame (ORF) of *ClMYB1* was amplified by PCR with gene-specific primers and then inserted into a pCAMBIA1302 vector by restriction enzyme ligation, yielding a *ClMYB1::sGFP* (synthetic green fluorescent protein) fusion expression cassette, which was transformed into onion epidermal cells via the particle bombardment method. The fluorescence signals were observed and imaged via a Zeiss LSM510 confocal microscope (Carl Zeiss, Oberkochen, Germany).

The ORF of *ClMYB1* without its termination codon was inserted into a pGBKT7 vector (Clontech, USA) harbouring GAL4-BD via Gateway technology. The recombinant plasmid was subsequently transformed into the yeast (*Saccharomyces cerevisiae*) strain Gold2. Positive clones were selected on SD/-Trp media, and the transformants were plated and cultured on SD/-Trp/-His/-Ade/X-a-gal media for transcriptional activation analysis.

*ClMYB1* was inserted into a pCAMBIA13011 vector downstream to the 35S promoter by restriction enzyme digestion and ligation. This *p35S::ClMYB1* expression cassette was transformed into *N. benthamiana* via the leaf disc method, mediated by *Agrobacterium tumefaciens* strain GV3101. Phenotypic analysis of two-month-old transgenic plants (T3) was performed. The DNA was extracted by the cetyl-trimethylammonium bromide (CTAB) method, and the transformation events of *ClMYB1* were analysed via PCR. Total RNA was extracted via a TRIzol kit (ThermoFisher Scientific, USA), and the expression of *ClMYB1* was measured via semiquantitative RT-PCR (semi-qRT-PCR) as described by [24]. The relevant primer sequences are listed in Supplementary Table S1. Phloroglucinol-HCl staining was applied to observe the lignin deposition in the stems, and the relative lignin content in the stems was determined according to the lignin-thioglycolic acid (LTGA) method [33].

# Results

# Identification and domain analysis of ClMYBs

A total of 116 candidate sequences with MYB domains were selected from the full-length transcriptomic library, including three 4R-MYB, seven 3R-MYB, 70 R2R3-MYB, and 36 1R-MYB proteins. After the removal of redundant sequences, we identified 52 R2R3-MYB with complete ORF, which were sequentially named as *ClMYB1-52* (Supplementary Table S2).

The number of amino acid residues in the 52 ClMYBs ranged from 215 to 596. The corresponding molecular weight and isoelectric points ranged between 24.3–64.1 kDa and 4.7–9.9 (Supplementary Table S2), respectively. Multiple sequence alignments showed that all the N-terminal regions of 52 ClMYBs contained conserved R2 and R3 domains. The R2 domains of ClMYBs contained three highly conserved Trp (W) residues. However, within the R3 domain, the first W residue is substituted by Leu (L), Ile (I) or Phe (F). In addition, some conserved amino acids were present around these W residues, including Gly (G2), Glu (E8), Asp (D9), Leu (L12), Gly (G20), Arg (R36), Gly (G38), Lys (K39), Cys (C41), Arg (R42), Arg (R44), Asn (N47) and Leu (L49) in the R2 domain; Glu (E8), His (H16), Gly (G20), Asn (N21), Gly (G32), Arg (R33), Thr (T34), Asp (D35), Asn (N36), Lys (K39) and Asn (N40) in the R3 domain (Fig. 1). These characteristics ensure that MYB domains fold into a stable helix-helix-turn-helix structure.

### A comparative phylogenetic study of R2R3-MYBs in C. lanceolata and other plant species

To analyse the evolutionary relationship of ClMYBs and known R2R3-MYBs from *A. thaliana, E. grandis, P. trichocarpa, O. sativa, P. taeda,* and *P. glauca,* we constructed the neighbour-joining phylogenetic tree (Fig. 2; Supplementary Figure S2). According to the topological structure, all the proteins could be divided

into 47 subgroups, 21 of which harboured R2R3-MYBs from *C. lanceolata*. Over 50% of the subgroups were supported by bootstrap value of >70%. Lower bootstrap values were found in the subgroups with mixed sequences of gymnosperm and angiosperm origin. This can be justified because many gene duplication events in the R2R3-MYBs occurred after divergence from a common ancestor [34] and rapid diversification of the family after the division [35]. Nomenclatures of the subgroups were previously reported for *A. thaliana* [8, 36], later these were modified by Soler et al., (2015) [37] following a genome-wide evolutionary analysis of the R2R3-MYBs from the woody species, *E. grandis*. The sequences of *A. thaliana* and *E. grandis* were correctly placed in the previously assigned subgroups. Three *C. lanceolata* specific subgroups were identified and they were named as subgroup 13b and ClMYB15, which harboured single candidate, ClMYB24 and ClMYB15, respectively. Subgroup ClMYB17 contained two candidates, ClMYB17 and ClMYB18. Furthermore, subgroup 9c containing ClMYB28 and PgMYB11 represented a gymnosperm specific lineage.

Only seven common subgroups (S4, SAtM103, SAtM46, S13a, SAtM85, S22, S21) contained sequences from all the selected plants (Fig. 2). For the MYBs that are important in regulating secondary wall development in *A. thaliana*, seventeen of them were distributed across nine subgroups (Table 1; Supplementary Figure S2). AtMYB46 and AtMYB83 (subgroup SAtM46) were the second layer master switches to secondary wall developmental growth [38, 39]. AtMYB52, AtMYB54 and AtMYB69 (subgroup S21) were proposed to be the activator of lignin, xylan and cellulose biosynthesis [36, 40]. AtMYB20, AtMYB42, AtMYB43 and AtMYB85 (subgroup SAtM85) and AtMYB58 and AtMYB63 (subgroup S2 & S3) were the activators of lignin biosynthesis [41, 42]. AtMYB26 (subgroup SAtM26) was involved in secondary cell wall thickening in anther endothecium [43, 44]. AtMYB61 (subgroup SAtMYB61) was proposed to act as a mediator to allocate resources to sink tissues such as seed coat [45], root and xylem [46] by modulating genes related cell wall biosynthesis. Besides, two subgroups, namely S4 and S6, were dominated by MYBs functioning as antagonists of cell wall development. For example, AtMYB4 and AtMYB32 (subgroup S4) were the suppressors of biosynthesis of phenylpropanoid derived compounds [47, 48].

These subgroups implicated for secondary cell wall development in *A. thaliana* also included those from woody species such as seven MYBs in *P. trichocarpa* (PtrMYB152, PtrMYB010, PtrMYB128, PtrMYB002, PtrMYB003, PtrMYB020 and PtrMYB021) [5, 13, 49], two *E. grandis* proteins (EgrMYB1 and EgrMYB2) [14, 15, 50], and three *P. taeda* proteins (PtMYB1, PtMYB4, and PtMYB8) [18, 19] (Table 1; Supplementary Figure S2). Not all ClMYBs were found in these secondary wall related subgroups. There were nine orthologous ClMYBs (ClMYB1, ClMYB2, ClMYB3, ClMYB4, ClMYB5, ClMYB26, ClMYB27, ClMYB49, and ClMYB51) segregated across six clades, suggesting they are the candidate regulators of secondary wall development during wood formation in *C. lanceolata* (Table 1). ClMYB49 belonging to subgroup SAtM46 was a homolog of AtMYB46 and AtMYB83, meanwhile ClMYB2 and ClMYB26 were clustered with subgroup S21 and SAtM103, respectively. For the subgroups associated with activation of lignin biosynthesis (subgroup SAtM85 and S2&S3), ClMYB1 and ClMYB51 belonged to SAtM85, while none were found in S2&S3. Meanwhile, ClMYB3, ClMYB4 and ClMYB5 were clustered with subgroup S4, which harbored the repressors of phenylpropanoid metabolism.

From the evolutionary analysis of R2R3-MYBs from *E. grandis*, MYBs of strictly woody species are detected in a total of five woody-preferential subgroups [12]. In the present study, we found that only ClMYB21 present in the woody-preferential subgroup III (Supplementary Figure S2). Subgroup 5 (S5) and AtMYB5 (SAtM5) contained a high proportion of MYBs from woody perennial plant species and were referred to as woody-expanded subgroups, with eight and six ClMYBs assigned to these two subgroups, respectively.

Table 1

Summary of R2R3-MYB transcription factors related to development of secondary cell wall.

| Function | R2R3-MYB related to SCW development | Group name | Homologous *ClMYBs* | Biological process | Reference |
|---|---|---|---|---|---|
| Activator | AtMYB52, AtMYB54, AtMYB69 | S21 | ClMYB2 | Lignin, xylan and cellulose biosynthesis | [36, 40] |
| | AtMYB20, AtMYB42, AtMYB43, AtMYB85 PtrMYB152 PtMYB1 | SAtM85 | ClMYB1 ClMYB51 | Lignin biosynthesis. Binds to AtMYB4 to repress flavonoid biosynthesis | [19, 42, 49] |
| | AtMYB103 PtrMYB010, PtrMYB128 | SAtM103 | ClMYB26 | Lignin biosynthesis | [5, 51] |
| | AtMYB46, AtMYB83 PtrMYB002, PtrMYB003, PtrMYB020, PtrMYB021 EgrMYB2 PtMYB4 | SAtM46 | ClMYB49 | Lignin, xylan and cellulose biosynthesis | [18, 39, 52–55] |
| | AtMYB26 | SAtM26 | | Secondary wall thickening in anther endothecium | [43, 44] |
| | AtMYB58, AtMYB63 | S2&S3 | | Lignin biosynthesis | [41] |
| | AtMYB61 PtMYB8 | S13a | ClMYB27 | Regulates stomatal aperture closure, seed coat mucilage biosynthesis, xylem and root development | [45, 46, 56] |
| Repressor | AtMYB4, AtMYB32 EgrMYB1 | S4 | ClMYB3, ClMYB4, ClMYB5 | Biosynthesis of phenylpropanoid derived compound | [47, 48, 50] |
| | AtMYB75 | S6 | | Anthocyanin biosynthesis | [58] |

**Motif analysis of R2R3-MYB proteins of C. lanceolata**

In addition to the R2 and R3 domains, 17 conserved motifs were identified via MEME; these motifs were designated as motifs 1 to 17 in increasing order of their corresponding E-values (Fig. 3; Supplementary Data S1). Motif 1 was upstream of the R2 domain, and the other motifs were downstream of the R3 domain (Fig. 3B). Among the different subgroups, the amount and type of motifs were significantly different. Most of the members belonging to the same subgroup shared one or more motifs (Fig. 3A). For example, ClMYB8, ClMYB9, ClMYB11 and ClMYB29 belonging to S22 contained motif 5, motif 8, and motif 17. Motif 15 carrying the core sequence of (E/I)LIHMADFFQ was unique to ClMYB3, ClMYB4 and ClMYB5 in the subgroup S4. Whereas, motif 17 with the sequence of (D/E)(I/V)NLDL(C/S)ISLP was the ERF-associated amphiphilic repression (EAR) motif that common between angiosperm and gymnosperm in subgroup S4 and dictated the repressive role of the members on the phenylpropanoid metabolism [59].

## Expression profiles of ClMYBs in different organs and tissues

To explore the involvement of ClMYB in the secondary wall development, expression profiles of the 52 ClMYB encoding genes in the six organs and tissues were analyzed using the available transcriptome data. Forty-three *ClMYB* were expressed in all the organs and tissues tested (FPKM > 0), eight of which (*ClMYB5, 8, 9, 11, 17, 23, 25,* and *47*) were constitutively expressed (FPKM > 2 in all samples) (Fig. 4; Supplementary Table S3). Compared with the other *ClMYBs, ClMYB9* showed relatively higher expression levels in the organs and tissues examined, with a corresponding mean FPKM value of 100.4. In some cases, the expression profiles of closely related members were similar. For example, the expression patterns of *ClMYB22, ClMYB23,* and *ClMYB25* in the subgroup SAtM5, one of the woody-expanded subgroups, they were clustered in two sub-clades that were highly expressed in female cone. However, in most cases, the expression profiles of the genes present in the same subgroup differed, suggesting their specific roles in different organs. This notion was manifested by *ClMYB14* and *ClMYB45*, though they also belonged to subgroup SAtM5, they were highly expressed in the male cone. Likewise, the expression of another SAtM5 ortholog, *ClMYB20* was highly expressed in root and bark tissues.

The expression profiles of the nine candidates that might involve in regulating secondary wall development in lignified tissues could be divided into three categories: *ClMYB1, ClMYB26, ClMYB27, ClMYB49,* and *ClMYB51* were dominantly expressed in the stem xylem. (*ClMYB3* and *ClMYB5*) and (*ClMYB2* and *ClMYB4*) were segregated in sub-clades that exhibited a high expression in the root and female cone, respectively. Since the R2R3-MYBs may act as an activator or repressor to secondary wall development, we followed their expression profile in different stem segments with maturation increasing from S1 to S5 (Fig. 5, Supplementary Table S3). For the xylem specific regulators (*ClMYB1, ClMYB26, ClMYB49,* and *ClMYB51*), their expression increased congruent with stem maturation; except ClMYB27 that its expression was indifferent in the developmental stem tissues. The expression of *ClMYB3, ClMYB4,* and *ClMYB5* in the stems tended to decrease with stem maturation, suggesting they may play negative regulatory roles in wood formation. Interestingly, we found that although ClMYB21 belonged to the woody-preferential subgroup III, its gene expression was not significantly different in the developmental stems. The RNA-Seq data were further verified by qRT-PCR experiments on eight representative samples for 12 selected MYB genes (Supplementary Figure S3). Correlation coefficients of

0.845−0.956 (p < 0.01) indicated that the gene expression patterns obtained from the RNA-Seq data were consistent with the qRT-PCR results.

### Expression patterns of ClMYBs involved in compression wood development

Trees develop reaction wood in response to a gravistimulus. In gymnosperms, reaction wood usually develops on the lower side of leaning stems or branches and is referred to as compression wood. Typical features of *C. lanceolate* compression wood could be observed when plants were mechanically bent at a 45° angle from the vertical axis. The tracheid walls of compression wood were significantly thicker than those of normal wood (Fig. 6A, Supplementary Data S2) and contained more lignin but less cellulose [60]. By analysing the expression of ClMYBs encoding genes associated with secondary wall development in the compression wood, we could follow how they were coordinated or reprogrammed that might cause perturbation of anatomical and chemical features of woody tissues in Chinese fir (Fig. 6B, Supplementary Data S3).

Xylem specific ClMYBs generally exhibited increased expression level in compression wood as the bending period prolonged albeit with minor differences. The expression level of *ClMYB26, ClMYB27* and *ClMYB49* exhibited similar expression patterns, as the expression level in the compression wood was significantly greater than that in the upright wood and reached its maximum after 60 days of bending. The expression level of *ClMYB1* in the compression wood was first increased but then decreased with the progression of the bending; also, these levels were significantly higher than those in the upright wood. No significant difference in expression levels between the opposite wood and upright wood, except that *ClMYB49* in opposite wood had shown a significant higher expression level than that in the upright wood. The expression level of *ClMYB51* in both the compression wood and opposite wood fluctuated and did not exhibit a distinct increment in the compression wood as that shown by other xylem specific MYBs.

The responses of the potential negative regulators *ClMYB3*, *ClMYB4*, and *ClMYB5* to bending were also different from each other during compression wood induction. The expression level of *ClMYB3* in the compression wood was significantly lower than that in the control, but the expression level in the opposite wood was not significantly different from that of upright wood. The expression level of *ClMYB4* in the compression wood decreased significantly during the first 10 days and then returned to values similar to that in the upright wood, while the corresponding values of the opposite wood were significantly lower than those of the control and were 0.45 and 0.44 at 30 and 60 days after bending, respectively. Additionally, the expression of *ClMYB5* in the compression wood gradually increased, reaching a maximum (2.11) after 60 days of bending. Only after ten days of bending, the expression of this gene in the opposite wood was significantly different from that in the control.

# Subcellular Localization And Transcriptional Activity Analysis Of Clmyb1

The phylogenetic and gene expression analysis suggested that the function of the xylem specific ClMYBs related to the regulation of secondary wall development might be conserved in *C. lanceolata*. To verify the protein function, we examined the functionality of *ClMYB1* (GenBank accession number: JQ904045) by subcellular localization and overexpression in tobacco plant. *ClMYB1* was highly homologous to *PtMYB1* and *PtrMYB152* (Fig. 7A) and was expressed predominantly in the developing stem, suggesting that this gene participated in secondary wall development in *C. lanceolata*. The subcellular localization showed that the control GFP was distributed throughout the whole epidermal cell, while the ClMYB1:GFP fusion protein accumulated exclusively in the nucleus (Fig. 7B). The results indicated that ClMYB1 is localised in the nucleus, which is consistent with its function as a transcription factor. As shown in Fig. 7C, the yeast transformants harboring the negative control plasmid pGBKT7 and the recombinant plasmid pGBKT7-ClMYB1 were both able to grow on SD/-Trp media, but only the yeast containing the recombinant plasmid grew well on the selection media and induced the expression of the α-galactosidase reporter gene, indicating that ClMYB1 is a transcriptional activator. Taken together, these results indicated that ClMYB1 functions as a transcriptional activator in the nucleus.

### Ectopic expression of ClMYB1 in N. benthamiana

To investigate its regulatory functions in secondary vascular development, *ClMYB1* gene was successfully transformed into *N. benthamiana*. Eight putative transgenic lines were obtained, and two *ClMYB1*-overexpressed lines were selected for in-depth characterization. PCR revealed that *ClMYB1* had been successfully introduced into the *N. benthamiana* genome and was transcribed in transformed lines (Fig. 8B). Compared with the wild-type (WT), the transgenic lines exhibited smaller leaves with an undulate margin and a smaller angle between the petiole of the upper leaves and the main stem (Fig. 8A). Cross-sections of the top, middle and bottom parts of the stems were prepared for phloroglucinol-HCl staining. No significant difference was observed from staining results between the upper stem segments of the transgenic lines and WT, but darker staining was observed in the xylem regions of the middle and bottom stem segments from transgenic lines (Fig. 8C). A subsequent assay showed that the lignin content of the stems from transgenic lines was significantly higher than that of wild type (Fig. 8D, Supplementary Data S4). These results indicated that *ClMYB1*-overexpression could promote lignin deposition in the xylem regions of stems.

## Discussion

Large genome size and limited gene information severely hinder the study of molecular mechanisms on critical biological processes related to organ development and stress resistance in many valuable coniferous timber species. Using an advanced sequencing technology, 52 R2R3-MYB encoding genes with complete ORFs were obtained from the full-length transcriptomic libraries of *C. lanceolata*. Although not as many as in *Arabidopsis* (126 *AtMYBs*) and *Populus* (187 *PtrMYBs*), the number of identified ClMYBs were distinctly greater than that previously reported in the conifers such as *P. taeda* and *P. glauca* [6, 59]. This work effectively updated and improved the collection of full-length gymnosperms specific R2R3-MYB transcription factors related to stem development.

## Evolutionary relationship of R2R3-MYBs from C. lanceolata

R2R3-MYBs are crucial in the regulation of plant development [8]. The number of R2R3-MYBs increased gradually in basal land plants and eventually expanded rapidly after the emergence of seed plants (spermatophyte) [21]. Depending on the mechanism of gene duplication, some duplicated genes may undergo sub- or neo-functionalization that implicated in the diversification of gene function [61]. The comparative analysis of gene mapping on the spruce genome with those of monocots and *A. thaliana* suggests that the number of gene duplication shared by angiosperm and gymnosperm is greater than that of gymnosperm specific gene expansion [62]. Similarly, the R2R3-MYBs from *C. lanceolata* belonged to most subgroups shared by monocot and dicots, indicating that these subgroups might share the common ancestor. Only a small number of subgroups were specific to gymnosperm lineage. There is an excellent scope for characterization of the candidate genes to hypothesize new functions related to the coniferous developmental growth. The dominant expression of gene encoding ClMYB24 of subgroup S13b in xylem tissues, inferred it a potential regulator of stem development in gymnosperms.

It has been shown that the divergence between monocot and dicotyledonous species, which accompanied by whole-genome duplications [63], resulted in the speciation of subgroups limited to monocots and/or dicotyledonous lineages [9, 12]. Out of the five woody-preferential subgroups identified by Soler et al., (2015)[12], only one of them harboured the ClMYB21 from *C. lanceolata*. In contrast, the woody-expanded subgroups that were typically enriched in R2R3-MYBs of woody species, but also carrying one or two sequences from *A. thaliana* or *O. sativa*, contained a high number of sequences from *C. lanceolata.* In particular, the woody-expanded subgroup SAtM5 contained up to six ClMYBs, with their expression being variably expressed in different part of organs. There are two possible reasons for the absence of ClMYBs in majority of the woody-preferential subgroup: 1) the corresponding transcripts were not detected in the present transcriptomic libraries; 2) most of the woody-preferential subgroups have emerged after the gymnosperm/angiosperm split. The overrepresentation of ClMYBs in the woody-expanded subgroups suggested that the subgroup appears predating the divergence of angiosperm-gymnosperm and underwent asymmetric gene duplication and diversification in woody species [20].

## Involvement of ClMYBs in the regulation of secondary wall development

The ClMYBs resided in some of the *A. thaliana* secondary cell wall-related subgroups based on the phylogeny analysis. We also observed their expression in different organ and developmental stem, as well as compression wood to justify their involvement in secondary wall development. It was hypothesized that the functional role of the xylem specific ClMYB1/26/49/51 might be directly involved in the upregulation of secondary wall development in the stem of *C. lanceolata*. Their categorisation with those of *A. thaliana* and woody species was supported by a high bootstrap value. ClMYB1 and ClMYB51 were clustered in subgroup SAtm85 (bootstrap value 83%), which included several MYBs related to secondary wall development, such as AtMYB85, PtrMYB152, and PtMYB1. AtMYB85, a target gene of AtMYB46/83, regulate lignin biosynthesis in *A. thaliana* [40] and is highly expressed in the xylem of inflorescence stems and roots [64]. PtrMYB152 is a transcriptional activator specific for lignin

biosynthesis in *P. trichocarpa* and is expressed mainly in lignified tissues [65]. MYB1 of *P. taeda* activates the transcription of genes encoding lignin biosynthetic enzymes by binding with AC elements and is mainly expressed in developing xylem [19]. *ClMYB1* and *ClMYB51* were expressed mainly in the developing xylem of the stem, and their expression levels in the stem increased with lignification. In response to the bending treatment, the expression of *ClMYB1* was upregulated, corresponding to the lignin content in the compression wood. To date, no available genetic transformation system has been established for *C. lanceolata*, so the biological functions of most candidates involved in the wood formation of this species remain mostly unexplored. In this study, we selected ClMYB1 for further functional characterization, based on subcellular localization and transcriptional activity analysis. Functional analysis by overexpression of *ClMYB1* in *N. banthamiana* indicated that it could increase lignin deposition in the stems. The results indicated the retention of function between ClMYB1, AtMYB85 and PtrMYB152, although the conservation of their activators and downstream targets remains to be studied further.

The expression patterns of *ClMYB49* in different organs, tissues and during the formation of compression wood were essentially consistent with those of *ClMYB1*, indicating that they may also be involved in the regulation of secondary wall development in *C. lanceolata*. The sequences of ClMYB49 were segregated in subgroup SAtm46 together with PtMYB4 and AtMYB46/83 (bootstrap value 91%). AtMYB46/83 activated the transcriptional activity of secondary wall biosynthesis and was directly controlled by the NAC domain-containing SND1 master regulator in *A. thaliana* [39, 52, 66]. *PtMYB4* was expressed mainly in the xylem and its protein can bind to AC elements. The overexpression of this gene in tobacco lead to an increase in lignin content [18]. ClMYB26 was clustered in the subgroup SAtM103 with AtMYB103, PtrMYB010, and PtrMYB128 (bootstrap value 97%). Intriguingly, AtMYB103 influenced syringyl-lignin synthesis by regulating the expression of ferulate-5-hydroxylase (F5H) gene [40, 67]. S-lignin is typically enriched in the secondary wall of dicots, while gymnosperms have only a small amount of S-lignin [68]. The regulatory process of ClMYB26 will be subjected to future investigation to identify the downstream targets and the influence on lignin composition. The specificity of the regulatory mechanism of ClMYB26 and AtMYB103 towards the biosynthesis of monolignol would clarify whether divergence of function has occurred after the angiosperm/gymnosperm split.

## Repressive Role Of Clmybs In Subgroup S4

The expression of *ClMYB3*, *ClMYB4*, and *ClMYB5* in the stems decreased with increasing lignification, and these genes clustered with AtMYB4 and EgrMYB1 in subgroup 4. Both AtMYB4 and EgrMYB1 are negative regulators of lignin biosynthesis. *EgrMYB1* is expressed predominantly in the secondary xylem of root and stem (Jin et al. 2000; Legay et al. 2007; Legay et al. 2010). Therefore, ClMYB3/4/5 could be the potential regulators of *C. lanceolata* wood formation by negatively regulating lignin biosynthesis. Unlike the xylem specific ClMYBs, ClMYB3/4/5 received only 46% support for their classification in subgroup S4. The probable reason of this might be the diversity of C-terminal region in the orthologs. Compared with angiosperm, subgroup S4 was asymmetrically expanded in gymnosperm species such as *P. glauca* and *P. taeda* [59]. ClMYB3/4/5 carried a unique motif with core sequence of (E/I)LIETADFFQ.

The similar motif was also detected when aligning with the *P. glauca* orthologs (Pg5/10/13). Still, none were found in the orthologs of *O. sativa*, *E. grandis*, and *P. trichocarpa*, suggesting this motif was specific to gymnosperm lineage. Since the role of this unique motif is unknown, the regulatory process involving ClMYB3/4/5 may en-route differently, though they share the repressive EAR motif responding to external stimuli and hormonal signals [69] with angiosperms. Moreover, in response to the bending treatment, *ClMYB3* and *ClMYB5* were down- and upregulated respectively with the development of compression wood. These all observations suggested that they might be recruited differently in the regulatory process, hence targeting different downstream targets, and playing different roles in wood formation responding to defence and stress stimulation.

## Conclusions

A total of 52 R2R3-MYB encoding genes with complete ORFs were obtained from the full-length transcriptomic libraries of *C. lanceolata* developmental stem. Though they were mostly segregated in the subgroups harbouring those from monocot and dicot species, a minor portion of the subgroups strictly carried members of gymnosperm origin. The asymmetric distribution of the ClMYBs among woody-preferential and woody-expanded subgroups indicated that woody species of angiosperm and gymnosperm underwent gene expansion in the woody-expanded subgroups. Nine candidates shared the same subgroups with *A. thaliana*, from which ClMYB1/2/5/26/27/49 may be directly involved in the upregulation of cell wall development. Functional analysis of ClMYB1 in *N. banthamiana* demonstrated its positive regulation in lignification. Our findings will contribute to a greater understanding of the evolutionary relationship of ClMYBs with those from the angiosperms. The information obtained will provide important clues for the selection of candidates for use in improving the wood quality of *C. lanceolata*.

## Declarations

### Authors' contribution

HZ performed the experiments, data analysis, prepared figures and table and wrote the manuscript text. SLC prepared figures and table, wrote and revised the manuscript text. BP revised the manuscript text. XH and EL analysed the data and prepared figures. ZT contributed reagents/materials/analysis tools. HH conceived and designed the experiments, performed the experiments, analysed the data, contributed reagents/materials/analysis tools, prepared figures and tables. All authors reviewed the manuscript and approved the final draft.

### Funding

no additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Availability of data and material

## Ethics approval and consent to participate

Not applicable

## Consent for publication

Not applicable

## Competing Interests

The authors declare that they have no competing interests.

## Author details

[1]State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Lin'an, Hangzhou 311300, China

# References

1. Shi D, Lebovka I, López-Salmerón V, Sanchez P, Greb T. Bifacial cambium stem cells generate xylem and phloem during radial plant growth. Development. 2019;146:dev171355.

2. Jokipii-Lukkari S, Delhomme N, Schiffthaler B, Mannapperuma C, Prestele J, Nilsson O, et al. Transcriptional roadmap to seasonal variation in wood formation of Norway spruce. Plant Physiol. 2018;176:2851–70.

3. Demura T, Fukuda H. Transcriptional regulation in wood formation. Trends Plant Sci. 2007;12:64–70.

4. Groover AT. What genes make a tree a tree? Trends Plant Sci. 2005;10:210–4.

5. Zhong R, Mccarthy RL, Lee C, Ye ZH. Dissection of the transcriptional program regulating secondary wall biosynthesis during wood formation in poplar. Plant Physiol. 2011;157:1452–68.

6. Bedon F, Grima-Pettenati J, Mackay J. Conifer R2R3-MYB transcription factors: Sequence analyses and gene expression in wood-forming tissues of white spruce (*Picea glauca*). BMC Plant Biol. 2007;7:1–17.

7. Karpinska B, Karlsson M, Srivastava M, Stenberg A, Schrader J, Sterky F, et al. MYB transcription factors are differentially expressed and regulated during secondary vascular tissue development in hybrid aspen. Plant Mol Biol. 2004;56:255–70.

8. Stracke R, Werber M, Weisshaar B. The R2R3-MYB gene family in *Arabidopsis thaliana*. Curr Opin Plant Biol. 2001;4:447–56.

9. Zhao K, Bartley LE. Comparative genomic analysis of the R2R3 MYB secondary cell wall regulators of Arabidopsis, poplar, rice, maize, and switchgrass. BMC Plant Biol. 2014;14.

10. Zhong R, McCarthy RL, Haghighat M, Ye ZH. The Poplar MYB master switches bind to the SMRE site and activate the secondary wall biosynthetic program during wood formation. PLoS One. 2013;8:18–21.

11. Zhong R, Ye ZH. Secondary cell walls: Biosynthesis, patterned deposition and transcriptional regulation. Plant Cell Physiol. 2015;56:195–214.

12. Soler M, Camargo ELO, Carocha V, Cassan-Wang H, San Clemente H, Savelli B, et al. The *Eucalyptus grandis* R2R3-MYB transcription factor family: Evidence for woody growth-related evolution and function. New Phytol. 2015;206:1364–77.

13. McCarthy RL, Zhong R, Fowler S, Lyskowski D, Piyasena H, Carleton K, et al. The poplar MYB transcription factors, PtrMYB3 and PtrMYB20, are involved in the regulation of secondary wall biosynthesis. Plant Cell Physiol. 2010;51:1084–90.

14. Goicoechea M, Lacombe E, Legay S, Mihaljevic S, Rech P, Jauneau A, et al. EgMYB2, a new transcriptional activator from Eucalyptus xylem, regulates secondary cell wall formation and lignin biosynthesis. Plant J. 2005;43:553–67.

15. Soler M, Plasencia A, Lepikson-Neto J, Camargo ELO, Dupas A, Ladouce N, et al. The woody-preferential gene EgMYB88 regulates the biosynthesis of phenylpropanoid-derived compounds in wood. Front Plant Sci. 2016;7 September.

16. Soler M, Plasencia A, Larbat R, Pouzet C, Jauneau A, Rivas S, et al. The Eucalyptus linker histone variant EgH1.3 cooperates with the transcription factor EgMYB1 to control lignin biosynthesis during wood formation. New Phytol. 2017;213:287–99. doi:https://doi.org/10.1111/nph.14129.

17. Bomal C, Bedon F, Caron S, Mansfield SD, Levasseur C, Cooke JEK, et al. Involvement of *Pinus taeda* MYB1 and MYB8 in phenylpropanoid metabolism and secondary cell wall biogenesis: A comparative in planta analysis. J Exp Bot. 2008;59:3925–39.

18. Patzlaff A, McInnis S, Courtenay A, Surman C, Newman LJ, Smith C, et al. Characterisation of a pine MYB that regulates lignification. Plant J. 2003;36:743–54.

19. Patzlaff A, Newman LJ, Dubos C, Whetten RW, Smith C, McInnis S, et al. Characterisation of PtMYB1, an R2R3-MYB from pine xylem. Plant Mol Biol. 2003;53:597–608.

20. Jiang CK, Rao GY. Insights into the diversification and evolution of R2R3-MYB transcription factors in plants. Plant Physiol. 2020;183:637–55.

21. Du H, Liang Z, Zhao S, Nan MG, Tran LSP, Lu K, et al. The evolutionary history of R2R3-MYB proteins across 50 eukaryotes: New insights into subfamily classification and expansion. Sci Rep. 2015;5

June:1−7.

22. Akiyoshi N, Nakano Y, Sano R, Kunigita Y, Ohtani M, Demura T. Involvement of VNS NAC-domain transcription factors in tracheid formation in *Pinus taeda*. Tree Physiol. 2020;40:704−16.

23. Wang G, Gao Y, Yang L, Shi J. Identification and analysis of differentially expressed genes in differentiating xylem of Chinese fir (*Cunninghamia lanceolata*) by suppression subtractive hybridization. Genome. 2007;50:1141−55.

24. Huang HH, Xu LL, Tong ZK, Lin EP, Liu QP, Cheng LJ, et al. De novo characterization of the Chinese fir (*Cunninghamia lanceolata*) transcriptome and analysis of candidate genes involved in cellulose and lignin biosynthesis. BMC Genomics. 2012;13:1−14.

25. Wang G, Gao Y, Wang J, Yang L, Song R, Li X, et al. Overexpression of two cambium-abundant Chinese fir (*Cunninghamia lanceolata*) alpha-expansin genes ClEXPA1 and ClEXPA2 affect growth and development in transgenic tobacco and increase the amount of cellulose in stem cell walls. Plant Biotechnol J. 2011;9:486−502.

26. Hellgren JM, Olofsson K, Sundberg B. Patterns of auxin distribution during gravitational induction of reaction wood in poplar and pine. Plant Physiol. 2004;135:212−20. doi:10.1104/pp.104.038927.

27. McLaughlin SP, Schuck SM. Fiber properties of several species of agavaceae from the Southwestern United States and Northern Mexico. Econ Bot. 1991;45:480−6. doi:10.1007/BF02930710.

28. 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.

29. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7:e1002195. doi:10.1371/journal.pcbi.1002195.

30. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23:2947−8. doi:10.1093/bioinformatics/btm404.

31. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870−4.

32. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: An integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020;13:1194−202. doi:https://doi.org/10.1016/j.molp.2020.06.009.

33. Doster MA, Bostock RM. Quantification of lignin formation in almond bark in response to wounding and infection by Phytophthora species. Phytopathology. 1988;78:473−7. doi:10.1094/phyto-78-473.

34. Rabinowicz PD, Braun EL, Wolfe AD, Bowen B, Grotewold E. Maize R2R3 Myb genes: Sequence analysis reveals amplification in the higher plants. Genetics. 1999;153:427−44.

35. Dias AP, Braun EL, McMullen MD, Grotewold E. Recently duplicated maize R2R3 MYB genes provide evidence for distinct mechanisms of evolutionary divergence after duplication. Plant Physiol. 2003;131:610−20.

36. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. MYB transcription factors in Arabidopsis. Trends Plant Sci. 2010;15:573−81. doi:10.1016/j.tplants.2010.06.005.

37. Soler M, Camargo EL, Carocha V, Cassan-Wang H, San Clemente H, Savelli B, et al. The *Eucalyptus grandis* R2R3-MYB transcription factor family: evidence for woody growth-related evolution and function. New Phytol. 2015;206:1364–77.

38. Zhong R, Richardson EA, Ye ZH. The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in Arabidopsis. Plant Cell. 2007;19:2776–92.

39. Ko JH, Jeon HW, Kim WC, Kim JY, Han KH. The MYB46/MYB83-mediated transcriptional regulatory programme is a gatekeeper of secondary wall biosynthesis. Ann Bot. 2014;114:1099–107.

40. Zhong R, Lee C, Zhou J, McCarthy RL, Ye ZH. A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis. Plant Cell. 2008;20:2763–82.

41. Zhou J, Lee C, Zhong R, Ye ZH. MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in Arabidopsis. Plant Cell. 2009;21:248–66.

42. Geng P, Zhang S, Liu J, Zhao C, Wu J, Cao Y, et al. MYB20, MYB42, MYB43, and MYB85 regulate phenylalanine and lignin biosynthesis during secondary cell wall formation. Plant Physiol. 2020;182:1272–83.

43. Yang C, Xu Z, Song J, Conner K, Barrena GV, Wilson ZA. Arabidopsis MYB26/MALE STERILE35 regulates secondary thickening in the endothecium and is essential for anther dehiscence. Plant Cell. 2007;19:534–48.

44. Yang C, Song J, Ferguson AC, Klisch D, Simpson K, Mo R, et al. Transcription factor MYB26 is key to spatial specificity in anther secondary thickening formation. Plant Physiol. 2017;175:333–50.

45. Penfield S, Meissner RC, Shoue DA, Carpita NC, Bevan MW. MYB61 is required for mucilage deposition and extrusion in the arabidopsis seed coat. Plant Cell. 2001;13:2777–91.

46. Romano JM, Dubos C, Prouse MB, Wilkins O, Hong H, Poole M, et al. AtMYB61, an R2R3-MYB transcription factor, functions as a pleiotropic regulator via a small gene network. New Phytol. 2012;195:774–86.

47. Jin H. Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. EMBO J. 2000;19:6150–61.

48. Preston J, Wheeler J, Heazlewood J, Song FL, Parish RW. AtMYB32 is required for normal pollen development in *Arabidopsis thaliana*. Plant J. 2004;40:979–95.

49. Wang S, Li E, Porth I, Chen JG, Mansfield SD, Douglas CJ. Regulation of secondary cell wall biosynthesis by poplar R2R3 MYB transcription factor PtrMYB152 in Arabidopsis. Sci Rep. 2014;4:3–9.

50. Legay S, Sivadon P, Blervacq AS, Pavy N, Baghdady A, Tremblay L, et al. EgMYB1, an R2R3 MYB transcription factor from eucalyptus negatively regulates secondary cell wall formation in Arabidopsis and poplar. New Phytol. 2010;188:774–86.

51. Ohman D, Demedts B, Kumar M, Gerber L, Gorzsas A, Goeminne G, et al. MYB103 is required for FERULATE-5-HYDROXYLASE expression and syringyl lignin biosynthesis in Arabidopsis stems. Plant J. 2013;73:63–76.

52. Zhong R, Richardson EA, Ye Z-H. The MYB46 Transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in Arabidopsis. Plant Cell Online. 2007;19:2776–92.

53. Goicoechea M, Lacombe E, Legay S, Mihaljevic S, Rech P, Jauneau A, et al. EgMYB2, a new transcriptional activator from Eucalyptus xylem, regulates secondary cell wall formation and lignin biosynthesis. Plant J. 2005;43:553–67.

54. Zhong R, Ye ZH. MYB46 and MYB83 bind to the SMRE sites and directly activate a suite of transcription factors and secondary wall biosynthetic genes. Plant Cell Physiol. 2012;53:368–80.

55. Zhong R, McCarthy RL, Haghighat M, Ye Z-H. The poplar MYB master switches bind to the SMRE site and activate the secondary wall biosynthetic program during wood formation. PLoS One. 2013;8:e69219. doi:10.1371/journal.pone.0069219.

56. Newman LJ, Perazza DE, Juda L, Campbell MM. Involvement of the R2R3-MYB, AtMYB61, in the ectopic lignification and dark-photomorphogenic components of the det3 mutant phenotype. Plant J. 2004;37:239–50.

57. Soler M, Plasencia A, Lepikson-Neto J, Camargo ELO, Dupas A, Ladouce N, et al. The woody-preferential gene EgMYB88 regulates the biosynthesis of phenylpropanoid-derived compounds in wood. Frontiers in Plant Science. 2016;7:1422. doi:10.3389/fpls.2016.01422.

58. Zheng T, Tan W, Yang H, Zhang L, Li T, Liu B, et al. Regulation of anthocyanin accumulation via MYB75/HAT1/TPL-mediated transcriptional repression. PLoS Genet. 2019;15:1–28. doi:10.1371/journal.pgen.1007993.

59. Bedon F, Bomal C, Caron S, Levasseur C, Boyle B, Mansfield SD, et al. Subgroup 4 R2R3-MYBs in conifer trees: Gene family expansion and contribution to the isoprenoid-and flavonoid-oriented responses. J Exp Bot. 2010;61:3847–64.

60. Zhang S, Liu J, Lou X, Liu Y, Tong Z, Huang H. Morphological characteristics of cells and main metabolic components in xylem of Cunninghamia lanceolata compression wood. J Beijing For Univ. 2015;37:126–33.

61. Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu S-H. Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. Plant Physiol. 2008;148:993 LP – 1003. doi:10.1104/pp.108.122457.

62. Pavy N, Pelgas B, Laroche J, Rigault P, Isabel N, Bousquet J. A spruce gene map infers ancient plant genome reshuffling and subsequent slow evolution in the gymnosperm lineage leading to extant conifers. BMC Biol. 2012;10 October.

63. Panchy N, Lehti-Shiu M, Shiu SH. Evolution of gene duplication in plants. Plant Physiol. 2016;171:2294–316.

64. Nakano Y, Nishikubo N, Goue N, Ohtani M, Yamaguchi M, Katayama Y, et al. MYB transcription factors orchestrating the developmental program of xylem vessels in Arabidopsis roots. Plant tissue Cult Lett. 2010;27:267–72. doi:10.5511/plantbiotechnology.27.267.

65. Li C, Wang X, Lu W, Liu R, Tian Q, Sun Y, et al. A poplar R2R3-MYB transcription factor, PtrMYB152, is involved in regulation of lignin biosynthesis during secondary cell wall formation. Plant Cell, Tissue

Organ Cult. 2014;119:553–63. doi:10.1007/s11240-014-0555-8.

66. Ko JH, Kim WC, Han KH. Ectopic expression of MYB46 identifies transcriptional regulatory genes involved in secondary wall biosynthesis in Arabidopsis. Plant J. 2009;60:649–65.

67. Öhman D, Demedts B, Kumar M, Gerber L, Gorzsás A, Goeminne G, et al. MYB103 is required for FERULATE-5-HYDROXYLASE expression and syringyl lignin biosynthesis in Arabidopsis stems. Plant J. 2013;73:63–76. doi:10.1111/tpj.12018.

68. Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W. Lignin Biosynthesis and Structure. Plant Physiol. 2010;153:895 LP – 905. doi:10.1104/pp.110.155119.

69. Liu J, Osbourn A, Ma P. MYB transcription factors as regulators of phenylpropanoid metabolism in plants. Mol Plant. 2015;8:689–708. doi:https://doi.org/10.1016/j.molp.2015.03.012.

70. Wang Y, Song F, Zhu J, Zhang S, Yang Y, Chen T, et al. GSA: Genome Sequence Archive*. In: Genomics, Proteomics and Bioinformatics. Beijing Genomics Institute; 2017. p. 14–8.

71. Database Resources of the BIG Data Center in 2019. Nucleic Acids Res. 2019;47:D8–14. doi:10.1093/nar/gky993.

# Supplementary Files

**Figure S1** Cross-sections stained with Phloroglucinol-HCl for the different lignification stems of *C. lanceolata*. Five stem segments from the top to bottom of a one-year-old branch were collected and marked sequentially S1, S2, S3, S4, and S5.

**Figure S2** Phylogenetic tree of 629 R2R3-MYB proteins from *Cunninghamia lanceolata* (Cl), *Eucalyptus grandis* (Egr), *Populus trichocarpa* (Ptr), *Arabidopsis thaliana* (At), *Oryza sativa* (Os), *Pinus taeda* (Pt), and *Picea glauca* (Pg). All proteins were divided into 47 subgroups.

**Figure S3** Expression profiles of 12 *ClMYB* genes in eight organs and tissues using qRT-PCR. Data were normalised to *ClActin* gene. The expression level in S1 was used as a reference and was set at 1. Error bars indicate standard deviation of three biological replicates. The corresponding FPKM values were listed. The *r* represented the correlation coefficient between qRT-PCR and RNA-Seq data, and the *p* was the P-value. L: leaves, S1-5: stem segments from top to bottom, X: xylem, B: bark.

**Table S1** Primers used in the analysis of qRT-PCR, subcellular localization, yeast hybridization, overexpression, and semi-qRT-PCR.

**Table S2** Physicochemical properties and protein-coding sequences of ClMYB transcription factors.

**Table S3** FPKM values of 52 *ClMYB* genes in different organs or tissues.

**Data S1** Descriptions and sequence logo plots of motifs in ClMYB proteins.

**Data S2** Tracheid wall thickness of different samples.

**Data S3** Quantification Cq results of 9 *ClMYB* genes in qRT-PCR.

**Data S4** Lignin contents of stems from transgenic lines and wide-type.

**Electrophoretogram 1** *ClMYB1* PCR detection of genomic DNA in transgenic *N. benthamiana*. M: DNA marker (DL2000, TaKARa), WT: wide-type, T3.1 and T3.2: two T3 generation lines.

**Electrophoretogram 2** *ClMYB1* transcripts detection by semi-quantitative RT-PCR in transgenic *N. benthamiana*. A, Transcript detections of *ClActin* (lane 2-7) and *ClUBC* (lane 8-13). B, Transcript detections of *ClGAPDH* (lane 2-7) and *ClMYB1* (lane 8-13). M: DNA marker (DL2000, TaKaRa), WT: wide-type, T3.1 and T3.2: two T3 generation lines. T2.1, T2.2 and T2.3: three T2 generation lines.