Journal of Experimental Botany, Vol. 65, No. 15, pp. 4255–4269, 2014
doi:10.1093/jxb/eru196 Advance Access publication 12 May, 2014

Journal of Experimental Botany
www.jxb.oxfordjournals.org

RESEARCH PAPER

R2R3-MYB gene pairs in *Populus*: evolution and contribution to secondary wall formation and flowering time

Guohua Chai[1,*], Zengguang Wang[1,*], Xianfeng Tang[1], Li Yu[1], Guang Qi[1], Dian Wang[1], Xiaofei Yan[1], Yingzhen Kong[2] and Gongke Zhou[1,†]

[1] Key Laboratory of Biofuels, Chinese Academy of Sciences, Shandong Provincial Key Laboratory of Energy Genetics, Qingdao Institute of BioEnergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao 266101, China
[2] Key Laboratory of Tobacco Gene Resource, Tobacco Research Institute, Chinese Academy of Agricultural Sciences, Qingdao 266101, China

* These authors contributed equally to this manuscript.
† To whom correspondence should be addressed. E-mail: zhougk@qibebt.ac.cn

Received 5 February 2014; Revised 31 March 2014; Accepted 7 April 2014

Abstract

In plants, the *R2R3-MYB* gene family contains many pairs of paralogous genes, which play the diverse roles in developmental processes and environmental responses. The paper reports the characterization of 81 pairs of *Populus R2R3-MYB* genes. Chromosome placement, phylogenetic, and motif structure analyses showed that these gene pairs resulted from multiple types of gene duplications and had five different gene fates. Tissue expression patterns revealed that most duplicated genes were specifically expressed in the tissues examined. qRT-PCR confirmed that nine pairs were highly expressed in xylem, of which three pairs (*PdMYB10*/*128*, *PdMYB90*/*167*, and *PdMYB92*/*125*) were further functionally characterized. The six PdMYBs were localized to the nucleus and had transcriptional activities in yeast. The heterologous expression of *PdMYB10* and *128* in *Arabidopsis* increased stem fibre cell-wall thickness and delayed flowering. In contrast, overexpression of *PdMYB90*, *167*, *92*, and *125* in *Arabidopsis* decreased stem fibre and vessel cell-wall thickness and promoted flowering. Cellulose, xylose, and lignin contents were changed in overexpression plants. The expression levels of several genes involved in secondary wall formation and flowering were affected by the overexpression of the six *PdMYB*s in *Arabidopsis*. This study addresses the diversity of gene duplications in *Populus R2R3-MYB*s and the roles of these six genes in secondary wall formation and flowering control.

Key words: *Arabidopsis*, evolution, flowering time, gene pair, *Populus*, *R2R3-MYB*, secondary wall thickening.

Introduction

The MYB transcription factors are at least one billion years old and are widely distributed in eukaryotes (Lipsick, 1996). They are characterized by the presence of a highly conserved MYB domain that consists of 1–4 imperfect MYB repeats located near the N-terminus. In plants, most MYB proteins contain the two-repeat R2R3-MYB domain and belong to the R2R3-MYB family (Martin and Paz-Ares, 1997). To date, *R2R3-MYB* genes have been identified in multiple plant species, including *Arabidopsis* (126 members, Stracke *et al.*, 2001), rice (84 members, Jiang *et al.*, 2004), poplar (192 members, Wilkins *et al.*, 2009), maize (157 members, Du *et al.*, 2012), and wheat (22 members, Zhang *et al.*, 2012). The extensive expansion of the *Arabidopsis R2R3-MYB* gene family suggests that its members play diverse roles in many physiological and biochemical processes, including control of cell morphogenesis and pattern formation (Higginson *et al.*, 2003; Baumann *et al.*, 2007; Tominaga *et al.*, 2007), regulation of primary and secondary metabolism (Teng *et al.*,

2005; Stracke *et al.*, 2007; Zhou *et al.*, 2009), participation in defence and response to various biotic and abiotic stresses (Zhao *et al.*, 2007; Jung *et al.*, 2008; Ramirez *et al.*, 2011), and hormone synthesis and signal transduction (Gocal *et al.*, 2001; Abe *et al.*, 2003; Cheng, 2007).

Because of the economic importance in pulp and biofuel production, study of the woody perennial plant *Populus* has been intensive for many years. Compared to the largely well-characterized MYBs in *Arabidopsis*, only a small number of MYB proteins have been functionally characterized in *Populus*. These MYBs have been implicated to participate mainly in wood formation (Legay *et al.*, 2010; McCarthy *et al.*, 2010; Tian *et al.*, 2013) and response to abiotic stresses (Mellway *et al.*, 2009; Ma *et al.*, 2013). It is known that *Populus* has undergone at least three rounds of genome-wide duplication followed by multiple segmental duplication, tandem duplication, and transposition events, which in turn has produced nearly 8000 gene pairs in its genome (Tuskan *et al.*, 2006). Since duplicated genes may develop non-, sub-, or neofunctions after gene duplication and may contribute to some of the evolutionary innovations in plants (Riechmann *et al.*, 2000), it is of interest to characterize *R2R3-MYB* gene pairs in *Populus*.

This study identified 81 pairs of paralogous *R2R3-MYB* genes in *Populus* and comprehensively characterized their gene duplication types and evolutionary outcomes based on their chromosome placement and motif structures. This work also shows the roles of *PdMYB10*/*128*, *PdMYB90*/*167*, and *PdMYB92*/*125* in secondary wall formation and flowering in transgenic *Arabidopsis*. The results provide valuable information for further exploration of the functions of these gene pairs in poplar.

Materials and methods

Identification of Populus R2R3-MYB *gene pairs*

The *Populus trichocarpa* genome database (version 3.0, http://www.phytozome.net/poplar.php) was searched to identify R2R3-MYB domain-containing proteins, using the Basic Local Alignment Search Tool algorithms BLASTP and TBLASTN with known R2R3-MYB proteins as query sequences and with E-value cut off set as 1e-005. The obtained genes were verified by the Hidden Markov Model of Pfam (http://pfam.sanger.ac.uk/search) and further compared with 192 *R2R3-MYB* genes from Wilkins *et al.* (2009).

Gene pairs were determined by aligning and phylogenetically analysing 194 full-length R2R3-MYB proteins. Multiple sequence alignment was conducted using ClustalX version 1.83 (Thompson *et al.*, 1997) and a phylogenetic tree was generated using MEGA version 4.0 (Tamura *et al.*, 2007) with the neighbour-joining method (Saitou and Nei, 1987). Bootstrap analysis with 1000 replicates was used to evaluate the significance of the nodes. Pairwise gap deletion mode was used to ensure that the divergent C-terminal residues could contribute to the topology of the neighbour-joining method tree.

Sequence properties and chromosomal location

The exon/intron organization was analysed using Gene Structure Display Server (Guo *et al.*, 2007). Protein motifs were identified using MEME version 4.9.1 (Bailey *et al.*, 2006) with the following parameter settings: distribution of motifs, 0 or 1 per sequence; maximum number of finds, 100; minimum width of motif, 6; maximum width of motif, 300. Only motifs with an e-value ≤1e-10 were kept for further analysis. The MAST program (Bailey and Gribskov, 1998) was used to search detected motifs in protein databases. The synonymous substitution rate (k_s) and the nonsynonymous substitution rate (k_a) were calculated following the method of Goldman and Yang (1994).

Homologous chromosome segments resulting from whole-genome duplication events were identified as described previously (Tuskan *et al.*, 2006). Tandem gene duplications were identified according to criteria in TIGR Rice Genome Annotation (http://compbio.dfci.harvard.edu/tgi/). Genes separated by five or fewer gene loci in the range of 100 kb were considered to be tandem duplicates.

Gene expression

The expression analysis of *Populus MYB* genes was performed based on genome-wide microarray data from the NCBI Gene Expression Omnibus obtained with accession number GSE13990. The expression data were gene-wise normalized based on Pearson coefficients with average linkage using Genesis version 1.75 (Sturn *et al.*, 2002).

Quantitative real-time PCR (qRT-PCR) was used to examine the expression levels of *PdMYB* genes in different poplar tissues and transgenic *Arabidopsis*. For tissue-specific expression analysis, roots, leaves, shoots, cortex, xylem, and phloem (including the cambial zones) were sampled from 1.5-m-high poplar (*Populus deltoides* cv. nanlin895) grown in a glasshouse. Total RNA was isolated from poplar and *Arabidopsis* as described previously (Chai *et al.*, 2012), treated with DNase I (Sigma), and used for synthesis of first-strand cDNA with a Supermo III reverse transcriptase kit (Bioteke, China). qRT-PCR was conducted on a LightCycler 480 Detection System (Roche) using SYBR Green Realtime PCR Master Mix (TOYOBA), according to the manufacturers' instructions. Gene-specific primers (Supplementary Table S3 available at *JXB* online) were designed using Beacon Designer version 7.0 (Premier Biosoft International) and confirmed by melting-curve analysis. Baseline and threshold cycles (Ct) were determined by the 2nd maximum derivative method using LightCycler 480 software (release 1.5.0). Quantification of gene expression relative to reference genes (*PtUBQ* BU879229 and *ACTIN2* At3g18780) was determined using the $2^{-\Delta CT}$ method (Pfaffl, 2001). The data shown are means of three biological replicates.

Semi-quantitative RT-PCR was used to detect the expression of *PdMYB*s in wild-type and overexpressing *Arabidopsis* plants. *ACTIN2* was used as an internal control. Various PCR cycles were performed to determine the linear range of amplification for each gene. At least three biological replications were used in each experiment.

Subcellular localization and transcriptional activation analysis

Subcellular localization of PdMYB proteins was examined using the tobacco leaf transient expression system (Sparkes *et al.* 2006). Full-length *PdMYB* cDNA was separately amplified from stems of *P. deltoides* cv. nanlin895 using the corresponding primers (Supplementary Table S3) and ligated between the CaMV 35S promoter and the 35S terminator in pK7FWG2 (Invitrogen). The resulting constructs encoded fusion proteins with PdMYBs located at the N-terminus and GFP at the C-terminus. After a 3-d postinfiltration period, transfected leaves were examined for green fluorescence signal using an FluoView FV1000 confocal microscope equipped with 488-nm argon laser (Olympus). Images were processed using Photoshop version 7.0 (Adobe).

For transcriptional activation analysis, full-length *PdMYB* cDNA was individually fused in frame with the GAL4 DNA-binding domain in pGBKT7 (Clontech). The recombinant vectors and the pGBKT7 empty vector (control) were transferred into *Saccharomyces cerevisiae* AH109 using the lithium acetate method. The transformed strains were cultured on minimal medium (Clotech) without His or Trp and the transactivation activity of each protein was evaluated according to growth status and α-galactosidase activity.

Vector construction and plant transformation

PdMYB overexpression constructs were created by ligating full-length cDNA downstream of the 35S promoter in pK2GW7 (Invitrogen). Wild-type *Arabidopsis* plants (Col-0) were used to produce the transgenic plants. T_1 transgenic seedlings were selected on MS medium containing 50 mg l^{-1} kanamycin and further confirmed by RT-PCR. Phenotypic characterization was performed on homozygous T_3 overexpressing plants with a single T-DNA insertion locus, as estimated by Mendelian segregation of kanamycin resistance.

Microscopy and histochemistry

Six-week-old *Arabidopsis* inflorescence stems were cut 0.5 cm above the base and fixed with 2.5% glutaraldehyde in phosphate buffer (pH 7.2) at room temperature for 6 h. After rinsing three times with intervals of 30 min, the material was fixed by 1% osmic acid for 1 h and subsequently washed twice in 1× phosphate-buffered saline for 15 min, dehydrated by series concentrations of acetone, and embedded in resin (SPI-Chem). For light microscopy (Olympus DX51), 1-µm-thick sections were cut with a EM UC6 microtome (Leica) and stained with toluidine blue-O (1%, w/v). For transmission electron microscopy (Hitachi H-7650), 70-nm-thick sections were cut. Wall thickness was measured in metaxylem vessels and in the interfascicular fibres next to the endodermis using SmileView (JEOL, Tokyo, Japan). For each construct, at least three transgenic plants with the most severe phenotypes were examined.

Cell-wall composition assay

Inflorescence stems were collected from 6-week-old *Arabidopsis* plants and ground into a fine powder in liquid nitrogen. Cell-wall alcohol-insoluble residues (AIRs) were prepared by treating the powder sequentially with 80% ethanol, 100% ethanol, and acetone. The resulting AIRs were dried under vacuum at 60 °C overnight and used for analysis of monosaccharide composition following the procedure described by Selvendran *et al.* (1979). Briefly, 3 mg AIRs was hydrolysed by trifluoroacetic acid at 120 °C for 2 h and the derivatives were analysed on a Thermo ODS-2 C18 column (4.6 × 250 mm) connected to a Waters HPLC system (2489 UV visible detector) at a wavelength at 245 nm. To determine the cellulose content (Updegraff, 1969), acid-resistant material was treated with Updegraff reagent (acetic acid/nitric acid/water, 8:1:2, v/v) at 100 °C for 30 min and the resulting pellet was then completely hydrolysed with H_2SO_4 (67%, v/v). The released glucose was measured using a glucose assay kit (Cayman Chemical) with a dehydration factor of 0.9. To determine the lignin content (Fukushima and Hatfield, 2001), 3 mg AIRs was solubilized with acetyl bromide solution, and then 2 M sodium hydroxide and 0.5 M hydroxylamine hydrochloride were added to stop the reaction. Absorbance at 280 nm was measured using an UC-visible spectrophotometer (VARIAN Cary 50).

Results

Identification and evolutionary analysis of Populus R2R3-MYB *gene pairs*

Using known R2R3-MYB protein sequences to query the recently updated *Populus* genome database, this work identified 194 full-length *R2R3-MYB* genes in *Populus* and manually verified their uniqueness (Supplementary Table S1), which covered 192 R2R3-MYB members reported by Wilkins *et al.* (2009). A neighbour-joining phylogenetic tree was constructed using the 194 full-length MYB protein sequences (Supplementary Fig. S1), which allowed this work to identify 81 pairs of paralogous genes at the terminal nodes (Table 1), with the percentage of paralogous genes (83.5%) similar to those of *Populus* Dof (78.1%) (Yang *et al.*, 2006) and CCCH (74.7%) (Chai *et al.*, 2012) gene families.

Based on sequence similarity and topology, the 81 pairs of *R2R3-MYB* genes were divided into 21 subgroups according to clades with at least 50% bootstrap support (Fig. 1). Within each subgroup, the motifs and exon/intron organization were relatively conserved (Table 1), suggesting strong evolutionary relationships among subgroup members. According to Yang's hypothesis (Yang *et al.*, 2006), changes of duplicated genes result in six evolutionary outcomes: RR, RD, RN, DD, DN, and NN, where R represents retention (a copy retaining the original motif organization and function), D represents degeneration (a copy degenerating or losing one or more motifs and functions), and N represents neo-functionalization (a copy acquiring one or more motifs and functions). These results indicate that five types of evolutionary outcomes existed in the 81 MYB paralogues: 61 RR, nine RD, eight RN, two DD, and one NN (Table 1). Similar motif organization in protein sequences was retained in 75% (61 RR) of the 81 pairs of paralogues, suggesting that the duplicates had similar functions. To examine the driving force for gene evolution, this study also performed nonsynonymous and synonymous substitution ratio (k_a and k_s) analysis of the duplicated genes. Seventy-five of the 81 gene pairs had a k_a/k_s ratio <0.5 (Table 1), implying that most *Populus R2R3-MYB* gene pairs had evolved mainly under the influence of purifying selection (Hurst, 2002).

The physical locations of 80 pairs of paralogous genes were assigned to 19 linkage groups (Fig. 2), while *PtrMYB057* remained on an unassembled scaffold fragment. Of the 80 mapped gene pairs, 30 remained in a conserved position on 13 duplicated blocks within the 163 recently identified duplicated blocks (Tuskan *et al.*, 2006), suggesting that they may be the result of whole-genome duplication. Four adjacent gene pairs were located within 32 kb on the duplication blocks and had high sequence similarities (>75%) with their counterpart, suggesting an origin from tandem duplication. Among the non-genome-duplicated gene pairs, 17 genes were located on duplicated segments with their counterparts not on any duplicated blocks, two counterparts of 25 paralogous pairs were located separately on divergent rather than homologous duplicated blocks, and four gene pairs were not on any duplicated blocks. It appears that diverse duplication events have contributed to the complexity of *R2R3-MYB* gene pairs in the *Populus* genome.

Expression patterns of Populus R2R3-MYB *gene pairs*

To examine the expression patterns of the 81 pairs of *MYB* genes in various tissues, a comprehensive analysis was conducted based on the *Populus* microarray data generated by Wilkins *et al.* (2009). Thirteen *MYB* genes do not have corresponding probe sets in the microarray dataset, therefore only 149 genes containing 69 gene pairs were analysed (Supplementary Table S1). Most genes displayed tissue-specific expression patterns except for mature leaves, where all had low transcriptional levels (Fig. 1). Of the 149 *MYB*s

Table 1. *Divergence between* R2R3-MYB *gene pairs in* Populus

Gene pairs were identified at the terminal nodes (>80% similarity) of the phylogenetic tree shown in Supplementary Fig. S1. k_s, synonymous substitution rate; k_a, nonsynonymous substitution rate.

No.	Gene 1	Gene 2	Exon–intron organization	k_s	k_a	k_a/k_s	Gene fate[a]	Duplication[b]	Gene expression[c]
1	*PtrMYB001*	*PtrMYB022*	3-2/3-2	0.3012	0.1314	0.4362	RR	W	AD
2	*PtrMYB002*	*PtrMYB021*	2-1/2-1	0.2990	0.1289	0.4310	RD	W	AD
3	*PtrMYB003*	*PtrMYB020*	2-1/2-1	0.1909	0.0679	0.3555	RR	W	AB
4	*PtrMYB005*	*PtrMYB094*	3-2/3-2	0.2709	0.0547	0.2018	RR	O	AC
5	*PtrMYB006*	*PtrMYB126*	2-1/2-1	0.3070	0.1276	0.4156	RR	O	AD
6	*PtrMYB007*	*PtrMYB124*	3-2/3-2	0.1616	0.0837	0.5182	RR	O	AB
7	*PtrMYB008*	*PtrMYB127*	2-1/2-1	0.2637	0.1219	0.4621	RR	W	AD
8	*PtrMYB009*	*PtrMYB123*	2-1/2-1	0.2508	0.1068	0.4259	RN	W	AD
9	*PtrMYB010*	*PtrMYB128*	3-2/3-2	0.2333	0.0358	0.1535	RR	W	AB
10	*PtrMYB011*	*PtrMYB171*	3-2/3-2	0.3278	0.1225	0.3738	RR	W	–
11	*PtrMYB012*	*PtrMYB024*	4-3/3-2	0.2156	0.1126	0.5224	RR	O	AC
12	*PtrMYB014*	*PtrMYB015*	3-2/3-2	0.0809	0.0146	0.1811	RR	O	–
13	*PtrMYB017*	*PtrMYB043*	3-2/3-2	0.3134	0.0706	0.2252	RR	O	–
14	*PtrMYB018*	*PtrMYB152*	3-2/3-2	0.2568	0.0925	0.3601	RD	O	AC
15	*PtrMYB019*	*PtrMYB176*	1-0/1-0	0.3206	0.0970	0.3027	RR	O	AD
16	*PtrMYB023*	*PtrMYB206*	3-2/3-2	0.2458	0.0646	0.2630	RR	W	–
17	*PtrMYB025*	*PtrMYB099*	3-2/3-2	0.2480	0.0731	0.2948	RR	O	AB
18	*PtrMYB026*	*PtrMYB031*	2-1/2-1	0.2607	0.0770	0.2955	RR	O	AB
19	*PtrMYB027*	*PtrMYB032*	3-2/3-2	0.2012	0.0545	0.2712	RR	O	AD
20	*PtrMYB028*	*PtrMYB192*	3-2/3-2	0.3002	0.1126	0.3751	RR	O	AC
21	*PtrMYB029*	*PtrMYB033*	1-0/1-0	0.5040	0.1226	0.2432	RD	O	AC
22	*PtrMYB030*	*PtrMYB109*	3-2/3-2	0.1606	0.0741	0.4614	RR	O	AD
23	*PtrMYB034*	*PtrMYB164*	3-2/3-2	0.2001	0.1044	0.5217	RR	O	AB
24	*PtrMYB036*	*PtrMYB212*	3-2/3-2	0.2003	0.0672	0.3356	RR	W	AB
25	*PtrMYB039*	*PtrMYB136*	3-2/3-2	0.2300	0.1168	0.5080	RR	W	AD
26	*PtrMYB040*	*PtrMYB137*	3-2/3-2	0.2453	0.0926	0.3773	RR	W	AD
27	*PtrMYB044*	*PtrMYB045*	3-2/3-2	0.2985	0.1165	0.3902	RN	O	AB
28	*PtrMYB047*	*PtrMYB229*	3-2/3-2	0.2166	0.0885	0.4083	RR	O	AD
29	*PtrMYB048*	*PtrMYB065*	3-2/3-2	0.2577	0.1002	0.3887	RN	O	AD
30	*PtrMYB049*	*PtrMYB063*	3-2/3-2	0.4035	0.1110	0.2751	RR	O	AD
31	*PtrMYB050*	*PtrMYB051*	2-1/2-1	0.1229	0.0414	0.3369	RR	O	AD
32	*PtrMYB052*	*PtrMYB062*	3-2/3-2	0.2529	0.0948	0.3748	RR	O	AC
33	*PtrMYB053*	*PtrMYB240*	3-2/3-2	0.2087	0.0853	0.4089	RR	O	–
34	*PtrMYB055*	*PtrMYB121*	3-2/3-2	0.2109	0.0661	0.3134	RR	W	AC
35	*PtrMYB056*	*PtrMYB111*	3-2/4-3	0.3615	0.1628	0.4505	RR	W	AD
36	*PtrMYB057*	*PtrMYB093*	3-2/3-2	0.2575	0.0861	0.3342	RN	O	AC
37	*PtrMYB060*	*PtrMYB061*	2-1/2-1	0.0154	0.0056	0.3623	RR	T	AB
38	*PtrMYB066*	*PtrMYB142*	3-2/3-2	0.3165	0.0876	0.2766	RR	O	AC
39	*PtrMYB070*	*PtrMYB198*	3-2/3-2	0.2324	0.1003	0.4317	RR	W	AD
40	*PtrMYB071*	*PtrMYB072*	3-2/4-3	0.2781	0.1294	0.4651	RR	T	AB
41	*PtrMYB073*	*PtrMYB147*	3-2/3-2	0.2818	0.1147	0.4070	DD	O	AC
42	*PtrMYB074*	*PtrMYB148*	3-2/3-2	0.1737	0.0487	0.2806	RR	O	AB
43	*PtrMYB075*	*PtrMYB199*	2-1/2-1	0.3173	0.0779	0.2456	RR	W	AB
44	*PtrMYB076*	*PtrMYB145*	3-2/3-2	0.3661	0.1805	0.4931	RD	O	AD
45	*PtrMYB077*	*PtrMYB146*	3-2/3-2	0.3138	0.1252	0.3990	RR	O	AD
46	*PtrMYB079*	*PtrMYB102*	3-2/3-2	0.1981	0.0773	0.3903	RR	O	–
47	*PtrMYB081*	*PtrMYB155*	3-2/3-2	0.2028	0.0714	0.3523	RD	O	AD
48	*PtrMYB084*	*PtrMYB135*	3-2/3-2	0.2565	0.0934	0.3642	RR	O	AC
49	*PtrMYB085*	*PtrMYB133*	3-2/3-2	0.2529	0.0909	0.3594	RN	O	AD
50	*PtrMYB086*	*PtrMYB087*	3-2/3-2	0.0198	0.0119	0.6004	RR	O	–
51	*PtrMYB088*	*PtrMYB100*	3-2/3-2	0.2003	0.0787	0.3931	RR	W	AC
52	*PtrMYB090*	*PtrMYB167*	3-2/3-2	0.3563	0.0679	0.1905	DD	W	AC
53	*PtrMYB091*	*PtrMYB154*	1-0/1-0	0.3611	0.0845	0.2341	RR	O	AD
54	*PtrMYB092*	*PtrMYB125*	2-1/2-1	0.3078	0.0410	0.1331	RR	W	AB

Table 1. *Continued*

No.	Gene 1	Gene 2	Exon–intron organization	k_s	k_a	k_a/k_s	Gene fate[a]	Duplication[b]	Gene expression[c]
55	*PtrMYB097*	*PtrMYB101*	4-3/4-3	0.3635	0.1699	0.4675	RD	O	–
56	*PtrMYB105*	*PtrMYB122*	1-0/1-0	0.3440	0.0793	0.2305	RR	W	AD
57	*PtrMYB112*	*PtrMYB184*	3-2/3-2	1.0288	0.3615	0.1193	NN	O	AC
58	*PtrMYB113*	*PtrMYB117*	3-2/3-2	0.1523	0.1092	0.7166	RR	O	AC
59	*PtrMYB115*	*PtrMYB201*	2-1/2-1	0.4041	0.1229	0.3043	RN	W	AD
60	*PtrMYB119*	*PtrMYB120*	3-2/3-2	0.1551	0.0897	0.5783	RR	T	AD
61	*PtrMYB130*	*PtrMYB149*	2-1/2-1	0.2259	0.0854	0.3779	RD	O	AA
62	*PtrMYB138*	*PtrMYB186*	3-2/3-2	0.0209	0.0293	1.4041	RR	T	AB
63	*PtrMYB139*	*PtrMYB188*	3-2/3-2	1.7366	0.6243	0.3595	RD	O	AD
64	*PtrMYB156*	*PtrMYB221*	2-1/2-1	0.3786	0.0562	0.1484	RR	O	–
65	*PtrMYB158*	*PtrMYB189*	2-1/2-1	0.2031	0.0801	0.3946	RD	W	AA
66	*PtrMYB161*	*PtrMYB175*	3-2/3-2	0.2790	0.1088	0.3901	RR	O	AD
67	*PtrMYB165*	*PtrMYB194*	3-2/3-2	0.3149	0.1097	0.3484	RN	W	AC
68	*PtrMYB168*	*PtrMYB180*	3-2/3-2	0.3175	0.0963	0.3034	RN	O	AD
69	*PtrMYB170*	*PtrMYB216*	3-2/3-2	0.1904	0.0617	0.3241	RR	W	AD
70	*PtrMYB173*	*PtrMYB177*	1-0/1-0	0.3531	0.0891	0.2524	RR	W	AC
71	*PtrMYB181*	*PtrMYB182*	3-2/3-2	0.2556	0.0664	0.2598	RR	O	AD
72	*PtrMYB187*	*PtrMYB209*	3-2/3-2	0.2152	0.0677	0.3146	RR	W	AC
73	*PtrMYB190*	*PtrMYB220*	3-2/3-2	0.2574	0.0988	0.3840	RR	O	AD
74	*PtrMYB195*	*PtrMYB214*	3-2/3-2	0.2319	0.0563	0.2428	RR	W	AD
75	*PtrMYB196*	*PtrMYB213*	12-11/12-11	0.1899	0.0669	0.3524	RR	W	AC
76	*PtrMYB197*	*PtrMYB219*	3-2/3-2	0.3964	0.1282	0.3235	RR	O	AD
77	*PtrMYB202*	*PtrMYB210*	3-2/3-2	0.3698	0.0981	0.2653	RR	W	–
78	*PtrMYB204*	*PtrMYB207*	3-2/3-2	0.3348	0.0754	0.2252	RR	O	AC
79	*PtrMYB208*	*PtrMYB218*	3-2/3-2	0.2631	0.0606	0.2303	RR	W	AD
80	*PtrMYB226*	*PtrMYB232*	4-3/4-3	0.2403	0.0409	0.1703	RR	O	AC
81	*PtrMYB227*	*PtrMYB235*	2-1/2-1	0.1807	0.0487	0.2696	RR	W	AD

[a] Possible evolutionary modes of paralogues in the protein coding region after gene duplication described in Yang et al (2006). R (retention), a copy retains the original motif organization and function; D (degeneration), a copy degenerates or loses one or more motifs and functions; N (neo-functionalization), a copy acquires one or more motifs and functions.
[b] Gene pairs created by whole-genome duplication (W), tandem duplication (T), or other (O) events.
[c] Gene expression patterns based on microarray data (GSE13990) were categorized into four classes: AA, duplicates were expressed in different tissues; AB, duplicates shared almost identical expression patterns with respect to the tissues examined; AC, tissues in which one duplicate was expressed included all those of the paralogous duplicate plus other tissues; AD, the sets of tissues in which duplicates were expressed overlapped; –, no data for one duplicate in the microarray.

examined, 32 showed highest transcript accumulation in roots, 23 in young leaves, 27 in female catkins, 41 in male catkins, and 26 in xylem. These distinct expression patterns were divided into four categories according to the 69 gene pairs (Table 1). In the first category (AA), which covered two gene pairs, duplicates were expressed in different tissues, suggesting different functions. In the second category (AB), both duplicates of all 13 gene pairs shared almost identical expression patterns with respect to the tissues examined. The third category (AC) covered 20 pairs of duplicate genes, in which the tissues where one duplicate was highly expressed included all those of the paralogous duplicate plus other tissues. The fourth category (AD) contained 34 gene pairs, in which the sets of tissues in which duplicates were expressed overlapped. It is noteworthy that most gene pairs created by the whole-genome duplication fell within the third and fourth categories, suggesting that the duplicates have undergone subfunctionalization during subsequent evolution. In contrast, three gene pairs created by tandem duplication (*PtrMYB60/61*, *PtrMYB71/72*, and *PtrMYB138/186*) belonged to the second category, implying similar functions of the duplicates.

Identification of genes predominantly expressed in xylem from microarray data by qRT-PCR provides an important clue for their functions during wood formation in *Populus*. This study verified nine pairs of paralogous genes that were highly expressed in xylem. Of them, seven pairs were specifically expressed in wood-associated tissues including xylem, phloem, and cambium, and two pairs (*PdMYB55/121* and *PdMYB57/93*) had higher expression levels in wood-associated tissues other than the tissues examined (Fig. 3), all of which were in good agreement with the microarray profiles. These results suggest a role of these 18 *MYB*s in *Populus* wood formation.

Subcellular localization and transcriptional activation analysis of six PdMYB proteins

Of the nine gene pairs verified, *PdMYB10/128* (*AtMYB103* orthologues), *PdMYB90/167* (*AtMYB52* orthologues), and

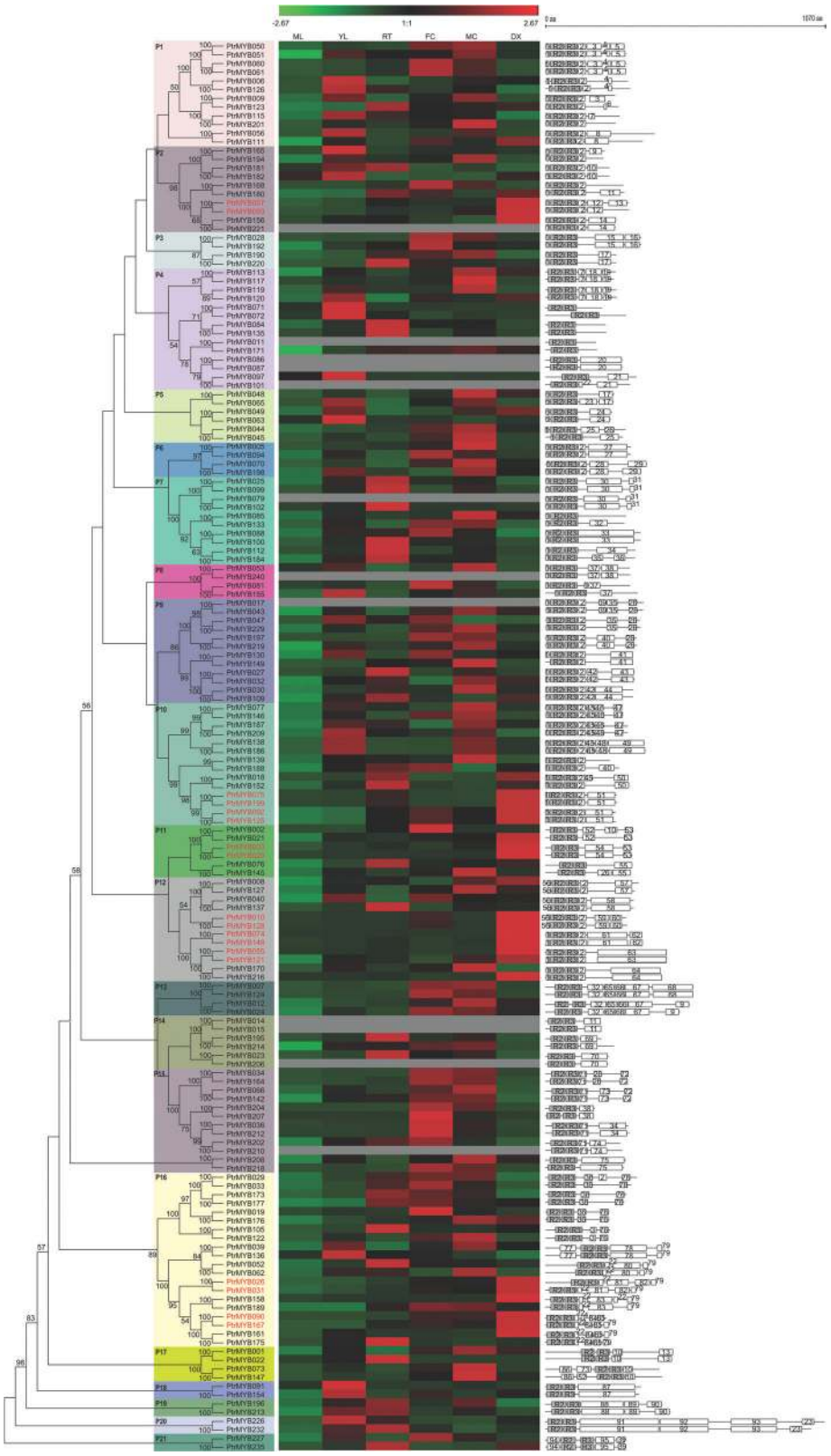

Fig. 1. Phylogenetic tree, expression profiles, and motif compositions of *Populus R2R3-MYB* gene pairs. (A) Phylogenetic tree showing major phylogenetic subfamilies (designated P1–P21) with high predictive value: multiple alignment of 162 full-length MYB proteins was conducted using Clustal X 1.83 and the tree was constructed using MEGA 4.0 with the neighbour-joining method and 1000 bootstrap replicates; bootstrap scores >50% are indicated at nodes. (B) Expression patterns of *Populus MYB* genes in different tissues; the genome-wide microarray data of Wilkins *et al.*, (2009) was reanalysed; colour scale at the top represents $\log_2$ expression values; genes highlighted in red were selected for validation by qRT-PCR. ML, mature leaves; YL, young leaves; Rt, roots; FC, female catkins; MC, male catkins; DX, differentiating xylem. (C) Schematic representation of the conserved motifs (see also Supplementary Table S2) in *Populus* R2R3-MYB proteins elucidated by MEME online; each box represents a motif in the protein; the length of the protein and motif can be estimated using the scale at top (this figure is available in colour at *JXB* online).

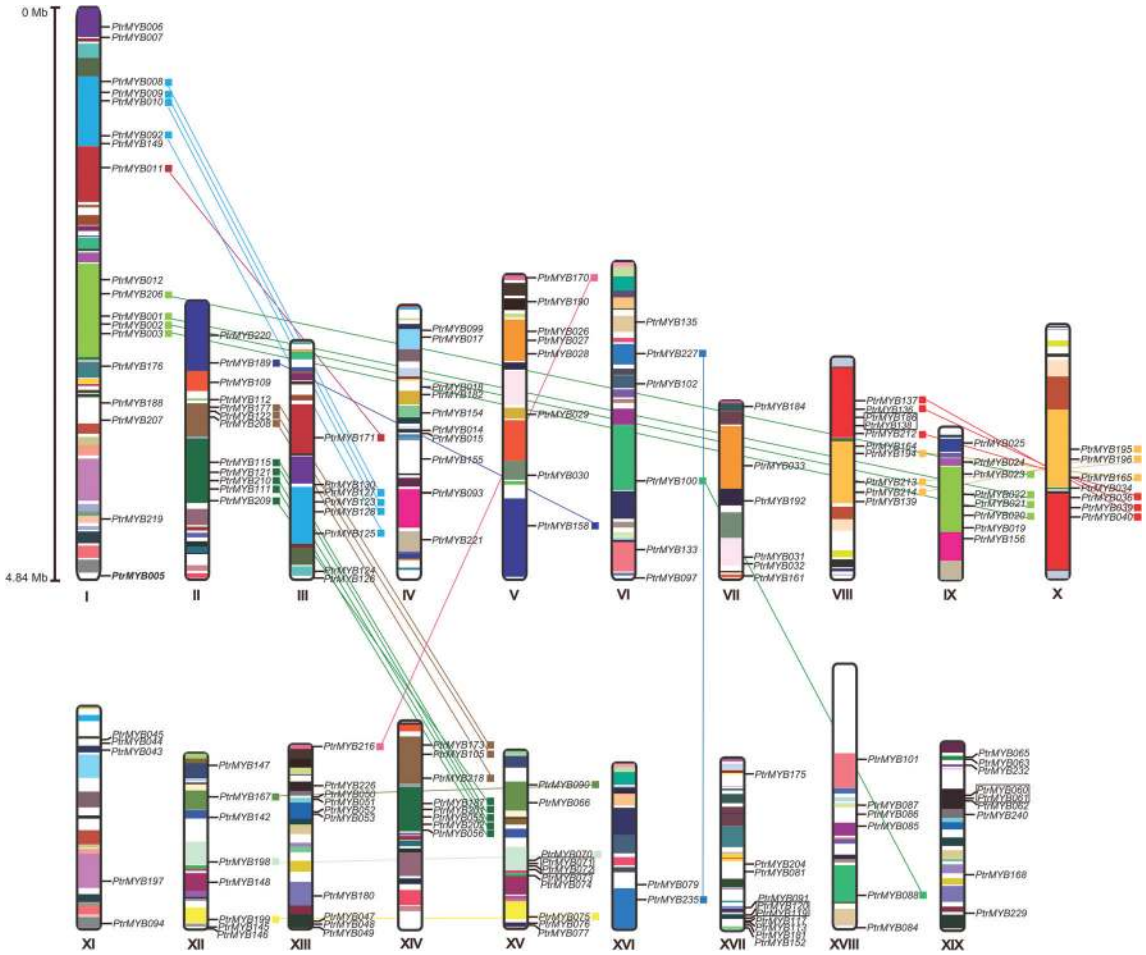

Fig. 2. Chromosomal locations of *R2R3-MYB* gene pairs in *Populus*. A total of 161 *R2R3-MYB* genes were mapped to 19 linkage groups, while only one gene (*PtrMYB057*) was shown to reside on an unassembled scaffold (not shown). The schematic diagram of *Populus* genome-wide chromosome organization arisen from the salicoid genome duplication event has been adapted from Tuskan *et al.*, (2006). Segmental duplicated homologous blocks are indicated with the same colour. Small boxes connected by coloured line indicate corresponding sister gene pairs, of which the genes connected by solid line locate in the duplicated blocks. Tandemly duplicated genes are encompassed in boxes. Scale represents the length (4.84 Mb) of chromosome I (this figure is available in colour at *JXB* online).

PdMYB92/*125* (*AtMYB42* orthologues) were specifically expressed in xylem and so were functionally analysed. Their subcellular localizations were first examined using a tobacco leaf transient expression system (Sparkes *et al.*, 2006). Fig. 4A shows that the six PdMYB:GFP fusion proteins were colocalized to DAPI-staining nuclei, indicating that these *PdMYB*s encode nuclear-localized proteins.

To elucidate transcriptional properties, six GAL4-full-length PdMYB fusion proteins were constructed. Fig. 4B shows that transcription of the *His* reporter gene was activated by the GAL4-full-length PdMYB fusion proteins, but not by the GAL4-MYB domains that cover R2R3 motifs for each PdMYB, implying that the six PdMYBs are transcriptional activators.

Morphological characteristics of Arabidopsis *overexpressing* PdMYB10, 128, 90, 167, 92, *or* 125

In order to gain insights into their biological roles, the six *PdMYB* genes were individually overexpressed under the control of the 35S promoter in wild-type *Arabidopsis*. At least 46 T_1 transgenic plants for each construct were generated and confirmed by RT-PCR (Fig. 5A–C). Phenotypic analyses were carried out on homozygous T_3 plants recovered from three independent transformants with high transcriptional levels. Transgenic *Arabidopsis* plants overexpressing *PdMYB10* and *128* displayed similar phenotypes (Fig. 5D and E and Table 2), with shorter and thinner inflorescence stems, smaller rosette leaves, severely inward-curling leaf blades, and lower fertility compared with the wild type. These phenotypes were similar to those obtained with plants overexpressing *Arabidopsis SND1* and *MYB46* (Zhong *et al.*, 2006, 2007). Ectopic expression of *PdMYB90*, *92*, *167*, and *125* in *Arabidopsis* resulted in similar phenotypes, with smaller rosette leaves and shorter and wrinkled siliques compared with the wild type (Fig. 5D and E and Table 2). It is notable that plants overexpressing *PdMYB90*, *92*, *167*, and *125* were no longer able to remain upright when they reached heights of >15 cm, possibly because of the loss of secondary walls in the stem cells (Fig. 5E).

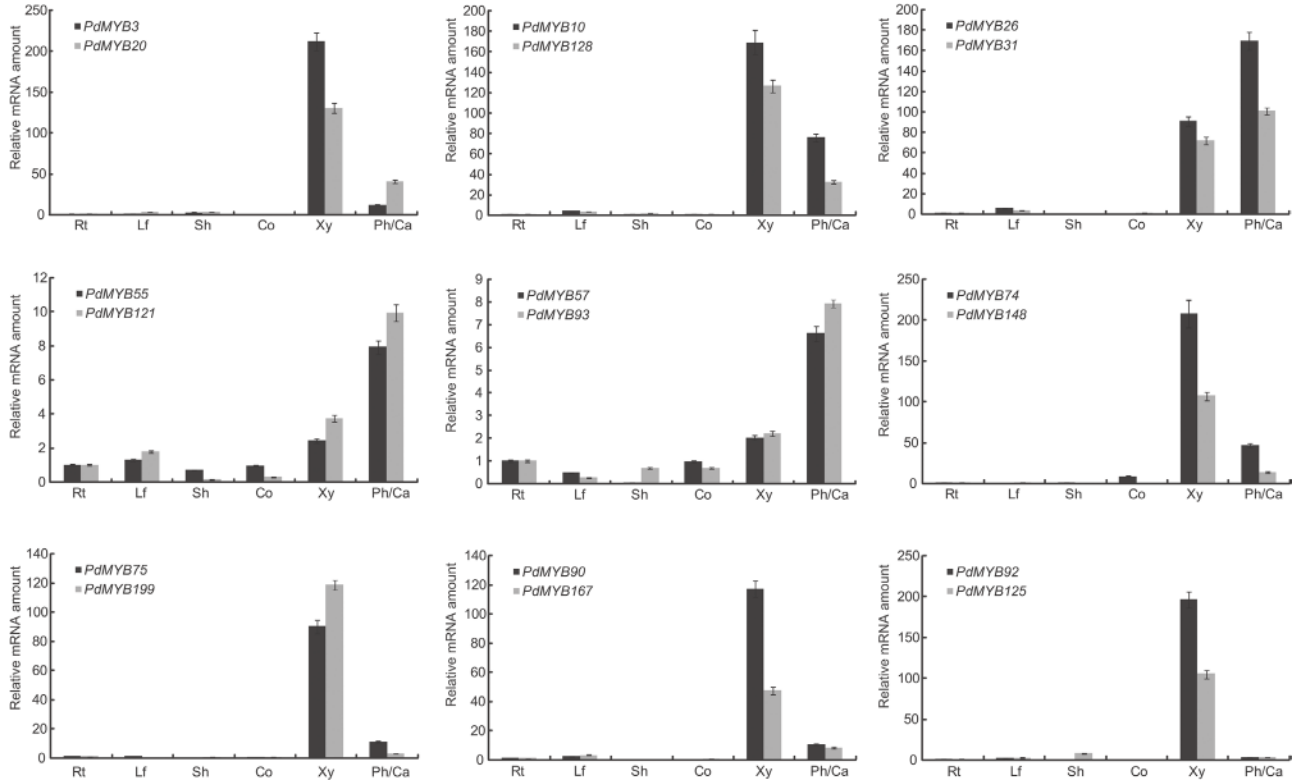

Fig. 3. Expression analysis of nine pairs of paralogous *R2R3-MYB* genes using qRT-PCR. The relative mRNA abundance of gene pair was normalized with respect to reference gene *PtUBQ* in different tissues. Data are mean±SD of three biological repeats. Rt, roots; Lf, mature leaves; Sh, shoots; Co, cortex; Xy, xylem; Ph/Ca, phloem and cambium.

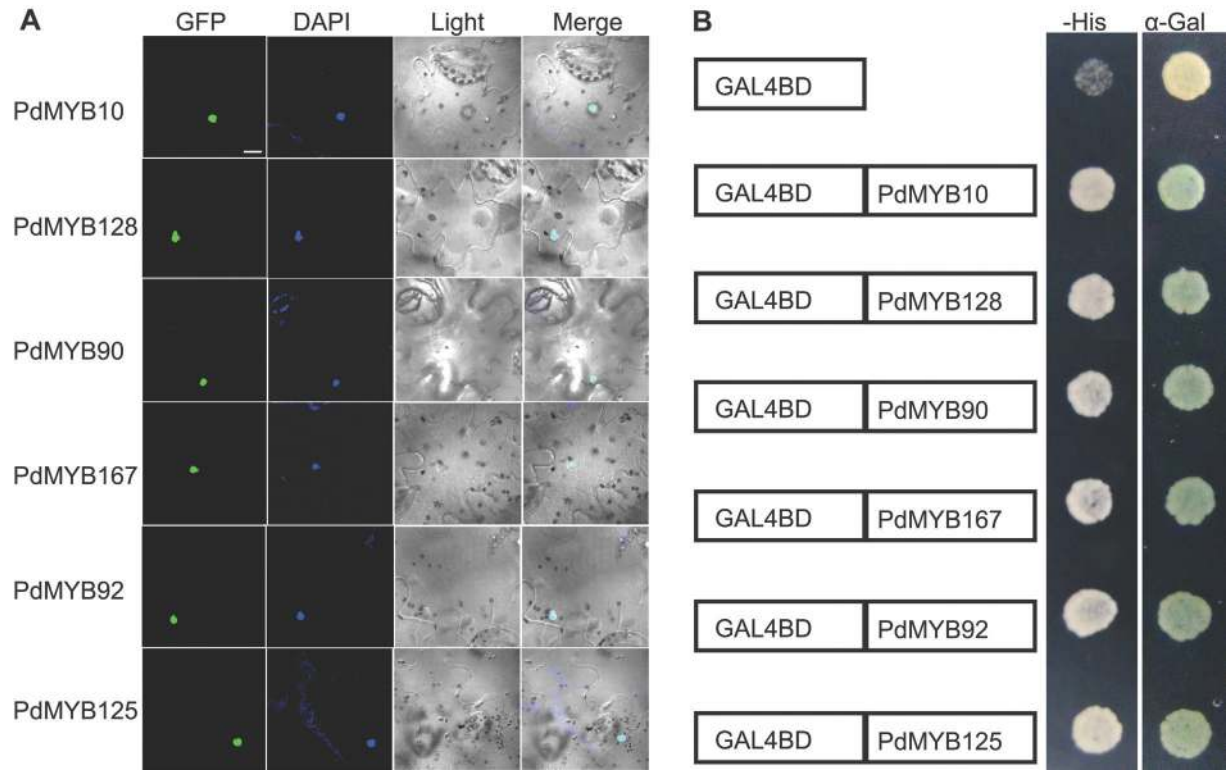

Fig. 4. Subcellular localization and transcriptional activation analysis of PdMYB10, 128, 90, 167, 92, and 125. (A) Confocal images of subcellular localizations of the six PdMYBs in tobacco leaf epidermal cells; expression of the *PdMYB-GFP* fusion gene was examined after 3 d by fluorescence and light microscopy; DAPI is used to stain the cell nucleus; bar, 10 μm. (B) Transcriptional activation analysis of the six *PdMYB*s fused with the GAL4 DNA-binding domain in yeast; each protein was able to activate the expression of *His3* and *MEL1* reporter genes (this figure is available in colour at *JXB* online).

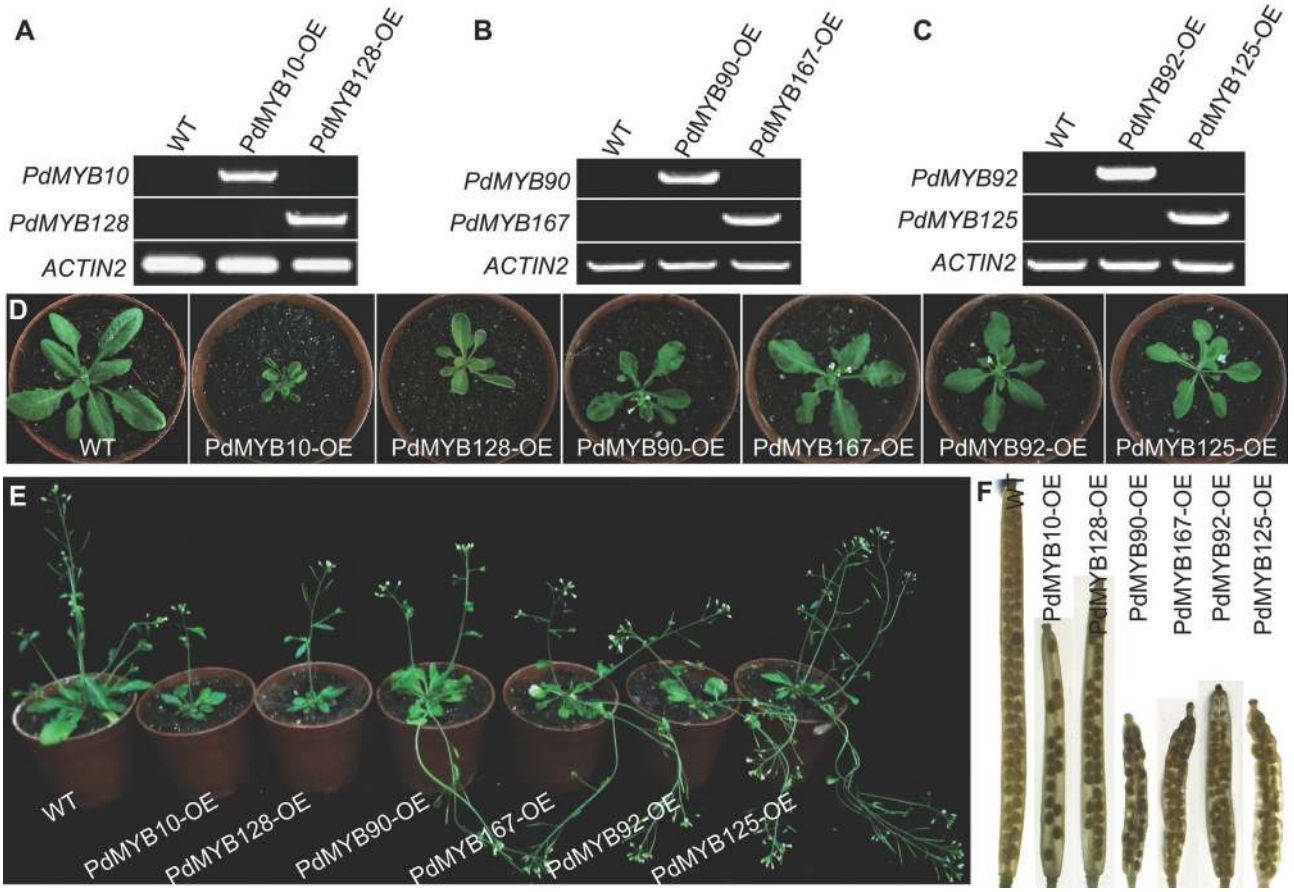

Fig. 5. Morphological analysis of *PdMYB10*, *128*, *90*, *167*, *92*, and *125* overexpressing *Arabidopsis* plants. (A–C) RT-PCR analysis of the six *PdMYB* representative overexpression lines. (D and E) Four- (D) and six- (E) week-old wild-type and *PdMYB* overexpression plants; arrows show upward-curling leaf blades. (F) Siliques of six *PdMYB* overexpression plants (this figure is available in colour at *JXB* online).

Table 2. *Measurements of growth parameters for wild-type and* PdMYB10*,* 128*,* 90*,* 167*,* 92*, and* 125 *overexpression plants*

Six-week-old plants were measured under long-day conditions (16/8 h light/dark cycle). Data are mean±SD ($n \geq 20$). The main stem was measured for stem diameter and the fourth pair of rosette leaves was measured for leaf blade dimensions. Significant differences from wild-type plants: **P*<0.05; ***P*<0.01 (Student's t-test).

Parameter	Wild type	*PdMYB10*	*PdMYB128*	*PdMYB90*	*PdMYB167*	*PdMYB92*	*PdMYB125*
Height (cm)	35.20±2.21	23.40±2.47**	25.73±3. 20**	35.60±2.82	38.40±2.97	34.47±3.91	33.40±3.38
Stem diameter (μm)	2097.4±109.23	1634.46±121.48**	1638.24±114.02**	1929.94±123.27	1989.7±311.34	2113.12±201.41	1958.75±224.64
Leaf blade length (cm)	4.13±0.21	2.94±0.27**	2.88±0.29**	3.53±0.15*	3.64±0.19*	3.87±0.19*	3.85±0.14*
Leaf blade width (cm)	2.02±0.15	1.4±0.14**	1.64±0.13**	1.54±0.15**	1.62±0.15**	1.62±0.07**	1.66±0.09**
Rosette leaf number	26.0±1.42	38.20±1.93**	37.80±1.31**	19.00±1.82**	20.60±1.34*	23.00±1.15*	20.00±1.41*
Flowering time (d)	34.64±2.13	39.07±1.95*	39.81±1.24*	26.54±2.06*	28.08±1.21*	30.23±1.28*	29.64±1.56*

Stem secondary wall formation in Arabidopsis *overexpressing* PdMYB10*,* 128*,* 90*,* 167*,* 92*, or* 125

To investigate the potential roles of the six *PdMYB*s in secondary wall formation, the basal stems of their overexpressing *Arabidopsis* plants were sectioned and observed. As shown in Fig. 6, overexpression of either duplicate of a gene pair had similar effects on secondary wall thickening in stems. *PdMYB10* and *128* overexpression lines displayed a thicker secondary wall of fibre cells but not of vessel cells. The cell-wall thickness of xylary fibres and interfascicular fibres was increased by at least 22 and 15%, respectively, in these overexpression lines compared with the wild type (Table 3). In contrast, overexpression of *PdMYB90*, *92*, *167*, or *125* in *Arabidopsis* caused a severe reduction in fibre and vessel cell-wall thickness. The cell-wall thickness of xylary fibres, interfascicular fibres, and vessels was decreased by at least 23, 15, and 11%, respectively, in these overexpression lines compared with the wild type (Table 3). It is notable that *PdMYB90* and *125* overexpression lines were often found

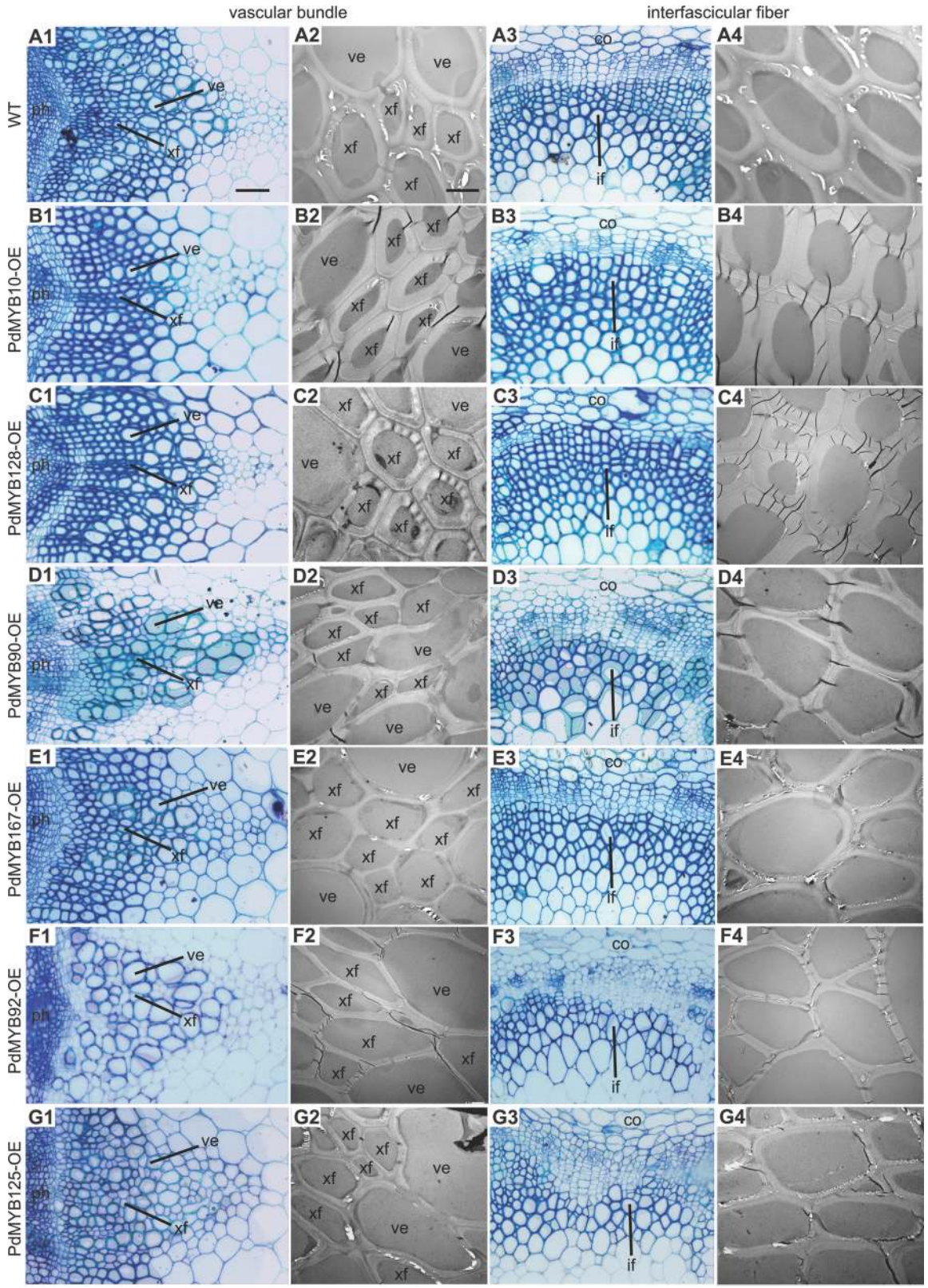

Fig. 6. Effects of overexpression of *PdMYB10*, *128*, *90*, *167*, *92*, and *125* in *Arabidopsis* on secondary wall thickening in stems. Cross sections of basal stems from 7-week-old wild type (A1–4) and *PdMYB* overexpression (B1–G4) plants; co, cortex; if, interfascicular fibre; ph, phloem; ve, vessel; xf, xylary fibre. Bar in A1 for the light micrographs, 50 μm; bar in A2 for the transmission electron micrographs, 5 μm (this figure is available in colour at *JXB* online).

Table 3. *Wall thickness of vessels and fibres in the stems of wild-type and* PdMYB10, 128, 90, 167, 92, *and* 125 *overexpression plants*

Six-week-old plants were used for analysis. Wall thickness was measured from transmission electron micrographs of fibres and vessels. Data are mean±SD from 30 cells. *P<0.05.

Parameter	Wild type	*PdMYB10*	*PdMYB128*	*PdMYB90*	*PdMYB167*	*PdMYB92*	*PdMYB125*
Xylary fibres (μm)	0.97±0.14	1.18±0.16*	1.21±0.11*	0.61±0.20*	0.76±0.14*	0.62±0.13*	0.75±0.09*
Vessels (μm)	0.99±0.06	0.97±0.11	1.01±0.10	0.72±0.04*	0.80±0.04*	0.74±0.02*	0.84±0.06*
Interfascicular fibres (μm)	1.37±0.22	1.58±0.26*	1.59±0.12*	1.19±0.21*	1.23±0.19*	0.96±0.13*	1.21±0.17*

Table 4. *Cell-wall composition analysis of the stems of wild-type and* PdMYB10, 128, 90, 167, 92, *and* 125 *overexpression plants*

Six-week-old plants were tested. Data are mean±SD (mg (g alcohol-insoluble residues)$^{-1}$) of three independent assays. *P<0.05, **P<0.01.

Cell-wall component	Wild type	*PdMYB10*	*PdMYB128*	*PdMYB90*	*PdMYB167*	*PdMYB92*	*PdMYB125*
Man	12.97±0.22	10.39±0.23	12.37±1.22	11.29±0.24	13.66±0.24	13.61±0.19	13.46±0.94
Rha	4.56±0.87	4.19±0.04	5.07±1.65	5.58±0.91	7.17±1.07	7.54±0.61	5.48±0.69
GlcA	2.15±0.50	1.52±0.22	2.33±0.41	1.95±0.30	1.87±0.71	2.70±0.47	1.89±0.53
GalA	13.70±1.11	12.71±0.45	16.75±5.51	18.61±0.59	20.63±1.67	21.49±0.97	22.08±0.62
Glc	11.19±0.94	14.38±0.14	12.14±2.35	13.54±0.09	12.09±0.69	11.58±0.19	13.15±0.62
Gal	9.92±0.39	8.52±0.28	8.66±1.83	11.96±0.33	12.09±0.97	12.35±0.71	10.62±1.77
Xyl	59.36±4.82	57.68±2.75	68.56±4.56*	41.30±0.53*	50.01±0.97*	50.52±4.16*	47.71±2.76*
Ara	8.32±0.35	6.36±0.81	7.40±0.48	10.72±0.21	8.53±0.22	9.37±0.79	10.59±1.86
Fuc	0.23±0.15	0.55±0.06	0.87±0.13	0.63±0.06	0.82±0.09	0.68±0.06	0.82±0.23
Cellulose	257.66±2.29	302.71±5.10*	310.62±7.10**	195.84±1.10**	216.99±2.11*	135.21±2.75**	186.72±4.15**
Lignin	247.80±3.67	243.23±1.85	269.12±5.85*	198.42±2.95*	220.29±2.12*	218.78±3.16*	217.61±1.23*

to have partially deformed vessels, probably because of the weakening of the vessels' secondary walls (Fig. 6D and G).

Given that the change of wall thickness may reflect differences in cell-wall composition, this work next examined the cell-wall composition of comparable tissues harvested from the stems of the six *PdMYB* overexpressing plants and the wild type. Compared with the wild type, the amounts of cellulose, xylose, and lignin were increased in *PdMYB128* overexpression lines, but were decreased in lines overexpressing *PdMYB90*, *92*, *125*, or *167* (Table 4). It is worthy to note in *PdMYB10* overexpression lines that cellulose content was slightly increased by 18%, whereas xylose and lignin contents were indistinguishable from those of the wild type. Together, these results demonstrate that the six *PdMYB*s are involved in the regulation of secondary wall formation in *Arabidopsis* stems.

Flowering in Arabidopsis *overexpressing* PdMYB10, 128, 90, 167, 92, *or* 125

Microarray analysis showed that *PdMYB10*, *128*, *90*, *167*, *92*, and *125* were highly expressed in floral organs in addition to in xylem (Fig. 1). Therefore, this study examined the effects of overexpression of the six *PdMYB*s in *Arabidopsis* on flower development (Fig. 7). Ectopic expression of *PdMYB10* or *128* led to increased rosette leaf numbers and a delay in flowering time of at least 5 days compared with the wild type (Table 2). Except for size, no visible morphological difference was observed in buds and open flowers between these overexpression lines and the wild type. By contrast, overexpression of *PdMYB90*, *92*, *167*, or *125* resulted in decreased rosette leaf numbers and an advance in flowering time of at least 4 days compared with the wild type (Table 2). Furthermore, lines overexpressing *PdMYB90*, *92*, *167*, or *125* displayed smaller flowers with wrinkled petals and abnormally short sepals (Fig. 7). Quantitative analysis indicated that the sepals of these overexpression lines grew more slowly than those of the wild type, from the budding stage to open flowers (Fig. 7V and W).

Expression of secondary wall- and flower-associated genes in Arabidopsis *overexpressing* PdMYB10, 128, 90, 167, 92, *or* 125

Given that overexpression of the six *PdMYB*s in *Arabidopsis* individually led to visible alteration in stem secondary wall formation and flowering, this study used the overexpression lines to examine the expression of the related genes by qRT-PCR. Expression of three cellulose synthase genes (*CESA4*, *CESA7*, and *CESA8*), two xylan synthetic genes (*IRX8* and *IRX9*), and two lignin synthetic genes (*4CL1* and *CCoAOMT1*) was induced in *PdMYB128* overexpression lines and repressed in *PdMYB90*, *92*, *167*, and *125* overexpression lines (Table 5). By contrast, expression of *CESA4*, *CESA7*, and *CESA8* was induced in *PdMYB10* overexpression lines. These were in good

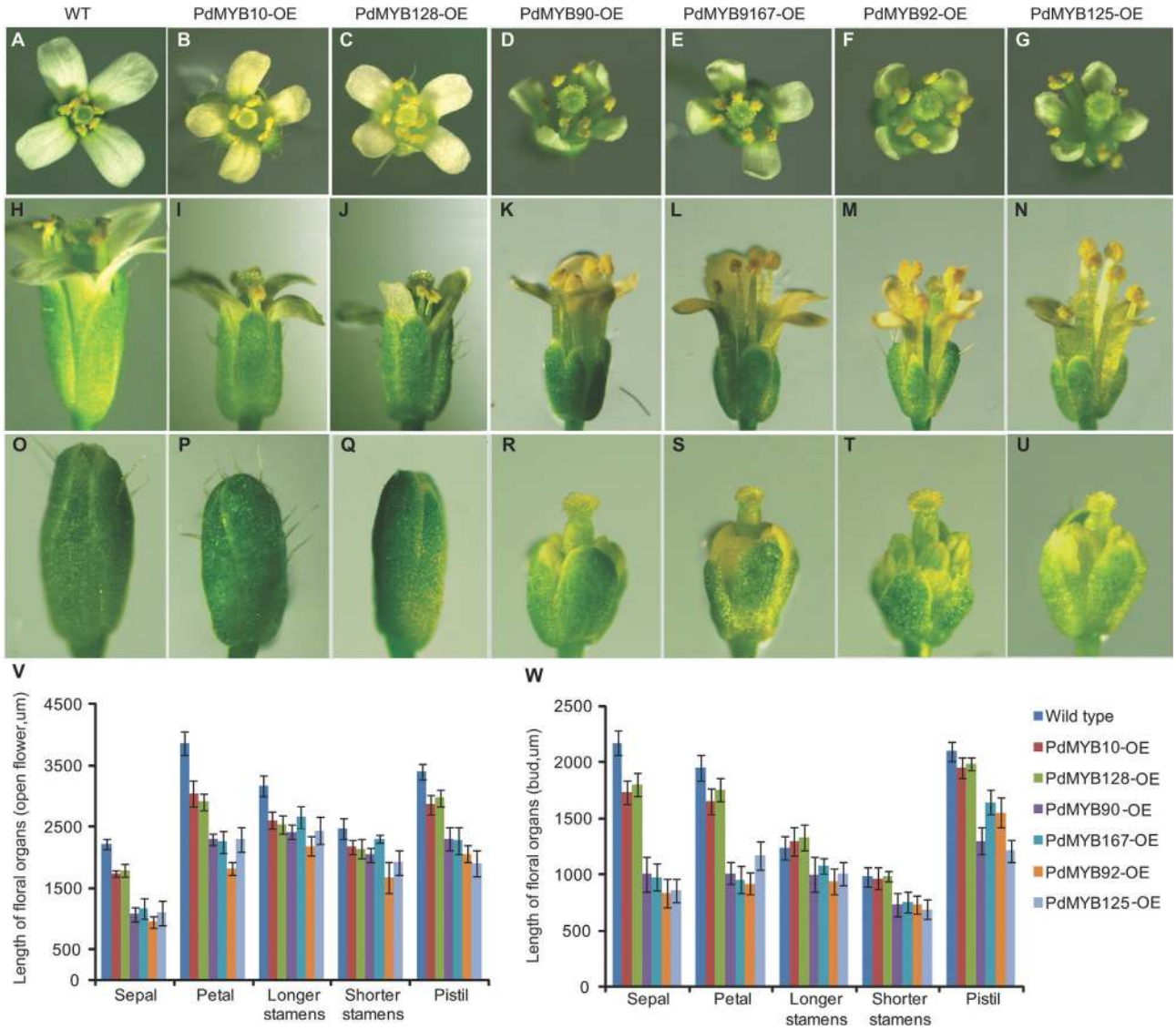

Fig. 7. Effects of overexpression of *PdMYB10*, *128*, *90*, *167*, *92*, and *125* in *Arabidopsis* on flower development. (A–U) Phenotypic comparison of open flowers (A–N) and buds (O–U) between the wild type and six *PdMYB* overexpression plants. (V and W) Floral organ lengths of open flowers (V) and buds (W) from the wild type and six *PdMYB* overexpression plants; measurements were taken using Olympus DP2-BSW software; data are mean±SD from at least six independent samples (this figure is available in colour at *JXB* online).

agreement with the alteration in cellulose, xylose, and lignin contents in the overexpression plants (Table 4). Of five key flowering genes, the expression of *GAI* and *LFY* was repressed in *PdMYB10* and *128* overexpression lines but induced in *PdMYB90*, *92*, *167*, and *125* overexpression lines (Table 5). In addition, expression of *AP1*, *AP3*, and *PI*, which are involved in sepal and petal development, was found to be upregulated in *PdMYB90*, *92*, *167*, and *125* overexpression lines.

Discussion

Evolution of Populus R2R3-MYB *gene pairs*

In plants, the *R2R3-MYB* gene family is one of the largest families of transcription factors and contains many pairs of paralogous genes (Martin and Paz-Ares, 1997). This study characterized 81 pairs of paralogous *R2R3-MYB* genes in *Populus*. Phylogenetic analysis showed these gene pairs were divided into 21 subgroups supported by bootstrap scores of >50%. Members of the same subgroups generally shared one or more identical motifs in the C-terminal region, outside the highly conserved R2R3-MYB domain, suggesting similar functions between different gene pairs. For example, in subgroup P11, PtrMYB002/021 and PtrMYB003/020 share an identical R2R3-MYB domain and motif53 and function as positive regulators of wood formation (McCarthy *et al.*, 2010; Zhong *et al.*, 2013). It is possible that identical motifs are required for similar roles in wood formation. It is widely accepted that gene duplication is a primary source of genetic novelty. These results indicate that the 81 *R2R3-MYB* gene pairs might have resulted from multiple gene duplication events and that they have five evolutionary outcomes in their coding regions: RR, RD, RN, DD, and NN.

Table 5. *Comparison of expression of secondary wall- and flower-associated genes in* PdMYB10*,* 128*,* 90*,* 167*,* 92*, and* 125 *overexpression plants with that in wild type*

Data are ratio of expression in overexpression line versus that in the wild type: +++, ≥10; ++, 5–10; +, 2–5; 0, 0.5–2; –, 0.2–0.5; —, 0.1–0.2; ——, ≤0.1. *ACTIN2* was used as an internal control. Three biological replicates were performed for each experiment

Gene	*PdMYB10*	*PdMYB128*	*PdMYB90*	*PdMYB167*	*PdMYB92*	*PdMYB125*
Secondary wall-associated genes						
CESA4	+	++	——	—	—	——
CESA7	+	+	–	–	–	–
CESA8	++	+	–	–	–	–
IRX8	0	++	–	–	–	–
IRX9	0	+	–	–	–	–
4CL1	0	+	–	–	–	–
CCoAMT1	0	+	–	–	–	–
Flower-associated genes						
CO	0	0	0	0	0	0
FLC	0	0	0	0	0	0
FT	0	0	0	0	0	0
GAI	–	–	+++	+++	++	++
LFY	—	—	++	++	++	++
AP1	0	0	++	++	+	+
AP3	0	0	+	+	+	+
PI	0	0	+	+	+	+

Duplicated genes that have resulted from whole-genome duplication (e.g. *PdMYB10* and *128*) or tandem duplication (e.g. *PtrMYB138* and *186*) generally have similar expression patterns, suggesting functional redundancy. By contrast, most non-genome-duplicated gene pairs have RD, RN, DD, and NN gene fates and different expression patterns (Table 1), implying that the duplicates may have been subject to sub- or neo-functionalization during the evolutionary process.

Effects of overexpression of PdMYB10*,* 128*,* 90*,* 167*,* 92*, and* 125 *in* Arabidopsis *on secondary wall formation*

Wood is mainly composed of secondary walls, which constitute the most abundant store of carbon produced by vascular plants. Understanding the molecular mechanisms controlling secondary wall deposition during wood formation is not only an important issue in plant biology but also critical for providing molecular tools to custom-design wood composition for a particular end use. Compared to the well-characterized *MYB*s in *Arabidopsis*, only some *Populus R2R3-MYB* genes have been shown to participate in the regulation of secondary wall biosynthesis during wood formation (Zhong *et al.*, 2013). The current work has provided evidence that three pairs of *Populus MYB* genes (*PdMYB10*/*128*, *PdMYB90*/*167*, and *PdMYB92*/*125*) may be involved in secondary wall formation. Overexpression of *PdMYB10* and *128* in *Arabidopsis* resulted in increased secondary wall thickness of fibre cells in stems. This phenotype was similar to that obtained with plants overexpressing the *Arabidopsis* orthologue *AtMYB103* (Zhong *et al.*, 2008). Since *Arabidopsis* has a regulatory network for secondary wall formation similar to that of poplar trees (Zhong *et al.*, 2013) and *AtMYB103* has been shown to positively control secondary wall formation in *Arabidopsis* (Zhong *et al.*, 2008), *PdMYB10* and *128* may share functions similar to that of *AtMYB103* during wood formation in poplar trees. The current work also demonstrated that overexpression of *PdMYB90*, *92*, *167*, and *125* in *Arabidopsis* individually led to decreased secondary cell-wall thickness of vessels and fibres. These were different from their *Arabidopsis* orthologue *AtMYB42* and *AtMYB52* overexpression phenotypes (no visible alteration in secondary cell walls), but resembled those of their dominant repression plants (Zhong *et al.*, 2008). This cannot be explained by cosuppression, due to the stable expression of *PdMYB90*, *92*, *167*, and *125* transcripts in overexpression plants. Interestingly, a similar paradox has been observed for *SND2* overexpression, where excess levels of this transcription activator has been reported to have little or an indirect repressive effect (Hussey *et al.*, 2011). This phenomenon could be attributed to gene dosage effects, where a stoichiometric increase in one MYB protein leads to decreased molar yield of a multiprotein complex and greater yield of incomplete intermediates (Birchler *et al.*, 2005). Further analysis is necessary to validate this assumption.

Effects of overexpression of PdMYB10*,* 128*,* 90*,* 167*,* 92*, and* 125 *in* Arabidopsis *on flowering time*

Populus has a very long juvenile phase before flowering, which significantly differs from the annual herbaceous plant *Arabidopsis*. However, comparative and functional genomic research has provided evidence on the conservation of flowering-regulatory pathways between the two species (Bohlenius *et al.*, 2006; Igasaki *et al.*, 2008). It is well known

that four major environmental and endogenous stimuli (photoperiod, vernalization, autonomous, and gibberellin) make *Arabidopsis* switch from vegetative growth to flowering. Among the floral regulators, CO is mainly involved in photoperiod, FLC in vernalization, and GAI in gibberellin. FT is promoted by CO but is repressed by FLC. LFY functions as an integrater in four flowering pathways (Irish, 2010). The current results show that ectopic expression of *PdMYB10* and *128* in *Arabidopsis* delays flowering, whereas *PdMYB90*, *92*, *167*, and *125* overexpression in *Arabidopsis* promotes flowering in long-day conditions. Accordingly, expression of *GAI* and *LFY* remained largely changed, but that of *CO*, *FLC*, and *FT* was not altered, in their overexpression plants. Considering that flowering is coordinately controlled by gibberellin and brassinosteroids in *Arabidopsis* (Domagalska *et al.*, 2010), that *AtMYB52* and *AtMYB42* are negatively regulated by DELLA (Cheng, 2007), and that *AtMYB103* is involved in brassinosteroids (Ye *et al.*, 2010), the current study suggests that *PdMYB10*, *128*, *90*, *167*, *92*, and *125* may participate in a gibberellin-mediated flowering pathway. Previous studies imply that flowering induction time might be associated with secondary wall formation in *Arabidopsis*. For instance, some major quantitative trait loci for secondary wall thickening during xylem expansion and fibre differentiation correlate tightly with a major flowering-time quantitative trait locus; furthermore, transient induction of flowering at the rosette stage promotes secondary wall thickening (Sibout *et al.*, 2008). A *soc1 ful* double mutant shows synergistically delayed flowering time and dramatically increased secondary wall thickening (Melzer *et al.*, 2008). The heterologous expression of *Miscanthus MIWRKY12* in *Arabidopsis* results in an increased pith cell-wall thickness and early flowering (Yu *et al.*, 2013). The current study showed that overexpression of six *Populus PdMYB*s in *Arabidopsis* affected stem secondary wall thickening and flowering time, which further supports this speculation. In addition, this work demonstrated that overexpression of *PdMYB90*, *92*, *167*, and *125* in *Arabidopsis* led to a visible alteration in sepal and petal morphologies. Since gibberellin is known to play a key role in petal and sepal development (Cheng, 2007), it is possible that these *PdMYB*s regulate petal and sepal development through gibberellin. It will be interesting to investigate the effects of these four genes on flower development in poplar trees because of the dramatic difference between flowers of *Populus* and *Arabidopsis*.

In conclusion, this paper has identified and evolutionarily analysed 81 pairs of paralogous *Populus R2R3-MYB* genes. Of them, nine pairs were determined by whole-genome microarray and qRT-PCR data to be highly expressed in xylem. Heterologous expression of *PdMYB10*, *128*, *90*, *167*, *92*, and *125* in *Arabidopsis* suggests potential roles in secondary wall formation and flowering in *Populus*. This work cannot absolutely exclude the possibilities that the six *PdMYB*s have different responses in *Populus*. Alternatively, ectopic expression of these genes may have off-target effects on these traits. However, these results provide valuable information for further studies on the roles of these genes in xylem formation and flowering of poplar trees.

Supplementary material

Supplementary data are available at *JXB* online.

Supplementary Fig. S1. Phylogenetic tree of 194 *Populus R2R3-MYB* genes.

Supplementary Table S1. Summary of the R2R3-MYB proteins in *Populus*.

Supplementary Table S2. Motifs of *Populus* R2R3-MYB proteins.

Supplementary Table S3. Primers.

Acknowledgements

This study was supported by the National Science and Technology Support Program (2013BAD22B01), the National Basic Research Program of China (2012CB114501), the National Natural Science Foundation of China (31300165), the Knowledge Innovation Program of Chinese Academy of Sciences (KSCX2-EW-J-10), the research project for the Application Foundation in Qingdao (12-1-4-9-(2)-jch), and the Youth Talent Plan of Chinese Academy of Agricultural Sciences to Y.Z.K. The authors thank the editor and the anonymous reviewers for constructive comments and suggestions on the revision of the manuscript.

References

Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. 2003. *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *The Plant Cell* **15,** 63–78.

Bailey T, Gribskov M. 1998. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14,** 48–54.

Bailey T, Williams N, Misleh C, Li W. 2006. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research* **34,** W369–373.

Baumann K, Perez-Rodriguez M, Bradley D, Venail J, Bailey P, Jin H, Koes R, Roberts K, Martin C. 2007. Control of cell and petal morphogenesis by R2R3 MYB transcription factors. *Development* **134,** 1691–1701.

Birchler JA, Riddle NC, Auger DL, Veitia RA. 2005. Dosage balace in gene regulation: biological implications. *Trends in Genetics* **21,** 219–226.

Bohlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, Strauss SH, Nilsson O. 2006. CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. *Science* **312,** 1040–1043.

Chai G, Hu R, Zhang D, Qi G, Zuo R, Cao Y, Chen P, Kong Y, Zhou G. 2012. Comprehensive analysis of CCCH zinc finger family in poplar (*Populus trichocarpa*). *BMC Genomics* **13,** 253.

Cheng H. 2007. Gibberellin regulates *Arabidopsis* floral development via suppresion of DELLA protein function, PhD thesis. Singapore: National University of Singapore.

Domagalska MA, Sarnowska E, Nagy F, Davis SJ. 2010. Genetic analysis of interactions among gibberellin, abscisic acid, and brassinosteroids in the control of flowering time in *Arabidopsis thaliana*. *PloS One* **17,** e14012.

Du H, Feng B, Yang S, Huang Y, Tang Y. 2012. The R2R3-MYB transcription factor gene family in maize. *PLoS One* **7,** e37463

Fukushima RS, Hatfield RD. 2001. Extraction and isolation of lignin for utilization as a standard to determine lignin concentration using the acetyl bromide spectrophotometric method. *Journal of Agricultural and Food Chemistry* **49,** 3133–3139.

Gocal GF, Sheldon CC, Gubler F, *et al*. 2001. GAMYB-like genes, flowering, and gibberellin signaling in *Arabidopsis*. *Plant Physiology* **127,** 1682–1693.

Goldman N, Yang, Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution* **11,** 725–736.

Guo A, Zhu Q, Chen X, Luo J. 2007. GSDS: a gene structure display server. *Yi Chuan* **29,** 1023–1026.

Higginson T, Li SF, Parish RW. 2003. AtMYB103 regulates tapetum and trichome development in *Arabidopsis thaliana*. *The Plant Journal* **35,** 177–192.

Hurst LD. 2002. The k_a/k_s ratio: diagnosing the form of sequence evolution. *Trends in Genetics* **18,** 486.

Hussey SG, Mizrach E, Spokevicius AV, Bossinger G, Berger DK, Myburg AA. 2011. SND2, a NAC transcription factor gene, regulates genes involved in secondary cell wall development in *Arabiodpsis* fibres and increases fibre cell area in *Eucalyptus*. *BMC Plant Biology* **11,** 173.

Igasaki T, Watanabe Y, Nishiguchi M, Kotoda N. 2008. The *FLOWERING LOCUS T/TERMINAL FLOWER 1* family in lombardy poplar. *Plant Cell Physiology* **49,** 291–300.

Irish VF. 2010. The flowering of *Arabidopsis* flower development. *The Plant Journal* **61,** 1014–1028.

Jiang C, Gu X, Peterson T. 2004. Identification of conserved gene structures and carboxy-terminal motifs in the Myb gene family of *Arabidopsis* and *Oryza sativa* L. ssp. *indica. Genome Biology* **5,** R46.

Jung C, Seo JS, Han SW, Koo YJ, Kim CH, Song SI, Nahm BH, Choi YD, Cheong JJ. 2008. Overexpression of AtMYB44 enhances stomatal closure to confer abiotic stress tolerance in transgenic *Arabidopsis*. *Plant Physiology* **146,** 623–635.

Legay S, Sivadon P, Blervacq A, *et al*. 2010. *EgMYB1*, an R2R3 MYB transcription factor from eucalyptus negatively regulates secondary cell wall formation in *Arabidopsis* and poplar. *New Phytologist* **188,** 744–786.

Lipsick JS. 1996. One billion years of Myb. *Oncogene* **13,** 223–235.

Ma T, Wang J, Zhou G, *et al*. 2013. Genomic insights into salt adaptation in a desert poplar. *Nature Communication* **4,** 2797.

Martin C, Paz-Ares J. 1997. MYB transcription factors in plants. *Trends in Genetics* **13,** 67–73.

McCarthy RL, Zhong R, Fowler S, Lyskowski D, Piyasena H, Carleton K, Spicer C, Ye ZH. 2010. The poplar MYB transcription factors, PtrMYB3 and PtrMYB20, are involved in the regulation of secondary wall biosynthesis. *Plant Cell Physiology* **51,** 1084–1090.

Mellway R, Tran L, Prouse M, Campbell M, Constabel C. 2009. The would-, pathogen-, and ultraviolet B-responsive MYB134 gene encodes an R2R3 MYB transcription factor that regulates proanthocyanidin synthesis in poplar. *Plant Physiology* **150,** 924–941.

Melzer S, Lens F, Gennen J, Vanneste S, Rohde A, Beeckman T. 2008. Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. *Nature Genetics* **40,** 1489–1492.

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29,** e45.

Ramirez V, Agorio A, Coego A, Garcia-Andrade J, Hernandez MJ, Balaguer B, Ouwerkerk P, Zarra I, Vera P. 2011. MYB46 modulates disease susceptibility to *Botrytis cinerea* in *Arabidopsis*. *Plant Physiology* **155,** 1920–1935.

Riechmann JL, Heard J, Martin G, *et al.* 2000. *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* **290,** 2105–2110.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4,** 406–425.

Selvendran RR, March JF, Ring SG. 1979. Determination of aldoses and uronic acid content of vegetable fiber. *Analytical Biochemistry* **96,** 282–292.

Sibout R, Plantegenet S, Hardtke CS. 2008. Flowering as a condition for xylem expansion in *Arabidopsis* hypocotyl and root. *Current Biology* **18,** 458–463.

Sparkes IA, Runions J, Kearns A, Hawes C. 2006. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nature Protocol* **1,** 2019–2025.

Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F, Niehaus K, Weisshaar B. 2007. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *The Plant Journal* **50,** 660–677.

Stracke R, Werber M, Weisshaar B. 2001. The *R2R3-MYB* gene family in *Arabidopsis thaliana*. *Current Opinion in Plant Biology* **4,** 447–456.

Sturn A, Quackenbush J, Trajanoski Z. 2002. Genesis: cluster analysis of orthologray data. *Bioinformatics*, **18,** 207–208.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24,** 1596–1599.

Teng S, Keurentjes J, Bentsink L, Koornneef M, Smeekens S. 2005. Sucrose-specific induction of anthocyanin biosynthesis in *Arabidopsis* requires the MYB75/PAP1 gene. *Plant Physiology* **139,** 1840–1852.

Thompson JD1, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25,** 4876–4882.

Tian Q, Wang X, Li C, Lu W, Yang L, Jiang Y, Luo K. 2013. Functional characterization of the poplar R2R3-MYB transcription factor PtoMYB216 involved in the regulation of lignin biosynthesis during wood formation. *PLoS ONE* **8,** e76369.

Tominaga R, Iwata M, Okada K, Wada T. 2007. Functional analysis of the epidermal-specific MYB genes CAPRICE and WEREWOLF in *Arabidopsis*. *The Plant Cell* **19,** 2264–2277.

Tuskan GA, Difazio S, Jansson S, *et al*. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313,** 1596–1604.

Updegraff DM. 1969. Semimicro determination of cellulose in biological materials. *Analytical Biochemistry* **32,** 420–424.

Wilkins O, Nahal H, Foong J, Provart NJ, Campbell MM. 2009. Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. *Plant Physiology* **149,** 981–993.

Yang X, Tuskan GA, Cheng Z. 2006. Divergence of the Dof gene families in poplar, *Arabidopsis*, and rice suggests multiple modes of gene evolution after duplication. *Plant Physiology* **142,** 820–830.

Ye Q, Zhu W, Li L, Zhang S, Yin Y, Ma H, Wang X. 2010. Brassinosteroids control male fertility by regulating the expression of key genes involved in *Arabidopsis* anther and pollen development. *Proceedings of the National Academy of Sciences, USA* **107,** 6100–6105.

Yu Y, Hu R, Wang H, Cao Y, He G, Fu C, Zhou G. 2013. MlWRKY12, a novel *Miscanthus* transcription factor, participates in pith secondary cell wall formation and promotes flowering. *Plant Science* **212,** 1–9.

Zhang L, Zhao G, Jia J, Liu X, Kong X. 2012. Molecular characterization of 60 isolated wheat MYB genes and analysis of their expression during abiotic stress. *Journal of Experimental Botany* **63,** 203–214.

Zhao J, Zhang W, Zhao Y, Gong X, Guo L, Zhu G, Wang X, Gong Z, Shumaker KS, Guo Y. 2007. SAD2, an importin β-like protein, is required for UV-B response in *Arabidopsis* by mediating MYB4 nuclear trafficking. *The Plant Cell* **19,** 3805–3818.

Zhong R, Demura T, Ye ZH. 2006. SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of *Arabidopsis*. *The Plant Cell* **18,** 3158–3170.

Zhong R, Lee C, Zhou J, McCarthy R, Ye ZH. 2008. A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. *The Plant Cell* **20,** 2763–2782.

Zhong R, McCarthy R, Haghighat M, Ye ZH. 2013. The poplar MYB master switches bind to the SMRE site and activate the secondary wall biosynthetic program during wood formation. *PLoS ONE* **8,** e69219.

Zhong R, Richardson EA, Ye ZH. 2007. The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in *Arabidopsis*. *The Plant Cell* **19,** 2776–2792.

Zhou J, Lee C, Zhong R, Ye ZH. 2009. MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in *Arabidopsis*. *The Plant Cell* **21,** 248–266.