

OPEN

# Mining MYB transcription factors from the genomes of orchids (*Phalaenopsis* and *Dendrobium*) and characterization of an orchid R2R3-MYB gene involved in water-soluble polysaccharide biosynthesis

Chunmei He<sup>1</sup>, Jaime A. Teixeira da Silva<sup>2</sup>, Haobin Wang<sup>1,3</sup>, Can Si<sup>1,3</sup>, Mingze Zhang<sup>1,3</sup>, Xiaoming Zhang<sup>1</sup>, Mingzhi Li<sup>4</sup>, Jianwen Tan<sup>5</sup> & Jun Duan<sup>1</sup>

Members of the MYB superfamily act as regulators in a wide range of biological processes in plants. Despite this, the MYB superfamily from the Orchidaceae has not been identified, and MYB genes related to bioactive water-soluble polysaccharide (WSP) biosynthesis are relatively unknown. In this study, we identified 159 and 165 MYB genes from two orchids, *Phalaenopsis equestris* and *Dendrobium officinale*, respectively. The MYB proteins were classified into four MYB classes in both orchids: MYB-related (MYBR), R2R3-MYB, 3R-MYB and atypical MYB proteins. The MYBR proteins in both orchids were classified into five subfamilies and 12 genes were strongly up-regulated in response to cold stress in *D. officinale*. The R2R3-MYB proteins were both divided into 31 clades in *P. equestris* and *D. officinale*. Among these clades, nine contained MYB TFs related to secondary cell wall biosynthesis or testa mucilage biosynthesis in *Arabidopsis thaliana*. In *D. officinale*, 10 candidate genes showed an expression pattern corresponding to changes in the WSP content. Overexpression of one of these candidate genes (*DoMYB75*) in *A. thaliana* increased seed WSP content by about 14%. This study provides information about MYB genes in two orchids that will further help to understand the transcriptional regulation of WSP biosynthesis in these orchids as well as other plant species.

Gene expression is regulated by various complex mechanisms, including modifications to DNA such as histone modification and DNA methylation, as well as various RNA-mediated processes. Transcription factors (TFs) regulate gene expression, and this is a well-known mechanism by which a TF binds to a specific nucleotide sequence upstream of target gene, ultimately controlling a range of biological processes<sup>1</sup>. MYB TFs exist widely in eukaryotes and are one of the largest and most diverse families of TFs in the plant kingdom, where they play an essential role in a wide range of physiological and biochemical processes<sup>2,3</sup>.

The first MYB gene (*v-myb*), which was isolated from avian myeloblastosis virus (AMV), encodes a MYB domain protein<sup>4</sup>. Ever since the first plant MYB gene (the *Zea mays* *COLORED1* (*C1*) gene) was cloned<sup>5</sup>, numerous MYB genes have been identified from plants as an increasing number of plant genomic sequences became available. For example, 198, 256, 127, 231 and 122 MYB genes have been identified in *Arabidopsis thaliana*<sup>6</sup>, *Brassica rapa*<sup>3</sup>, *Solanum lycopersicum*<sup>7</sup>, *Pyrus bretschneideri*<sup>8</sup> and *Brachypodium distachyon*<sup>9</sup>, respectively. MYB proteins share a highly conserved DNA-binding domain (the MYB domain), which ranges from one to four

<sup>1</sup>Key Laboratory of South China Agricultural Plant Molecular Analysis and Gene Improvement, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, 510650, China. <sup>2</sup>P. O. Box 7, Miki-cho post office, Ikenobe 3011-2, Miki-cho, Kita-gun, Kagawa-ken, 761-0799, Japan. <sup>3</sup>University of the Chinese Academy of Sciences, Beijing, 100049, China. <sup>4</sup>Biodata Biotechnology Co. Ltd, Hefei, 230031, China. <sup>5</sup>College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou, 510642, China. Correspondence and requests for materials should be addressed to J.D. (email: [duanj@scib.ac.cn](mailto:duanj@scib.ac.cn))

imperfect amino acid sequence repeats (R)<sup>10</sup>. Based on the number of adjacent repeats, the MYB genes in plants have been divided into four distinct groups: MYB-related (MYBR, only contains one R1- or R2-like repeat), R2R3-MYB (containing two R2/3R-like repeats), 3R-MYB (containing three R1/R2/R3-like repeats) and atypical MYB proteins (4R-MYB, four R1/R2-like repeats; CDC5-like)<sup>6,10</sup>. The atypical MYB group of proteins is the smallest class, and contains one or two genes in several higher plant genomes. For example, one and two atypical MYB genes were found in *O. sativa* and *A. thaliana*, respectively<sup>11</sup>. The 3R-MYB group is the second smallest class, and contains about five members in plants such as *A. thaliana*, *Populus trichocarpa* and *Vitis vinifera*<sup>3</sup>. The MYBR proteins contain only a single repeat and fall into five subclasses, CCA1/R-R-like, CPC-like, I-box-like, TBP-like and TRF-like, that contain 30 to 70 genes in plant genomes<sup>2</sup>. The R2R3-MYB group is the largest group in plants with more than 100 members that were extensively amplified approximately 500 million years ago after the appearance of land plants<sup>12</sup>. For example, 113, 126 and 244 genes encode R2R3-MYB proteins in *O. sativa*<sup>11</sup>, *A. thaliana*<sup>6</sup> and *Glycine max*<sup>13</sup>, respectively. The *A. thaliana* R2R3-MYB proteins have been divided into 33 clades, while those in *P. bretschneideri* have been divided into 37 clades<sup>8</sup>, which suggests evolutionary diversity of the R2R3-MYB family.

All four groups of MYB proteins were found after genome-wide analyses of MYB TFs in plants and while the function of a number of MYB genes have been characterized in many plants, the function of 4R-MYB proteins remains unclear. The 3R-MYBs play a role in cell cycle control<sup>14</sup>. MYBR proteins are involved in cellular morphogenesis, secondary metabolism, organ morphogenesis, phosphate starvation, chloroplast development and circadian regulation<sup>15</sup>. More recently, two *A. thaliana* MYBR genes (*MYBS1* and *MYBS2*) have been shown to have opposite roles in sugar signaling mechanisms<sup>16</sup>. Over the past two decades, the R2R3-MYB TFs have been extensively exploited and many R2R3-MYB proteins have been shown to play roles in several biological processes, such as development, response to biotic and abiotic stresses, and metabolism<sup>15,17</sup>. Among the metabolic processes, R2R3-MYB genes are acutely involved in phenylpropanoid metabolism<sup>18</sup>. Very recently, two R2R3-MYB genes from *Marchantia polymorpha*, *MpMYB14* and *MpMYB02*, were found to act as essential regulators in the biosynthesis of riccionidins and marchantins, respectively<sup>19</sup>.

Secondary cell walls (SCWs), which have a critically important function by supporting plants and are a major source of plant biomass, are composed of cellulose, lignin and hemicellulose<sup>20</sup>. An increasing body of studies has demonstrated that R2R3-MYB proteins are critical for SCW biosynthesis<sup>21</sup>. In *A. thaliana*, *AtMYB46* binds to the promoter of *AtCSLA9*, which is involved in glucomannan biosynthesis, and regulates its expression<sup>22</sup>. Testa mucilage, which is composed of polysaccharides, is regarded as a useful model for exploring the biosynthesis of cell wall polysaccharides<sup>23,24</sup>. Previous studies have demonstrated that R2R3-MYB genes, such as *AtMYB5* and *AtMYB61*, are required for the production of seed mucilage, and the polysaccharide content of seed mucilage in the *A. thaliana myb61* mutant was significantly reduced<sup>25,26</sup>. These results indicate that R2R3-MYB members play roles in the biosynthesis of plant polysaccharides.

Water-soluble polysaccharides (WSPs) play an important role in plants' stress response. For example, tolerant genotypes of wheat (*Triticum aestivum* L.) seedlings accumulated more water-soluble carbohydrates, including glucose, fructose, sucrose, and fructan than sensitive genotypes under drought and salt stress<sup>27</sup>. Recent studies showed that WSPs isolated from plants may increase immunity<sup>28</sup> and have an antitumor function<sup>29,30</sup>. The Orchidaceae is one of the largest plant families in the world and has about 25,000 species<sup>31</sup>. The *Dendrobium* genus, which belongs to the Orchidaceae, has several important species that are used as herbal medicines<sup>32</sup>. WSPs isolated from *Dendrobium* species such as *D. huoshanense* and *D. officinale* are regarded as major active ingredients and display immunomodulating activities<sup>33</sup>. Several genes involved in the biosynthesis of WSPs have been identified and characterized in *D. officinale*, whose stems contain an abundance of bioactive WSPs<sup>34–36</sup>. However, the TFs that regulate the biosynthesis of these WSPs are still unknown. In this study, MYB proteins were identified from two orchids, *Phalaenopsis equestris* and *D. officinale*, in a genome-wide process, and the putative R2R3-MYB genes related to SCW or mucilage biosynthesis were identified based on phylogenetic analysis. One R2R3-MYB gene involved in the biosynthesis of WSPs was characterized. This work provides novel information that would allow for a better understanding of the functional diversity of MYB genes in plants and would aid in revealing the molecular mechanisms underlying the biosynthesis of bioactive WSPs in *D. officinale* or in other plants.

## Methods

**Plant materials and treatments.** *D. officinale* plants were grown as described previously<sup>34</sup>. The stems of five developmental stages were harvested to determine WSPs and for gene expression analysis. S1 is about 4 months after sprouting in April, while S2, S3, S4, and S5 are about 9, 10, 12, and 13 months after sprouting, respectively. Roots, leaves and stems that were collected from plants grown in a growth chamber when they were 10 cm in height, were used to analyze the expression pattern of different organs. *D. officinale* seed capsules were surface sterilized in 0.1% mercuric chloride (HgCl<sub>2</sub>), sown on half-strength Murashige and Skoog medium (half the macronutrients; ½MS)<sup>37</sup> supplemented with 1 g/L activated charcoal, 20 g/L sucrose and 6 g/L agar and 0.5 mg/L 1-naphthalene-acetic acid (NAA), and cultivated at 26 ± 1 °C, 40 μmol m<sup>-2</sup> s<sup>-1</sup>, and a 12-h photoperiod. Seedlings (1 cm in height and about 3 months after sowing) were transferred to liquid ½MS medium supplemented with 20 g/L sucrose and 0.5 mg/L NAA for three days. After adapting, 30 seedlings were used to perform abiotic stress bioassays in liquid medium containing 150 g/L polyethylene glycol (PEG) 6000 (Sigma-Aldrich, Shanghai, China), 300 mM mannitol (Sigma-Aldrich), or 250 mM NaCl (Guangzhou Chemical Reagent Factory, Guangzhou, China). Seedlings were transferred to fresh ½MS medium supplemented with 20 g/L sucrose and 0.5 mg/L NAA as the control. There were three replicates with 36 seedlings in each treatment. After 6 h, seedlings were harvested, frozen in liquid nitrogen and RNA was isolated immediately.

*A. thaliana* plants (Col-0) were grown in soil under a 16-h photoperiod at 22 °C. For screening resistant lines, *A. thaliana* seeds were surface sterilized and sown on ½MS medium supplemented with 15 g/L sucrose and 8 g/L agar, stratified in the dark at 4 °C for 2 d, and then cultivated under a 16-h photoperiod at 22 °C.

Species	MYBR	2R-MYB	3R-MYB	Atypical MYB genes	Total
<i>A. thaliana</i> <sup>*</sup>	64	126	5	2	197
<i>P. equestris</i>	40	115	3	1	159
<i>D. officinale</i>	42	117	4	2	165

**Table 1.** Number of members in the four groups of MYB transcription factors in *A. thaliana*, *P. equestris* and *D. officinale*. <sup>\*</sup>The MYB genes from *A. thaliana* provided by Dubos *et al.*<sup>10</sup> and Stracke *et al.*<sup>74</sup>.

**Identification of MYB transcription factors in orchids.** The protein sequences of *P. equestris* and *D. officinale* were downloaded in a FASTA format from orchidbase (<http://orchidbase.itsps.ncku.edu.tw/EST/releaseSummary2012.aspx>) and the National Center for Biotechnology Information (NCBI) provided by Zhang *et al.*<sup>38</sup>. HMMER 3.0 software (<http://hmm.janelia.org/>) was used to identify the putative MYB TFs under default parameters. The putative MYB TFs were annotated by Pfam (Protein family)<sup>39</sup>, Swissprot<sup>40</sup> and nr (NCBI non-redundant protein sequences)<sup>41</sup>. The putative MYB TFs, which were confirmed to be MYB TFs by annotation, were regarded as MYB TFs. The classification of MYBR, R2R3-MYB, 3R-MYB and atypical MYB proteins were based on the annotation and BLAST against the *A. thaliana* MYB TFs.

**Phylogenetic analysis.** MYB proteins from *A. thaliana* (At), *P. equestris* (Pe) and *D. officinale* (Do) were aligned using MAFFT software version 7<sup>42</sup>. A phylogenetic tree of R2R3-MYBs was constructed using the Neighbor-Joining (NJ) method and 1,000 bootstraps with Clustalx<sup>43</sup>. The phylogenetic trees of MYBR and C2 (S6) R2R3-MYBs were constructed using MEGA 7<sup>44</sup> with the NJ method using 1,000 bootstraps.

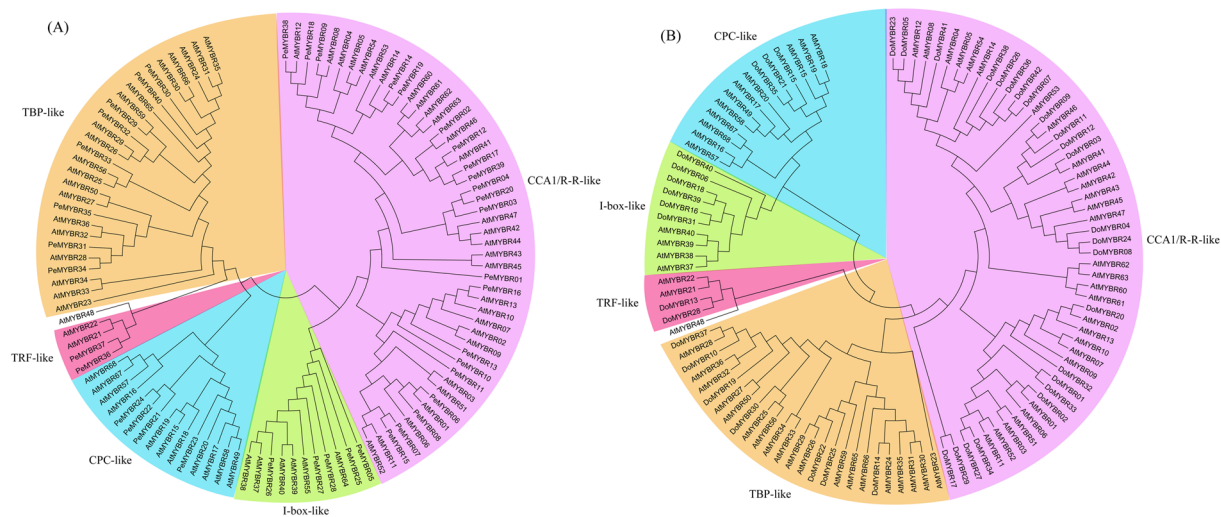
**Calculation of Ks and Ka of MYB family genes in the two orchids.** Orthologous gene pairs of MYB genes between *P. equestris* and *D. officinale* were identified by Orthofinder v2.2.6 with the BLAST method under default parameters<sup>45</sup>. The orthologous gene pairs were used to calculate synonymous (Ks) and nonsynonymous (Ka) values using the KaKs\_Calculator2.0<sup>46</sup>.

**Prediction of cis-responsive elements on the promoters of orchid MYB genes.** The 2000 bp genomic DNA sequences upstream of the initiation codon (ATG) of orchid MYB genes were obtained and used for predicting cis-acting regulatory DNA elements (cis-responsive elements). The PlantCARE database<sup>47</sup> and PLACE database<sup>48</sup> were adopted to identify the putative cis-responsive elements.

**Expression profiling of MYB genes from *D. officinale* under cold stress.** For the expression profiles of *D. officinale* under cold stress (4 °C), the transcriptome sequencing data of the control condition (SRR3210630, SRR3210635 and SRR3210636) and cold stress treatment (SRR3210613, SRR3210621 and SRR3210626) were obtained from the NCBI Sequence Read Archive (SRA) database<sup>49</sup>. The clean reads were obtained by filtering out low quality reads and were mapped to the nucleotide sequences of MYB genes using TopHat version 2.0.8<sup>50</sup>. The expression level of MYB genes was calculated by the fragments per kilobase of exon per million fragments mapped (FPKM) method using HTSeq<sup>51</sup>. The heatmap of expression profiling was drawn by a green-red gradient in R version 3.4.1 (<https://www.r-project.org/>). The genes with a FPKM value >5 in the control or cold stress treatment were regarded as sense, then were used to calculate fold change (mean of FPKM cold/mean of FPKM control). Genes with a ≥1.5-fold change were defined as up-regulated genes, and those with a ≤0.66-fold change were regarded as down-regulated genes.

**Quantitative RT-PCR (qRT-PCR) analysis.** Total RNA from *D. officinale* organs (roots, stems and leaves) and seedlings, as well as *A. thaliana* seedlings, was extracted using an RNA extraction kit (Column Plant RNAout2.0, Tiandz, Inc., Beijing, China). RNA was purified by excluding genomic DNA using the DNase I digestion kit (Takara Bio Inc., Dalian, China). The integrity and content of purified RNA was determined by 1% agarose gel electrophoresis and a NanoDrop 2000c Spectrophotometer (Thermo Scientific, Wilmington, NC, USA), respectively. Total RNA was reversed transcribed into cDNA by M-MLV reverse transcriptase (Promega, Madison, WI, USA) according to the manufacturer's protocol. The cDNA of each sample was diluted to 200 ng/mL and 1 μL was used as template for the qRT-PCR reaction. Three PCR reactions were performed using the SoAdvanced™ Universal SYBR® Green Supermix detection system (Bio-Rad, Hercules, CA, USA) in an ABI 7500 Real-time system (ABI, Foster City, CA, USA) with the following amplification regime: 95 °C for 2 min, and 40 cycles of 95 °C for 15 s and 60 °C for 30 s. Actin from *D. officinale* (NCBI accession number: JX294908) was used to normalize the expression of genes. The 2<sup>-ΔΔCT</sup> method<sup>52</sup> was used to calculate the relative gene expression level. All the primers of *DoMYB* genes and actins for qRT-PCR were designed by an online web tool (<http://www.idtdna.com/Primerquest/Home/Index>) and are listed in Supplementary Table 1.

**Generation of *DoMYB75* transgenic lines.** The coding sequence (CDS) of *DoMYB75* without a termination codon was amplified using the KOD FX High Success-rate DNA polymerase Kit (Toyobo Biotechnology Co. Ltd., Shanghai, China) and cloned into the pCAMBIA 1302 vector (Cambia, Canberra, Australia) at the *NcoI* site. The construct was verified by DNA sequencing at the Beijing Genomics Institute (Shenzhen, China). *A. thaliana* was transformed by the floral dip method<sup>53</sup> using about 15 independent plants. Twenty five resistant lines were identified by screening in ½MS medium supplemented with 25 mg/L hygromycin B (Roche Diagnostics, Mannheim, Germany). Three resistant lines were randomly selected to extract genomic DNA and verified as



**Figure 1.** Phylogenetic trees of MYBR proteins. **(A)** Unrooted phylogenetic tree of *P. equestris* and *A. thaliana* MYBR proteins. **(B)** Unrooted phylogenetic tree of *D. officinale* and *A. thaliana* MYB proteins. The trees were generated by MEGA 7<sup>44</sup> using the Neighbor-Joining method and aligned by MAFFT<sup>42</sup>.

transgenic lines using PCR. The 2 × BlueStar™ PCR Master Mix kit (Tingke Biotechnology Co. Ltd., Beijing, China) was used to perform PCR with 95 °C for 2 min and 35 cycles of 98 °C for 10 s, 58 °C for 30 s, and 72 °C for 30 s, followed by a final extension at 72 °C for 10 min. The primers used to construct the overexpression vector are listed in Supplementary Table 1.

**Analysis of *DoMYB75* transcript level in wild type and *DoMYB75* transgenic *A. thaliana* plants.** Total RNA from one-week-old *A. thaliana* seedlings were extracted, purified and reverse transcribed as indicated above. The 2 × BlueStar™ PCR Master Mix kit was used for semi-quantitative RT-PCR analysis. One microliter of cDNA sample (about 400 ng/μL) was used for each independent PCR reaction using the following thermocycling conditions: 95 °C for 2 min, followed by 40 cycles of 98 °C for 10 s, 55 °C for 30 s, and a final extension at 72 °C for 60 s. The *DoMYB75* primer pair was the same as that used for vector construction. The primers UBQ10F/R for the *A. thaliana* ubiquitin gene (*AtUBQ10*) are listed in Supplementary Table 1.

**Analysis of water-soluble polysaccharide content.** The WSPs in *D. officinale* stems were extracted and determined as previously described<sup>34</sup>. Whole mature and dry *A. thaliana* seeds were ground to a fine powder using a tissue lyser (TL2020, Beijing Haoyuan Technology Co. Ltd., Beijing, China). Twenty mg of powder was weighed precisely, pre-extracted twice with 1 mL 80% (v/v) hot ethanol for 20 min in each extraction step, and centrifuged by a Centrifuge 5424 R (Eppendorf, Hamburg, Germany) at 10,000 rpm for 10 min at 16 °C. The supernatant was discarded. The pellet was suspended with 2 mL of distilled water, then incubated in an ultrasonic bath (VCX600, Sonics and Materials Inc., Newtown, CT, USA) for 2 h at 60 °C to extract the WSPs. After centrifugation at 10,000 rpm for 10 min at 16 °C, the supernatant was collected and used to analyze WSPs by the phenol-sulfuric acid method<sup>54</sup>, as described in He *et al.*<sup>34</sup>.

**Statistical analyses.** Data were analyzed using SigmaPlot12.3 software (Systat Software Inc., San Jose, CA, USA) by one-way analysis of variance (ANOVA) followed by Duncan's multiple range test (DMRT) or Dunnett's test.  $P < 0.05$  was considered to be statistically significant.

## Results

**Identification of MYB superfamily genes in orchids.** In this study, a total of 159 and 165 MYB TFs were identified in two orchids, *P. equestris* and *D. officinale*, respectively. In *P. equestris*, 40 MYBR genes, 115 R2R3-MYB genes, three 3R-MYB genes and one atypical MYB gene (CDC5-type) were found (Table 1). A similar number of MYB TF family members was found in *D. officinale*: 42 MYBR genes, 117 R2R3-MYB genes, four 3R-MYB genes and two atypical MYB genes (one 4RMYP and one CDC5-type) were identified in the *D. officinale* genome (Table 1). All the details of MYB genes in both orchids are listed in Supplementary Table 2.

**Classification of MYBR and R2R3-MYB proteins in two orchids.** In plants, MYBR proteins can be divided into five subgroups: CCA1/RR-like, CPC-like, I-box-like, TBP-like and TRF-like<sup>2</sup>. Genes in these five subgroups were also found in *P. equestris* and *D. officinale*, 21 and 24 CCA1/RR-like, 4 and 3 CPC-like, 5 and 6 I-box-like, 8 and 7 TBP-like, and 2 and 2 TRF-like, respectively. The CCA1/RR-like subfamily is the largest of the five subfamilies, while the TRF-like family is the smallest, with just two members in each orchid (Fig. 1A,B).

Compared with the MYB-related family, the R2R3-MYB family contained nearly three times as many genes as MYBR proteins in both orchids. The R2R3-MYB proteins are classified into 25 clades based on conservation

Clade	<i>D. officinale</i>	<i>P. equestris</i>	<i>A. thaliana</i>	Functions in <i>A. thaliana</i>	References
C1 ( <i>AtMYB5</i> )	<i>DoMYB01</i>	<i>PeMYB42</i>	<i>AtMYB5</i>	Testa mucilage synthesis	25
	<i>DoMYB06</i>	<i>PeMYB91</i>			
	<i>DoMYB42</i>	<i>PeMYB93</i>			
		<i>PeMYB95</i>			
	<i>DoMYB117</i>	<i>PeMYB104</i>	<i>AtMYB113</i>	Anthocyanin biosynthesis	73
C2 (S6)	<i>DoMYB74</i>	<i>PeMYB30</i>	<i>AtMYB114</i>	Anthocyanin biosynthesis	73
	<i>DoMYB75</i>	<i>PeMYB37</i>	<i>AtMYB75</i>	Cell wall thickening, testa, anthocyanin biosynthesis	70–72
	<i>DoMYB86</i>	<i>PeMYB54</i>	<i>AtMYB90</i>		
		<i>AtMYB0</i>			
C3 (S15)			<i>AtMYB23</i>	Trichome morphogenesis	75
			<i>AtMYB66</i>		
C4 (S5)			<i>AtMYB123</i>		
	<i>DoMYB22</i>	<i>PeMYB40</i>	<i>AtMYB108</i>	Stamen maturation	76
	<i>DoMYB96</i>	<i>PeMYB75</i>	<i>AtMYB112</i>	Flavonoid biosynthesis	77
C5 (S20)		<i>PeMYB86</i>	<i>AtMYB116</i>		
		<i>PeMYB87</i>	<i>AtMYB2</i>	Phosphate-starvation responses	78
			<i>AtMYB62</i>	Phosphate-starvation responses	79
		<i>AtMYB78</i>			
C6 (S19)	<i>DoMYB44</i>	<i>PeMYB18</i>	<i>AtMYB21</i>		
	<i>DoMYB62</i>	<i>PeMYB70</i>	<i>AtMYB24</i>	Stamen maturation	76
	<i>DoMYB108</i>	<i>PeMYB02</i>	<i>AtMYB121</i>		
	<i>DoMYB34</i>	<i>PeMYB27</i>	<i>AtMYB27</i>		
	<i>DoMYB53</i>	<i>PeMYB34</i>	<i>AtMYB48</i>		
C7 (S17)	<i>DoMYB66</i>	<i>PeMYB41</i>	<i>AtMYB59</i>	Root growth; cell cycle progression	80
	<i>DoMYB77</i>	<i>PeMYB44</i>	<i>AtMYB71</i>		
	<i>DoMYB91</i>	<i>PeMYB49</i>			
		<i>PeMYB94</i>			
	<i>DoMYB02</i>	<i>PeMYB01</i>	<i>AtMYB105</i>		
	<i>DoMYB111</i>	<i>PeMYB04</i>	<i>AtMYB110</i>		
	<i>DoMYB16</i>	<i>PeMYB26</i>	<i>AtMYB117</i>		
	<i>DoMYB26</i>	<i>PeMYB32</i>	<i>AtMYB52</i>	Secondary cell wall biosynthesis	81
	<i>DoMYB27</i>	<i>PeMYB36</i>	<i>AtMYB54</i>	Secondary cell wall biosynthesis	81
C8 (S21)	<i>DoMYB31</i>	<i>PeMYB45</i>	<i>AtMYB56</i>		
	<i>DoMYB36</i>	<i>PeMYB48</i>	<i>AtMYB69</i>		
	<i>DoMYB45</i>	<i>PeMYB64</i>			
	<i>DoMYB52</i>	<i>PeMYB65</i>			
	<i>DoMYB64</i>	<i>PeMYB67</i>			
	<i>DoMYB88</i>	<i>PeMYB80</i>			
	<i>DoMYB97</i>	<i>PeMYB82</i>			
	<i>DoMYB98</i>	<i>PeMYB84</i>			
C9 ( <i>AtMYB124/AtMYB88</i> )	<i>DoMYB106</i>	<i>PeMYB29</i>	<i>AtMYB124</i>		
			<i>AtMYB88</i>		
C10 ( <i>AtMYB91</i> )	<i>DoMYB18</i>	<i>PeMYB57</i>	<i>AtMYB91</i>		
		<i>PeMYB105</i>			
		<i>PeMYB106</i>			
C11		<i>PeMYB112</i>			
		<i>PeMYB115</i>			
		<i>PeMYB89</i>			
	<i>DoMYB03</i>	<i>PeMYB31</i>	<i>AtMYB44</i>	Root development	82
	<i>DoMYB24</i>	<i>PeMYB38</i>	<i>AtMYB77</i>		
C12 (S22)	<i>DoMYB30</i>	<i>PeMYB73</i>	<i>AtMYB70</i>		
	<i>DoMYB84</i>		<i>AtMYB73</i>		
	<i>DoMYB94</i>				
		<i>AtMYB1</i>	Salt stress response	83	
C13 (S23)			<i>AtMYB109</i>		
			<i>AtMYB35</i>		
Continued					

Clade	<i>D. officinale</i>	<i>P. equestris</i>	<i>A. thaliana</i>	Functions in <i>A. thaliana</i>	References	
C14 ( <i>AtMYB125</i> )	<i>DoMYB114</i>	<i>PeMYB09</i>	<i>AtMYB125</i>			
	<i>DoMYB70</i>	<i>PeMYB88</i>	<i>AtMYB36</i>	Root development	84	
	<i>DoMYB109</i>	<i>PeMYB110</i>	<i>AtMYB100</i>			
	<i>DoMYB13</i>	<i>PeMYB28</i>	<i>AtMYB115</i>	Benzoyloxy glucosinolate pathway	85	
C15 ( <i>S25</i> )	<i>DoMYB15</i>	<i>PeMYB90</i>	<i>AtMYB118</i>	Benzoyloxy glucosinolate pathway	86	
	<i>DoMYB83</i>		<i>AtMYB119</i>	Gametogenesis	86	
	<i>DoMYB87</i>		<i>AtMYB22</i>			
			<i>AtMYB64</i>	Gametogenesis	86	
	<i>DoMYB107</i>	<i>PeMYB107</i>	<i>AtMYB101</i>	Fertilization	87	
	<i>DoMYB12</i>	<i>PeMYB114</i>	<i>AtMYB104</i>			
C16 ( <i>S18</i> )	<i>DoMYB25</i>	<i>PeMYB23</i>	<i>AtMYB120</i>	Fertilization	87	
	<i>DoMYB50</i>	<i>PeMYB59</i>	<i>AtMYB33</i>			
	<i>DoMYB55</i>		<i>AtMYB65</i>			
	<i>DoMYB72</i>		<i>AtMYB81</i>			
C17			<i>AtMYB97</i>	Fertilization	87	
	<i>DoMYB73</i>	<i>PeMYB50</i>				
		<i>PeMYB83</i>				
	<i>DoMYB04</i>	<i>PeMYB05</i>	<i>AtMYB36</i>			
	<i>DoMYB11</i>	<i>PeMYB08</i>	<i>AtMYB37</i>			
	<i>DoMYB35</i>	<i>PeMYB102</i>	<i>AtMYB38</i>			
	<i>DoMYB38</i>	<i>PeMYB103</i>	<i>AtMYB68</i>			
	<i>DoMYB49</i>	<i>PeMYB12</i>	<i>AtMYB84</i>			
	C18 ( <i>S14</i> )	<i>DoMYB56</i>	<i>PeMYB13</i>	<i>AtMYB87</i>	Cell wall organization	88
		<i>DoMYB68</i>	<i>PeMYB46</i>			
<i>DoMYB79</i>		<i>PeMYB51</i>				
<i>DoMYB89</i>		<i>PeMYB53</i>				
<i>DoMYB95</i>		<i>PeMYB66</i>				
		<i>PeMYB74</i>				
C19 ( <i>AtMYB80/AtMYB35</i> )		<i>PeMYB76</i>				
	<i>DoMYB102</i>	<i>PeMYB24</i>	<i>AtMYB80</i>			
	<i>DoMYB113</i>	<i>PeMYB69</i>	<i>AtMYB35</i>			
	<i>DoMYB19</i>					
	<i>DoMYB05</i>	<i>PeMYB33</i>	<i>AtMYB30</i>			
	<i>DoMYB37</i>	<i>PeMYB78</i>	<i>AtMYB31</i>			
C20 ( <i>S1</i> )	<i>DoMYB57</i>	<i>PeMYB97</i>	<i>AtMYB60</i>			
		<i>PeMYB98</i>	<i>AtMYB94</i>	Cuticular wax biosynthesis	89	
			<i>AtMYB96</i>	Cuticular wax biosynthesis	89	
	<i>DoMYB23</i>	<i>PeMYB15</i>	<i>AtMYB107</i>	Suberin deposition	90	
	C21 ( <i>S10</i> )	<i>DoMYB47</i>	<i>PeMYB22</i>	<i>AtMYB39</i>		
<i>DoMYB82</i>			<i>AtMYB9</i>	Suberin deposition	90	
<i>DoMYB104</i>		<i>PeMYB79</i>	<i>AtMYB53</i>			
C22 ( <i>S24</i> )	<i>DoMYB61</i>		<i>AtMYB92</i>			
	<i>DoMYB76</i>		<i>AtMYB93</i>	Root development	91	
	<i>DoMYB39</i>	<i>PeMYB21</i>	<i>AtMYB102</i>			
	<i>DoMYB40</i>	<i>PeMYB99</i>	<i>AtMYB41</i>	Osmotic stress responses	92	
C23 ( <i>S11</i> )	<i>DoMYB41</i>		<i>AtMYB74</i>	Salt stress responses	93	
	<i>DoMYB51</i>					
	<i>DoMYB58</i>					
	<i>DoMYB32</i>	<i>PeMYB25</i>	<i>AtMYB106</i>			
C24 ( <i>S9</i> )	<i>DoMYB60</i>	<i>PeMYB52</i>	<i>AtMYB16</i>	Cuticle formation	94	
	<i>DoMYB69</i>	<i>PeMYB56</i>	<i>AtMYB17</i>			
	<i>DoMYB100</i>	<i>PeMYB06</i>	<i>AtMYB122</i>			
C25 ( <i>S12</i> )	<i>DoMYB43</i>		<i>AtMYB51</i>			
			<i>AtMYB34</i>			
			<i>AtMYB29</i>			
			<i>AtMYB76</i>			
	<i>DoMYB10</i>	<i>PeMYB03</i>	<i>AtMYB13</i>			

Continued

Clade	<i>D. officinale</i>	<i>P. equestris</i>	<i>A. thaliana</i>	Functions in <i>A. thaliana</i>	References
	<i>DoMYB28</i>	<i>PeMYB109</i>	<i>AtMYB14</i>	Testa polymer biosynthesis	25
C26 (S2)	<i>DoMYB54</i>	<i>PeMYB47</i>	<i>AtMYB15</i>		
	<i>DoMYB71</i>				
	<i>DoMYB78</i>				
	<i>DoMYB99</i>				
	<i>DoMYB46</i>	<i>PeMYB14</i>	<i>AtMYB58</i>	Secondary cell wall biosynthesis	95
C27 (S3)	<i>DoMYB92</i>	<i>PeMYB19</i>	<i>AtMYB63</i>	Secondary cell wall biosynthesis	95
		<i>PeMYB85</i>			
	<i>DoMYB116</i>	<i>PeMYB113</i>	<i>AtMYB11</i>	Flavonoid biosynthesis	96
G28 (S7)	<i>DoMYB59</i>	<i>PeMYB43</i>	<i>AtMYB111</i>	Flavonoid biosynthesis	96
		<i>PeMYB68</i>	<i>AtMYB12</i>	Flavonoid biosynthesis	96
C29 ( <i>AtMYB47/AtMYB95</i> )	<i>DoMYB20</i>	<i>PeMYB63</i>	<i>AtMYB47</i>		
			<i>AtMYB95</i>		
	<i>DoMYB07</i>	<i>PeMYB17</i>	<i>AtMYB20</i>		
	<i>DoMYB103</i>	<i>PeMYB20</i>	<i>AtMYB40</i>	Secondary cell wall biosynthesis	81
C30 (S8)	<i>DoMYB67</i>	<i>PeMYB62</i>	<i>AtMYB42</i>	Secondary cell wall biosynthesis	81
		<i>PeMYB96</i>	<i>AtMYB43</i>	Secondary cell wall biosynthesis	81
			<i>AtMYB85</i>	Secondary cell wall biosynthesis	81
			<i>AtMYB99</i>		
	<i>DoMYB21</i>	<i>PeMYB07</i>	<i>AtMYB3</i>	Phenylpropanoid biosynthesis	97
	<i>DoMYB33</i>	<i>PeMYB35</i>	<i>AtMYB32</i>		
C31 (S4)	<i>DoMYB48</i>	<i>PeMYB39</i>	<i>AtMYB4</i>		
	<i>DoMYB80</i>	<i>PeMYB58</i>	<i>AtMYB7</i>		
	<i>DoMYB85</i>	<i>PeMYB60</i>			
	<i>DoMYB93</i>				
	<i>DoMYB08</i>		<i>AtMYB18</i>		
C32 (S16)			<i>AtMYB19</i>		
			<i>AtMYB45</i>		
	<i>DoMYB09</i>	<i>PeMYB100</i>	<i>AtMYB50</i>		
	<i>DoMYB110</i>	<i>PeMYB101</i>	<i>AtMYB55</i>		
	<i>DoMYB112</i>	<i>PeMYB108</i>	<i>AtMYB61</i>	Testa mucilage synthesis	26,98
	<i>DoMYB115</i>	<i>PeMYB11</i>	<i>AtMYB86</i>		
C33 (S13)	<i>DoMYB14</i>	<i>PeMYB111</i>			
	<i>DoMYB29</i>	<i>PeMYB55</i>			
	<i>DoMYB65</i>	<i>PeMYB71</i>			
	<i>DoMYB81</i>	<i>PeMYB77</i>			
		<i>PeMYB81</i>			
C34 ( <i>AtMYB26/AtMYB67/AtMYB103</i> )	<i>DoMYB105</i>	<i>PeMYB16</i>	<i>AtMYB26</i>	Secondary cell wall thickening	99
	<i>DoMYB63</i>	<i>PeMYB61</i>	<i>AtMYB67</i>		
	<i>DoMYB90</i>	<i>PeMYB72</i>	<i>AtMYB103</i>	Lignin biosynthesis; secondary cell wall thickening	100
		<i>PeMYB92</i>			
C35 ( <i>AtMYB46/AtMYB83</i> )	<i>DoMYB17</i>	<i>PeMYB10</i>	<i>AtMYB46</i>	Secondary wall biosynthesis	69,101
			<i>AtMYB83</i>	Secondary wall biosynthesis	69

**Table 2.** Classification and putative functions of R2R3-MYB transcription factors.

of the MYB domain and C terminal amino acid motifs in *A. thaliana*<sup>10</sup>. To survey the classification within the R2R3-MYB gene family, we conducted a phylogenetic analysis of *A. thaliana* (126 members), *P. equestris* (115 members) and *D. officinale* (117 members) R2R3-MYB proteins. Based on a phylogenetic tree, all the MYB proteins could be grouped into 35 clades (C1-C35) (Table 2 and Supplementary Fig. 1). The C3 (S15), C4 (S5) and C13 (S23) clade genes were absent in both orchids, while C11 subfamily genes were only found in *P. equestris* and C17 clade genes were only present in the two orchids (Table 2 and Supplementary Fig. 1). This result suggests that the C3 (S15), C4 (S5) and C13 (S23) proteins might have been lost in orchids after divergence from the most recent common ancestor.

**Non-synonymous (Ka) and synonymous (Ks) substitutions in orthologous gene pairs between *P. equestris* and *D. officinale*.** The Ka/Ks value is regarded as a pointer to assess selective pressure on a protein-coding gene. A Ka/Ks ratio less than 1 indicates a negative or purifying selection, a Ka/Ks ratio equal to 1 indicates neutral evolution, while a Ka/Ks ratio greater than 1 indicates positive or adaptive evolution. In total,

Paralogous gene pairs		Ka	Ks	Ka/Ks	P-Value (Fisher)
<i>D. officinale</i>	<i>P. equestris</i>				
DoMYBR22	PeMYBR29	0.252639	0.226031	1.11772	0.359576
DoMYBR01	PeMYBR11	0.229767	0.264664	0.868147	4.37E-01
DoMYB30	PeMYBCDC	0.540245	0.678633	0.796079	3.34E-01
DoMYB113	PeMYB24	0.281005	0.395265	0.710927	9.42E-02
DoMYBR35	PeMYBR22	0.174475	0.302301	0.577156	1.47E-01
DoMYBR02	PeMYBR06	0.187853	0.333361	0.563513	1.95E-05
DoMYBR28	PeMYBR37	0.288114	0.517813	0.556406	0.002783
DoMYBR30	PeMYBR33	0.587592	1.18628	0.495324	2.48E-05
DoMYBR33	PeMYBR08	0.14577	0.324747	0.448873	4.66E-08
DoMYB50	PeMYB108	0.1481	0.330785	0.447722	8.97E-07
DoMYB94	PeMYB73	0.181577	0.411152	0.441631	0.001206
DoMYB35	PeMYB102	0.129134	0.294916	0.437867	0.000129
DoMYB27	PeMYB80	0.135674	0.311095	0.436116	3.87E-05
DoMYB727	PeMYB23	0.268111	0.652068	0.41117	5.06E-09
DoMYB22	PeMYB40	0.195574	0.487158	0.401458	1.66E-05
DoMYBR25	PeMYBR32	0.174571	0.461248	0.378476	4.33E-08
DoMYB20	PeMYB63	0.216538	0.5768	0.375412	9.05E-06
DoMYB106	PeMYB29	0.124663	0.350774	0.355394	3.95E-09
DoMYBR13	PeMYBR36	0.160839	0.461041	0.34886	8.74E-08
DoMYB65	PeMYB11	0.127844	0.376824	0.339266	1.8E-10
DoMYBR14	PeMYBR30	0.148252	0.460324	0.322059	1.47E-18
DoMYB02	PeMYB26	0.139964	0.434743	0.321947	2.6E-07
DoMYB46	PeMYB14	0.143144	0.445163	0.321555	3.64E-09
DoMYB87	PeMYB28	0.172107	0.540112	0.31865	6.7E-10
DoMYB3R1	PeMYB3R1	0.115854	0.367231	0.31548	4.86E-14
DoMYB63	PeMYB61	0.1622	0.485588	0.30112	7.53E-11
DoMYB97	PeMYB67	0.122021	0.419452	0.290905	4.14E-09
DoMYBR32	PeMYBR10	0.07528	0.263304	0.285905	3.16E-08
DoMYBR20	PeMYBR16	0.078868	0.286058	0.275706	6.34E-11
DoMYB95	PeMYB66	0.099102	0.362533	0.273361	2.15E-08
DoMYB52	PeMYB32	0.156645	0.578849	0.270614	4.84E-12
DoMYB36	PeMYB48	0.123647	0.465682	0.265519	1.93E-07
DoMYBR15	PeMYBR21	0.237202	0.901119	0.263231	0.003537
DoMYB06	PeMYB91	0.131802	0.517651	0.254616	1.24E-10
DoMYBR19	PeMYBR35	0.077925	0.310476	0.250985	2.15E-10
DoMYB09	PeMYB81	0.100041	0.398857	0.250819	8.1E-10
DoMYBR09	PeMYBR12	0.098051	0.392184	0.250012	1.16E-12
DoMYB77	PeMYB34	0.099167	0.399335	0.248331	4.09E-09
DoMYBR40	PeMYBR01	0.125086	0.51066	0.24495	4.89E-06
DoMYB91	PeMYB41	0.079636	0.333581	0.238729	2.76E-08
DoMYB38	PeMYB53	0.095368	0.4069	0.234376	7.07E-14
DoMYB04	PeMYB74	0.083609	0.36295	0.230358	4.17E-10
DoMYB43	PeMYB06	0.109957	0.48052	0.228829	3.95E-13
DoMYB68	PeMYB51	0.072732	0.325264	0.223608	3.18E-10
DoMYBR10	PeMYBR31	0.07664	0.349733	0.219138	4.03E-08
DoMYB111	PeMYB45	0.093899	0.434686	0.216015	2.4E-08
DoMYB19	PeMYB69	0.092868	0.459706	0.202017	2.53E-16
DoMYB92	PeMYB85	0.105135	0.526037	0.199862	1.51E-13
DoMYB108	PeMYB44	0.092366	0.469872	0.196577	1.14E-12
DoMYBR17	PeMYBR05	0.060215	0.308167	0.195397	1.47E-21
DoMYBR11	PeMYBR02	0.093764	0.48181	0.194608	1.11E-15
DoMYB26	PeMYB04	0.132161	0.687219	0.192312	6.56E-12
DoMYBR24	PeMYBR20	0.061053	0.343438	0.177769	5.88E-16
DoMYB23	PeMYB22	0.078265	0.460628	0.16991	3.87E-21
DoMYB07	PeMYB20	0.060264	0.356788	0.168906	1.25E-13
DoMYB84	PeMYB31	0.199724	1.27411	0.156756	5.73E-22

Continued



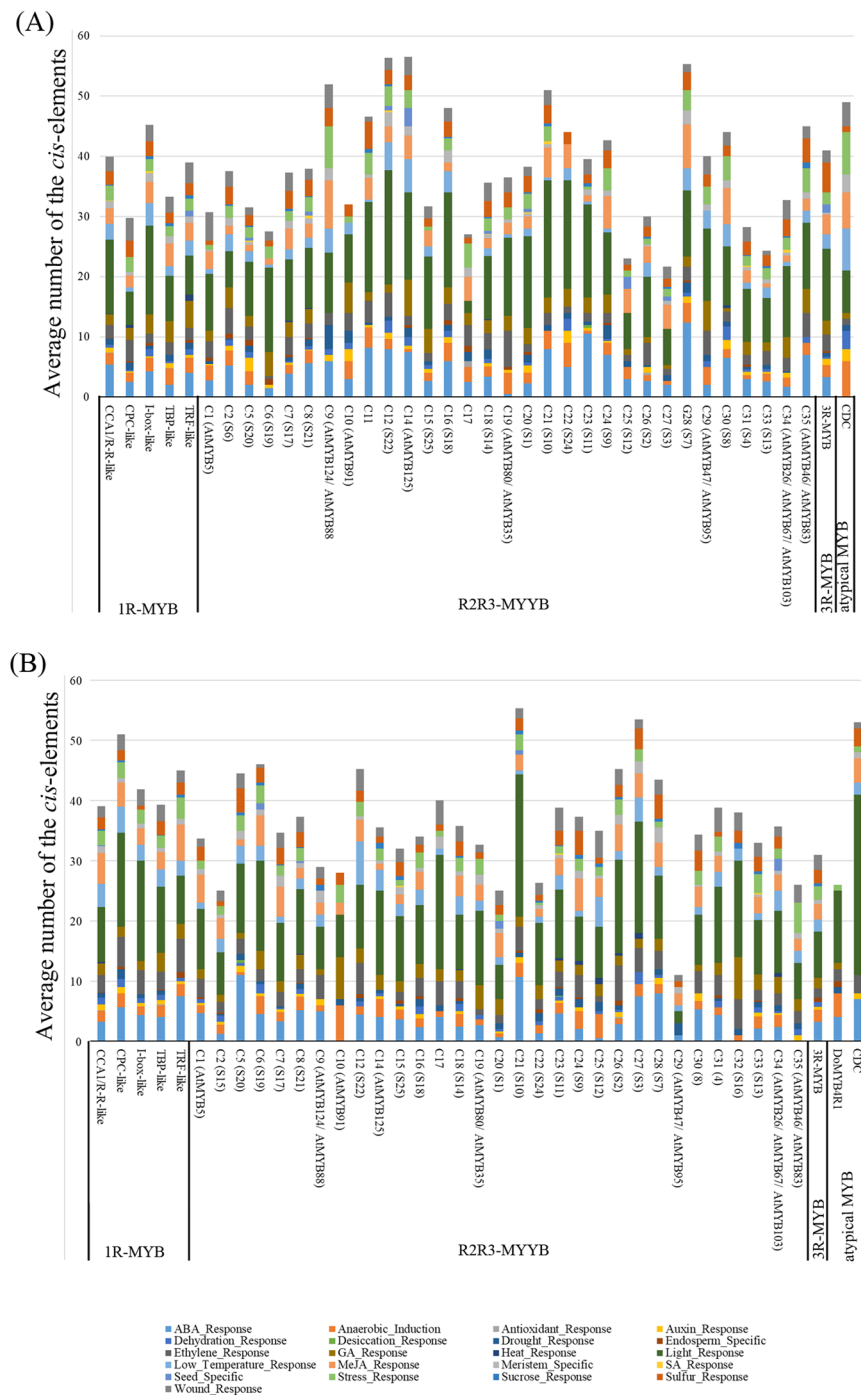
Paralogous gene pairs		Ka	Ks	Ka/Ks	P-Value (Fisher)
<i>D. officinale</i>	<i>P. equestris</i>				
DoMYB59	PeMYB114	0.125627	0.887715	0.141517	1.47E-17
DoMYB13	PeMYB90	0.123877	0.879398	0.140866	8.44E-18
DoMYB17	PeMYB10	0.103123	0.733765	0.14054	1.39E-24
DoMYB01	PeMYB42	0.09443	0.685337	0.137786	5.07E-19
DoMYB96	PeMYB86	0.062697	0.46096	0.136014	2.21E-19
DoMYBR37	PeMYBR34	0.05965	0.454563	0.131225	6.45E-23
DoMYB57	PeMYB33	0.044033	0.349604	0.125952	1.68E-16
DoMYB64	PeMYB36	0.049998	0.39719	0.12588	1.82E-17
DoMYB66	PeMYB02	0.093345	0.755975	0.123477	3.24E-21
DoMYB24	PeMYB38	0.109397	0.886578	0.123393	3.53E-32
DoMYBR03	PeMYBR17	0.067015	0.544493	0.123078	2.42E-25
DoMYBR31	PeMYBR26	0.098098	0.816452	0.120151	1.76E-10
DoMYB61	PeMYB79	0.058864	0.495516	0.118792	1.14E-22
DoMYB103	PeMYB62	0.064227	0.566367	0.113402	4.68E-22
DoMYBR39	PeMYBR28	0.060879	0.555284	0.109635	1.13E-08
DoMYB51	PeMYB21	0.042387	0.400007	0.105965	2.99E-26
DoMYB18	PeMYB57	0.064086	0.618519	0.103613	5.75E-32
DoMYB47	PeMYB15	0.061099	0.602174	0.101465	7.01E-31
DoMYB34	PeMYB49	0.072019	0.786213	0.091603	5.96E-25
DoMYB69	PeMYB56	0.071401	0.785038	0.090952	1.5E-34
DoMYB44	PeMYB18	0.036016	0.429379	0.083879	1.45E-18
DoMYB112	PeMYB99	0.218163	2.74748	0.079405	2.25E-12
DoMYB12	PeMYB107	0.231365	3.25787	0.071017	5.2E-11
DoMYBR06	PeMYBR25	0.037052	0.562053	0.065922	1.61E-12
DoMYB09	PeMYB16	0.039861	0.637587	0.062519	5.85E-34
DoMYB80	PeMYB35	0.045443	0.806638	0.056336	7.78E-28
DoMYB116	PeMYB113	0.136094	2.97652	0.045722	4.68E-23

**Table 3.** Ka/Ks analysis and estimated selective pressure for orthologous gene pairs between *P. equestris* and *D. officinale*. Ka non-synonymous substitutions per non-synonymous site, Ks synonymous substitutions per synonymous site); Ka/Ks the ratio.

84 orthologous gene pairs between *P. equestris* and *D. officinale* were found (Table 3). In other words, about 50% of orchid MYB genes appeared to be duplicated. This suggests that most orchid MYB genes underwent functional diversity and expansion during evolution. In our research, most of these orthologous gene pairs were deduced to be under negative selection with a Ka/Ks ratio less than 1, except for *DoMYBR22* and *PeMYBR29*, which had a Ka/Ks ratio greater than 1 (Table 3).

**Cis-responsive element analysis of MYB genes from *P. equestris* and *D. officinale*.** All the 2000 bp upstream regions of the initiation codon of MYB genes from *P. equestris* and *D. officinale* were obtained from their respective genomes. The stress response elements, tissue-specific activation, hormone responsive elements, and other responsive elements were identified and analyzed (Fig. 2). Various stress responsive elements, including anaerobic induction, and response to antioxidant, dehydration, desiccation, drought, heat, low temperature, stress, and wound elements, were analyzed. Only anaerobic induction, low temperature response and wound response elements were widely present in the MYB gene promoters of both orchids (Fig. 2 and Supplementary Table 3). Hormone responsive elements such as ABA, ethylene, GA and MeJA response were abundant in the putative promoters of MYB genes, especially the ABA response element (Fig. 2). ABA, ethylene and MeJA are related to stress response in plants<sup>55</sup>. These results suggested that the MYB genes may play an important role in stress response in orchids.

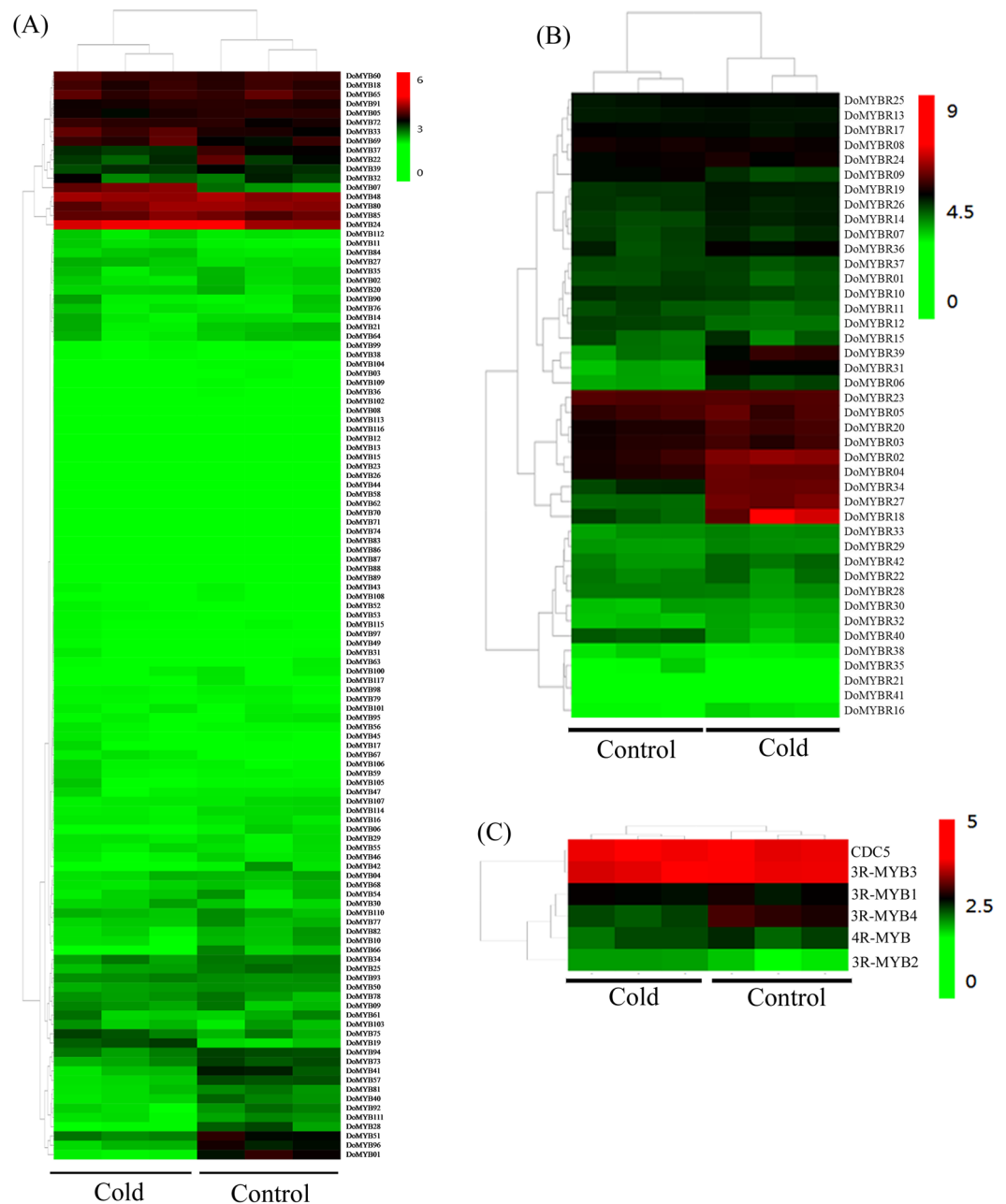
**Expression analyses of MYB genes under cold stress in *D. officinale*.** MYB genes are involved in a plant's response to stress, such as drought and cold stress<sup>56–58</sup>. To gain insight into *D. officinale* MYB proteins in stress responses, the expression of MYB genes were evaluated under the control condition (20 °C) and cold stress (4 °C) by comparing the FPKM values for each gene at 20 °C and 4 °C. Only nine of 117 R2R3-MYB genes were modulated by cold stress, consisting of three up-regulated genes (*DoMYB07*, –33, and –69) and six down-regulated genes (*DoMYB01*, –22, –37, –41, –51, and –96) (Fig. 3A and Supplementary Table 4). A total of 15 out of the 42 MYBR genes (*DoMYBR02*, –04, –06, –09, –12, –14, –18, –20, –27, –31, –34, –36, –39, –40, and –42) were modulated by cold stress, nine of which were up-regulated by at least two-fold while three genes were down-regulated (less than 0.5-fold, Fig. 3B). Four up-regulated genes (*DoMYBR06*, –18, –31, and –39) were from the I-box-like subfamily, seven up-regulated genes (*DoMYBR02*, –04, –20, –27, –34, –36



**Figure 2.** Average number of *cis*-responsive elements of orchid MYB genes from each group. The *cis*-responsive elements were analyzed in the 2 kb upstream promoter region of the initiation codon using the PlantCARE database and the PLACE database.

and −42) were from the CCA1/R-R-like subfamily, and one gene (*DoMYBR14*) was from the TBP-like clade (Fig. 3 and Supplementary Table 4). Three 3R-MYBs, all 4R-MYBs and only one CDC5-type genes of *D. officinale* showed no differences between the control and cold stress (Fig. 3 and Supplementary Table 4). One 3R-MYB (*DoMYB3R4*) gene was down-regulated (Fig. 3 and Supplementary Table 4).

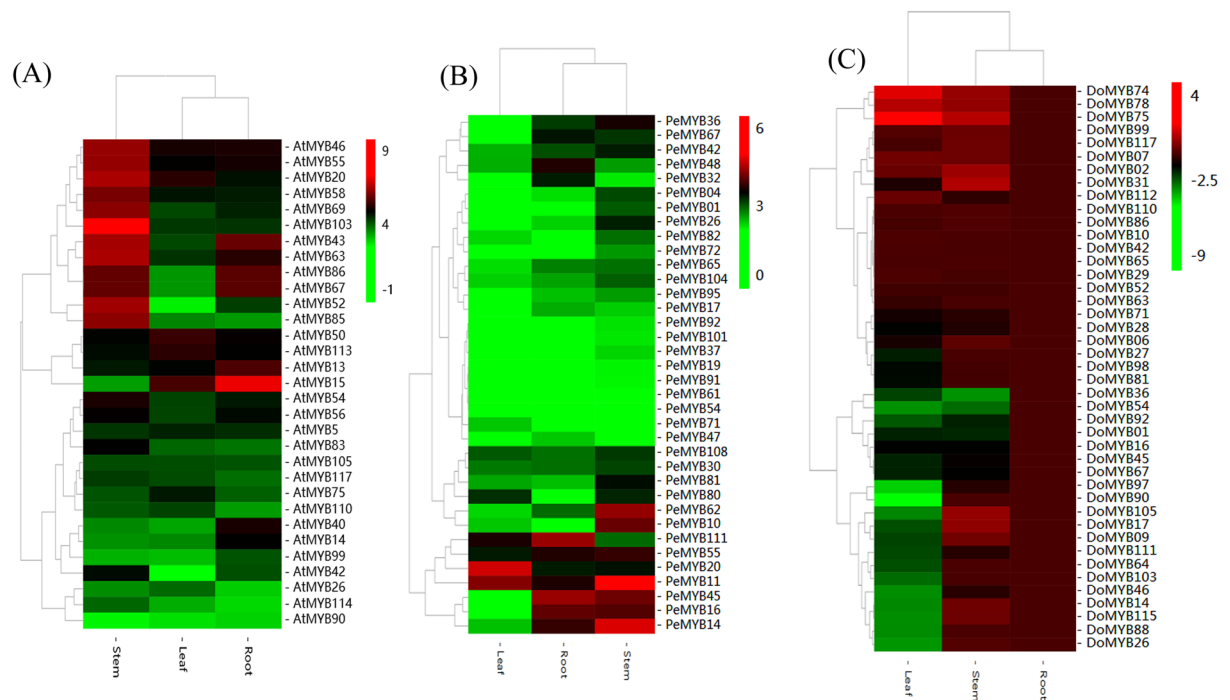
**Identification and expression analysis of R2R3-MYB genes related to polysaccharide biosynthesis.** Polysaccharides are abundant in the secondary cell wall and testa mucilage in *A. thaliana*<sup>24,59</sup>. Members of nine clades, namely C1 (AtMYB5), C2 (S6), C8 (S21), C26 (S2), C27 (S3), C30 (S8), C33 (S13), C34 (AtMYB26/AtMYB67/AtMYB103) and C35 (AtMYB46/AtMYB83), are involved in secondary cell wall or testa mucilage biosynthesis in *A. thaliana* (Table 2). The genes from these clades are probably involved in polysaccharide



**Figure 3.** Expression patterns of the four groups of MYB genes from *D. officinale* under cold stress (4 °C). Expression profiles of R2R3-MYB genes (A), MYBR genes (B), and 3R-MYB and atypical MYB genes (C). The heatmap was generated using R version 3.4.1 with a color scale according to the gene expression level [ $\log_2$  (FPKM + 1)]. Red indicates high gene expression level while green indicates a low level of expression. Each column indicates a discrete biological sample. All treatments consisted of three biological replicates.

biosynthesis and are regarded as putative MYB genes of polysaccharide biosynthesis. A total of 32, 45 and 43 putative MYB polysaccharide biosynthesis genes were identified from *A. thaliana*, *P. equestris* and *D. officinale*, respectively.

To comparatively analyze their expression patterns, the mRNA steady state levels of the putative MYB genes related to polysaccharide biosynthesis were monitored in the roots, stems and leaves of *A. thaliana*, *P. equestris* and *D. officinale*. Interestingly, 13 of the putative MYB genes from *A. thaliana* (*AtMYB20*, *-42*, *-43*, *-46*, *-52*, *-54*, *-55*, *-58*, *-63*, *-69*, *-83*, *-85*, and *-103*) were highly expressed in stems, and most of these genes were also involved in secondary cell wall biosynthesis (Fig. 4A, Supplementary Table 5). Eight (*PeMYB10*, *-11*, *-14*, *-26*, *-36*, *-42*, *-62*, and *-81*) and nine (*DoMYB02*, *-09*, *-14*, *-17*, *-31*, *-99*, *-105*, *-115*, and *-117*) MYB genes were highly expressed in the stems of *P. equestris* and *D. officinale*, respectively (Fig. 4B,C, Supplementary Tables 5 and 6). All C35 (*AtMYB46/AtMYB83*) genes were abundantly expressed in stems (Fig. 4).



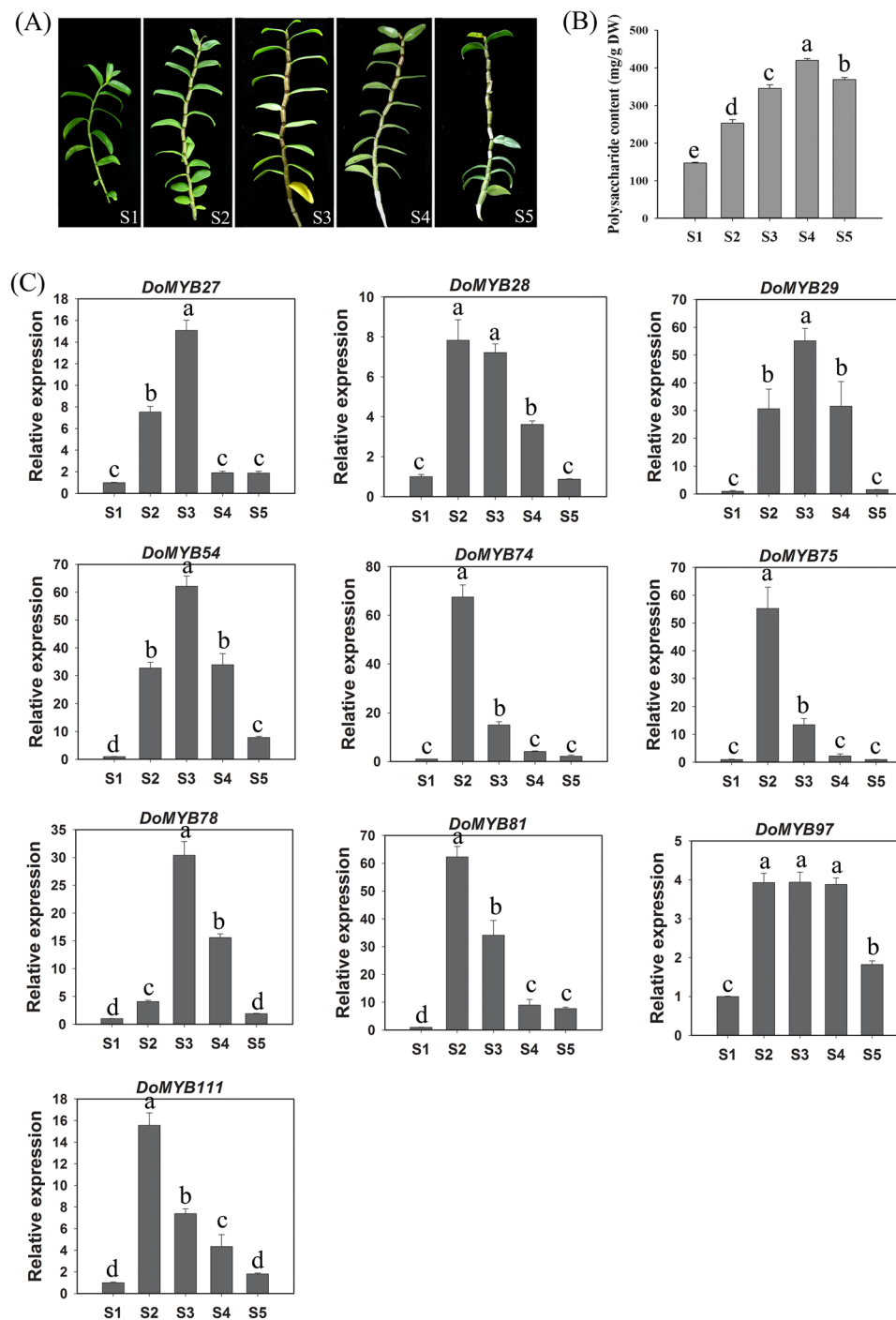
**Figure 4.** Heat map displaying the expression pattern of nine clades of R2R3-MYB genes in roots, stems and leaves. (A) Expression profile of 32 R2R3-MYB genes in different tissues or organs of *A. thaliana* from microarray data sets available at Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>). GEO accessions included: GSM131558, GSM131559 and GSM131560 (roots); GSM131655, GSM131656 and GSM131657 (stems); GSM131528, GSM131529 and GSM131530 (leaves). The color scale represents  $\log_2$  of the mean of gene expression. (B) Expression profile of 36 R2R3-MYB genes from *P. equestris*. Expression data were downloaded from Orchidbase (<http://orchidbase.itps.ncku.edu.tw>). The color scale was based on the gene expression level, which was measured as  $\log_2$  (gene expression value + 1). (C) Analysis of the 43 R2R3-MYB genes from *D. officinale* based on an analysis of qRT-PCR results, and visualized as a heatmap. The color scale represents  $\log_2$  of the mean of gene expression. At least two biological replicates were performed.

**Screening R2R3-MYB genes involved in biosynthesis of water-soluble polysaccharides in *D. officinale* and their expression under water deficit stress.** *D. officinale* is one of the most precious Chinese herbs with abundant WSPs in its stems<sup>33</sup>. The accumulation of WSPs in stems changes with plant growth<sup>34</sup>. To further screen the genes involved in the biosynthesis of WSPs, the expression of the 43 *D. officinale* genes belonging to the clusters C1 (AtMYB5), C2 (S6), C8 (S21), C26 (S2), C27 (S3), C30 (S8), C33 (S13), C34 (AtMYB26/AtMYB67/AtMYB103) and C35 (AtMYB46/AtMYB83) was analyzed across five plant developmental stages. In the S1 stage, plants were in the vegetative stage and had lowest WSP content (Fig. 5A,B). In stages S2–4, plants stopped growing and accumulated WSPs rapidly (Fig. 5A,B). In the S5 stage, plants started to undergo senescence and degradation of WSPs (Fig. 5A,B).

Ten genes, out of the 43 tested, showed an expression pattern that mirrored the accumulation pattern of WSP. Among these genes, six peaked at S2 (*DoMYB28*, –74, –75, –81, –97 and –111) and four at S3 (*DoMYB27*, –29, –54 and –78). All these genes were poorly expressed at S1, and most of them at S5 returned to the levels of S1 (Fig. 5C). The 10 MYB genes included two (*DoMYB74* and –75) from C2 (S6), three (*DoMYB27*, –97, and –111) from C8 (S21), three (*DoMYB28*, –54, and –78) from C26 (S2) and two (*DoMYB29* and –81) from C33 (S13). The remaining genes were either not detected or were inconsistent with changes in polysaccharide content (Supplementary Fig. 2).

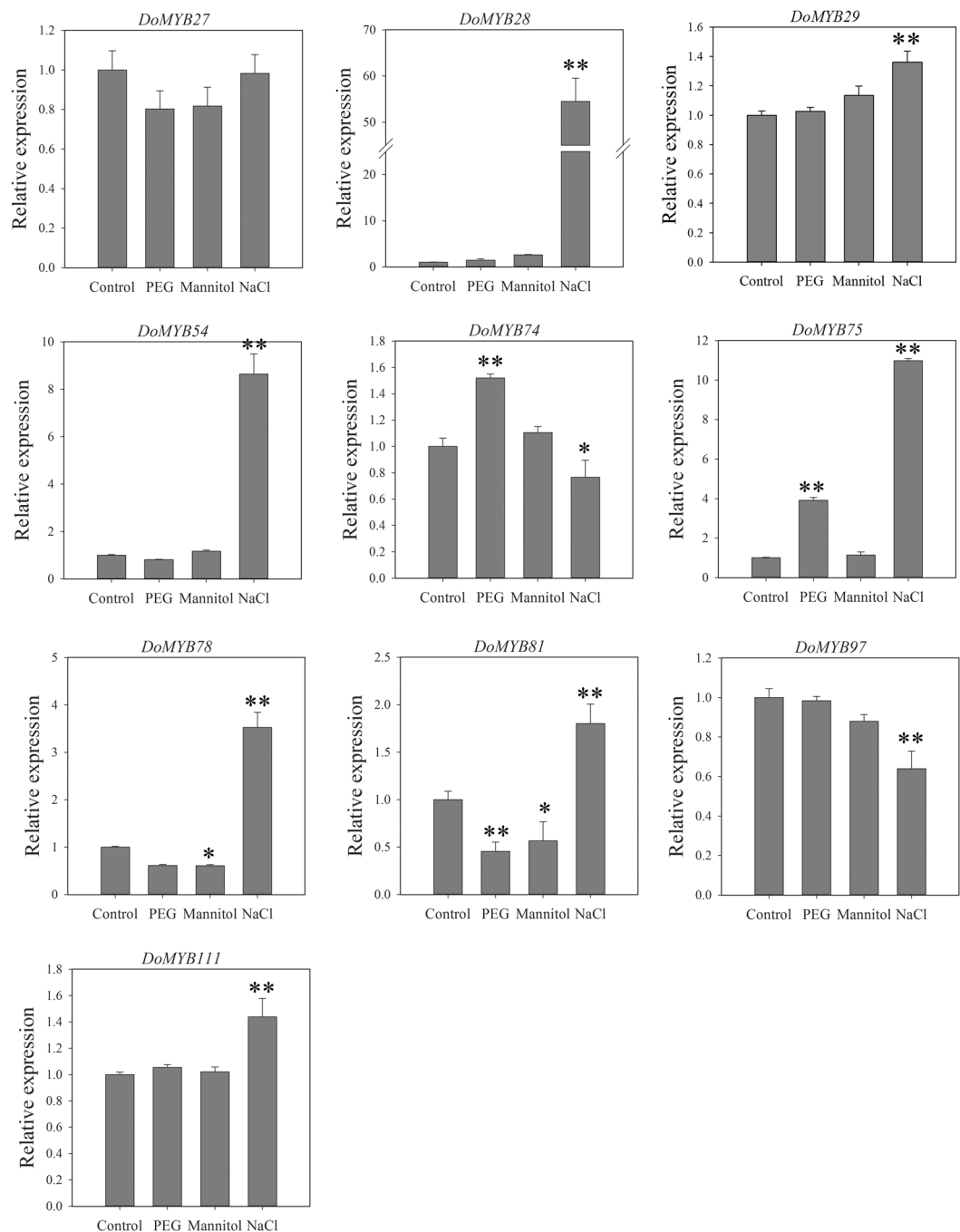
In our previous studies in *D. officinale*, the genes of the WSP biosynthetic pathway were shown to be involved in abiotic stresses influenced by PEG and NaCl treatment<sup>34,36</sup>. Thus, the expression of candidate genes under 150 g/L PEG, 300 mM mannitol and 250 mM NaCl treatments was analyzed. Our results indicate that most of the candidate genes were up-regulated under salinity stress, i.e. *DoMYB28*, –29, –54, –75, –78, –81 and –111. Only two genes (*DoMYB74* and –75) were up-regulated after exposure to PEG while only one gene (*DoMYB81*) was down-regulated after exposure to both PEG and mannitol (Fig. 6). *DoMYB27* expression showed no significant difference between stress treatments and the control (Fig. 6). *DoMYB75* was up-regulated in response to PEG and NaCl, similar with other WSP biosynthetic pathway genes such as *DoPMM*<sup>36</sup> and *DoCSLAs*<sup>34</sup>.

**Overexpression of *DoMYB75* in *A. thaliana* confirms its role in the biosynthesis of water-soluble polysaccharides.** *DoMYB75* may be involved in WSP biosynthesis, so it was used in further analyses. In a phylogenetic tree, the C2 (S6) clade was divided into two branches: cluster I included proteins of monocots (*P. equestris* and *D. officinale*) while cluster II only contained the proteins of *A. thaliana*, a dicot (Supplementary



**Figure 5.** Screening candidate R2R3-MYB genes related to the biosynthesis of WSPs in *D. officinale* by qRT-PCR. **(A)** Five stages of *D. officinale* stems were used to analyze gene expression. **(B)** Polysaccharide content in the stems of five stages. DW, dry weight. **(C)** Candidate genes identified using qRT-PCR. Details pertaining to S1-S5 can be found in the materials and methods and results sections. Bars represent mean  $\pm$  SD ( $n = 3$ ). Three biological replicates were performed. Different letters in each bar are significantly different at  $P < 0.05$  (Duncan's multiple range test).

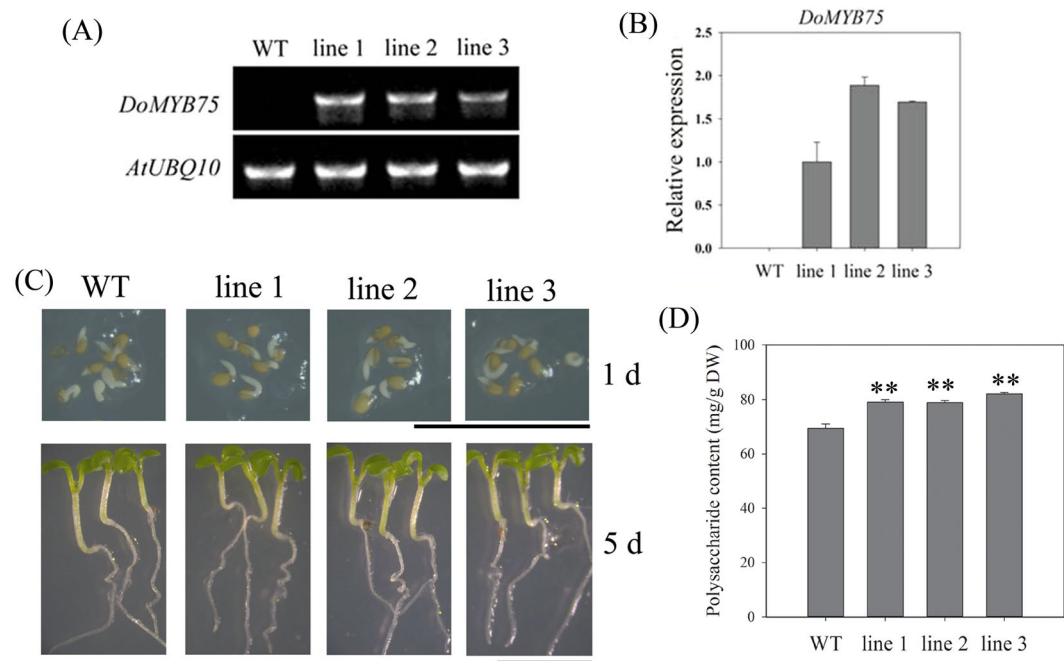
Fig. 3A). The complete CDS of *DoMYB75* without a termination codon was cloned into an over-expression vector (pCABIA 1302 vector) driven by the CaMV-35S promoter (Supplementary Fig. 3B). *DoMYB75* was detected and expressed in *DoMYB75* transgenic Arabidopsis lines, but not in WT plants (Fig. 7A,B). Overexpression of *DoMYB75* showed no difference in germinating seeds or seedlings (Fig. 7C). The average WSP content of seeds of all transgenic plants (around 79 mg/g DW) was significantly higher than WT plants (~69 mg/g DW, Fig. 7D). This indicates that *DoMYB75* plays a role in WSP biosynthesis in *A. thaliana* seeds.



**Figure 6.** Analysis of the expression level of ten R2R3-MYB genes by qRT-PCR after *D. officinale* seedlings were subjected to PEG (150 g/L), mannitol (300 mM) and salt (NaCl, 250 mM) treatments. Control are seedlings treated with  $\frac{1}{2}$ MS medium supplemented with 20 g/L sucrose (pH 5.4). Bars represent mean  $\pm$  SD of three technical replicates. Three biological replicates of each treatment were performed. \*indicates  $P < 0.05$ ; \*\*indicates  $P < 0.001$  between control and stress treatments following Dunnett's test.

## Discussion

**Identification and classification of MYB proteins.** In this study, we identified 159 and 165 MYB genes from *P. equestris* and *D. officinale* genomes, respectively. Although these orchid species belong to different genera, the number of MYB genes in their genomes is similar. Four groups of MYB proteins were found in *D. officinale*, similar to previous studies in other plant species such as *A. thaliana*<sup>6</sup>, *G. max*<sup>2,13</sup> and *P. bretschneideri*<sup>8</sup>. However, no 4R-MYB gene was found in the genome of *P. equestris*, which had only one atypical MYB gene, namely the CDC5-type gene. Another monocot, rice, contained one CDC5-type gene but not a 4R-MYB gene in its genome<sup>11</sup>, similar to *P. equestris* in our study. This suggests that 4R-MYB genes are not found in all higher plants and might not play essential roles in all plants.



**Figure 7.** Characterization of the *DoMYB75* gene in the biosynthesis of WSPs. **(A)** Analysis of the *DoMYB75* gene in wild type (WT) and transgenic lines by semi-quantitative PCR. **(B)** Analysis of the *DoMYB75* gene in WT and transgenic lines by qRT-PCR. Expression levels were calculated relative to transgenic line 1. **(C)** Germinating seeds (one day after stratification) and seedlings (5 days after stratification) of WT and transgenic lines showed no obvious phenotypic changes. **(D)** Content of WSPs in mature dry seeds of *A. thaliana*. \*\*Indicates  $P < 0.01$  between WT and transgenic lines following Dunnett's test.

The MYBR proteins in both orchids could be divided into five subfamilies, with CCA1/RR-like as the largest subfamily, and containing two TRB-like genes in their genomes. Five MYBR subfamilies were also found in other higher plants such as *A. thaliana*, *S. lycopersicum*, *V. vinifera*, and *B. distachyon*, with the number of CCA1/RR-like genes ranging from 21 to 42, and with all plants just containing only one or two TRF-like genes in their genomes<sup>2</sup>. The R2R3-MYB group is largest group of the MYB family, with more than 100 members having been found in the genomes of both monocots and dicots.

The two orchids contain 115 and 117 R2R3-MYB genes in their genomes, respectively. Feller *et al.*<sup>12</sup> predicted that the plant R2R3-MYB group underwent extensive amplification before the separation of monocots and dicots but after the separation of plants and animals. This explains why many members of R2R3-MYB genes have been found in orchids. The majority of clades of the R2R3-MYB family are present in both orchids and in the model plant *A. thaliana* (Supplementary Fig. 1), suggesting that they were present before the divergence between monocots and dicots. However, there is also an orchid-specific (C17) and a *P. equestris*-specific clade (C11), while in both orchids, C3 (S15), C4 (S5) and C13 (S23) are missing (Table 2). It can be deduced that several R2R3-MYB genes might have been lost in orchids during evolution, or experienced an amplification process that might have caused a change in their function in dicots.

**The MYB genes play a role in abiotic stress responses.** The MYB genes involved in abiotic stress responses have been widely investigated in plants and have mainly focused on the R2R3-MYB group. For example, R2R3-MYB genes were involved in high temperature stress by increasing the levels of cellular abscisic acid<sup>60</sup>, cold stress by regulating *CBF* genes or ascorbic acid synthesis<sup>56,61</sup>, drought stress<sup>62,63</sup>, and salt stress by regulating ABA signaling<sup>64,65</sup>. Among the 117 R2R3-MYB *D. officinale* genes, only nine that were modulated by low temperature were found in this study (Fig. 3). *DoMYB28*, -29, -54, -75, -78, -81, and -111 were up-regulated under salinity stress (Fig. 6). The *DoMYB74* and *DoMYB75* homologues, which were recognized as the C2 (S6) clade of R2R3-MYB genes, displayed different expression patterns in response to the two osmotic stresses, PEG and mannitol (Fig. 6). This may due to different *cis*-responsive elements among their transcriptional regulatory regions. For example, the putative promoter of *DoMYB75* contained two ABA response elements, three dehydration response elements, one drought response element, three ethylene response elements, and six MeJA response elements, while *DoMYB74* had only one ABA response element, one dehydration response element, two ethylene response elements, and four MeJA response elements, but no drought response element in its putative promoter (Supplementary Table 3).

MYBR genes are also involved in stress responses. For example, the MYB-related gene *AQUILO* improves cold tolerance in transgenic *A. thaliana* and caused the accumulation of oligosaccharides<sup>66</sup>, *OsMYB48-1* is involved in salt stress by regulating the expression of stress-related genes<sup>67</sup>, and *OsMYBR1* promoted drought stress in transgenic rice by increasing the free proline and soluble sugar content and up-regulated the expression

of stress-related genes under drought treatment<sup>68</sup>. In this work, nine *DoMYBR* genes were up-regulated by cold stress, similar to the above studies, suggesting that the MYB genes in the MYBR or R2R3-MYB groups play roles in plant abiotic stress responses.

**The involvement of R2R3-MYB genes in polysaccharide biosynthesis.** Secondary cell walls (SCWs), which are mainly found in plant stems, are primarily composed of cellulose, lignin and hemicelluloses (xylan and glucomannan)<sup>20</sup>. Several TFs are involved in the regulation of SCW biosynthesis. MYB TFs make up the vast majority of TFs in transcriptional regulation of SCW biosynthesis<sup>21</sup>. The R2R3-MYB genes involved in plant SCW biosynthesis are thought to regulate the biosynthesis of SCW polysaccharides. For example, *AtMYB46* acts as a regulator in SCW formation and directly regulates the expression of *CSLA9*, which encodes mannan synthase in *A. thaliana*<sup>22,69</sup>. *AtMYB75* in the C2 (S6) group acts as a regulator in cell wall thickening, testa, as well as biosynthesis of anthocyanins<sup>70–72</sup>. Another MYB gene, *AtMYB113*, in the C2 (S6) subgroup increases pigment production and results in strongly visible anthocyanin pigmentation<sup>73</sup>. *35S::DoMYB75* transgenic *Arabidopsis* lines showed no anthocyanin, possibly due to the low levels of sucrose present in the medium, but increased WSPs in seeds of transgenic lines (Fig. 7).

In conclusion, 159 and 165 MYB genes were identified from *P. equestris* and *D. officinale* genomes, respectively. They could be classified into four groups in both orchids: MYBR, R2R3-MYB, 3R-MYB and atypical MYB proteins. Only three R2R3-MYB genes and 12 MYBR genes from *D. officinale* were up-regulated under low temperature, suggesting that MYB genes may play a role in the cold stress response in this orchid. Ten R2R3-MYB genes with an expression pattern corresponding to WSP accumulation were identified and regarded as the candidate genes involved in WSP biosynthesis. Over-expression of one candidate gene (*DoMYB75*) in *A. thaliana* caused the accumulation of WSPs in *A. thaliana* seeds.

## Data Availability

The datasets used during the current study are available from the corresponding author on reasonable request.

## References

- Riechmann, J. L. *et al.* *Arabidopsis* transcription factors: Genome-wide comparative analysis among eukaryotes. *Science* **290**, 2105–2110 (2000).
- Du, H. *et al.* Genome-wide identification and evolutionary and expression analyses of MYB-related genes in land plants. *DNA Research* **20**, 437–448 (2013).
- Wang, Z. *et al.* Genome-wide analysis of the R2R3-MYB transcription factor genes in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*) reveals their stress and hormone responsive patterns. *BMC Genomics* **16**, 17 (2015).
- Klempnauer, K. H. *et al.* Nucleotide sequence of the retroviral leukemia gene *v-myb* and its cellular progenitor *c-myb*: the architecture of a transduced oncogene. *Cell* **31**, 453–463 (1983).
- Paz-Ares, J. *et al.* The regulatory *c1* locus of *Zea mays* encodes a protein with homology to *myb* proto-oncogene products and with structural similarities to transcriptional activators. *The EMBO Journal* **6**, 3553–3558 (1987).
- Chen, Y. *et al.* The MYB transcription factor superfamily of Arabidopsis: Expression analysis and phylogenetic comparison with the rice MYB family. *Plant Molecular Biology* **60**, 107–124 (2006).
- Li, Z. *et al.* Genome-wide identification and analysis of the MYB transcription factor superfamily in *Solanum lycopersicum*. *Plant & Cell Physiology* **57**, 1657–1677 (2016).
- Li, X. *et al.* Genome-wide identification, evolution and functional divergence of MYB transcription factors in Chinese white pear (*Pyrus bretschneideri*). *Plant & Cell Physiology* **57**, 824–847 (2016).
- Chen, S., Niu, X., Guan, Y. & Li, H. Genome-wide analysis and expression profiles of the MYB genes in *Brachypodium distachyon*. *Plant & Cell Physiology* **58**, 1777–1788 (2017).
- Dubos, C. *et al.* MYB transcription factors in Arabidopsis. *Trends in Plant Science* **15**, 573–581 (2010).
- Katiyar, A. *et al.* Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis. *BMC Genomics* **13**, 544 (2012).
- Feller, A., Machemer, K., Braun, E. L. & Grotewold, E. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *The Plant Journal* **66**, 94–116 (2011).
- Du, H. *et al.* Genome-wide analysis of the MYB transcription factor superfamily in soybean. *BMC Plant Biology* **12**, 106 (2012).
- Haga, N. *et al.* R1R2R3-Myb proteins positively regulate cytokinesis through activation of *KNOLLE* transcription in *Arabidopsis thaliana*. *Development* **134**, 1101–1110 (2007).
- Ambawat, S., Sharma, P., Yadav, N. R. & Yadav, R. C. MYB transcription factor genes as regulators for plant responses: an overview. *Physiology and Molecular Biology of Plants* **19**, 307–321 (2013).
- Chen, Y. S. *et al.* Two MYB-related transcription factors play opposite roles in sugar signaling in *Arabidopsis*. *Plant Molecular Biology* **93**, 299–311 (2017).
- Allan, A. C. *et al.* MYB transcription factors that colour our fruit. *Trends in Plant Science* **13**, 99–102 (2008).
- Liu, J. *et al.* MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Molecular Plant* **8**, 689–708 (2015).
- Kubo, H. *et al.* Biosynthesis of riccionidins and marchantins is regulated by R2R3-MYB transcription factors in *Marchantia polymorpha*. *Journal of Plant Research* **131**, 849–864 (2018).
- Kumar, M. *et al.* Secondary cell walls: biosynthesis and manipulation. *Journal of Experimental Botany* **67**, 515–531 (2016).
- Zhong, R. & Ye, Z. H. Secondary cell walls: biosynthesis, patterned deposition and transcriptional regulation. *Plant & Cell Physiology* **56**, 195–214 (2015).
- Kim, W.-C. *et al.* Transcription factors that directly regulate the expression of *CSLA9* encoding mannan synthase in *Arabidopsis thaliana*. *Plant Molecular Biology* **84**, 577–587 (2014).
- Haughn, G. W. & Western, T. L. *Arabidopsis* seed coat mucilage is a specialized cell wall that can be used as a model for genetic analysis of plant cell wall structure and function. *Frontiers in Plant Science* **3**, 64 (2012).
- Yu, L. *et al.* CELLULOSE SYNTHASE-LIKE A2, a glucomannan synthase, is involved in maintaining adherent mucilage structure in *Arabidopsis* seed. *Plant Physiology* **164**, 1842–1856 (2014).
- Li, S. F. *et al.* The *Arabidopsis* MYB5 transcription factor regulates mucilage synthesis, seed coat development, and trichome morphogenesis. *The Plant Cell* **21**, 72–89 (2009).
- Penfield, S., Meissner, R. C., Shoue, D. A., Carpita, N. C. & Bevan, M. W. *MYB61* is required for mucilage deposition and extrusion in the *Arabidopsis* seed coat. *The Plant Cell* **13**, 2777–2791 (2001).
- Kerepesi, I. & Galiba, G. Osmotic and salt stress-induced alteration in soluble carbohydrate content in wheat seedlings. *Crop Science* **40**, 482–487 (2000).



28. Schepetkin, I. A. & Quinn, M. T. Botanical polysaccharides: Macrophage immunomodulation and therapeutic potential. *International Immunopharmacology* **6**, 317–333 (2006).
29. Chen, Y. *et al.* Structural characterization and antitumor activity of a polysaccharide from *Ramulus mori*. *Carbohydrate Polymers* **190**, 232–239 (2018).
30. Fan, L. *et al.* Antitumor and immunomodulatory activity of water-soluble polysaccharide from *Inonotus obliquus*. *Carbohydrate Polymers* **90**, 870–874 (2012).
31. Givnish, T. J. *et al.* Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proceedings of the Royal Society B: Biological Sciences* **282**, 1814 (2015).
32. Teixeira da Silva, J. A. & Ng, T. B. The medicinal and pharmaceutical importance of *Dendrobium* species. *Applied Microbiology and Biotechnology* **101**, 2227–2239 (2017).
33. Ng, T. B. *et al.* Review of research on *Dendrobium*, a prized folk medicine. *Applied Microbiology and Biotechnology* **93**, 1795–1803 (2012).
34. He, C. *et al.* Identification of genes involved in biosynthesis of mannan polysaccharides in *Dendrobium officinale* by RNA-seq analysis. *Plant Molecular Biology* **88**, 219–231 (2015).
35. He, C. *et al.* DoGMP1 from *Dendrobium officinale* contributes to mannose content of water-soluble polysaccharides and plays a role in salt stress response. *Scientific Reports* **7**, 41010 (2017).
36. He, C. *et al.* Molecular cloning and functional analysis of the *phosphomannomutase* (PMM) gene from *Dendrobium officinale* and evidence for the involvement of an abiotic stress response during germination. *Protoplasma* **254**, 1693–1704 (2017).
37. Murashige, T. & Skoog, F. A revised medium for rapid growth and bio-assays with tobacco tissue cultures. *Plant Physiology* **15**, 473–497 (1962).
38. Zhang, G.-Q. *et al.* The *Dendrobium catenatum* Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. *Scientific Reports* **6**, 19029 (2016).
39. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* **44**, D279–D285 (2016).
40. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research* **28**, 45–48 (2000).
41. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research* **33**, D501–504 (2005).
42. Kazutaka, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology & Evolution* **30**, 772–780 (2013).
43. Thompson, J. D., Gibson, T. J. & Higgins, D. G. Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics* **00**, 2.3.1–2.3.22 (2003).
44. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**, 1870–1874 (2016).
45. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**, 157 (2015).
46. Wang, D., Zhang, Y., Zhang, Z., Zhu, J. & Yu, J. KaKs\_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics, Proteomics & Bioinformatics* **8**, 77–80 (2010).
47. Rombauts, S., Déhais, P., Van Montagu, M. & Rouzé, P. PlantCARE, a plant *cis*-acting regulatory element database. *Nucleic Acids Research* **27**, 295–296 (1999).
48. Higo, K. *et al.* Plant *cis*-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Research* **27**, 297–300 (1999).
49. Wu, Z.-G. *et al.* Insights from the cold transcriptome and metabolome of *dendrobium officinale*: global reprogramming of metabolic and gene regulation networks during cold acclimation. *Frontiers in Plant Science* **7**, 1653 (2016).
50. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**, R36, <https://doi.org/10.1186/gb-2013-14-4-r36> (2013).
51. Anders, S. *et al.* HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
52. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* **25**, 402–408 (2001).
53. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal* **16**, 735–743 (1998).
54. Dubois, M. *et al.* Colorimetric method for determination of sugars and related substances. *Analytical Chemistry* **28**, 350–356 (1956).
55. Wani, S. H. *et al.* Phytohormones and their metabolic engineering for abiotic stress tolerance in crop plants. *The Crop. Journal* **4**, 162–176 (2016).
56. Agarwal, M. *et al.* A R2R3 type MYB transcription factor is involved in the cold regulation of *CBF* genes and in acquired freezing tolerance. *The Journal of Biological Chemistry* **281**, 37636–37645 (2006).
57. Yang, A. *et al.* A R2R3-type MYB gene, *OsMYB2*, is involved in salt, cold, and dehydration tolerance in rice. *Journal of Experimental Botany* **63**, 2541–2556 (2012).
58. Zhang, L. *et al.* A wheat R2R3-MYB gene, TaMYB30-B, improves drought stress tolerance in transgenic *Arabidopsis*. *Journal of Experimental Botany* **63**, 5873–5885 (2012).
59. Handford, M. G. *et al.* Localisation and characterisation of cell wall mannan polysaccharides in *Arabidopsis thaliana*. *Planta* **218**, 27–36 (2003).
60. Zhao, Y. *et al.* Characterization of wheat MYB genes responsive to high temperatures. *BMC Plant Biology* **17**, 208 (2017).
61. Xing, C. *et al.* A novel MYB transcription factor regulates ascorbic acid synthesis and affects cold tolerance. *Plant, Cell & Environment* **42**, 832–845 (2019).
62. Su, L. T. *et al.* A novel MYB transcription factor, GmMYBJ1, from soybean confers drought and cold tolerance in *Arabidopsis thaliana*. *Gene* **538**, 46–55 (2014).
63. Wei, Q. *et al.* A wheat MYB transcriptional repressor TaMyb1D regulates phenylpropanoid metabolism and enhances tolerance to drought and oxidative stresses in transgenic tobacco plants. *Plant Science* **265**, 112–123 (2017).
64. Fang, Q. *et al.* A salt-stress-regulator from the poplar R2R3 MYB family integrates the regulation of lateral root emergence and ABA signaling to mediate salt stress tolerance in *Arabidopsis*. *Plant Physiology and Biochemistry* **114**, 100–110 (2017).
65. Yu, Y., Ni, Z., Chen, Q. & Qu, Y. The wheat salinity-induced R2R3-MYB transcription factor TaSIM confers salt stress tolerance in *Arabidopsis thaliana*. *Biochemical and Biophysical Research Communications* **491**, 642–648 (2017).
66. Sun, X. *et al.* The GARP/MYB-related grape transcription factor AQUILO improves cold tolerance and promotes the accumulation of raffinose family oligosaccharides. *Journal of Experimental Botany* **69**, 1749–1764 (2018).
67. Xiong, H. *et al.* Overexpression of *OsMYB48-1*, a novel MYB-related transcription factor, enhances drought and salinity tolerance in rice. *PLOS ONE* **9**, e92913 (2014).
68. Yin, X., Cui, Y., Wang, M. & Xia, X. Overexpression of a novel MYB-related transcription factor, *OsMYBRI*, confers improved drought tolerance and decreased ABA sensitivity in rice. *Biochemical and Biophysical Research Communications* **490**, 1355–1361 (2017).

69. Ko, J. H. *et al.* The MYB46/MYB83-mediated transcriptional regulatory programme is a gatekeeper of secondary wall biosynthesis. *Annals of Botany* **114**, 1099–1107 (2014).
70. Bhargava, A. *et al.* The interacting MYB75 and KNAT7 transcription factors modulate secondary cell wall deposition both in stems and seed coat in *Arabidopsis*. *Planta* **237**, 1199–1211 (2013).
71. Bhargava, A., Mansfield, S. D., Hall, H. C., Douglas, C. J. & Ellis, B. E. MYB75 functions in regulation of secondary cell wall formation in the *Arabidopsis* inflorescence stem. *Plant Physiology* **154**, 1428–1438 (2010).
72. Gatica-Arias, A. *et al.* Flavonoid production in transgenic hop (*Humulus lupulus* L.) altered by PAP1/MYB75 from *Arabidopsis thaliana* L. *Plant Cell Reports* **31**, 111–119 (2012).
73. Gonzalez, A. *et al.* Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *The Plant Journal* **53**, 814–827 (2008).
74. Stracke, R., Werber, M. & Weisshaar, B. The R2R3-MYB gene family in *Arabidopsis thaliana*. *Current Opinion in Plant Biology* **4**, 447–456 (2001).
75. Khosla, A., Boehler, A. P., Bradley, A. M., Neumann, T. R. & Schrick, K. HD-Zip proteins GL2 and HDG11 have redundant functions in *Arabidopsis* trichomes, and GL2 activates a positive feedback loop via MYB23. *The Plant Cell* **26**, 2184–2200 (2014).
76. Mandaokar, A. MYB108 acts together with MYB24 to regulate jasmonate-mediated stamen maturation in *Arabidopsis*. *Plant Physiology* **149**, 851–862 (2009).
77. Lotkowska, M. E. *et al.* The *Arabidopsis* transcription factor MYB112 promotes anthocyanin formation during salinity and under high light stress. *Plant Physiology* **169**, 1862–1880 (2015).
78. Baek, D. *et al.* Regulation of *miR399f* transcription by AtMYB2 affects phosphate starvation responses in *Arabidopsis*. *Plant Physiology* **161**, 362–373 (2013).
79. Devaiah, B. N. *et al.* Phosphate starvation responses and gibberellic acid biosynthesis are regulated by the MYB62 transcription factor in *Arabidopsis*. *Molecular Plant* **2**, 43–58 (2009).
80. Mu, R.-L. *et al.* An R2R3-type transcription factor gene AtMYB59 regulates root growth and cell cycle progression in *Arabidopsis*. *Cell Research* **19**, 1291–1304 (2009).
81. Zhong, R. *et al.* A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. *The Plant Cell* **20**, 2763–2782 (2008).
82. Zhao, Q. *et al.* AtMYB44 positively regulates the enhanced elongation of primary roots induced by *n*-3-oxo-hexanoyl-homoserine lactone in *Arabidopsis thaliana*. *Molecular Plant-Microbe Interactions* **29**, 774–785 (2016).
83. Tohge, T. *et al.* Salt-related MYB1 (SRM1) coordinates abscisic acid biosynthesis and signaling during salt stress in *Arabidopsis*. *Plant Physiology* **169**, 1027–1041 (2015).
84. Fernández-Marcos, M. *et al.* Control of *Arabidopsis* lateral root primordium boundaries by MYB36. *New Phytologist* **213**, 105–112 (2017).
85. Zhang, Y. *et al.* The conserved transcription factors, MYB115 and MYB118, control expression of the newly evolved benzoyloxy glucosinolate pathway in *Arabidopsis thaliana*. *Frontiers in Plant Science* **6**, 343 (2015).
86. Rabiger, D. S. & Drews, G. N. MYB64 and MYB119 are required for cellularization and differentiation during female gametogenesis in *Arabidopsis thaliana*. *PLOS Genetics* **9**, e1003783 (2013).
87. Liang, Y. *et al.* MYB97, MYB101 and MYB120 function as male factors that control pollen tube-synergid interaction in *Arabidopsis thaliana* fertilization. *PLOS Genetics* **9**, e1003933 (2013).
88. Fujiwara, S. *et al.* Chimeric repressor analysis identifies MYB87 as a possible regulator of morphogenesis via cell wall organization and remodeling in *Arabidopsis*. *Biotechnology Letters* **36**, 1049–1057 (2014).
89. Lee, S. B. *et al.* MYB94 and MYB96 additively activate cuticular wax biosynthesis in *Arabidopsis*. *Plant and Cell Physiology* **57**, 2300–2311 (2016).
90. Lashbrooke, J. *et al.* MYB107 and MYB9 homologs regulate suberin deposition in angiosperms. *The Plant Cell* **28**, 2097–2116 (2016).
91. Gibbs, D. J. *et al.* AtMYB93 is a novel negative regulator of lateral root development in *Arabidopsis*. *New Phytologist* **203**, 1194–1207 (2014).
92. Lippold, F. *et al.* AtMyb41 regulates transcriptional and metabolic responses to osmotic stress in *Arabidopsis*. *Plant Physiology* **149**, 1761–1772 (2009).
93. Xu, R. *et al.* Salt-induced transcription factor MYB74 is regulated by the RNA-directed DNA methylation pathway in *Arabidopsis*. *Journal of Experimental Botany* **66**, 5997–6008 (2015).
94. Oshima, Y. & Mitsuda, N. The MIXTA-like Transcription factor MYB16 is a major regulator of cuticle formation in vegetative organs. *Plant Signaling & Behavior* **8**, e26826 (2013).
95. Zhou, J. *et al.* MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in *Arabidopsis*. *The Plant Cell* **21**, 248–266 (2009).
96. Pandey, A. *et al.* Constitutive expression of *Arabidopsis* MYB transcription factor, AtMYB11, in tobacco modulates flavonoid biosynthesis in favor of flavonol accumulation. *Plant Cell Reports* **34**, 1515–1528 (2015).
97. Zhou, M. *et al.* LNK1 and LNK2 corepressors interact with the MYB3 transcription factor in phenylpropanoid biosynthesis. *Plant Physiology* **174**, 1348–1358 (2017).
98. Newman, L. J. *et al.* Involvement of the R2R3-MYB, AtMYB61, in the ectopic lignification and dark-photomorphogenic components of the *det3* mutant phenotype. *The Plant Journal* **37**, 239–250 (2004).
99. Yang, C. *et al.* *Arabidopsis* MYB26/MALE STERILE35 regulates secondary thickening in the endothecium and is essential for anther dehiscence. *The Plant Cell* **19**, 534–548 (2007).
100. Öhman, D. *et al.* MYB103 is required for FERULATE-5-HYDROXYLASE expression and syringyl lignin biosynthesis in *Arabidopsis* stems. *The Plant Journal* **73**, 63–76 (2013).
101. Ko, J. H. *et al.* Ectopic expression of MYB46 identifies transcriptional regulatory genes involved in secondary wall biosynthesis in *Arabidopsis*. *The Plant Journal* **60**, 649–665 (2009).

## Acknowledgements

This research was funded by the National Natural Science Foundation of China (grant number 31800204 and 31871547), and the Natural Science Foundation of Guangdong Province Projects (grant numbers 2016A030310012 and 2018A030313603).

## Author Contributions

J.D. supervised the project. C.H. conceived the research and designed the experiments. C.S. performed qRT-PCR. H.W. and M.Z. constructed the vector and generated the transgenic lines. X.Z. provided the plant materials. M.L. performed the bioinformatics analyses. C.H., J.T. and JATdS collectively interpreted the results and wrote all drafts of the manuscript. All authors approved the final draft for submission and take full public responsibility for the content of the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-49812-8>.

**Competing Interests:** The authors declare no competing interests.

**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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019