



OPEN

Evaluation of reference genes and characterization of the *MYBs* in xylem radial change of Chinese fir stem

Kui-Peng Li¹, Wei Li², Gui-Yun Tao³ & Kai-Yong Huang^{1✉}

The radial change (RC) of tree stem is the process of heartwood formation involved in complex molecular mechanism. Chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook.), an evergreen species, is an important fast-growing timber tree in southern China. In this study, the top four stable genes (*IDH*, *UBC2*, *RCA* and *H2B*) were selected in RC tissues of 15 years old Chinese fir stem (RC15) and the genes (*H2B*, *18S*, *TIP41* and *GAPDH*) were selected in RC tissues of 30 years old Chinese fir stem (RC30). The stability of the reference genes is higher in RC30 than in RC15. Sixty-one *MYB* transcripts were obtained on the PacBio Sequel platform from woody tissues of one 30 years old Chinese fir stem. Based on the number of MYB DNA-binding domain and phylogenetic relationships, the *CIMYB* transcripts contained 21 transcripts of MYB-related proteins (1R-MYB), 39 transcripts of R2R3-MYB proteins (2R-MYB), one transcript of R1R2R3-MYB protein (3R-MYB) belonged to 18 function-annotated clades and two function-unknown clades. In RC woody tissues of 30 years old Chinese fir stem, *CIMYB22* was the transcript with the greatest fold change detected by both RNA-seq and qRT-PCR. Reference genes selected in this study will be helpful for further verification of transcript abundance patterns during the heartwood formation of Chinese fir.

Gymnosperms and angiosperms are categorized into woody plants and herbaceous plants. The stem of woody plants is mainly composed of secondary xylem¹. Wood formation in tree species is performed with the complex coordination of cell differentiation and secondary cell wall (SCW) thickening, which are two highly ordered processes initiated from the vascular cambium². The structure of gymnosperms wood is relatively simple, consisting of tracheids, rays and parenchyma. On the other hand, angiosperm wood is structurally more complex, composed of vessels, fibers, rays, and parenchyma³.

The vascular bundle contains phloem cells, cambium cells, and xylem cells from the outer layer to the inner layer of tree stem⁴. Wood developed from several major steps is mainly composed by secondary xylem cells of trees⁵. Sapwood (SW) is derived from vascular cambium has living cells and is located in the outer layers of wood. Heartwood (HW) formation is the process of sapwood senescence with the death of parenchyma cells and the cease of all physiological activity. The transition zone (TZ) is located in between the SW and HW⁶. Parenchyma tissue contained ray and axial parenchyma cells in secondary xylem represents the bulk of living cells. Ray parenchyma cells are involved in HW formation. Programmed cell death (PCD) plays an important role in secondary xylem cells and HW formation in woody plants⁴.

Gene expression level analysis plays an essential role in molecular biology⁷. Real-time quantitative polymerase chain reaction (qRT-PCR) is a widely applied technique for precise quantification of gene expression and is widely used in modern biology due to its prominent advantages of high sensitivity, specificity, and repeatability^{8,9}. Stable internal reference genes are essential to normalize gene expression for obtaining accurate measurement results in the qRT-PCR assay⁷. Reference genes are usually derived from housekeeping genes, which maintain the basic processes of cell cycle^{10,11}. However, the housekeeping genes expression levels vary greatly in species, tissues, and under certain experimental conditions¹². Therefore, selecting stable reference genes according to

¹Guangxi Key Laboratory of Superior Timber Trees Resource Cultivation and Key Laboratory of Central South Fast-Growing Timber Cultivation of Forestry Ministry of China, Guangxi Forestry Research Institute, Nanning 530002, People's Republic of China. ²College of Landscape Architecture and Forestry, Qingdao Agricultural University, Qingdao 266109, People's Republic of China. ³Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, People's Republic of China. ✉email: huangky73@163.com

Gene symbol	Gene description	Primer sequence (forward/reverse)	Product (bp)	Amplification efficiency (%)	R ²
<i>GAPDH</i>	Glyceraldehyde-3-phosphate dehydrogenase	GGTCACTGGTTCTGCCAAAT/TGACAACGAGTG GGGATACA	101	88.1	0.999
<i>H2B</i>	Histone H2B	TATCGGAATTCCAGCAAGG/ATACCTGGCCAA TCTGGATG	97	92.0	0.999
<i>TIP41</i>	Transcription and replication	CAAGCCAGTCTCTCCAAAG/GAGAGCAGGACA TGGAGGAG	199	94.1	0.994
<i>18S</i>	18S ribosomal protein	GCTTCTTGCTCTACCGGATG/AATGCAACATCA AGCATGGA	144	90.5	0.998
<i>IDH</i>	Isocitrate dehydrogenase	CTTTTCATGCAGTCCCAGGT/TTGCGCTAGCTG AAGCTGTA	122	87.7	0.999
<i>PXMP2</i>	Peroxisomal membrane protein	TAGTGCAGGCTTGAGGCTTT/AGTTTCCAGTTT GCCACCAC	159	101.4	0.999
<i>RCA</i>	Rubisco activase	GCTTGCCAATGCTCTCTACC/TTTTATTGGGCT CCAACCAG	208	102.7	0.996
<i>UBC2</i>	Ubiquitin-conjugating enzyme	TTGTTTTGGCAGTCTGCTTG/GCTCGTTTCTGA TGGCTTTC	216	84.6	0.998
<i>RPL2</i>	Large subunit ribosomal protein	CCGCTGCTCTTTATCCTCAG/GCATCCGAGAGG GATTATGA	223	100.9	0.999
<i>DnaJ</i>	Chaperone protein dnaJ	TGGACCTGGGGATATGTA/CATGCCAACTGA AGCAAAGA	98	95.9	0.999
<i>TUBα</i>	Tubulin alpha	CGGAGACTTTGTGCAGTGA/TCCTGAATGTCG TGCTTGAG	118	87.4	0.994
<i>TUBβ</i>	Tubulin beta	TGCAGACGAGGATGCTTATG/GCAATTGCAGAA GCACAGAA	97	106.0	0.998

Table 1. Primer characteristics of 12 candidate reference genes for qRT-PCR.

the different sample types and various experimental conditions is the primary step for a reliable gene expression analysis by qRT-PCR assay¹³.

Transcription factors (TFs) are proteins capable of controlling the expression of the target genes and play an important role in regulating of multiple biological processes. According to conserved domains, TFs can be divided into different gene families. The MYB family members contain a conserved MYB binding domain consisting of one to four imperfect tandem repeats (R) of about 52 amino acids at the N-terminus. Depending on the number of the adjacent MYB repeats, MYB genes can be classified into four major classes: MYB-related (1R-MYB), R2R3-MYB (2R-MYB), R1R2R3-MYB (3R-MYB), and 4R-like MYB (4RMYB)¹⁴. MYB proteins, especially R2R3-MYB proteins, are the key factors regulating plant development, primary and secondary metabolism, responses to biotic and abiotic stresses, lignin biosynthesis, xylem secondary cell wall (SCW) formation, and PCD^{15,16}.

The radial change (RC) of tree stem is the process of heartwood formation and the process involved in PCD of living cells including ray and axial parenchyma cells. Up to our knowledge, there were few previous studies related to the selection of suitable reference genes for target gene expression in RC of tree species stem. Chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook) is a fast-growing native, allogamous, long-lived tree species mainly distributed in southern China and has been planted for timber production for over 1000 years¹⁷. Chinese fir accounts for 20–30% of the total commercial timber production in China¹⁸. The tree species is a gymnosperm in the cypress family Taxodiaceae. Taxodiaceae traditionally defined includes 10 genera and 16 species¹⁹. The somatic chromosome number of Chinese fir is $2n = 2x = 22$ which is in concurrence with other diploid members of Taxodiaceae²⁰. The diploid genome sizes of Chinese fir (28.34 pg/2C) is the largest among Taxodiaceae²¹.

In the present study, expression stability of 12 candidate reference genes were assessed by qRT-PCR in RC of Chinese fir stem. Phylogenetic tree was constructed to predict the biological functions of *CIMYB* transcripts. The expression profile of 61 *CIMYB* transcripts in the RC were analyzed by RNA-seq. Furthermore, the expression patterns of 25 *CIMYB* transcripts were validated by qRT-PCR using the four selected reference genes for normalization. This research provides a new understanding of suitable reference genes in RC of tree species and expression pattern and biological functions of *CIMYB* transcripts.

Results

Expression stability of candidate reference genes. The primers of 12 candidate reference genes were listed in Table 1. The specificity of each primer pair was verified by 2.0% agarose gel electrophoresis (Fig. S1) and melting curves analysis (Fig. S2). The primer pair with only a single qRT-PCR product was selected. Amplification efficiency and R² were counted by the standard curves (Fig. S3). BestKeeper ranks the candidate reference genes according to the Pearson correlation coefficient (r). The reference gene with high stability has a higher r value. In the radial change of 15 years old Chinese fir stem (RC15), *IDH*, *RCA*, *UBC2*, and *RPL2* ranked as the top four stable reference genes (Table 2). In radial change of 30 years old Chinese fir stem (RC30), the results indicated that the *18S*, *H2B*, *IDH*, and *GAPDH* were the top four stable genes (Table 3).

In the GeNorm analysis, *H2B* and *UBC2* had the lowest M values and thus the highest expression stability, followed by *Rpl2* and *IDH* in RC15 tissues (Fig. 1A). In GeNorm analysis of the RC30 tissues, *H2B* and *18S* were

Gene	Geometric mean (Ct)	Arithmetic mean (Ct)	Minimum (Ct)	Maximum (Ct)	Coefficient of Variance ± Standard deviation (CV ± SD)	Pearson correlation coefficient (r)
IDH	25.47	25.59	23.25	32.02	7.87 ± 2.01	0.990
RCA	26.28	26.52	22.81	34.19	10.98 ± 2.91	0.985
UBC2	22.86	23.08	20.02	31.52	10.96 ± 2.53	0.972
RPL2	20.18	20.43	17.68	29.27	12.89 ± 2.63	0.964

Table 2. Expression stability of top four stable reference genes in RC15 evaluated by BestKeeper.

Gene	Geometric mean (C _t)	Arithmetic mean (C _t)	Minimum (C _t)	Maximum (C _t)	Coefficient of Variance ± Standard deviation (CV ± SD)	Pearson correlation coefficient (r)
18S	27.24	27.30	25.25	32.33	5.56 ± 1.52	0.989
H2B	23.28	23.36	21.23	28.19	6.74 ± 1.57	0.987
IDH	24.75	24.78	23.31	28.27	4.12 ± 1.02	0.985
GAPDH	24.30	24.35	22.63	28.23	5.18 ± 1.26	0.984

Table 3. Expression stability of top four stable reference genes in RC30 evaluated by BestKeeper.

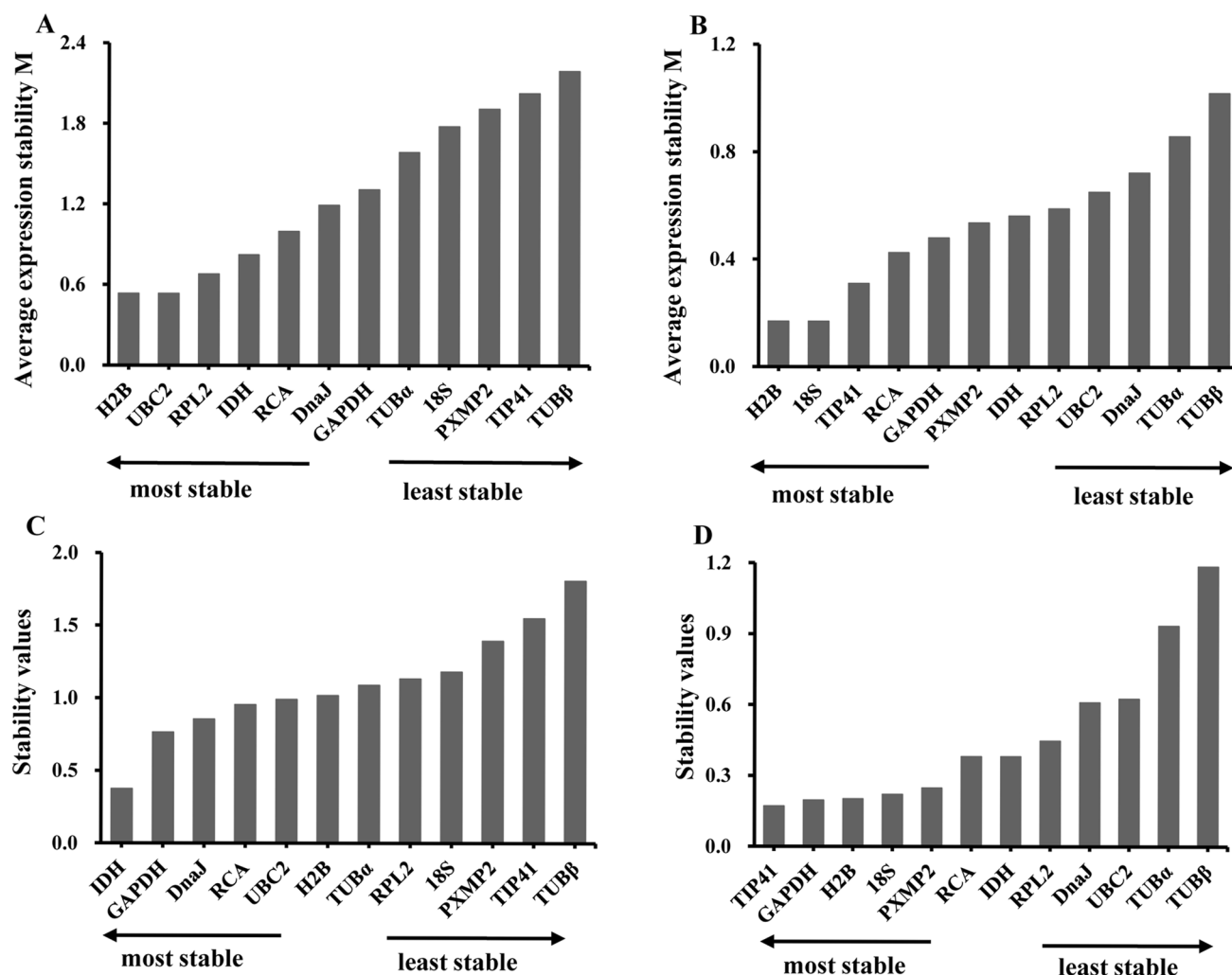


Figure 1. Ranking and expression stability values of the reference genes in RC15 and RC30 tissues by geNorm and NormFinder. (A) Reference genes in RC15 evaluated by geNorm. (B) Reference genes in RC30 evaluated by geNorm. (C) Reference genes in RC15 evaluated by NormFinder. (D) Reference genes in RC30 evaluated by NormFinder.

Ranking order	GeNorm	NormFinder	BestKeeper	Comprehensive ranking (mean rank value)
1	<i>H2B/UBC2</i>	<i>IDH</i>	<i>IDH</i>	<i>IDH</i> (1.7)
2	<i>RPL2</i>	<i>GAPDH</i>	<i>RCA</i>	<i>UBC2</i> (3.0)
3	<i>IDH</i>	<i>DnaJ</i>	<i>UBC2</i>	<i>RCA</i> (3.3)
4	<i>RCA</i>	<i>RCA</i>	<i>RPL2</i>	<i>H2B</i> (4.0)

Table 4. Comprehensive ranking of 12 reference genes in RC15.

Ranking order	GeNorm	NormFinder	BestKeeper	Comprehensive ranking (mean rank value)
1	<i>H2B/18S</i>	<i>TIP41</i>	<i>18S</i>	<i>H2B</i> (2.0)
2	<i>TIP41</i>	<i>GAPDH</i>	<i>H2B</i>	<i>18S</i> (2.0)
3	<i>RCA</i>	<i>H2B</i>	<i>IDH</i>	<i>TIP41</i> (2.7)
4	<i>GAPDH</i>	<i>18S</i>	<i>GAPDH</i>	<i>GAPDH</i> (3.3)

Table 5. Comprehensive ranking of 12 reference genes in RC30.

the most stable genes, followed by *TIP41* and *RCA* (Fig. 1B). The NormFinder analysis ranked *IDH*, *GAPDH*, *DnaJ* and *RCA* in the top positions for RC15 samples (Fig. 1C), and in NormFinder analysis of the RC30 tissues, *TIP41* exhibited the most stable expression, followed by *GAPDH*, *H2B* and *18S* (Fig. 1D).

Comprehensive ranking of the reference genes. The ranking order of each reference gene was obtained from 1 (most stable) to 11 (least stable) by geNorm and from 1 (most stable) to 12 (least stable) by BestKeeper and NormFinder. Different ranking of candidate genes among algorithms were revealed in other woody plants^{22–24}. According to the arithmetic mean value of the ranking order of each gene, the top 4 stable genes were obtained. *IDH*, *UBC*, *RCA*, and *H2B* exhibited higher stability of gene expression in XRC15, and the mean rank value of the genes were 1.7, 3.0, 3.3, and 4.0 respectively (Table 4). *H2B* (2.0), *18S* (2.0), *TIP41* (2.7) and *GAPDH* (3.3) were the top 4 stable genes in XRC30 (Table 5).

Phylogenetic analysis and functional prediction of CIMYB transcripts. To evaluate the evolutionary relationships of the *CIMYB* transcripts, an NJ unrooted phylogenetic tree was constructed using amino acid sequences of 61 *CIMYB*s and 135 *AtMYB*s (*Arabidopsis thaliana*) (Fig. 2). Based on the number of MYB DNA-binding domain and phylogenetic relationships, we identified 21MYB-related proteins (1R-MYB), 39 R2R3-MYB proteins (2R-MYB), 1R1R2R3-MYB proteins (3R-MYB). According to the alignment result, all MYB members from the two species were subdivided into twenty subgroups (designated as C1–C20). All subgroups were common to the two species. 59 *CIMYB*s belonging to 18 function-annotated subgroups, and 2 *CIMYB*s belonging to two function-unknown subgroups were found. Because these *CIMYB* transcripts were captured from different woody tissues in RC of Chinese fir stem, *CIMYB* transcripts were involved in divergent functions including stresses response, cell differentiation, lignin biosynthesis, and secondary cell wall formation during heartwood formation of Chinese fir.

R2R3-MYB family genes have been conducted in depth studies and the function of genes are relatively clear^{25–28}. To explore the phylogenetic relationships of the R2R3-MYB transcripts, two unrooted phylogenetic trees were constructed with 41 R2R3-MYBs (Chinese fir), 135 R2R3-MYBs (*Arabidopsis thaliana*), 194 R2R3-MYBs (*Populus trichocarpa*), and 142 R2R3-MYBs (*Eucalyptus grandis*) by NJ and ML method. A similar topology was obtained between NJ tree (Fig. S4) and ML tree (Fig. S5). Some homologs between Chinese fir and *Arabidopsis thaliana* were clustered within the same subgroup may share similar functions. For example, *CIMYB* transcripts and *AtMYB* genes of the C7 subgroup in Fig. 2 (*CIMYB18*, *CIMYB24*, *CIMYB31*, *CIMYB41*, *AtMYB44*, and *AtMYB73*) and C8 subgroup (*CIMYB13*, *CIMYB54*, *CIMYB59*, *AtMYB33*, *AtMYB65*, and *AtMYB101*) were clustered within the same subgroup by different methods.

***CIMYB*s expression profile in RC30 tissues.** *CIMYB* transcript abundance was analysed by RNA-seq in 4 woody tissues of Chinese fir (Fig. 3). Based on RNA-seq expression, 61 *CIMYB* transcripts revealed 12 expression clusters by *K-means*. Thirty-five *CIMYB* transcripts had different expression patterns (fold change > 2, $P < 0.05$) in woody tissues (Table S1). Expression of 25 *CIMYB* transcripts validated by qRT-PCR in the same woody tissues as RNA-seq. Normalized to *H2B*, *18S*, *TIP41*, and *GAPDH*, a heatmap was constructed with the expression level of *CIMYB* transcripts (Fig. 4). Based on qPCR-based expression, 25 *CIMYB* transcripts revealed six expression clusters by *K-means*.

Three *CIMYB* transcripts (*CIMYB10*, *CIMYB33*, and *CIMYB41*) had the same expression pattern with the *K-means* cluster by both qRT-PCR and RNA-seq. Based on six pairwise comparisons in the four woody tissues (X1, X2, X3, and X4), a total of 18 differentially expressed transcripts were identified by RNA-seq (Table S1). Eight differentially expressed transcripts identified by qRT-PCR were also identified by RNA-seq. *CIMYB22* was the transcript with the greatest fold change based on RNA-seq data and it was up-regulated with a 6525-fold change

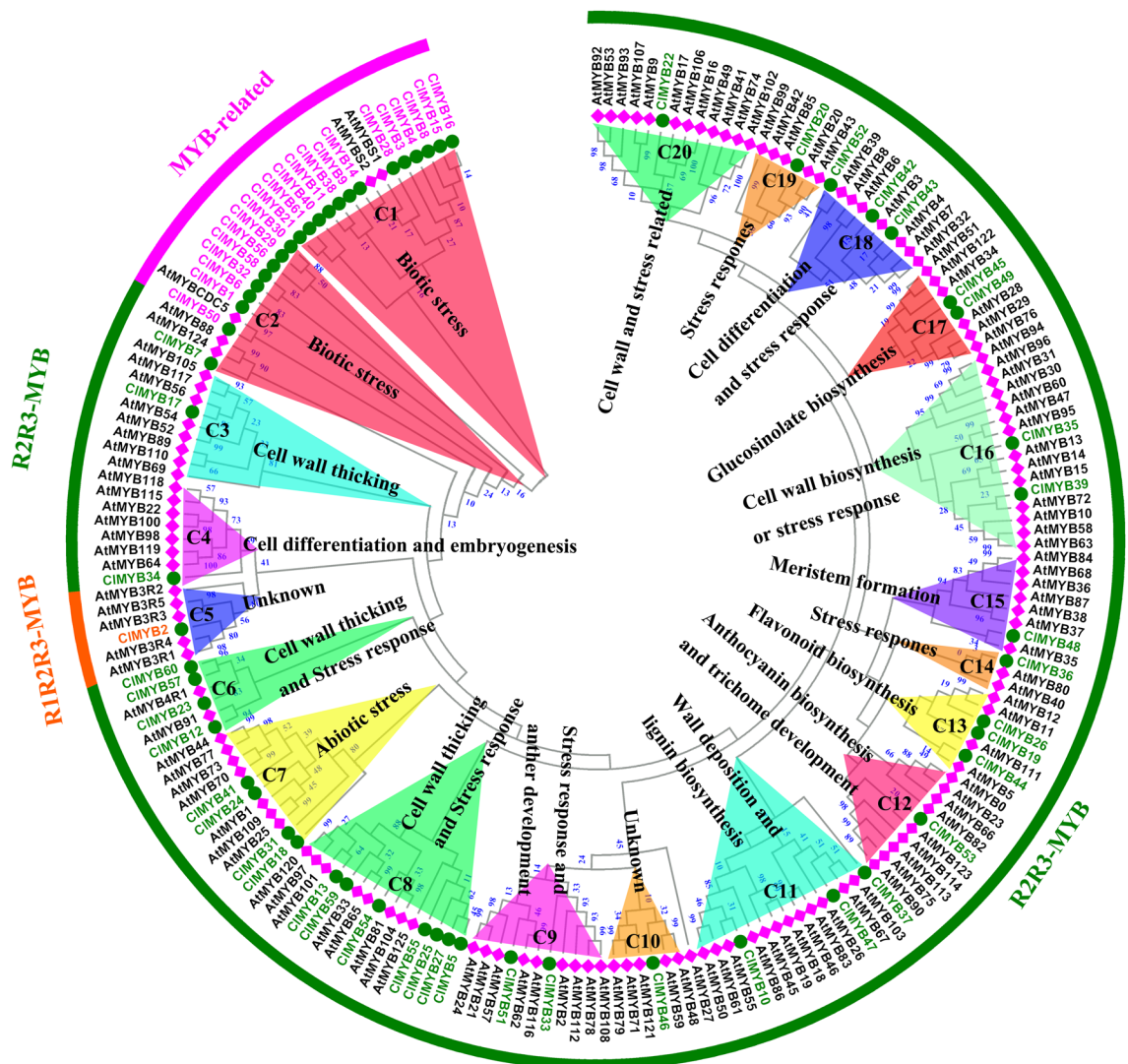


Figure 2. Evolutionary relationships and putative functions of the MYB proteins in *Cunninghamia lanceolata* (starting as C1) based on the phylogenetic tree with MYB proteins in *Arabidopsis thaliana* (At). The circular unrooted tree was generated by Neighbor-Joining method in MEGA 7.0 with 1,000 bootstrap replicates, Jones-Taylor-Thornton (JTT) model and pairwise deletion treatment. The analysis involved all domain sequences of respective R2R3-MYB protein with 61 CIMYBs and 135 AtMYBs.

in X1 compared to X2. *CIMYB22* was also the transcript with the greatest fold change based on qRT-PCR data, but the greatest fold change was up-regulated with an 8714-fold in X1 compared to X4. Nine *CIMYB* transcripts (*CIMYB19*, *CIMYB21*, *CIMYB22*, *CIMYB24*, *CIMYB26*, *CIMYB35*, *CIMYB36*, *CIMYB42*, *CIMYB45*, *CIMYB49*, and *CIMYB52*) had same preferential expression detected by two methods. These transcripts all highly expressed in X1, which indicate the results of transcript abundance detected by qRT-PCR were not highly consistent with the RNA-seq in specific xylem tissues (X2, X3, and X4) representing the heartwood formation. Because qRT-PCR and RNA-Seq were conducted with the RNA taken from different increment cores, it reduced the consistency of gene expression patterns detected by the two methods.

Discussion

In molecular biological research, transcript expression analysis has been an effective basic strategy to understand the establishment of cellular states and predict gene function²⁹. qRT-PCR is a commonly used method to quantify gene expression with appropriate reference genes as internal controls and it is unreliable to apply a universal reference gene for different tissues and treatments³⁰. HW formation is described as a form of PCD⁴. Expression of genes involved in secondary metabolite biosynthesis and PCD change greatly from SW and TZ toward HW regions²⁴. At present, little attention has been paid to selecting stable and appropriate reference genes for the gene expression normalization in RC of gymnosperm.

We determined the expression stability of 12 reference genes in RC of Chinese fir stem by three statistical algorithms. Pearson coefficient of correlation (r) was used as the main reference value for gene stability in Best-keeper. GeNorm recommends selecting genes with an M value < 1.5. NormFinder provides a stability value of

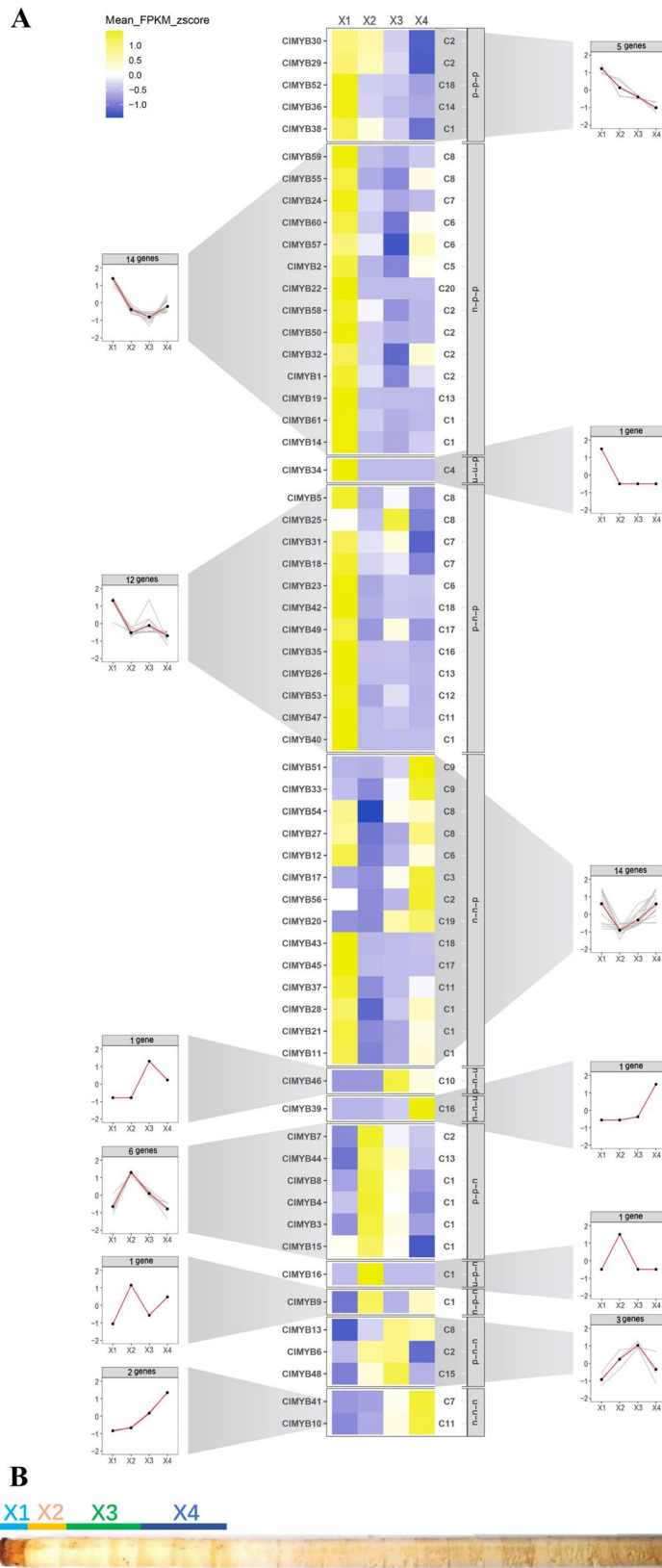


Figure 3. Heatmap of the RNaseq transcript abundance pattern of the 61 MYB transcripts from *Cunninghamia lanceolata* in four woody tissues clustered in 12 expression groups by *K-means*. (A) Heatmap of the RNaseq transcript abundance pattern. Transcript name is included to the left of the heatmap and the short name of the phylogenetic subgroup is shown to the right. *K-means* clusters were performed based on fragments per kilobase of exon per million fragments mapped (FPKM) values adjusted through z-score standardization. (B) Localization of different woody tissues used for RNA-seq and qRT-PCR experiment of 30 years old Chinese fir stem. X1, cambium zone; X2, outer sapwood; X3, inner sapwood; X4, transition zone.

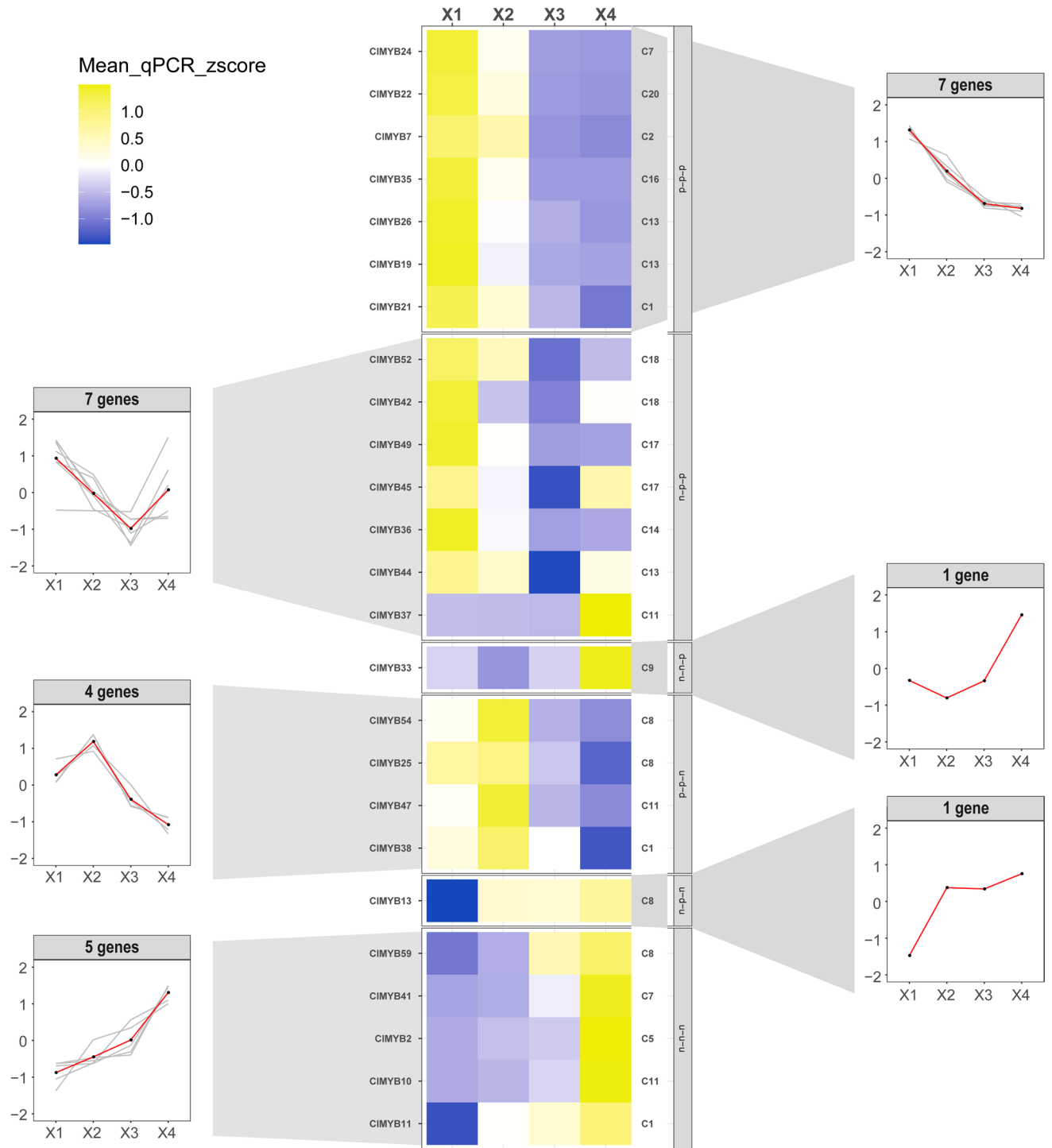


Figure 4. The heatmap of the transcript abundance patterns validated by qRT-PCR of the 25 *MYB* transcripts from *Cunninghamia lanceolata* in four woody tissues. Twenty-five *CIMYB* transcripts were clustered in 6 expression groups by *K-means*. Data was analyzed by the $2^{-\Delta\Delta C_t}$ method and adjusted through z-score standardization. Transcript name is included to the left of the heatmap and the short name of the phylogenetic subgroup is shown to the right.

each reference gene by estimating both intra- and inter- group expression variation^{12,32,33}. According to the three algorithms, these results show that the expression of the candidate genes is more variable in RC15 than in RC30. The accuracy of target gene expression level may be influenced by using just a single reference gene in qRT-PCR⁷. Integrated ranking of reference genes stability from the three software and utilizing multiple reference genes for the normalization is necessary and shows up as a more reliable accuracy of target gene expression level in

qRT-PCR³⁴. Compared to the calculation of geometric mean, arithmetical mean of ranking values by the different algorithms is a more comprehensive way to create a consensus ranking³⁵.

In our study, *IDH*, *UBC2*, *RCA*, and *H2B* were the most stable reference genes in RC15. *H2B*, *18S*, *TIP41*, and *GAPDH* were the most stable reference genes in RC30. *H2B* was considered as the stable reference genes from adventitious rooting in *Eucalyptus globulus*³⁶. *IDH* was used as reference genes in gene expression experiments to analyze flowers of *Pinus massoniana* L. in different developmental phases³⁷. *GAPDH* was the suitable reference gene for xylem tissues in *Camellia oleifera*³⁸. *18S* was one of stable reference genes in different tissues of *Taxus spp*³⁹. *UBC2* was identify the best reference genes for *Osmanthus fragrans*⁴⁰. *RCA* was upregulated in response to salt stress for cucumber⁴¹. *TIP41* was the stable reference gene for low-temperature stress in *Populus trichocarp*⁴².

Vascular plants developed specialized conductive tissues in xylem vessels characterized by SCW deposition to increase water conductive activity⁴³. MYBs play critical roles in controlling xylem cell differentiation and SCW biosynthesis pathway⁴⁴. Homology classification is important to predict the gene function and genes in the same subgroup based on the sequence are thought to have relatively similar roles⁴⁵. Three MYB subgroups (S5, S6 and SAtM5) are identified by phylogenetic approach to belonging to putative woody-expanded subgroups from the woody perennial species. These woody-expanded subgroups are involved in the phenylpropanoid pathway²⁷. *CIMYB26* and *CIMYB49* were identify to belong to S5. *CIMYB19* and *CIMYB36* belong to SAtM5. These four MYB transcripts from Chinese fir were preferentially expressed in the cambium-enriched region of Chinese fir and may be involved in regulating of the phenylpropanoid pathway.

CIMYB transcript and *AtMYB* genes of the C3 subgroup (*CIMYB17*, *AtMYB56*, *AtMYB105*, and *AtMYB117*) were clustered into the same subgroup with different methods. These *MYB* genes of Arabidopsis are involved in the cell wall metabolism^{46,47} and organ patterning^{48,49}. *CIMYB17* may have similar functions and was preferentially expressed in the transition zone of Chinese fir. Some *CIMYB* transcripts and *AtMYB* genes of C8 subgroup (*CIMYB13*, *CIMYB54*, *CIMYB59*, *AtMYB33*, *AtMYB65*, and *AtMYB101*) were clustered into the same subgroup. These *MYB* genes of Arabidopsis are involved in the programmed cell death process⁵⁰. Although *CIMYB13*, *CIMYB54*, and *CIMYB59* maybe involved in programmed cell death process, the transcripts with different expression patterns in radial change of Chinese fir stem have different functions.

Some *CIMYB* transcripts and *AtMYB* genes of C2 subgroup (*CIMYB7*, *AtMYB88*, and *AtMYB124*), C7 subgroup (*CIMYB18*, *CIMYB24*, *CIMYB31*, *CIMYB41*, *AtMYB44*, and *AtMYB73*), C9 subgroup (*CIMYB33*, *CIMYB51*, *AtMYB62*, and *AtMYB108*), C16 subgroup (*CIMYB39*, *AtMYB13*, *AtMYB14*, and *AtMYB15*), C18 subgroup (*CIMYB42*, *CIMYB43*, *CIMYB52* and *AtMYB4*), C19 subgroup (*CIMYB20* and *AtMYB20*) were clustered into each subgroup. *AtMYB* genes of these subgroups involve in transducing or conferring resistance to abiotic stresses^{51–56}. *CIMYB* transcripts of these subgroups may have similar functions. *CIMYB18*, *CIMYB20*, *CIMYB24*, *CIMYB31*, *CIMYB52*, *CIMYB42*, and *CIMYB43* were preferentially expressed in the cambium zone of Chinese fir. *CIMYB7* was highly expressed in outer sapwood. *CIMYB33*, *CIMYB39*, *CIMYB41*, and *CIMYB51* showed higher expression in the transition zone. The different expression pattern of these *CIMYBs* indicates the transcripts involved in stress responses at the different stages during heartwood formation of Chinese fir.

Methods

Collection of plant material. Samples excised at breast height (1.3 m) of 26 years old Chinese fir stem were successfully used to study cellular changes from cambium to heartwood in Chinese fir⁵⁷. Increment cores were taken from 35 years old Taiwanina (*Taiwania cryptomerioides* Hayata) stems to analyze the expression profiling of the heartwood formation process³¹. In this study, increment cores were sampled from breast height of three well-grown individuals (of each) of 15 years old and 30 years old Chinese fir stems in summer (August 2019). The radial change tissues of outer sapwood, inner sapwood, and the transition zone were taken from 15 years old Chinese fir stems. The tissues of outer sapwood, middle sapwood, inner sapwood, and the transition zone were taken from 30 years old Chinese fir stems. While these above samples were collected, all samples were immediately frozen in liquid nitrogen and later stored in -80°C for subsequent experiments. Increment cores were taken from our test plantation in Rongshui county of China ($25^{\circ}04'48.6''\text{N}$, $108^{\circ}15'16.3''\text{E}$). Since the collected materials were taken from our (own) test plantation, there will be no permission needed. The study complies with relevant institutional, national, and international guidelines and legislation.

Total RNA extraction and cDNA synthesis. Total RNA from different samples was isolated by RNAPrep Pure Plant Kit (Tiangen, China) without DNA contamination following the manufacturer's protocol. RNA integrity was checked on 1% agarose gel electrophoresis. The quantity and quality of total RNA were measured by a NanoDrop 2000 Spectrophotometer (Thermo Scientific, US). RNA was homogenized with RNase-free water. Total RNA (500 ng per samples) was reversely transcribed to first-strand of cDNA by PrimeScript RT reagent Kit (TaKaRa, Japan). All cDNA samples were stored at -20°C .

qRT-PCR. The full-length transcriptome from equal amounts of mixed RNA obtained from 4 woody tissues (X1, X2, X3, and X4) of one 30 years old Chinese fir was sequenced on the PacBio Sequel platform from our laboratory (SRA accession: PRJNA760952). The consensus sequence was obtained. A total of 12 RNA-seq libraries, with 4 woody tissues (X1, X2, X3, and X4) per tree for three 30 years old individuals were sequenced by Illumina HiSeqX platform (SRA accession: PRJNA760952). According to the literature of reference genes selection^{36,58,59} and our previous RNA-Seq data of 4 woody tissues (Novogene, China), 12 candidate reference genes were screened based on rules of $q\text{-value} \geq 0.05$, $\text{FPKM} \geq 5$, and $\text{Fold Change} < 2$. Specificity of primer pair of 12 candidate reference genes and 25 *CIMYB* transcripts were controlled by performing melting curves of the qRT-PCR products (Tianlong, China). Melting curves showed a single peak indicates the absence of side products and primer dimers.

Nucleotide sequence data of 12 candidate reference genes and the primers of 25 *CIMYBs* were listed in Table S2. The conserved domains of their mRNA sequences were identified using the NCBI database and local blast of PacBio-Seq data, the primers for qRT-PCR amplification were then designed on the web using Primer 3.0 (http://www.primer3plus.com/primer3web/primer3web_input.htm) with the following criteria: the primers should be with limited lengths of 20–24 bp, melting temperatures (T_m) in the range of 55–62 °C, GC contents varying from 45 to 55%, and product lengths of 80–250 bp, other parameters were set as default. qRT-PCR assay was performed in a volume of 20 μ L with three technical replicates on Gentier 32R real-time PCR system (Tian-long, China) using TB Green Premix Ex Taq (TaKaRa, Japan). The reaction was performed at the following conditions: initial denaturation at 95 °C for 30 s, followed by 45 cycles of 95 °C for 5 s, 55 °C for 10 s, and 72 °C for 10 s.

Phylogenetic analysis of the CIMYB proteins. The full-length transcriptome of 4 woody tissues (X1, X2, X3, and X4) from one 30-year-old Chinese fir was obtained on the PacBio Sequel platform (SRA accession: PRJNA760952). To derive annotation information and identify the *MYB* transcripts, the unique isoforms were mapped to Swiss-Prot (<https://www.uniprot.org/uniprot/>). The online software SMART (<http://smart.embl.de/>) were used to identify the domain sequences of all *MYB* proteins. The names and accession numbers of *MYBs* in the other three species were based on previous research^{25–28}. According to accession numbers, the protein sequence of *MYBs* was available in *Arabidopsis thaliana* (Plant Transcription Factor Database: <http://plantfdb.gao-lab.org/>), *Populus trichocarpa* (PoplarGene: <http://bioinformatics.caf.ac.cn/PoplarGene/>), and *Eucalyptus grandis* (NCBI Genbank: <https://www.ncbi.nlm.nih.gov/genbank/>). The online SMART software (<http://smart.embl.de/>) were used to identify the amino acid sequence of domains of each *MYB* protein. The protein sequence of 61 *CIMYBs* and the whole domain of each *MYB* protein are shown in Table S3.

Based on the whole domain sequence of each *MYB* protein, neighbor-joining (NJ) phylogenetic trees were generated by MEGA 7.0 with 1,000 bootstrap replicates, Jones-Taylor-Thornton (JTT) model and pairwise deletion treatment. Based on the whole domain sequence of each *MYB* protein, Maximum Likelihood (ML) phylogenetic tree was generated by MEGA 7.0 with 1,000 bootstrap replicates, Jones-Taylor-Thornton (JTT) model and complete deletion treatment. The biological functions of *CIMYBs* were predicted by homologous *MYBs* with validated function within the same phylogenetic tree clade.

RNAseq and qRT-PCR expression analysis. The correlation coefficient (R^2) and amplification efficiency (E) of primer pairs were counted by the linear regression model with a series of fivefold dilution of first-strand cDNA (0.5 μ g/ μ L). The PCR efficiency was calculated by the equation $E = (10^{[-1/\text{slope}] - 1}) \times 100\%$. The expression stability of candidate reference genes was calculated using geNorm, NormFinder, and Bestkeeper^{12,55,56}. We used geometric mean of 4 selected reference genes of Chinese fir for the normalization of data. The relative expression levels were evaluated by $2^{-\Delta\Delta CT}$.

RNA-seq data from 4 different woody tissues (X1, X2, X3, and X4) per tree for three 30 years old individuals were sequenced by Illumina HiSeqX platform (SRA accession: PRJNA760952). Mean fragments per kilobase of exon per million fragments mapped (FPKM) values per transcript were calculated for each tissue. *K-means* clusters performed with Perl were based on FPKM values adjusted through z-score standardization. Six pairwise comparisons were performed to identify DEGs: (1) X1 versus X2; (2) X1 versus X3; (3) X1 versus X4; (4) X2 versus X3; (5) X2 versus X4; and (6) X3 versus X4. Differential expression analysis by RNA-seq and qRT-PCR were based on rules of fold change > 2 and $P < 0.05$. Analysis of z-score standardization and ANOVA and drawing of heatmap and trendline was performed by R package.

Received: 6 April 2021; Accepted: 14 December 2021

Published online: 07 January 2022

References

- Li, X., Wu, H. X. & Southerton, S. G. Comparative genomics reveals conservative evolution of the xylem transcriptome in vascular plants. *BMC Evol. Biol.* **10**, 1–14 (2010).
- Zhao, Y. *et al.* Intron-mediated alternative splicing of PtrWIND1B regulates cell wall thickening during fiber development in Populus. *Plant Physiol.* **164**, 765–776 (2014).
- Zhang, M. *et al.* Comparison of wood physical and mechanical traits between major gymnosperm and angiosperm tree species in China. *Wood Sci. Technol.* **51**, 1405–1419 (2017).
- Spicer, R. Senescence in secondary xylem: heartwood formation as an active developmental program. In *Vascular transport in plants* (eds Holbrook, N. M. & Zwieniecki, M. A.) 457–475 (Elsevier Academic Press, 2005).
- Plomion, C., Leprovost, G. & Stokes, A. Wood formation in trees. *Plant Physiol.* **127**, 1513–1523 (2001).
- Kampe, A. & Magel, E. New insights into heartwood and heartwood formation. In *Cellular aspects of wood formation* 71–95 (Springer, Berlin/Heidelberg, 2013).
- Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, 1–11 (2002).
- Kulcheski, F. R. *et al.* The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean. *Anal. Biochem.* **406**, 185–192 (2010).
- Li, Q., Fan, C. M., Zhang, X. M. & Fu, Y. F. Validation of reference genes for real-time quantitative PCR normalization in soybean developmental and germinating seeds. *Plant Cell Rep.* **31**, 1789–1798 (2012).
- Plachetka-Božek, A. & Augustyniak, M. Evaluation of candidate reference genes for quantitative gene expression analysis in *Spodoptera exigua* after long-time exposure to cadmium. *Sci. Rep.* **7**, 8338 (2017).
- Sun, H. P., Li, F., Ruan, Q. M. & Zhong, X. H. Identification and validation of reference genes for quantitative real-time PCR studies in *Hedera helix* L. *Plant Physiol. Biochem.* **108**, 286–294 (2016).
- Radonic, A. *et al.* Guideline to reference gene selection for quantitative real-time PCR. *Biochem. Biophys. Res. Commun.* **313**, 856–862 (2004).

13. Saha, P. & Blumwald, E. Assessing reference genes for accurate transcript normalization using quantitative real-time PCR in pearl millet [*Pennisetum glaucum* (L.) R. Br]. *PLoS ONE* **9**, e106308 (2014).
14. Liu, C., Hao, J., Qiu, M., Pan, J. & He, Y. Genome-wide identification and expression analysis of the MYB transcription factor in Japanese plum (*Prunus salicina*). *Genomics* **112**, 4875–4886 (2020).
15. Hirai, R. *et al.* The progression of xylem vessel cell differentiation is dependent on the activity level of VND7 in *Arabidopsis thaliana*. *Plants (Basel)* **9**, 39 (2019).
16. Xie, M. *et al.* Regulation of lignin biosynthesis and its role in growth-defense tradeoffs. *Front. Plant Sci.* **28**, 1427 (2018).
17. Ma, X. *et al.* Nutrient cycling and distribution in different-aged plantations of Chinese fir in southern China. *Forest Ecol. Manage.* **243**, 61–74 (2007).
18. Bian, L. *et al.* Spatial analysis increases efficiency of progeny testing of Chinese fir. *J. For. Res.* **28**, 445–452 (2017).
19. Brunsfeld, S. J. *et al.* Phylogenetic relationships among the genera of *Taxodiaceae* and *Cupressaceae*: evidence from rbcL sequences. *Syst. Bot.* **19**, 253–262 (1994).
20. Khoshoo, T. N. Chromosome numbers in gymnosperms. *Silvae Genetica* **10**, 1–9 (1961).
21. Hizume, M. *et al.* Flow cytometric determination of genome size in the *Taxodiaceae* Cupressaceae sensu stricto and Sciadopytaceae. *Cytologia* **66**, 307–311 (2001).
22. Imai, T. H. *et al.* Evaluation of reference genes for accurate normalization of gene expression for real time quantitative PCR in *Pyrus pyrifolia* using different tissue samples and seasonal conditions. *PLoS ONE* **9**, 1–11 (2014).
23. Lin, Y. L. & Lai, Z. X. Reference gene selection for qPCR analysis during somatic embryogenesis in longan tree. *Plant Sci.* **178**, 359–365 (2010).
24. Li, H. P. *et al.* Screening of valid reference genes for real-time RT-PCR data normalization in *Hevea brasiliensis* and expression validation of a sucrose transporter gene *HbSUT3*. *Plant Sci.* **181**, 132–139 (2011).
25. Chen, Y. H. *et al.* The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Mol Biol.* **60**, 107–124 (2006).
26. Stracke, R., Werber, M. & Weisshaar, B. The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* **4**, 447–456 (2001).
27. Soler, M. *et al.* The *Eucalyptus grandis* R2R3-MYB transcription factor family: evidence for woody growth-related evolution and function. *New Phytol.* **206**, 1364–1377 (2015).
28. Wilkins, O., Nahal, H., Foong, J., Provart, N. J. & Campbell, M. M. Expansion and diversification of the Populus R2R3-MYB family of transcription factors. *Plant Physiol.* **149**, 981–993 (2009).
29. Gonzales, K. A. & Ng, H. H. Biological networks governing the acquisition, maintenance, and dissolution of pluripotency: insights from functional genomics approaches. *Cold Spring Harb Symp. Quant. Biol.* **80**, 189–198 (2015).
30. Dheda, K. *et al.* The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. *Anal. Biochem.* **344**, 141–153 (2005).
31. Yeh, T. F., Chu, J. H., Liu, L. Y. & Chen, S. Y. Differential gene profiling of the heartwood formation process in *Taiwania cryptomerioides* Hayata xylem tissues. *Int. J. Mol. Sci.* **21**, 960 (2020).
32. Pfaffl, M. W., Tichopad, A., Prgomet, C. & Neuvians, T. P. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper-Excel-based tool using pair-wise correlations. *Biotechnol. Lett.* **26**, 509–515 (2004).
33. Andersen, C. L., Jensen, J. L. & Orntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* **64**, 5245–5250 (2004).
34. Yao, S. C. *et al.* Identification and validation of reference genes for real-time qPCR normalization during Al-induced programmed cell death in peanut. *Biol. Plant.* **63**, 237–246 (2019).
35. Kotrade, P., Sehr, E. M., Wischnitzki, E. & Brüggemann, W. Comparative transcriptomics-based selection of suitable reference genes for normalization of RT-qPCR experiments in drought-stressed leaves of three European *Quercus* species. *Tree Genet. Genomes* **15**, 1–12 (2019).
36. De-Almeida, M. R. *et al.* Reference gene selection for quantitative reverse transcription-polymerase chain reaction normalization during in vitro adventitious rooting in *Eucalyptus globulus* Labill. *BMC Mol. Biol.* **11**, 73 (2010).
37. Chen, H. *et al.* Reference genes selection for quantitative gene expression studies in *Pinus massoniana* L. *Trees* **30**, 685–696 (2016).
38. Yang, Z. Y., Zhang, R. & Zhou, Z. Identification and validation of appropriate reference genes for gene expression analysis in *Schima superba*. *Genes* **2**, 732 (2021).
39. Delporte, M. *et al.* Selection and validation of reference genes for quantitative real-time PCR analysis of gene expression in *Cichorium intybus*. *Front. Plant Sci.* **6**, 651 (2015).
40. Zhang, C. *et al.* Identification of suitable reference genes for gene expression normalization in the quantitative real-time PCR analysis of sweet Osmanthus (*Osmanthus fragrans* Lour.). *PLoS ONE* **10**, e0136355 (2015).
41. Shu, S. *et al.* Effects of exogenous spermidine on photosynthetic capacity and expression of Calvin cycle genes in salt-stressed cucumber seedlings. *J. Plant. Res.* **127**, 763–773 (2014).
42. Pettengill, E. A., Parmentier-Line, C. & Coleman, G. D. Evaluation of qPCR reference genes in two genotypes of Populus for use in photoperiod and low-temperature studies. *BMC. Res. Notes* **5**, 1–12 (2012).
43. Sperry, J. S. Evolution of water transport and xylem structure. *Int. J. Plant Sci.* **164**, S115–S127 (2003).
44. Nakano, Y., Yamaguchi, M., Endo, H., Rejab, N. A. & Ohtani, M. NAC-MYB-based transcriptional regulation of secondary cell wall biosynthesis in land plants. *Front. Plant Sci.* **6**, 288 (2015).
45. Jiang, C., Gu, X. & Peterson, T. Identification of conserved gene structures and carboxy-terminal motifs in the Myb gene family of *Arabidopsis* and *Oryza sativa* L. ssp. *indica*. *Genome Biol.* **5**, R46 (2004).
46. Zhang, Y. *et al.* MYB56 encoding a R2R3 MYB transcription factor regulates seed size in *Arabidopsis thaliana*. *J. Integr. Plant Biol.* **55**, 1166–1178 (2013).
47. Vilarrasa-Blasi, J. *et al.* Regulation of plant stem cell quiescence by a brassinosteroid signaling module. *Dev. Cell.* **30**, 36–47 (2014).
48. Lee, D. K., Geisler, M. & Springer, P. S. LATERAL ORGAN FUSION1 and LATERAL ORGAN FUSION2 function in lateral organ separation and axillary meristem formation in *Arabidopsis*. *Development* **136**, 2423–2432 (2009).
49. Gomez, M. D. *et al.* Characterization of constricted fruit (ctf) mutant uncovers a role for *AtMYB117/LOF1* in ovule and fruit development in *Arabidopsis thaliana*. *PLoS ONE* **16**, e18760 (2011).
50. Alonso-Peral, M. M. *et al.* The microRNA159-regulated GAMYB-like genes inhibit growth and promote programmed cell death in *Arabidopsis*. *Plant Physiol.* **154**, 757–771 (2010).
51. Xie, Z. *et al.* Role of the stomatal development regulators FLP/MYB88 in abiotic stress responses. *Plant J.* **64**, 731–739 (2010).
52. Kim, J. H. *et al.* Loss of the R2R3 MYB, AtMyb73, causes hyper-induction of the SOS1 and SOS3 genes in response to high salinity in *Arabidopsis*. *J. Plant Physiol.* **170**, 1461–1465 (2013).
53. Jung, C. *et al.* Overexpression of AtMYB44 enhances stomatal closure to confer abiotic stress tolerance in transgenic *Arabidopsis*. *Plant Physiol.* **146**, 623–635 (2008).
54. Kirik, V. *et al.* Two novel MYB homologues with changed expression in late embryogenesis-defective *Arabidopsis* mutants. *Plant Mol. Biol.* **37**, 819–827 (1998).
55. Mengiste, T. *et al.* The BOTRYTIS SUSCEPTIBLE1 gene encodes an R2R3MYB transcription factor protein that was required for biotic and abiotic stress responses in *Arabidopsis*. *Plant Cell* **15**, 2551–2565 (2003).

56. Galli, M. *et al.* Soluble interleukin-2 receptor decrease in the sera of HIV-infected patients treated with zidovudine. *AIDS* **5**, 1231–1235 (1991).
57. Song, K. *et al.* Cellular changes of tracheids and ray parenchyma cells from cambium to heartwood in *Cunninghamia lanceolata*. *J. Trop. For. Sci.* **23**, 478–487 (2011).
58. Chen, L. *et al.* Validation of reference genes for RT-qPCR studies of gene expression in banana fruit under different experimental conditions. *Planta* **234**, 377–390 (2011).
59. Gao, Z. H. *et al.* Selection and validation of reference genes for studying stress-related agarwood formation of *Aquilaria sinensis*. *Plant Cell. Rep.* **31**, 1759–1768 (2012).

Author contributions

K.P.L designed the methodology, conceived the study and wrote the manuscript. W.L and G.Y.T conducted the statistical analysis and revised the manuscript. K.Y.H provided project support. All authors read and approved the manuscript.

Funding

This study is supported by Innovation-driven development project of Guangxi Zhuang Autonomous Region, China (GK AA17204087-2), National Natural Science Foundation of China (32060352), National Natural Science Foundation of Guangxi Zhuang Autonomous Region, China (2018GXNSFAA294083), Department of Human Resources and Social Security of Guangxi Zhuang Autonomous Region, China (GuiCaiSheHan[2018]112).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-04406-1>.

Correspondence and requests for materials should be addressed to K.-Y.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022