JSE Journal of Systematics and Evolution

doi: 10.1111/jse.12595

Research Article

Functional divergence of *Populus MYB158* and *MYB189* gene pair created by whole genome duplication

Peng-Fei Jiang[1,2], Hui Xu[1,2], Chao-Nan Guan[3,4], Xiao-Xia Wang[3,4], Ai-Min Wu[5], Yan-Jing Liu[3], and Qing-Yin Zeng[3]*

[1]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Beijing 100091, China
[4]College of Biological Sciences and Biotechnology, Beijing Forestry University, Beijing 100083, China
[5]Guangdong Key Laboratory for Innovative Development and Utilization of Forest Plant Germplasm, College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou 510642, China
*Author for correspondence. E-mail: qingyin.zeng@ibcas.ac.cn
Received 28 February 2020; Accepted 23 April 2020; Article first published online 27 April 2020

Abstract Whole genome duplication (WGD) provides new genetic material for genome evolution. After a WGD event, some duplicates are lost, while other duplicates still persist and evolve diverse functions. A particular challenge is to understand how this diversity arises. This study identified two WGD-derived duplicates, *MYB158* and *MYB189*, from *Populus tomentosa*. *Populus MYB158* and *MYB189* had expression divergence. *Populus tomentosa* overexpressing *MYB158* or *MYB189* had similar phenotypes: creep growth, decreased width of xylem and secondary cell wall thickness. Compared to wild-type, neither *myb158* mutant nor *myb158 myb189* double mutant showed obvious phenotypic variation in *P. tomentosa*. Although MYB158 and MYB189 proteins could repress the same structural genes involved in lignin, cellulose, and xylan biosynthesis, the two proteins had their own specific regulatory targets. *Populus MYB158* could act as the upstream regulator of secondary cell wall NAC master switch and directly represses the expression of the *SND1-B2* gene. Taken together, *Populus MYB158* and *MYB189* have retained similar functions in negatively regulating secondary cell wall biosynthesis, but have evolved partially distinct functions in direct regulation of NAC master switch, with *MYB158* playing a more crucial role. Our findings provide new insights into the evolutionary and functional divergence of WGD-derived duplicate genes.

Key words: duplicate gene, functional divergence, MYB transcription factor, regulatory networks, secondary cell wall biosynthesis, whole genome duplication.

1 Introduction

Gene duplication is one of the main driving forces in the evolution of genomes and genetic systems. The accumulation of genomic data revealed that gene duplication was common in plant genomes. In particular, whole-genome duplication (WGD) was ubiquitous in plants. All angiosperms shared at least two WGD events in their evolutionary history (Jiao et al., 2011). Many plant species have undergone recent WGD events, such as soybean, cotton, and wheat (Schranz et al., 2012). Immediately following a WGD, all the genes that previously existed in the genome are present in duplicate. Most duplicate genes created by WGD will be lost over the course of subsequent genome evolution (Liu et al., 2015). However, some duplicate genes still persist. Why were some polyploidy-derived duplicate genes retained? The neofunctionalization or subfunctionalization models are frequently used to explain the retention of duplicate genes created by WGD (Force et al., 1999). In the neofunctionalization model, one copy mutates to a function that does not exist in the preduplication gene, while another copy retains the original function. In the subfunctionalization model, each duplicate copy partitions the ancestral gene function. Although many studies showed that functional divergences (subfunctionalization or neofunctionalization) have indeed taken place between the retained duplicates formed by WGD (Blanc & Wolfe, 2004; Moghe et al., 2014), evolutionary mechanisms responsible for the retention and functional divergence of duplicate genes remain largely unknown.

Plant cells have two types of cell wall, primary cell wall and secondary cell wall. Secondary cell wall conducts water and/or provide mechanical support, such as xylem vessels and fibers (Nakano et al., 2015). A hierarchical network of transcription factors, including mainly NAC and MYB family members, has been proposed to regulate secondary cell wall biosynthesis (Zhong et al., 2008, 2010a). The secondary cell wall-related NAC transcription factors function as master switches that activated the process of the secondary cell wall biosynthesis (Zhong et al., 2010a). These NAC genes were also named secondary wall-associated NAC domain (SNDs) or vascular-related NAC domain (VNDs) (Kubo et al., 2005; Zhong et al., 2006). SNDs play a specific role in fiber cell wall

differentiation, and VNDs activate the formation of vessel cell walls. Most of the secondary wall-related MYB transcription factors act downstream of NAC transcription factors, and function as the second layer of master switches to regulate downstream genes in secondary wall formation (Wang & Dixon, 2012). For example, *Arabidopsis MYB46* was a direct target of *SND1* in secondary wall formation (Zhong et al., 2007). Repression of *MYB46* results in a drastic reduction in secondary cell wall thickening in fibers and vessels, and overexpression of *MYB46* causes ectopic deposition of secondary cell walls in *Arabidopsis* (Zhong et al., 2007, 2008). However, in the secondary cell wall formation, transcription factors that directly regulate the *SND* or *VND* genes are rarely reported. The identification of transcription factors directly regulating *SND* or *VND* genes facilitates understanding of the molecular mechanisms of secondary cell wall biosynthesis.

To identify transcription factors related to secondary cell wall biosynthesis, we undertook coexpression analyses between 16 secondary cell wall-related NAC transcription factors and all *Populus MYB* genes based on the AspWood database (Sundell et al., 2017). An R2R3-MYB transcription factor, *MYB158*, was identified as a key candidate regulator responsible for secondary cell wall formation. In addition, we found that *Populus MYB158* and *MYB189* were a duplicate gene pair created by the Salicaceae WGD event. Based on functional analyses, this study showed functional divergence between *Populus* duplicated *MYB158* and *MYB189* genes, and found that *Populus MYB158* directly negatively regulates *SND1* gene involved in secondary cell wall formation.

2 Materials and Methods

2.1 Coexpression and phylogenetic analysis

Coexpression analyses between secondary cell wall associated NACs and all MYB genes in *Populus trichocarpa* (Torr. & Gray) were undertaken using the AspWood tool (http://aspwood.popgenie.org/aspwood-v3.0/) with a threshold value of 4 (Sundell et al., 2017). Coexpressed genes were linked by lines as in Fig. 1. All the members of the MYB family of *Arabidopsis thaliana* (thale cress) and *P. trichocarpa* were downloaded from PlantTFDB 4.0 database (http://planttfdb.cbi.pku.edu.cn/). The amino acid sequences were aligned using the MAFFT program (https://www.ebi.ac.uk/Tools/msa/mafft/). The optimal substitution model of amino acid substitution was JTT + G selected by the ModelGenerator version 0.84 program (Keane et al., 2006). The phylogenetic tree was constructed by the maximum-likelihood procedure in RAxML software. One hundred replicates were undertaken in each analysis to obtain the confidence support.

2.2 Plant material and grown conditions

Populus tomentosa Carr. (Clone 741) was cultured in woody plant medium (WPM) (pH 5.8) supplemented with 1.5% (w/v) sucrose, and rooted plants were grown in the glasshouse at 25 °C under a 14:10-h light:dark cycle with supplemental light (4500 lux). Tobacco was cultivated in potting soil at 22 °C in a growth chamber (16:8 h, light:dark).

2.3 Molecular cloning of *P. tomentosa MYB158* and *MYB189*

Total RNA was isolated from *P. tomentosa* different tissues (including bark, stem, root, petiole, xylem, and leaf) using RNAprep Pure Plant Kit (Tiangen, Beijing, China). Total RNA was then treated with RNase-free DNase I (Promega, Madison, WI, USA) and reverse transcribed into cDNA with the RNA PCR Kit (AMV) version 3.0 (Takara, Dalian, China). The full-length cDNA fragment of *MYB158* and *MYB189* were amplified with gene-specific primers (Table S1) designed according to the reference genomic sequence of *P. trichocarpa*. The PCR products were amplified using ExTaq polymerase (Takara) and cloned into pEASY-Blunt vector (TransGen, Beijing, China) for sequencing.

2.4 RNA extraction and real-time quantitative polymerase chain reaction analysis

Total RNA was isolated from bark, root, stem, xylem, leaf, and petiole tissues of 3-month-old *P. tomentosa* using RNAprep Pure Plant Kit (Tiangen). Among which, the total RNA of stem and xylem was extracted from the 4th to 8th internodes, and the total RNA of leaf was isolated from the 4th leaves. Total RNA was then reverse transcribed into cDNA as described above. For real-time quantitative polymerase chain reaction (RT-qPCR), primers that specially anchored to the differential regions between *MYB158* and *MYB189* were designed (Table S1). The PrimerScript RT Reagent Kit (Perfect Real Time; Takara) was used to carry out RT-qPCR with ROX as a reference dye in a StepOne Plus Real-Time PCR instrument (Applied Biosystems, Foster city, CA, USA). The relative expression level of the target gene was calculated by the $2^{-\Delta\Delta CT}$ method (Pfaffl, 2001) and the *Populus* actin gene (GenBank No. XM_002316253) was used as an internal control. Three biological replicates were carried out for each gene.

2.5 Construction of *35S:MYB158* and *35S:MYB189* vectors and generation of transgenic *P. tomentosa* trees

The full coding sequences of *MYB158* and *MYB189* were cloned into a modified *pCAMBIA1302* vector (Δ*pCAMBIA1302*, Fig. S1), respectively, to generate *35S:MYB158* and *35S:MYB189* constructs. These overexpression constructs were introduced into *Agrobacterium tumefaciens* strains EHA105 by freeze-thaw method. The explants, including leaf discs and stem segments of *P. tomentosa*, were infected with *Agrobacterium tumefaciens* carrying the recombinant plasmids. Subsequently, the leaf and stem explants were cultivated on WPM medium containing 100 μmol/L acetosyringone, 1 mg/L 1-naphthaleneacetic acid (NAA), 2 mg/L zeatin, and 0.6% (w/v) agar for 48 h in the dark at 24 °C. These explants were then transferred onto callus-inducing medium (WPM supplemented with 1 mg/L NAA, 2 mg/L zeatin, 10 mg/L hygromycin, 200 mg/L cefotaxime, and 0.6% (w/v) agar). After cultivation in darkness for 3–4 weeks, during which the medium was refreshed every 10 days, the calli were visible from the cut edges of the explants. Subsequently, the explants were transferred to selective medium (WPM supplemented with 0.1 mg/L NAA, 2 mg/L zeatin, 10 mg/L hygromycin, 200 mg/L cefotaxime and 0.6% (w/v) agar) to induct and select the transformed shoots under light. The generated shoots were excised from calli when they were approximately 1 cm long, and were transferred to rooting medium (WPM supplemented with 0.1 mg/L indole-3-butyric acid, 10 mg/L

hygromycin, 200 mg/L cefotaxime, and 0.6% (w/v) agar). The identification of positive transformed plants was undertaken by PCR amplification of the *HPT* gene using gene-specific primers (Table S1).

2.6 Isolation and analysis of *MYB158* and *MYB189* promoters

The promoter fragments (2300 bp) of *MYB158* and *MYB189* were isolated from *P. tomentosa* genomic DNA using the specific primers, which were designed based on the reference genomic sequence of *P. trichocarpa* (Table S1). These two promoter fragments were cloned into Δ*pCAMBIA1302* vectors by replacing the cauliflower mosaic *35S* promoters. Subsequently, the complete *GUS* coding sequences were inserted into the *Kpn*I and *Pml*I sites of Δ*pCAMBIA1302* vectors to generate *MYB158pro:GUS* and *MYB189pro:GUS* constructs. These constructs were transformed into *Agrobacterium tumefaciens* strain EHA105 for *P. tomentosa* transformation. Histochemical GUS assays of transgenic seedlings were carried out as described by Jefferson et al. (1987).

2.7 Microscopic and transmission electron microscopy analysis

Stem segments were sampled from the ninth internode of 2-month-old OE-*MYB158*, OE-*MYB189* and wild-type trees, and then fixed in 0.1 M phosphate buffer (PB) (pH 7.2) containing 3% glutaraldehyde with vacuum infiltration. Samples were then thoroughly washed with 0.1 M PB and fixed by 1% osmic acid for 2 h. After extensive rinsing with 0.1 M PB, samples were dehydrated in a graded ethanol series (10, 30, 50, 70, 80, 95, and 100%). The samples were treated with ethanol/acetone (1:1, v/v) for 30 min, and then were treated with pure acetone for another 30 min. Samples were progressively infiltrated with acetone/resin of 3:1, 1:1, and 3:1 (v/v) for 4 h each, then replaced with pure resin for more than 8 h. Finally, samples were embedded in pure resin and polymerized at 60 °C for more than 24 h. Semi-thin cross-sections (1 μm) were cut from embedded blocks using a Leica Ultra-Cut microtome (Leica, Wetzlar, Germany) equipped with a glasscutter. After staining with 0.2% toluidine blue solution, sections were analyzed using Zeiss Axio Imager A1 microscopy (Carl Zeiss, Oberkochen, Germany). Ultra-thin cross sections (70 nm) were cut from a diamond knife on a Leica Ultra-Cut microtome and observed with an HT7700 transmission electron microscope (Hitachi, Tokyo, Japan) at 80 kV. Cell wall thickness was measured from transmission electron micrographs by software. Two transgenic lines for each gene (L5, L7 for OE-*MYB158* and L5, L11 for OE-*MYB189*) were used to measure the cell wall thickness, and each line with two independent plants. Finally, for each gene, the cell wall thickness of at least 250 xylem fiber cells and 35 vessel cells from transgenic lines and wild-type plants were measured and analyzed, respectively.

2.8 Measurement of cellulose, hemicellulose, and lignin contents

Stem tissues of OE-*MYB158*, OE-*MYB189*, and wild-type trees were dried and then finely ground. Dewaxing powders were extracted from crude powders with toluene-ethanol (2:1, v/v) in Soxhlet for 6 h at reflux (92 °C). Dewaxing powders were subjected to sulfuric acid hydrolysis as specified in standard TappiT222 om-02 for acid-insoluble lignin. Acid-soluble lignin was measured by absorption of UV radiation ($\varepsilon 205 = 110$ L $(gcm)^{-1}$). Dewaxing powders were delignified in sodium chlorite (pH 4.5, 76 °C) for 4 h, leaving holocelluloses (hemicellulose and cellulose). The cellulose content was examined following the Kurschne–Hofner method. Dewaxing powders (1 g, dry weight) were hydrolyzed in nitric acid-ethanol solution and heated with boiling water at the same time. The nitric acid-ethanol solution was refreshed several times until the fibers whitened. Subsequently, the cellulose was washed, dried, and weighed. The hemicellulose content was calculated from the difference between holocellulose and cellulose content.

2.9 Determination of monosaccharide composition

Alcohol insoluble residues were prepared by extraction of the stem powder with 96% (v/v) ethanol at 70 °C for 30 min. Alcohol insoluble residues were hydrolyzed for 2 h at 120 °C with 2N trifluoroacetic acid. Separation and quantification of sugars was carried out by high performance anion exchange chromatography with pulsed amperometric detection as described by Øbro et al. (2004).

2.10 Fourier transform infrared analysis

The KBr disk standard technique was used to prepare the samples for infrared detection. Briefly, KBr pellets of the samples were obtained by mixing 1 mg (dry weight) of stem powders with 100 mg KBr. Then the mixture was squeezed for 1–2 min using a tablet press machine with 10 ton pressure. Fourier transform infrared (FT-IR) spectroscopy was carried out on a Bruker Tensor 27 spectrometer (Bruker, Ettlingen, Germany) at the transmittance mode.

2.11 RNA sequencing analysis of overexpressors

Total RNAs were isolated from the stems (1st to 8th internodes) of 2-month-old OE-*MYB158*, OE-*MYB189*, and wild-type trees following the method described above. RNA concentration was measured using a Qubit RNA Assay Kit in a Qubit 2.0 Flurometer (Life Technologies, Carlsbad, CA, USA) and integrity using RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, Palo Alto, CA, USA). Three overexpressed lines for each gene (L5, L7, L19 for OE-*MYB158* and L5, L11, L20 for OE-*MYB189*) and three wild-type trees were used for construction of cDNA libraries, and each transgenic line with three biological replicates. These cDNA libraries were sequenced by the Illumina HiSeq 2500 platform (Novogene, Beijing, China). Raw sequence reads corresponding to OE-*MYB158* and OE-*MYB189* trees have been submitted to the National Center for Biotechnology Information Sequence Read Archive (accession Nos. PRJNA579854 and PRJNA582494).

The raw reads were first filtered to obtain clean reads and then mapped to the *P. trichocarpa* genome using Hisat2 version 2.0.4. HTSeq version 0.9.1 was used to counter the read numbers mapped to each gene, and FPKM (expected number of fragments per kilobase of transcript sequence per millions base pairs sequenced) was used to estimate gene expression levels. Differential expression analysis between transgenic lines and wild-type trees was carried out using the DESeq R package (1.18.0). Genes with an adjusted *P*-value <0.05 identified by DESeq were assigned as differentially expressed. We finally chose the differentially expressed

genes (DEGs) shared by the three transgenic lines of each gene for analysis.

Gene Ontology (GO) enrichment analysis was carried out using the Cytoscape plugin ClueGO with an adjusted *P*-value <0.05. For this, the closest *Arabidopsis* homologues of all DEGs were identified based on the AspWood database (http://aspwood.popgenie.org/aspwood-v3.0/), and the corresponding gene list used as the input. Heat maps were plotted by HemI software.

2.12 CRISPR/Cas9-mediated mutagenesis of *MYB158* and *MYB189* in *P. tomentosa*

The full-length DNA sequences of *MYB158* and *MYB189* were analyzed in the online tool CRISPRdirect (http://crispr.dbcls.jp/) (Naito et al., 2014). Five putative target sites (T1–5) located at exon regions of *MYB158* and *MYB189* were selected for designing single guide RNA (sgRNA) sequences. Among which, T1 and T2 were unique to *MYB158*, T3 was unique to *MYB189*, and T4 and T5 were common to *MYB158* and *MYB189*. Five pairs of oligos (Table S1) harboring the sgRNA target sequence were annealed and subsequently introduced into the *Bsa*I sites of pYLsgRNA-AtU3d/LacZ or pYLsgRNA-AtU6-29 to generate sgRNA cassettes. These sgRNA cassettes were assembled into binary *pYLCRISPR/Cas9* vectors. Ultimately, three CRISPR/Cas9 vectors (I, II, and III) were constructed. Among them, vector I, harboring T1 and T2, targeted *MYB158*. Vector II, harboring T3, targeted *MYB189*. Vector III, harboring T4 and T5, targeted both *MYB158* and *MYB189*. These CRISPR/Cas9 vectors were transformed into *Agrobacterium tumefaciens* strains EHA105 for *P. tomentosa* transformation.

To determine the mutation pattern of target genes in transgenic plants, gene-specific primers (Table S1) spanning sgRNA target sites were used for PCR amplification of genomic DNA extracted from stems of wild-type and gene-edited plants. The PCR products were cloned into pEASY-Blunt vectors (TransGen) and more than 40 clones for each transgenic line were randomly selected for sequencing.

The 10th internodes of 3-month-old *P. tomentosa* knockout mutants were sectioned for wood morphology analysis. For each mutant, at least 200 xylem cells and 50 vessel cells from two independent plants were measured and analyzed.

2.13 Dual luciferase assays

Luciferase promoter activation assays were undertaken using *Agrobacterium* transient transformation of *Nicotiana benthamiana* leaves as previously described by Hellens et al. (2005). The promoter of *SND1-B2* (1500 bp) was amplified from *P. tomentosa* genomic DNA and cloned into the pGreenII 0800-LUC vector. The full coding sequences of *MYB158* and *MYB189* were inserted into pGreenII 62-SK vector. Empty vector pGreenII 62-SK acted as a control. The firefly and *Renilla* luciferase activity was quantified using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions. Primers for all constructs are listed in Table S1.

2.14 Chromatin immunoprecipitation–qPCR assay

The full coding sequences of *MYB158* and *MYB189* were cloned into the pCAMBIA1305-FLAG vectors. The resulting constructs were introduced into *P. tomentosa* and *35S:MYB158-FLAG* and *35S:MYB189-FLAG* transgenic plants were used for chromatin immunoprecipitation (ChIP) assay. Primer sequences used for cloning are listed in Table S1.

The ChIP assays were carried out as described by Saleh et al. (2008). The stem tissues of 2-month-old wild-type and transgenic plants were harvested and cross-linked with 1% formaldehyde for chromatin isolation. Monoclonal anti-FLAG antibody (Sigma-Aldrich) was used for immunoprecipitation. Both input samples and immunoprecipitated samples were analyzed by ChIP-qPCR. *Actin* was amplified as a negative control. Three biological replicates were carried out with similar results. The primer sequences used for enrichment analysis are listed in Table S1.

3 Results

3.1 *Populus MYB158* could be involved in biosynthesis of secondary cell walls

A previous study identified 16 NAC transcription factors involved in the biosynthesis of secondary walls in *P. trichocarpa* (Sundell et al., 2017). We undertook coexpression analyses between these 16 NAC transcription factors and all MYB genes from *P. trichocarpa* based on the AspWood database (http://aspwood.popgenie.org/aspwood-v3.0/). The AspWood database showed that 19 *P. trichocarpa* MYB genes were closely adjacent to these 16 NAC genes (Fig. 1A) (Z-score threshold of four). Among these 19 MYB genes, *MYB011* and *MYB158* were adjacent to five and four NAC genes, respectively, and other MYB genes were adjacent to less than four NAC genes (Fig. 1A). The expression patterns of *MYB011* and *MYB158* genes in *P. trichocarpa* stems were further examined using the AspWood database. *MYB011* was not expressed in the expanded xylem (Fig. 1B), indicating that *MYB011* might not be involved in secondary cell wall biosynthesis. However, *MYB158* was expressed throughout the xylem (Fig. 1B), indicating that *MYB158* might be a novel regulator of secondary cell wall biosynthesis.

3.2 *Populus MYB158* and *MYB189* are a duplicate gene pair created by the Salicaceae WGD event

We undertook a joint phylogenetic analysis of MYB gene families from *P. trichocarpa* and *A. thaliana*. Phylogenetic tree showed that *P. trichocarpa MYB158* and *MYB189* genes were grouped together with 100% bootstrap support (Fig. 2A). *P. trichocarpa MYB158* and *MYB189* genes were located in chromosomes 5 and 2, respectively. *P. trichocarpa MYB158* and *MYB189* are each located in a pair of paralogous segments created by the WGD event in Salicaceae (Fig. 2B). Collinearity analysis showed that the flanking regions of *P. trichocarpa MYB158* and *MYB189* contained more than five paralogous gene pairs created by the Salicaceae WGD provided by PLAZA version 4.0 (Fig. 2C). These analyses indicated that *P. trichocarpa MYB158* and *MYB189* were a duplicate gene pair created by the Salicaceae WGD event.

This study used several strategies to identify orthologous genes of *P. trichocarpa MYB158* in *P. tomentosa*. First, using *P. trichocarpa MYB158* as templates, we designed up to six primer pairs to amplify the *MYB158* gene from *P. tomentosa*. Second, we amplified *MYB158* from different *P. tomentosa* tissues including bark, stem, root, petiole, xylem, and leaf

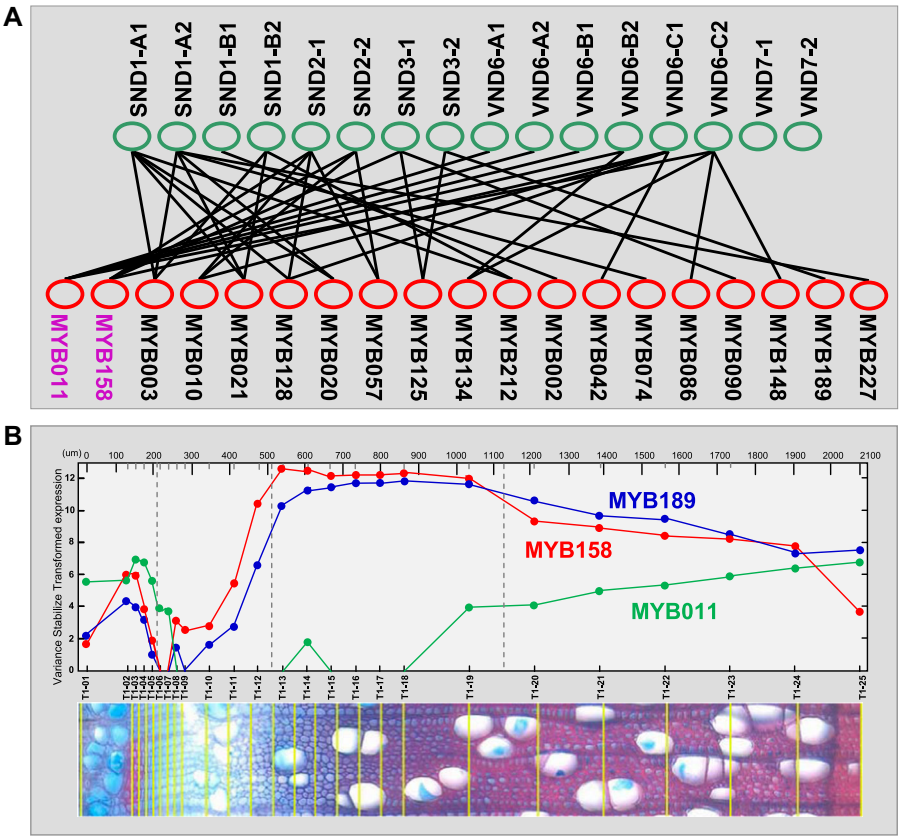

Fig. 1. Identification of secondary cell wall-related MYB transcription factors. **A,** Coexpression analyses between these *Populus* NAC and MYB transcription factors based on the AspWood database. **B,** Expression patterns of *Populus MYB158*, *MYB189*, and *MYB011*.

tissues. Third, we also amplified *MYB158* from *P. tomentosa* using total genomic DNA as a template. Finally, all PCR products amplified from RNA or genomic DNA were cloned into pEASY-Blunt vector and fully sequenced. As a result, we only amplified the *P. tomentosa MYB158* without amplifying any other DNA sequences, indicating that the *MYB158* gene did not expand in *P. tomentosa*. The protein sequences of *P. tomentosa MYB158* and *P. trichocarpa MYB158* showed 97.6% identity. Phylogenetic tree of MYB gene families from *P. trichocarpa* and *A. thaliana* showed *P. tomentosa MYB158* and *P. trichocarpa MYB158* were grouped together with high bootstrap support (Fig. S2). Based on the above results, *P. tomentosa MYB158* and *P. trichocarpa MYB158* should be orthologous genes. Using the same method, we found that *P. tomentosa MYB189* and *P. trichocarpa MYB189* are also orthologous genes.

We proposed that *P. tomentosa MYB158* and *MYB189* genes were a duplicate gene pair created by the Salicaceae WGD event for the following reasons: (i) *P. tomentosa MYB158* and *P. trichocarpa MYB158* are orthologous genes, as are *P. tomentosa MYB189* and *P. trichocarpa MYB189*; (ii) *P. trichocarpa MYB158* and *MYB189* genes were a duplicate gene pair created by the Salicaceae WGD event; and (iii) all *Populus* species shared a WGD event that occurred in the ancestors of Salicaceae, and after this WGD event, no polyploid events occurred in all *Populus* species.

In the following sections, this study focused on the functional analysis of *P. tomentosa MYB158* and *MYB189* genes.

3.3 Differential expression patterns of *MYB158* and *MYB189* in *P. tomentosa*

The expression patterns of *P. tomentosa MYB158* and *MYB189* genes were investigated by RT-qPCR. *MYB158* and *MYB189* genes were expressed in the bark, root, stem, xylem, leaf, and petiole tissues of *P. tomentosa* (Fig. 3A). Among these six tissues, the *MYB158* gene showed the highest expression level in xylem tissue, whereas *MYB189* gene showed the highest expression in root tissue. Compared to the *MYB189* gene, *MYB158* had a higher expression level in bark, stem, xylem, leaf, and petiole tissues, and had a lower expression level in root tissue.

To further investigate the expression patterns of *MYB158* and *MYB189* genes, we generated transgenic *P. tomentosa* trees that express GUS protein driven by *MYB158* and *MYB189* promoters, respectively. GUS staining showed that *MYB158* and *MYB189* were expressed in shoot tip, stem, leaf, and petiole (Fig. 3B). Compared to *MYB189pro:GUS* transgenic trees, *MYB158pro:GUS* transgenic trees showed stronger GUS staining in stem and leaf tissues. The cross-section of internode 2 showed that *MYB158* and *MYB189* were mainly

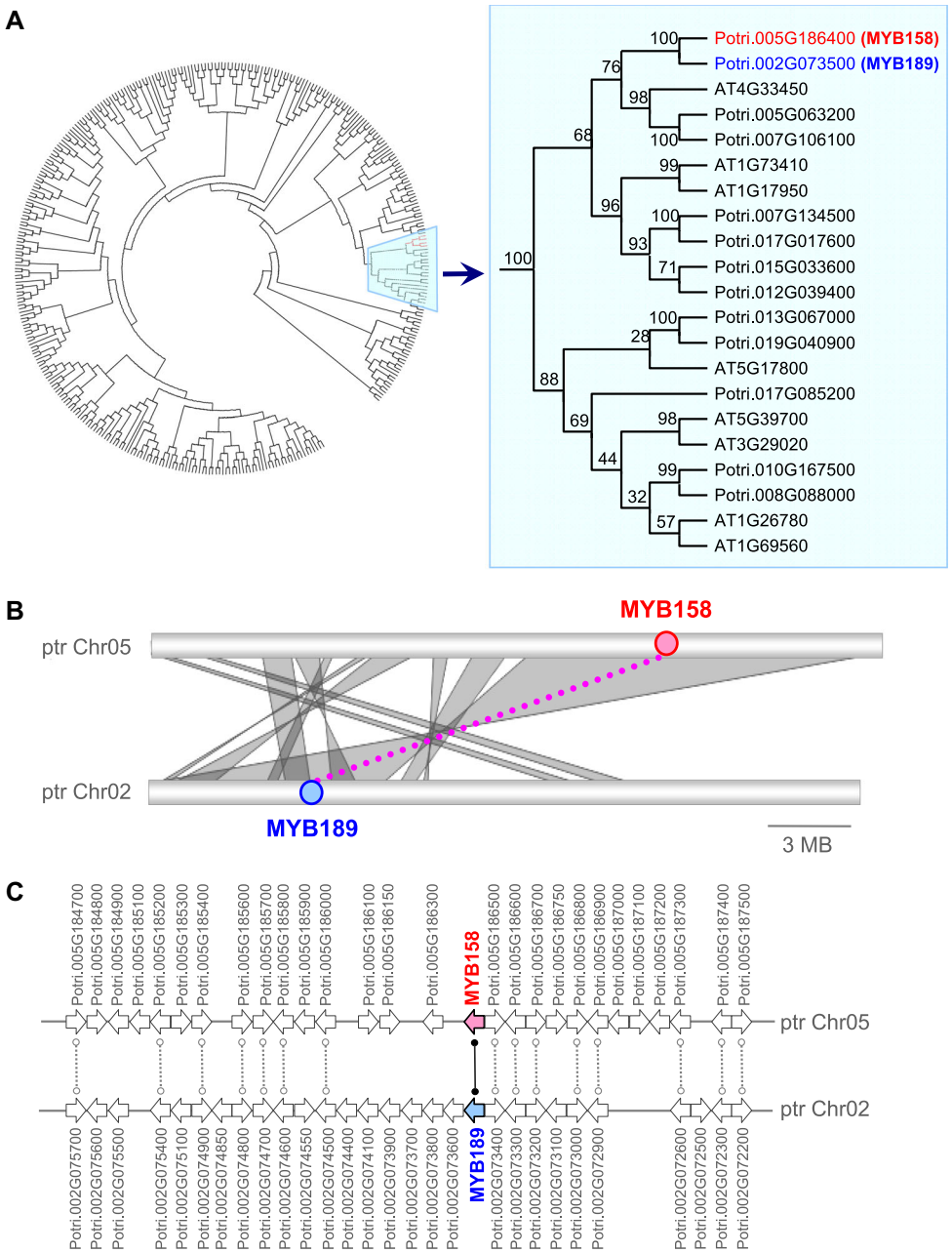

Fig. 2. Identification of duplicated *MYB158* and *MYB189* gene pair created by a whole-genome duplication event. **A,** Phylogenetic tree of MYB gene families from *Populus* and *Arabidopsis.* **B,** Genomic locations of *Populus MYB158* and *MYB189* genes. **C,** Collinearity analysis of *Populus MYB158* and *MYB189* genes.

expressed in xylem tissues (Figs. 3C, 3D). Compared to *MYB189pro:GUS* transgenic trees, *MYB158pro:GUS* transgenic trees showed a stronger GUS staining in xylem tissues (Figs. 3C, 3D).

3.4 Overexpression of *MYB158* or *MYB189* causes pleiotropic phenotypes in transgenic *P. tomentosa*

To investigate the biological functions of *P. tomentosa MYB158* and *MYB189* genes, this study generated transgenic *P. tomentosa* trees overexpressing *MYB158* (*35S:MYB158*) and *MYB189* (*35S:MYB189*), respectively. For *35S:MYB158* and *35S:MYB189* transgenic trees, we obtained at least 15 transgenic lines. The RT-qPCR analyses showed the 35S promoter-driven mRNA accumulations of *MYB158* and *MYB189* in the stems of corresponding transgenic *P. tomentosa* trees (Fig. 4A).

All transgenic *P. tomentosa* trees overexpressing *MYB158* (OE-*MYB158*) or *MYB189* (OE-*MYB189*) were unable to grow straight (Fig. 4B). After 2 months of growth, we measured the tree height, stem diameter, and stem internode length of transgenic trees. The heights of OE-*MYB158* and OE-*MYB189* trees were significantly lower than that of wild-type

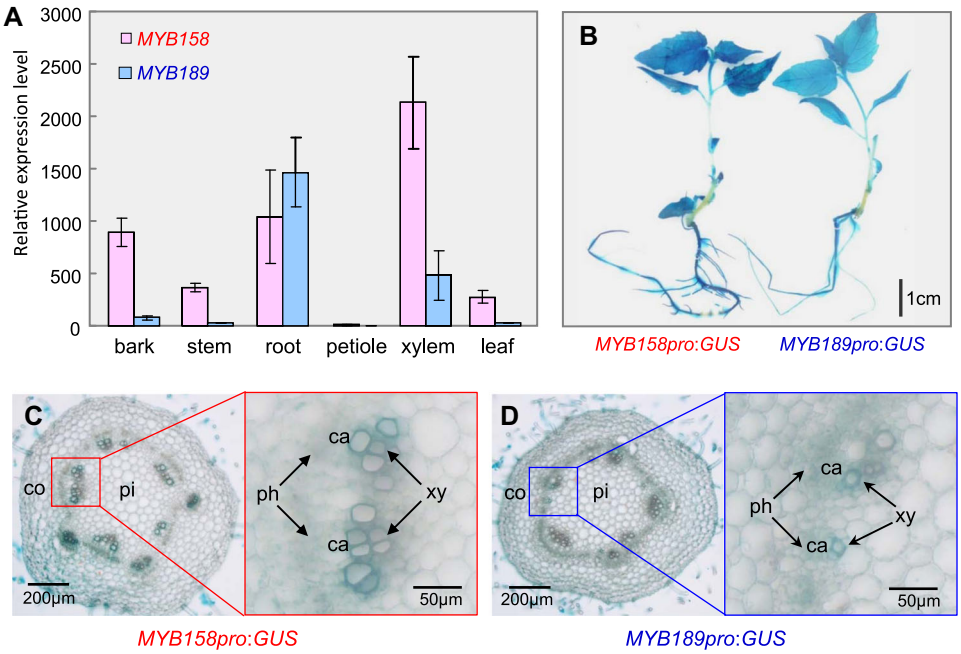

Fig. 3. Expression patterns of *MYB158* and *MYB189* genes in *P. tomentosa*. **A,** Relative expression levels of *MYB158* and *MYB189* genes in *P. tomentosa*. **B,** Histochemical staining of GUS activity in transgenic *P. tomentosa* trees harboring *MYB158pro:GUS* (left image) and *MYB189pro:GUS* (right image). **C, D,** Detection of GUS activity in the stems of *MYB158pro:GUS* transgenic *Populus* and *MYB189pro:GUS* transgenic *Populus*, respectively. ca, cambium; co, cortex; ph, primary phloem; pi, pith; xy, xylem.

(independent sample *t*-test, $P < 0.001$) (Fig. 4C). Compared to wild-type, the average heights of OE-*MYB158* and OE-*MYB189* trees were reduced by 26% and 32%, respectively. The stem diameters of OE-*MYB158* and OE-*MYB189* trees were significantly smaller than that of wild-type (independent sample *t*-test, $P < 0.002$) (Fig. 4D). Compared to wild-type, the average stem diameters of OE-*MYB158* and OE-*MYB189* trees were reduced by 27% and 21%, respectively. The internode lengths of OE-*MYB158* and OE-*MYB189* trees were significantly shorter than that of wild-type (independent sample *t*-test, $P < 0.001$) (Figs. 4E, 4F). We did not observe significant differences in height, stem diameter, or stem length between OE-*MYB158* and OE-*MYB189* trees (independent sample *t*-test, $P > 0.18$).

3.5 Overexpression of *MYB158* or *MYB189* represses secondary cell wall thickening in transgenic *P. tomentosa*

The 9th internode of the stem was sectioned to examine the wood morphology. The xylem widths of OE-*MYB158* and OE-*MYB189* trees were significantly narrower than that of wild-type (independent sample *t*-test, $P < 0.001$) (Figs. 5A, 5C). There was no significant difference in the width of xylems between OE-*MYB158* and OE-*MYB189* trees (independent sample *t*-test, $P > 0.61$). The vessel cell wall thickness of the OE-*MYB158* and OE-*MYB189* trees was significantly smaller than that of the wild-type (independent sample *t*-test, $P < 0.002$) (Figs. 5B, 5C). We did not observe significant differences between the vessel cell wall thickness of OE-*MYB158* and OE-*MYB189* trees (independent sample *t*-test, $P > 0.61$). The thickness of xylem fiber cell walls of OE-*MYB158* and OE-*MYB189* trees was significantly smaller than that of the wild-type (independent sample *t*-test, $P < 0.001$), which were 40% and 33% smaller, respectively (Figs. 5B, 5C). OE-*MYB158* trees showed much thinner xylem fiber cell wall thickness than OE-*MYB189* trees (independent sample *t*-test, $P < 0.005$) (Figs. 5B, 5C).

3.6 Overexpression of *MYB158* or *MYB189* changes cell wall compositions in transgenic *P. tomentosa*

This study examined the cell wall compositions in the stems of OE-*MYB158*, OE-*MYB189*, and wild-type trees. Compared to wild-type, the hemicellulose and lignin content in the stems of OE-*MYB158* and OE-*MYB189* trees were significantly decreased (Table 1). The cellulose content in the stems of the OE-*MYB158* trees was significantly reduced compared to the wild-type trees (Mann–Whitney *U*-test, $P < 0.05$). We did not observe a significant difference in cellulose content between OE-*MYB189* and wild-type tree stems (Mann–Whitney *U*-test, $P > 0.05$). Compared to wild-type, the content of glucose and xylose in the stems of OE-*MYB158* and OE-*MYB189* trees were significantly decreased (Mann–Whitney *U*-test, $P < 0.05$) (Table 2). However, the content of fucose, arabinose, and galactose in the stems of OE-*MYB158*, OE-*MYB189*, and wild-type trees did not show significantly difference (Mann–Whitney *U*-test, $P > 0.05$).

This study used FT-IR spectroscopy to detect changes in the chemical structures of cell wall components in transgenic trees. Compared to wild-type, lower absorbance at 1517 cm^{-1}, 1462 cm^{-1}, 1165 cm^{-1}, and 1045 cm^{-1} were found in the spectra of OE-*MYB158* and OE-*MYB189* trees (Fig. 6). The peak at 1517 cm^{-1} corresponds to C = C aromatic symmetrical

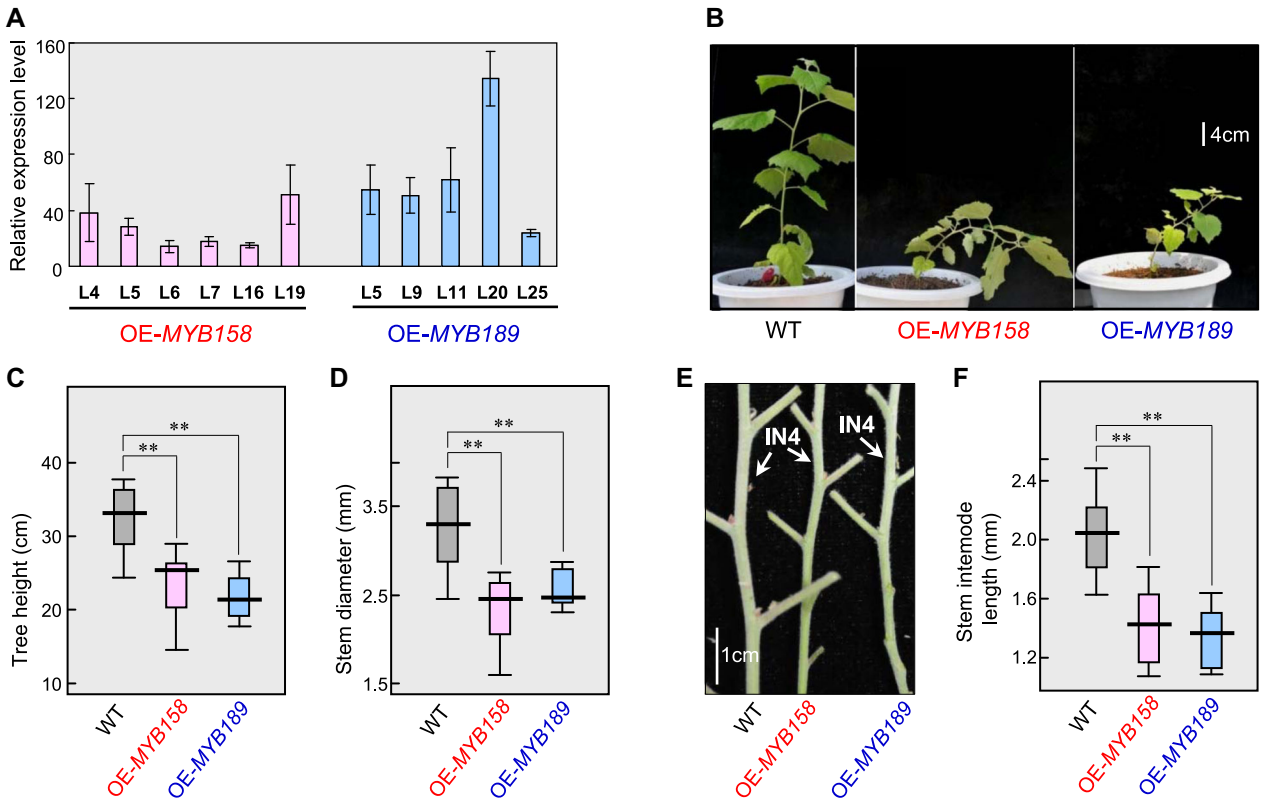

Fig. 4. Phenotypes and gene expression in transgenic *MYB158* or *MYB189* trees. **A,** Gene expression of *MYB158* and *MYB189* in the stems of the corresponding transgenic lines. **B,** Two-month-old wild-type, OE-*MYB158*, and OE-*MYB189* trees. **C–F,** Comparison of tree height (C), stem diameter (D), and stem internode length (E, F) of wild-type, OE-*MYB158*, and OE-*MYB189* trees. $^{**}P < 0.01$.

stretching in lignin. The peak at 1462 cm^{-1} is mainly assigned to C–H deformation in lignin. The peak at 1165 cm^{-1} is a typical ester bond absorption peak. The peak at 1045 cm^{-1} is assigned to C–O and C–C stretch in hemicelluloses. Compared to wild-type, the lower absorbance peaks at 1517 cm^{-1}, 1462 cm^{-1}, 1165 cm^{-1}, and 1045 cm^{-1} indicated reduced contents of hemicellulose and lignin in OE-*MYB158* and OE-*MYB189* trees. Except for the changes in absorption peak intensity, we did not observe any change in the infrared absorption spectra of the OE-*MYB158* and OE-*MYB189* trees compared to the wild-type. These results indicated that overexpression of *MYB158* or *MYB189* in *P. tomentosa* had no significant effect on the chemical structure of cell wall components.

3.7 Overexpression of *MYB158* or *MYB189* leads to transcriptomic reprogramming involved in secondary cell wall formation

This study compared the stem transcriptomes between overexpressed trees and wild-type. Compared with wild-type, 3248 DEGs were identified from OE-*MYB158* trees, of which 1386 were upregulated and 1862 were downregulated (Table S2). Compared with wild-type, 414 DEGs were found in OE-*MYB189* trees, including 80 upregulated and 334 downregulated genes (Table S3). Most of these downregulated genes in the OE-*MYB158* and OE-*MYB189* trees were involved in the formation of secondary cell walls, such as the phenylpropanoid biosynthetic pathway and xylan biosynthetic pathway (Figs. S3B, S3D; Tables S4, S5).

Cellulose is synthesized by cellulose synthase complexes, which consist of synthase protein isoforms (CesA). In this study, five *CesA* genes (*CesA1-B*, *CesA4*, *CesA7-A*, *CesA7-B*, and *CesA8-A*) were downregulated in OE-*MYB158* trees (Fig. 7). Four *CesA* genes (*CesA4*, *CesA7-A*, *CesA7-B*, and *CesA8-A*) were downregulated in OE-*MYB189* trees (Fig. 7).

For the 12 enzymatic steps involved in xylan biosynthesis, we identified 23 DEGs in OE-*MYB158* trees and 14 DEGs in OE-*MYB189* trees (Fig. 7). All of these DEGs were downregulated in the transgenic trees, and belonged to seven gene families. OE-*MYB158* and OE-*MYB189* trees shared 13 DEGs (*GT43B*, *GT43C*, *GT47C*, *GAUT12-A*, *GUX1-A*, *GUX1-B*, *GUX2*, *DUF579-1*, *DUF579-9*, *DUF579-10*, *XOAT1*, *XOAT2*, and *XOAT10*). Ten genes (*GT43D*, *GT4A-2*, *GT8E*, *GT8F*, *DUF579-4*, *RWA-A*, *XOAT5*, *XOAT8*, *XOAT9*, and *XOAT11*) were downregulated in OE-*MYB158* trees, but not in OE-*MYB189* trees. In contrast, *DUF579-2* was downregulated in OE-*MYB189* trees, but not in OE-*MYB158* trees.

In this study, we found that 17 DEGs in OE-*MYB158* trees and 10 DEGs in OE-*MYB189* trees were involved in monolignol biosynthesis (Fig. 8). All of these DEGs were downregulated in the transgenic trees, and belonged to nine gene families. OE-*MYB158* and OE-*MYB189* trees shared 10 DEGs (*PAL2*, *PAL3*, *C4H1*, *HCT1*, *4CL3*, *C3H3*, *CAD1*, *COMT2*, *CAld5H1*, and *CAld5H1*). However, seven genes (*PAL4*, *PAL5*, *C4H2*, *HCT6*,

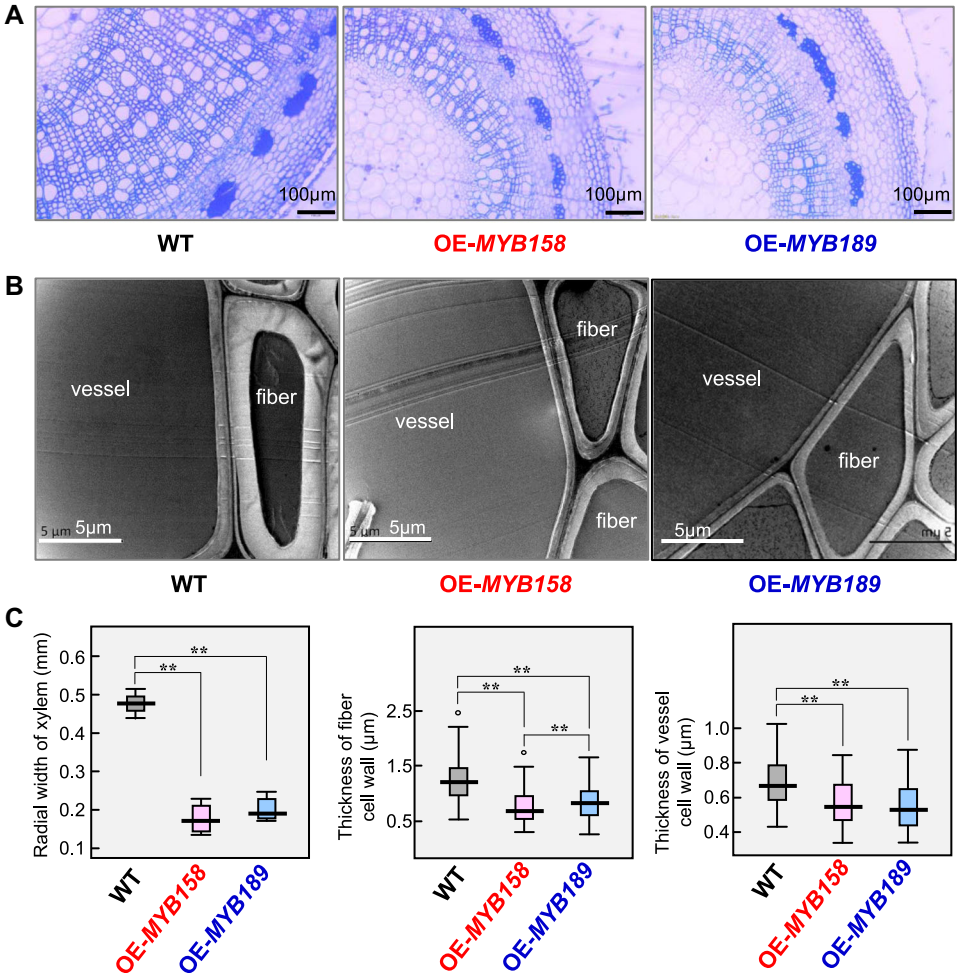

Fig. 5. Reduction of secondary cell wall thickening in fibers and vessels by overexpressing *MYB158* or *MYB189* genes. **A,** Cross-sections of the stems in wild-type (WT), OE-*MYB158*, and OE-*MYB189* trees. **B,** Transmission electron micrographs of xylem fiber and vessel walls of the stems in wild-type, OE-*MYB158*, and OE-*MYB189* trees. **C,** Radial width of xylem and thickness of fiber and vessel cell wall in the stems of wild type, OE-*MYB158*, and OE-*MYB189* trees.

4CL5, *CCoAOMT1*, and *CCoAOMT2*) were downregulated in OE-*MYB158* trees, but not in OE-*MYB189* trees.

This study identified 288 and 52 differentially expressed transcription factors (DETFs) in OE-*MYB158* and OE-*MYB189* trees, respectively (Fig. S4; Tables S6, S7). Among these DETFs, the numbers of transcription factors belonging to the MYB and NAC families were the largest. Functional analyses of these DETFs showed that the downregulated transcription factors were significantly enriched in GO terms related to secondary cell wall biosynthesis in OE-*MYB158* and OE-*MYB189* trees (Tables S8, S9). Compared with OE-*MYB189* trees, OE-*MYB158* trees enriched

Table 1 Compositions of cellulose, hemicellulose, and lignin in the stems of transgenic *MYB158*, *MYB189*, and wild-type *P. tomentosa*

	Composition (%, w/w)		
	Wild-type	OE-*MYB158*	OE-*MYB189*
Cellulose	57.24 ± 3.32	52.65 ± 1.10*	55.09 ± 6.44
Hemicellulose	13.37 ± 0.21	10.48 ± 0.92*	6.15 ± 0.44*
Lignin	28.73 ± 0.20	22.93 ± 1.62*	19.47 ± 0.24*

*Significant difference ($P < 0.05$).

Table 2 Monosaccharide composition in the stems of transgenic *MYB158*, *MYB189*, and wild-type *P. tomentosa*

	Composition (%, w/w)		
	Wild-type	OE-*MYB158*	OE-*MYB189*
Fucose	0.31 ± 0.00	0.29 ± 0.23	0.25 ± 0.01
Arabinose	3.04 ± 0.05	2.89 ± 0.13	2.14 ± 0.25
Galactose	2.13 ± 0.26	5.52 ± 0.65	1.55 ± 1.26
Glucose	57.24 ± 4.70	52.65 ± 1.10*	55.25 ± 3.14*
Xylose	10.32 ± 0.35	7.59 ± 1.05*	3.70 ± 0.33*

*Significant difference ($P < 0.05$).

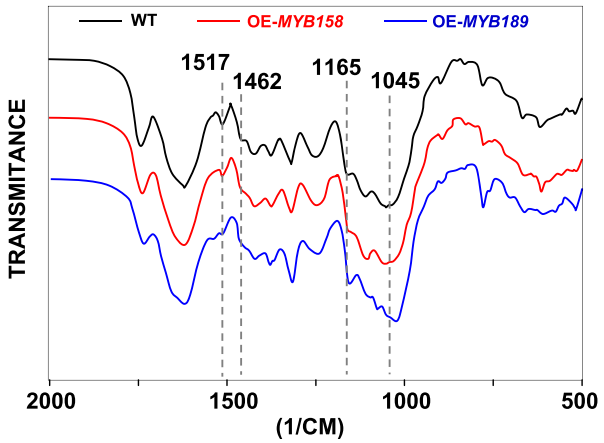

Fig. 6. Fourier transform infrared spectra of the stems of transgenic *MYB158* or *MYB189* trees.

more GO terms related to the regulation of secondary cell wall biosynthesis.

3.8 *myb158* mutant and *myb158 myb189* double mutant did not show obvious phenotypic variation in *P. tomentosa*

This study used the CRISPR/Cas9 system to generate knockout mutants of *P. tomentosa MYB158* or *MYB189*. Based on the sequence characteristics of *P. tomentosa MYB158* and *MYB189* genes, five 20 bp sequences (target sites T1, T2, T3, T4, and T5) followed by atrinucleotide (5′-NGG-3′) protospacer adjacent motif were selected as sgRNA complementary sites (Fig. S5A). Target sites T1 and T2 were designed to knock out the *P. tomentosa MYB158* gene, target site T3 to *P. tomentosa MYB189* gene, and target sites T4 and T5 to both *P. tomentosa MYB158* and *MYB189* genes. We constructed three CRISPR/Cas9 vectors (vectors I, II, and III in Fig. S5B). Vector I, for the specific target sites T1 and T2, was used to edit the *P. tomentosa MYB158* gene. Vector II, for the specific target site T3, was used to edit the *P. tomentosa MYB189* gene. Vector III, for specific target sites T4 and T5, was used to edit both *P. tomentosa MYB158* and *MYB189* genes.

We obtained 25 candidate *myb158* mutant trees. The regions covering the target sites T1 and T2 were amplified and sequenced. A 69 bp fragment deletion in the target site T2 was observed in the *myb158* line L19 (Fig. S6A). In addition, fragment insertions, deletions, and inversions were also detected at the target sites T1 or T2 in the *myb158* line L19. Thus, we successfully obtained *myb158* mutant tree (*myb158* [L19]). We obtained 10 candidate *myb158 myb189* mutant trees. Among these 10 candidate trees, the insertion/deletion of single nucleotide and the 32 bp fragment deletion in the *P. tomentosa MYB158* gene were detected only in the *myb158 myb189* line L38 (Fig. S6B). In the *myb158 myb189* line L38, we also observed fragment deletions and insertions in the *P. tomentosa MYB189* gene (Fig. S6B). Thus, we successfully obtained *myb158 myb189* double mutant tree (*myb158 myb189* [L38]). Unfortunately, we did not obtain a knockout mutant tree of the *P. tomentosa MYB189* gene.

Compared to wild-type trees, no obvious phenotypic variations were observed in *myb158* (L19) and *myb158 myb189* (L38) mutant trees (Fig. 9A). Among the wild-type, *myb158* (L19) and *myb158 myb189* (L38) trees, no significant differences were observed in the tree height, stem diameter, or stem internode length (independent sample *t*-test, $P > 0.05$) (Fig. 9B).

3.9 *myb158 myb189* double mutant shows increased thickness of vessel cell walls in *P. tomentosa*

Among the wild-type, *myb158* (L19) and *myb158 myb189* (L38) trees, no significant differences were observed in the radial width of xylem and the thickness of xylem fiber cell walls (independent sample *t*-test, $P > 0.05$) (Fig. 10). We did not observe significant differences in the thickness of vessel cell walls between wild-type and *myb158* (L19) trees (independent sample *t*-test, $P > 0.05$). However, compared to wild-type and *myb158* (L19) trees, the thickness of vessel cell walls were significantly increased in *myb158 myb189* (L38) mutant trees (independent sample *t*-test, $P < 0.05$) (Fig. 10C).

We examined the gene expression involved in secondary cell wall biosynthesis by RT-qPCR (Fig. S7). Compared to wild-type, the gene expression abundances involved in secondary cell wall formation were increased in *myb158* (L19) and *myb158 myb189* (L38) mutant trees. For example, compare to wild-type, the expression levels of coniferaldehyde 5-hydroxylase 1 (*CAld5H1*) gene involved in monolignol biosynthesis were increased 1.7- and 2.3-fold in *myb158* (L19) and *myb158 myb189* (L38) mutant trees, respectively.

3.10 *Populus MYB158* acts as the upstream regulator of secondary wall NAC master switches and directly represses expression of *SND1-B2* gene

Populus SND1-B2 (*Potri.002G178700*; also named *PtrWND2B*) was identified as the master switch regulating secondary wall biosynthesis (Zhong et al., 2010b). In this study, based on transcriptome analysis, we found that the *SND1-B2* gene was downregulated in OE-*MYB158* trees, but not in OE-*MYB189* trees (Fig. 11A). To verify whether *P. tomentosa MYB158* and *MYB189* genes can directly regulate the expression of *SND1-B2*, we constructed two sets of experiments for functional assays *in vivo*.

First, a dual luciferase assay system was used for the determination of *P. tomentosa MYB158*- and *MYB189*-mediated transcriptional repression. The promoter of *SND1-B2* was fused to the firefly luciferase (LUC) gene to construct reporter plasmid. *P. tomentosa MYB158* and *MYB189* were fused to the 35S promoter to construct effector plasmids. The reporter plasmids and effector plasmids were coinfiltrated into *Nicotiana benthamiana* leaves, and the relative LUC activities were measured. Under the expression of *P. tomentosa MYB158*, the relative LUC activity was decreased by 51% (Mann–Whitney *U*-test, $P < 0.01$) (Fig. 11B). However, compared to negative control, no significant difference in the relative LUC activity was observed under the expression of *P. tomentosa MYB189*.

Second, ChIP assays were carried out to detect the affinities of *P. tomentosa MYB158* and *MYB189* for the promoter of *SND1-B2 in vivo*, respectively. The transgenic

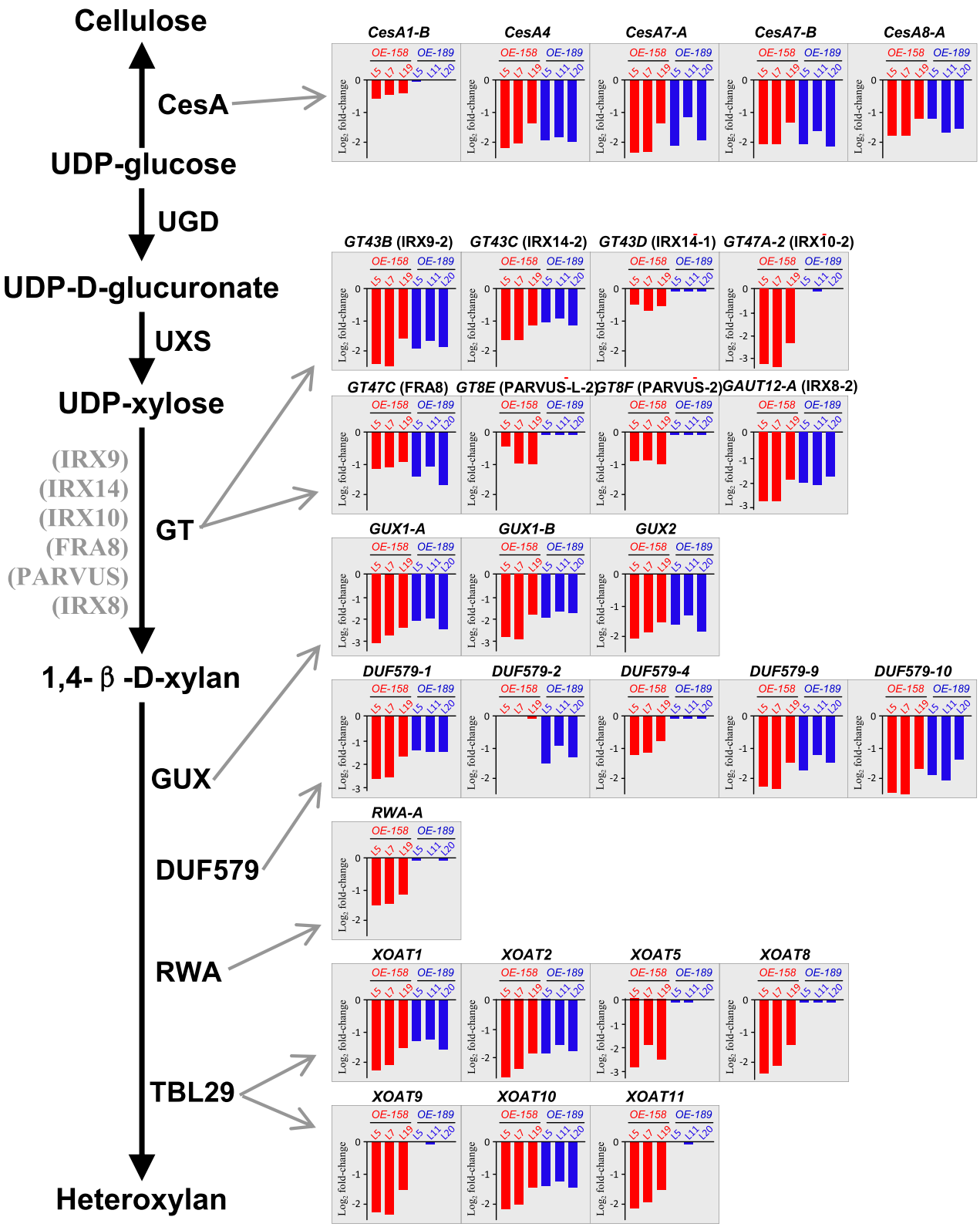

Fig. 7. Overexpression of *MYB158* or *MYB189* leads to decreased expression of cellulose and xylan biosynthetic genes. Relative expression levels of differentially expressed genes were calculated by Log_2fold-change. Red and blue columns show the relative expression levels of the differentially expressed genes in three *MYB158* transgenic lines (L5, L7, and L19) and three *MYB189* transgenic lines (L5, L11, and L20), respectively.

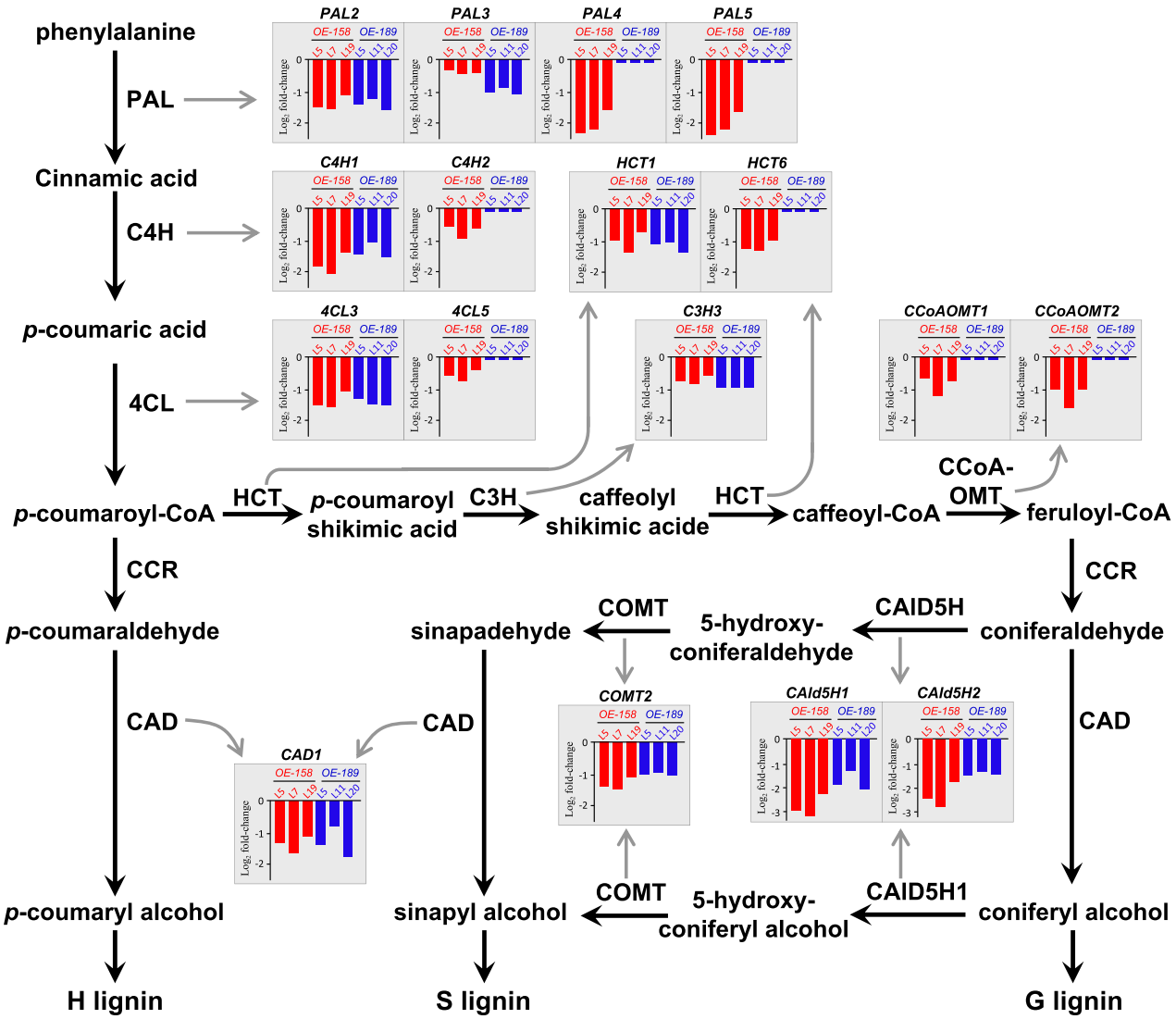

Fig. 8. Overexpression of *MYB158* or *MYB189* leads to decreased expression of monoligno biosynthetic genes. Relative expression levels of differentially expressed genes were calculated by Log_2fold-change. Red and blue columns show the relative expression levels of the differentially expressed genes in three OE-*MYB158* transgenic lines (L5, L7, and L19) and three OE-*MYB189* transgenic lines (L5, L11, and L20), respectively.

P. tomentosa overexpressing *MYB158-FLAG* or *MYB189-FLAG* were prepared, and the wild-type was used as a negative control. The MYB158-FLAG proteins or MYB189-FLAG proteins were immunoprecipitated using an anti-FLAG antibody, and the enrichment of the target promoters were detected by ChIP-qPCR. The results showed that *P. tomentosa* MYB158 and MYB189 proteins could directly bind to the F2 fragment of the *SND1-B2* promoter (Fig. 11C). However, compared to MYB189 protein, MYB158 protein showed 3.8-fold stronger affinity to the F2 fragment of the *SND1-B2* promoter.

4 Discussion

4.1 Expression divergence between duplicated *MYB158* and *MYB189*

A previous study found that the duplicate genes created by WGD tend to maintain similar expression patterns, whereas the duplicate genes created by tandem duplication rapidly diverged (Ganko et al., 2007). To explain the mechanism of this pattern, it was proposed that WGD could result in the duplication of genes and their promoter elements, whereas tandem replication might result in partial duplication of their promoter elements (Casneuf et al., 2006). However, this study revealed the clear divergence in expression patterns between *P. tomentosa MYB158* and *MYB189* genes created by a WGD event. Compared to the *MYB189* gene, *MYB158* had a higher expression level in the bark, stem, xylem, leaf, and petiole tissues, and had a lower expression level in root tissue (Fig. 3A). Compared to *MYB189pro:GUS* transgenic trees, *MYB158pro:GUS* transgenic trees showed stronger GUS staining in stem and leaf tissues (Fig. 3B). A possible explanation is that the promoter regions of *P. tomentosa MYB158* and *MYB189* genes could be under

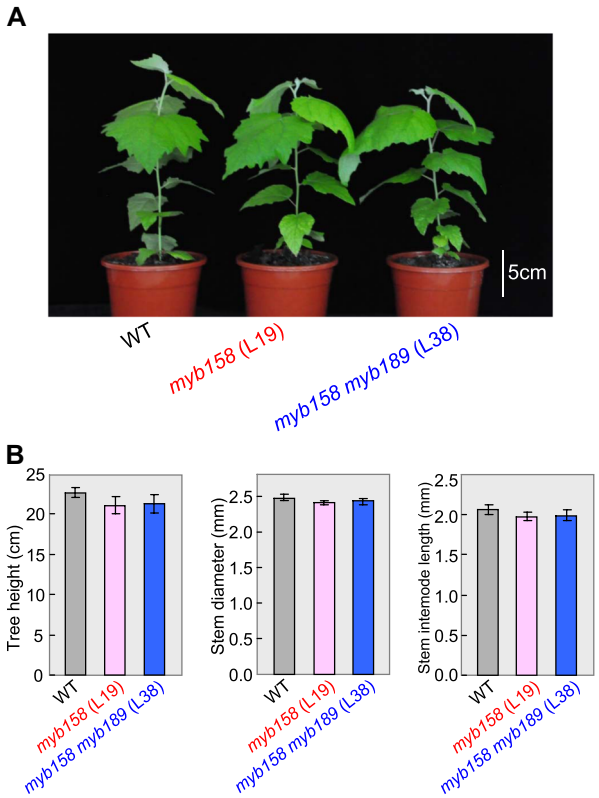

Fig. 9. The *myb158* mutant and *myb158 myb189* double mutant did not show phenotypic differences. **A,** Two-month-old wild-type (WT), *myb158* mutant, and *myb158 myb189* double mutant trees. **B,** Comparison of tree height, stem diameter, and stem internode length among WT, *myb158*, and *myb158 myb189* trees.

relaxed selective constraint, which might result in higher mutation rates in their promoter regions. In fact, compared to the coding regions of these two genes (they have a nucleic acid sequence identity of 89%), the 2300 bp promoter regions of the *P. tomentosa MYB158* and *MYB189* genes show much lower nucleic acid sequence identity (51%). These indicated that, compared to coding regions, promoter regions of the *P. tomentosa MYB158* and *MYB189* had undergone rapid evolution.

4.2 Functional divergence between MYB158 and MYB189 proteins

Although the coding protein regions of *P. tomentosa MYB158* and *MYB189* genes show much higher nucleic acid sequence identity than the promoter regions, we observed functional divergence between these two MYB proteins. For example, 10 genes involved in xylan biosynthesis were downregulated in OE-*MYB158* trees, but not in OE-*MYB189* trees; in contrast, *DUF579-2* was downregulated in OE-*MYB189* trees, but not in OE-*MYB158* trees (Fig. 7). Compared to the wild-type trees, the cellulose content in the stems of the OE-*MYB158* trees was significantly reduced, but not in the OE-*MYB189* trees (Table 1). More importantly, *P. tomentosa MYB158* could directly repress the expression of the *SND1-B2* gene, which was a master switch regulating secondary wall biosynthesis, whereas *MYB189* had weak affinity to the *SND1-B2* gene (Fig. 11). *Populus tomentosa* MYB158 and MYB189 proteins belonged to the R2R3 MYB subfamily proteins (Wilkins et al., 2009). R2R3 MYB proteins have three distinct regions: two N-terminal conserved MYB DNA-binding domains (R2 and R3 domains), and a diverse C-terminal region responsible for the regulatory activity of the protein. R2 and R3 domains were highly conserved in the *Populus* R2R3 MYB subfamily (Wilkins et al., 2009). In the R2 and R3 domains of *P. tomentosa* MYB158 and MYB189 proteins, there were only one and two amino acid differences, respectively (Fig. S8). However, the C-terminal region of *P. tomentosa* MYB158 and MYB189 proteins contained many amino acid variations. Recent study identified a repressive motif (GDDYGNHGMIKKE) in the C-terminal region of MYB189 protein (Jiao et al., 2019). In this 13-amino-acid region, *P. tomentosa* MYB158 and MYB189 contained two amino acid differences (Fig. S8). The variations in the C-terminal region of the protein might result in functional divergence between *P. tomentosa* MYB158 and MYB189 proteins.

4.3 Partial functional redundancy between duplicated *MYB158* and *MYB189* genes

This study observed that overexpression of *MYB158* or *MYB189* could repress secondary cell wall thickening in transgenic *P. tomentosa* (Fig. 5). and could downregulate the expression of some identical genes involved in the biosynthesis of cellulose, xylan, and lignin (Figs. 7, 8). These results indicated that *Populus* duplicated *MYB158* and *MYB189* genes had partial functional redundancy. After a WGD event, some duplicate genes will become pseudogenes by accumulating deleterious mutants, whereas others will persist and evolve diverse functions. A particularly interesting phenomenon was that many duplicate genes have partial functional similarities during the process of evolving diversified functions. For example, in *Arabidopsis*, two Kinesin-14 genes (*AtKIN14a* and *AtKIN14b*) created by segmental duplication have retained similar functions in gametophyte development, but have evolved partially different functions in male meiosis (Quan et al., 2008). Why do duplicate genes created by the WGD still retain partial functional redundancy? When WGD occurred, two duplicate genes might have redundant functions. After a long evolutionary time, two duplicate genes have at least partial functional divergence, otherwise, at least one member will become a pseudogene which may be deleted from the genome. Some of the duplicate genes might need a long time to achieve full functional divergence. During this process, the partial functional redundancy of these duplicate genes was still observed.

4.4 *Populus* duplicated *MYB158* and *MYB189* genes negatively regulate secondary cell wall biosynthesis

The formation of secondary cell walls is essential for woody plants to grow upright and adapt to terrestrial life (Xu et al., 2014). Secondary cell wall biosynthesis was mainly regulated by transcriptional factors. Previous studies identified many transcriptional factors that could activate secondary cell wall biosynthesis. The high

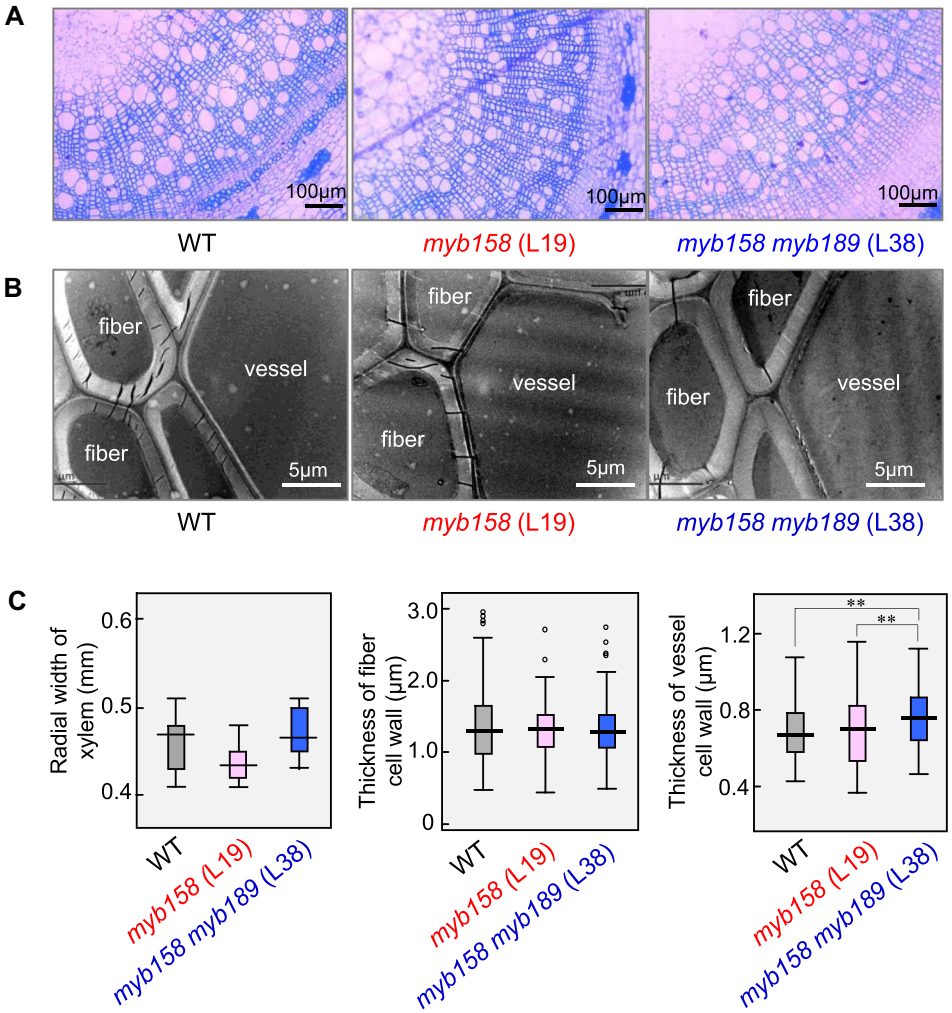

Fig. 10. Secondary cell wall analysis of *myb158* mutant and *myb158 myb189* double mutants. **A,** Cross sections of the stems in wild-type (WT), *myb158* mutant, and *myb158 myb189* double mutant trees. **B,** Transmission electron micrographs of xylem fiber and vessel walls of the stems in WT, *myb158*, and *myb158 myb189* trees. **C,** Radial width of xylem and thickness of fiber and vessel cell wall in the stems of WT, *myb158*, and *myb158 myb189* trees. $^{**}P < 0.01$.

expression of transcriptional activators might cause abnormal secondary cell wall biosynthesis. For example, overexpression of the *AtSND1* gene driven by 35S promoter in *Arabidopsis* can result in high expression of secondary cell wall biosynthetic genes (such as *CesA7*, *CesA8*, *FRA8*, *4CL1*, and *CCoAOMT*), and result in massive ectopic deposition of secondary cell walls in normally parenchymatous cells (Zhong et al., 2006). Thus, plants need transcriptional repressors to safeguard the normal biosynthesis of secondary cell walls. In this study, we found that *P. tomentosa MYB158* and *MYB189* are transcriptional repressors and negatively regulate secondary cell wall biosynthesis. A similar function has been observed for poplar *MYB221* gene, which was predominantly expressed in developing wood cells and function as the transcriptional repressor to directly repress the expression of poplar *CESA8*, *GT47C*, and *COMT2* (Tang et al., 2015). In addition, this study found that overexpression of *P. tomentosa MYB158* or *MYB189* in transgenic trees only resulted in a decrease in the content of cell wall components, but did not alter the chemical structure of cell wall components (Fig. 6; Table 1), indicating that these two proteins might act by preventing the over-deposition of secondary cell walls.

This study showed that *MYB158* and *MYB189* had similar functions in repressing secondary cell wall thickening. Compared to wild-type trees, *myb158 myb189* double mutant showed increased thickness of vessel cell walls in *P. tomentosa*, whereas the *myb158* mutant did not. Unfortunately, we did not get knockout mutant tree of the *P. tomentosa MYB189* gene. If we obtained *myb189* mutant, the phenotype of *myb189* mutant might be similar to *myb158* mutant, and of course it might be similar to *myb158 myb189* double mutant. For the duplicate gene pair, compared to wild-type, one duplicate gene mutant had no phenotypic changes, whereas the double mutant of two duplicate genes showed phenotype changes. Functional redundancy between duplicate genes might result in this phenomenon.

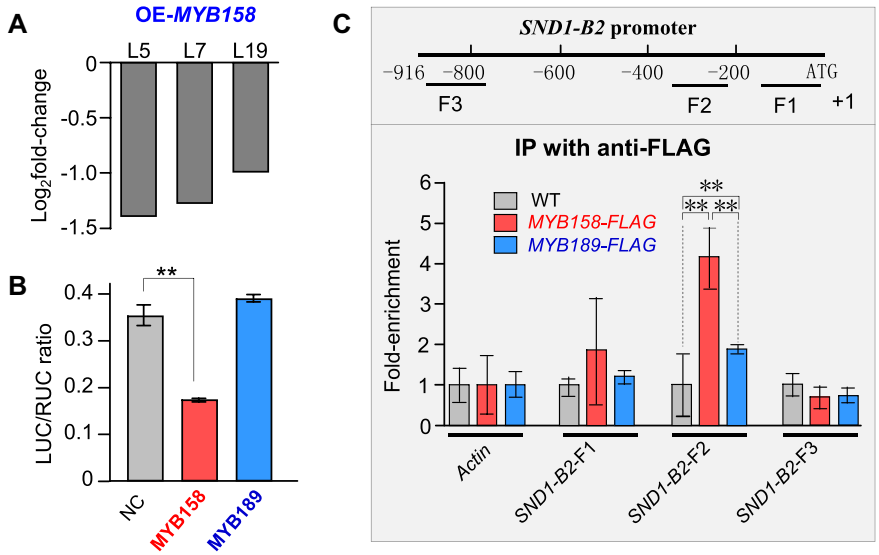

Fig. 11. MYB158 binds directly to the *SND1-B2* promoter and inhibits its expression. **A,** The expression of the *SND1-B2* gene was reduced in overexpressing *MYB158* trees. **B,** Transcriptional repressive activities of *MYB158 and MYB189* on the promoter of *SND1-B2* by dual luciferase assay. **C,** ChIP-qPCR analysis of MYB158 and MYB189 proteins binding to the promoter of *SND1-B2*. ***P* < 0.01. WT, wild-type.

One possible explanation is that when one duplicate gene is knocked out, the other duplicate gene can complement the function of the deleted duplicate. When both duplicate genes are knocked out, no other gene can complement the function of the deleted duplicate genes. Thus, the double mutant plants will show phenotypic changes compared to wild-type.

4.5 *MYB158* directly represses expression of *SND1-B2* gene and regulates secondary cell wall biosynthesis

In the secondary cell wall biosynthesis regulatory network, *SND* genes function as master switches that activate the process of secondary cell wall biosynthesis (Zhong et al., 2010a). *Arabidopsis* and *Populus* contain five and eight *SND* genes, respectively (Sundell et al., 2017). Recently, *Populus SND1-A2* and *VND6-C1* were found to be capable of generating splice variants *SND1-A2*^{IR} and *VND6-C1*^{IR}, respectively. Both *Populus SND1-A2*^{IR} and *VND6-C1*^{IR} could repress the transcript abundance of *SND* genes (Li et al., 2012, 2017). In this study, we found that *Populus MYB158* could directly repress the expression of *SND1-B2*. Previous studies and this study have shown that the precise expression of the *SND* gene was regulated by many transcription factors. Small changes in the expression level of one transcription factor could make it difficult to significantly change the expression level of the *SND* gene. However, a significant increase in the expression level of one transcription factor might disrupt the stable expression of the *SND1* gene, resulting in a change in the expression of the *SND* gene. In this study, the expression level of the *MYB158* gene in transgenic trees (OE-*MYB158*) was 14-fold higher than that of wild-type. This resulted in a significant downregulation of the expression level of the *SND1-B2* gene in OE-*MYB158* trees (Fig. 11A). Taken together, this study identified a new MYB transcription factor that regulates the *SND* gene, which is important for understanding secondary cell wall biosynthesis.

Acknowledgements

This study was supported by the National Science Fund for Distinguished Young Scholars (31425006) and the Chinese Academy of Forestry (CAFYBB2018ZX001).

References

Blanc G, Wolfe KH. 2004. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *The Plant Cell* 16: 1679–1691.

Casneuf T, De Bodt S, Raes J, Maere S, Van de Peer Y. 2006. Nonrandom divergence of gene expression following gene and genome duplications in the flowering plant *Arabidopsis thaliana*. *Genome Biology* 7: R13.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151: 1531–1545.

Ganko EW, Meyers BC, Vision TJ. 2007. Divergence in expression between duplicated genes in *Arabidopsis*. *Molecular Biology and Evolution* 24: 2298–2309.

Hellens RP, Allan AC, Friel EN, Bolitho K, Grafton K, Templeton MD, Karunairetnam S, Gleave AP, Laing WA. 2005. Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. *Plant Methods* 1: 13.

Jefferson RA, Kavanagh TA, Bevan MW. 1987. GUS fusions: Beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO Journal* 6: 3901–3907.

Jiao B, Zhao X, Lu W, Guo L, Luo K. 2019. The R2R3 MYB transcription factor MYB189 negatively regulates secondary cell wall biosynthesis in *Populus*. *Tree Physiology* 39: 1187–1200.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97–100.

Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McLnerney JO. 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evolutionary Biology* 6: 29.

Kubo M, Udagawa M, Nishikubo N, Horiguchi G, Yamaguchi M, Ito J, Mimura T, Fukuda H, Demura T. 2005. Transcription switches for protoxylem and metaxylem vessel formation. *Genes Development* 19: 1855–1860.

Li Q, Lin YC, Sun YH, Song J, Chen H, Zhang XH, Sederoff RR, Chiang VL. 2012. Splice variant of the SND1 transcription factor is a dominant negative of SND1 members and their regulation in *Populus trichocarpa*. *Proceedings of the National Academy of Sciences USA* 109: 14699–14704.

Lin YJ, Chen H, Li Q, Li W, Wang JP, Shi R, Tunlaya-Anukit S, Shuai P, Wang Z, Ma H, Li H, Sun YH, Sederoff RR, Chiang VL. 2017. Reciprocal cross-regulation of VND and SND multigene TF families for wood formation in *Populus trichocarpa*. *Proceedings of the National Academy of Sciences USA* 114: E9722–E9729.

Liu HJ, Tang ZX, Han XM, Yang ZL, Zhang FM, Yang HL, Liu YJ, Zeng QY. 2015. Divergence in enzymatic activities in the soybean GST supergene family provides new insight into the evolutionary dynamics of whole-genome duplicates. *Molecular Biology and Evolution* 32: 2844–2859.

Moghe GD, Hufnagel DE, Tang H, Xiao Y, Dworkin I, Town CD, Conner JK, Shiu SH. 2014. Consequences of whole-genome triplication as revealed by comparative genomic analyses of the wild radish *Raphanus raphanistrum* and three other *Brassicaceae* species. *The Plant Cell* 26: 1925–1937.

Naito Y, Hino K, Bono H, Ui-Tei K. 2014. CRISPRdirect: Software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics* 31: 1120–1123.

Nakano Y, Yamaguchi M, Endo H, Rejab NA, Ohtani M. 2015. NAC-MYB-based transcriptional regulation of secondary cell wall biosynthesis in land plants. *Frontiers in Plant Science* 6: 288.

Øbro J, Harholt J, Scheller HV, Orfila C. 2004. Rhamnogalacturonan I in *Solanum tuberosum* tubers contains complex arabinogalactan structures. *Phytochemistry* 65: 1429–1438.

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

Quan L, Xiao R, Li W, Oh SA, Kong H, Ambrose JC, Malcos JL, Cyr R, Twell D, Ma H. 2008. Functional divergence of the duplicated *AtKIN14a* and *AtKIN14b* genes: critical roles in Arabidopsis meiosis and gametophyte development. *Plant Journal* 53: 1013–1026.

Saleh A, Alvarez-Venegas R, Avramova Z. 2008. An efficient chromatin immunoprecipitation (ChIP) protocol for studying histone modifications in *Arabidopsis* plants. *Nature Protocols* 3: 1018–1025.

Schranz ME, Mohammadin S, Edger PP. 2012. Ancient whole genome duplications, novelty and diversification: The WGD radiation lag-time model. *Current Opinion in Plant Biology* 15: 147–153.

Sundell D, Street NR, Kumar M, Mellerowicz EJ, Kucukoglu M, Johnsson C, Kumar V, Mannapperuma C, Delhomme N, Nilsson O, Tuominen H, Pesquet E, Fischer U, Niittylä T, Sundberg B, Hvidsten TR. 2017. AspWood: High-spatial-resolution transcriptome profiles reveal uncharacterized modularity of wood formation in *Populus tremula*. *The Plant Cell* 29: 1585–1604.

Tang X, Zhuang Y, Qi G, Wang D, Liu H, Wang K, Chai G, Zhou G. 2015. Poplar PdMYB221 is involved in the direct and indirect regulation of secondary wall biosynthesis during wood formation. *Scientific Reports* 5: 12240.

Wang HZ, Dixon RA. 2012. On-off switches for secondary cell wall biosynthesis. *Molecular Plant* 5: 297–303.

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.

Xu B, Ohtani M, Yamaguchi M, Toyooka K, Wakazaki M, Sato M, Kubo M, Nakano Y, Sano R, Hiwatashi Y, Murata T, Kurata T, Yoneda A, Kato K, Hasebe M, Demura T. 2014. Contribution of NAC transcription factors to plant adaptation to land. *Science* 343: 1505–1508.

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, Ye ZH. 2010a. Evolutionary conservation of the transcriptional network regulating secondary cell wall biosynthesis. *Trends in Plant Science* 15: 625–632.

Zhong R, Lee C, Ye ZH. 2010b. Functional characterization of poplar wood-associated NAC domain transcription factors. *Plant Physiology* 152: 1044–1055.

Zhong R, Lee C, Zhou J, McCarthy RL, 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, 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.

Supplementary Material

The following supplementary material is available online for this article at http://onlinelibrary.wiley.com/doi/10.1111/jse.12595/suppinfo:

Fig. S1. Modified pCAMBIA1302 vector.

Fig. S2. Phylogenetic tree of MYB gene families from *Populus* and *Arabidopsis*.

Fig. S3. Identification and functional analysis of DEGs in transgenic *MYB158* or *MYB189* trees.(A) and (C), Heat maps of DEGs between wild type and transgenic *MYB158* and *MYB189* trees, respectively. Relative expression was calculated by Log_2Ratio. The heatmaps showed the common DEGs among the three transgenic lines. (B) and (D), GO enrichment of the biological process of DEGs in transgenic *MYB158* and *MYB189* trees, respectively.

Fig. S4. Differentially expressed transcription factors in transgenic *MYB158* (A) or *MYB189* (B) trees.

Fig. S5. Target sites for knocking out *P. tomentosa MYB158* and *MYB189* genes. (A) Five sgRNAcomplementary sites (T1, T2, T3, T4 and T5) were selected in the exons of *MYB158* and *MYB189*. (B) Strategy of vector construction for editing *MYB158* and *MYB189* genes. T1, T2, T3, T4 and T5 represent target sites for knocking out *MYB158* or *MYB189* genes.

Fig. S6. Detection of *myb158* mutant and *myb158 myb189* double mutant. In (A), the gel image shows PCR amplification from genomic DNA of *myb158* mutant. The red arrows indicate deletion mutation of the *MYB158* gene. Sequence

alignment of the mutation profiles of the *myb158* mutant. In (B), the gel image shows PCR amplification from genomic DNA of *myb158 myb189* double mutant. The red arrows indicate deletion mutation of the *MYB158* and *MYB189* genes. Sequence alignment of the mutation profiles of the *myb158 myb189* double mutant.

Fig. S7. Gene expression of secondary cell wall biosynthetic genes in the stems of wild-type, *myb158* mutant and *myb158 myb189* double mutant.

Fig. S8. Sequence alignment of *P. tomentosa MYB158* and *MYB189* proteins. Different residues in *P. tomentosa MYB158* and *MYB189* proteins are shaded red.

Table S1. Primers used in this study.

Table S2. Up- and down-regulated genes in transgenic *MYB158* trees (Padj < 0.05)

Table S3. Up- and down-regulated genes in transgenic *MYB189* trees (Padj < 0.05)

Table S4. GO Enrichment of the biological process of DEGs in transgenic *MYB158* trees. Cluster #1 and Cluster #2 represent up-regulated and down-regulated genes, respectively.

Table S5. GO enrichment of the biological process of DEGs in transgenic *MYB189* trees. Cluster #1 and Cluster #2 represent up-regulated and down-regulated genes, respectively.

Table S6. Differentially expressed transcription factors in transgenic *MYB158* trees.

Table S7. Differentially expressed transcription factors in transgenic *MYB189* trees.

Table S8. GO enrichment of down-regulated transcription factors in transgenic *MYB158* trees.

Table S9. GO enrichment of down-regulated transcription factors in transgenic *MYB189* trees.