



Genome-Wide Comparative Analysis of R2R3 MYB Gene Family in *Populus* and *Salix* and Identification of Male Flower Bud Development-Related Genes

Fangwei Zhou, Yingnan Chen, Huaitong Wu* and Tongming Yin

Key Laboratory for Tree Breeding and Germplasm Improvement, Southern Modern Forestry Collaborative Innovation Center, College of Forestry, Nanjing Forestry University, Nanjing, China

OPEN ACCESS

Edited by:

Boas Pucker,
University of Cambridge,
United Kingdom

Reviewed by:

Hongxia Zhang,
Ludong University, China
Liangsheng Zhang,
Zhejiang University, China

*Correspondence:

Huaitong Wu
wuhuaitong@163.com

Specialty section:

This article was submitted to
Plant Systematics and Evolution,
a section of the journal
Frontiers in Plant Science

Received: 07 June 2021

Accepted: 17 August 2021

Published: 14 September 2021

Citation:

Zhou F, Chen Y, Wu H and Yin T
(2021) Genome-Wide Comparative
Analysis of R2R3 MYB Gene Family
in *Populus* and *Salix* and Identification
of Male Flower Bud
Development-Related Genes.
Front. Plant Sci. 12:721558.
doi: 10.3389/fpls.2021.721558

The MYB transcription factor (TF) family is one of the largest plant transcription factor gene family playing vital roles in plant growth and development, including defense, cell differentiation, secondary metabolism, and responses to biotic and abiotic stresses. As a model tree species of woody plants, in recent years, the identification and functional prediction of certain MYB family members in the poplar genome have been reported. However, to date, the characterization of the gene family in the genome of the poplar's sister species willow has not been done, nor are the differences and similarities between the poplar and willow genomes understood. In this study, we conducted the first genome-wide investigation of the R2R3 MYB subfamily in the willow, identifying 216 R2R3 MYB gene members, and combined with the poplar R2R3 MYB genes, performed the first comparative analysis of R2R3 MYB genes between the poplar and willow. We identified 81 and 86 pairs of R2R3 MYB paralogs in the poplar and willow, respectively. There were 17 pairs of tandem repeat genes in the willow, indicating active duplication of willow R2R3 MYB genes. A further 166 pairs of poplar and willow orthologs were identified by collinear and synonymous analysis. The findings support the duplication of R2R3 MYB genes in the ancestral species, with most of the R2R3 MYB genes being retained during the evolutionary process. The phylogenetic trees of the R2R3 MYB genes of 10 different species were drawn. The functions of the poplar and willow R2R3 MYB genes were predicted using reported functional groupings and clustering by OrthoFinder. Identified 5 subgroups in general expanded in woody species, three subgroups were predicted to be related to lignin synthesis, and we further speculate that the other two subgroups also play a role in wood formation. We analyzed the expression patterns of the GAMYB gene of subgroup 18 (S18) related to pollen development in the male flower buds of poplar and willow at different developmental stages by qRT-PCR. The results showed that the GAMYB gene was specifically expressed in the male flower bud from pollen formation to maturity, and that the expression first

increased and then decreased. Both the specificity of tissue expression specificity and conservation indicated that GAMYB played an important role in pollen development in both poplar and willow and was an ideal candidate gene for the analysis of male flower development-related functions of the two species.

Keywords: *Populus*, *Salix*, R2R3 MYB transcription factors, GAMYB, pollen

INTRODUCTION

The initiation of transcription in eukaryotes is highly complex and often requires the assistance of a variety of protein factors (Shi et al., 2017). Transcription factors and RNA polymerase II form a transcription initiation complex to jointly participate in the transcription initiation process and control gene expression (Kamenova et al., 2019). MYB superfamily is one of the largest groupings of transcription factors and is found in both animals and plants. These proteins usually contain a highly conserved MYB domain at the N-terminus (Frampton et al., 1989; Stracke et al., 2001; Dubos et al., 2010; Jiang et al., 2011). Recently, it has been found that some MYB proteins also contain a MYB domain at the C-terminus (Linger and Price, 2009). The MYB domain is 51–52 amino acids in length and contains four imperfect amino acid repeats (R) each forming a helix-turn-helix (HTH) structure with three tryptophan residues acting as a hydrophobic core in each helix. These HTH structures are responsible for the specific binding of the MYB proteins to DNA (Cheng et al., 2016). According to the number of amino acids and tryptophan conservation, the MYB domain is divided into three subtypes, R1, R2, and R3 (Jin and Martin, 1999; Dubos et al., 2010). Based on the number of R repeats within the domain, MYB superfamily members are divided into four main groups: 1R-MYB (MYB-related proteins), 2R-MYB (R2R3), 3R-MYB (R1R2R3), and 4R-MYB (R1R2R2R1/2-MYB) (Frampton et al., 1989; Li et al., 2019). The R2R3 MYB subfamily is the most abundant type in plants (Jin and Martin, 1999; Millard et al., 2019b).

With the release of the whole-genome sequences of increasing numbers of plant species, R2R3 MYB proteins have been identified and analyzed in numerous species based, especially, on the structures of the conserved MYB domain. During plant evolution, the number of R2R3 MYB family members has expanded considerably due to a large number of gene duplication events. Studies have found that some gymnosperms have relatively few members of the R2R3 MYB family, for example, *Ginkgo biloba* (68 members) (Yang et al., 2021). And whole-genome analysis of dicotyledonous plants shows that most species have between 70 and 200 family members (Du et al., 2015), for example, *Arabidopsis thaliana* (126 members) (Stracke et al., 2001), *Zea mays* (157 members) (Du et al., 2012a), *Oryza sativa* (102 members) (Chen et al., 2006), *Vitis vinifera* (134 members) (Wong et al., 2016), *Camellia sinensis* (122 members) (Chen et al., 2021), and *Populus trichocarpa* (192 members) (Chai et al., 2014). Each pedigree continued to expand after the divergence. In addition to these expansion and diversification events, the tandem duplication of subgroups has also been predicted. Therefore, some species have more than 200 gene family members, for example, *Glycine max*

(244 members) (Du et al., 2012b) and *Gossypium raimondii* (205 members) (He et al., 2016), while *Brassica rapa* ssp. *pekinensis* has 256 members, the largest known plant R2R3 MYB gene family (Wang et al., 2015). The differences in the numbers of R2R3 MYB genes between species, in addition to gene amplification, may also be caused by environmental adaptation, indicating specificity after separation from the last common ancestor (Chang and Duda, 2012).

Research on the functions of plant R2R3 MYBs indicates that the functions of these transcription factors are diverse (Yang et al., 2012). They can participate in plant growth and development in multiple ways (Ambawat et al., 2013). In *A. thaliana*, *AtMYB16* and *AtMYB106* are mainly involved in regulating epidermal hair differentiation (Baumann et al., 2007; Jakoby et al., 2008), while *AtMYB58* and *AtMYB63* can specifically activate lignin biosynthesis (Noda et al., 2015; Wang F. P. et al., 2018). R2R3 MYBs can also regulate the synthesis of flavonoids and other secondary metabolites by regulating the expression of genes related to plant secondary metabolism (Aharoni et al., 2001; Jiang and Rao, 2020). For example, *AtMYB75* (*PAP1*), *AtMYB90* (*PAP2*), *AtMYB113*, and *AtMYB114* in *A. thaliana* can regulate the accumulation of anthocyanins (Borevitz et al., 2000; Bhargava et al., 2010; Ma and Constabel, 2019). *AtMYB123* (*TT2*) function in proanthocyanidin synthesis in plants (Nesi et al., 2001). R2R3 MYBs also play important roles in abiotic and biotic stress responses (Baldoni et al., 2015), such as *AtMYB15*, *AtMYB30*, *AtMYB60*, and *AtMYB96* from *A. thaliana* (Seo and Park, 2010; Cominelli et al., 2015). Additionally, R2R3 MYBs can also control cell morphogenesis and regulate the growth and development of flowers, fruits, and seeds (Lv et al., 2017). For example, *AtMYB33*, *AtMYB65*, and *AtMYB101* in *A. thaliana* play important roles in flower development, especially in pollen development (Gocal et al., 2001; Li et al., 2016). Genes regulating the development of stamens have also been found in other species with the particular gene identified as GAMYB (Aya et al., 2011). For example, the *HvGAMYB* gene of *Hordeum vulgare* plays an important role in the development of flowers, especially in the development of pollen (Murray et al., 2003), mutation of the rice GAMYB gene results in abnormal stamen development, shortened filaments, and sterile pollen (Ko et al., 2021), while the GAMYB gene *PtrMYB012* in the poplar heterologous transformation of *A. thaliana* also leads to male sterility (Zhang et al., 2020).

Poplar and willow belong to the Salicaceae family, which is widely distributed in the northern hemisphere (Wei et al., 2020). Both species adapt easily to different ecological environments, grow rapidly, and reproduce without difficulty. They are important tree species for timber forest, shelter forest, roadside trees, and greening and have important economic, ecological,

and social value (Zhang J. et al., 2017). However, when the male tree reaches sexual maturity, it produces a great deal of pollen, which causes allergic reactions in some people (Futamura et al., 2000; Costache et al., 2021). Therefore, studying the reproductive mechanism of poplar and willow is valuable for the improvement of forest trees. At present, both the poplar and willow whole-genome sequences have been sequenced and are publicly available (Tuskan et al., 2006; Dai et al., 2014). Although R2R3 MYB genes have been identified and analyzed in the poplar (Chai et al., 2014), this is not the case for the willow. To date, most of our understanding of the function of the R2R3 MYB genes in plants is derived from the analysis of the model plant *A. thaliana*. In the poplar and willow, only a few R2R3 MYB genes have been cloned and functionally verified (Fang et al., 2020; Tang et al., 2020; Wang et al., 2020; Liu C. et al., 2021). As these genes play important roles in multiple growth and developmental pathways, it is urgent to characterize their roles in the Salicaceae, to identify and classify the genes, and thus to provide a reference for subsequent functional verification of the R2R3 MYB gene family. In this study, the R2R3 MYB gene family of willow was comprehensively identified and analyzed for the first time. By comparing the willow genes with the poplar R2R3 MYB genes, the specific poplar and willow family members after differentiation of the species were identified. Further, the function of the R2R3 MYB gene family members was comprehensively predicted, and the expression patterns of family members related to stamen development were verified, providing a reference for the functional verification of the R2R3 MYB gene family members in the Salicaceae.

MATERIALS AND METHODS

Genome-Wide Identification and Sequence Analysis of *Populus*, *Salix*, and *A. thaliana* R2R3 MYB Genes

The consensus R2R3 MYB DNA-binding domain (containing 106 amino acid residues) (Ogata et al., 1994) sequence was used to identify homologous genes in the *Populus trichocarpa* genome database (version 3.1)¹ and *Salix purpurea* genome database (version 1.0)² (Zhou et al., 2018; Zhou R. et al., 2020; Costache et al., 2021). We also conducted further BLAST searches in the protein database using an *E*-value cut-off of 1e-005 (Chen et al., 2021). The Hidden Markov Model (HMM) profile for the R2R3 MYB binding domain was downloaded from Pfam³ to verify the data, and protein domains were manually identified using both Pfam and Smart⁴ (Letunic and Bork, 2018; Potter et al., 2018; El-Gebali et al., 2019), ensuring that the putative R2R3 MYB genes contained two MYB DNA-binding domains. Finally, 192 *Populus trichocarpa* and 216 *Salix purpurea* R2R3 MYB proteins were identified. Using these protein sequences, we obtained the corresponding cDNA and genomic sequences,

chromosomal locations, intron distribution patterns, phases, and intron/exon boundaries at the Joint Genome Institute (JGI).⁵ In addition, 126 *A. thaliana* R2R3 MYB protein sequences were downloaded from the *A. thaliana* Information Resource (TAIR)⁶ (Dubos et al., 2010). The R2R3 MYB protein sequences of seven other species were downloaded, including *Physcomitrella patens* (50 members) (Du et al., 2013), *Chlamydomonas reinhardtii* (10 members) (Du et al., 2013), *Vitis vinifera* L. (134 members) (Wong et al., 2016), *Zea mays* (157 members) (Du et al., 2012a), *Brassica napus* (249 members) (Hajiebrahimi et al., 2017), *Malus domestica* (228 members) (Cao et al., 2013), and *Ginkgo biloba* L. (69 members) (Yang et al., 2021).

Chromosomal Distribution of R2R3 MYB Genes

The chromosomal location information of the R2R3 MYB gene family was extracted using the annotation files of the poplar and willow genomes, and the length of each chromosome was obtained from the NCBI website⁷; the gene positions were drafted to chromosomes by using MapChart software (Voorrips, 2002). As described previously, homologous chromosomal segments generated by genome-wide duplication events were identified. Paralogous genes and tandem duplicated genes were identified based on criteria in the poplar and willow genome annotations. Paralogous genes were determined by aligning and phylogenetically analyzing full-length R2R3-MYB proteins (He et al., 2016; Zhou W. et al., 2020). Genes with a distribution on the chromosome in the range of 100 kb and sequences with a sequence similarity greater than 70% were considered to be tandem duplicates (Chai et al., 2014). To further analyze gene duplication events, the paralogous and orthologous gene pairs were aligned using ClustalX2.1 and analyzed using KaKs_calculator software to estimate their synonymous (ks) of evolution (Thompson et al., 1994; Zhang et al., 2006).

Phylogenetic Analysis and Functional Prediction of R2R3 MYB Genes

ClustalX 2.1 was used to perform multiple sequence alignments on the full-length protein sequences of poplar and willow R2R3 MYBs under the default parameters, following which the alignments were manually adjusted before constructing the phylogenetic tree (Thompson et al., 1994). The maximum likelihood (ML) and neighbor-joining (NJ) phylogenetic trees were constructed using MAGE X64 using Poisson correction, pairwise deletion, and bootstrap analysis with 1,000 replicates (Kumar et al., 2018). The R2R3 MYB genes were then classified according to their phylogenetic relationship with the corresponding *A. thaliana* R2R3 MYB genes (Dubos et al., 2010).

All R2R3 MYB protein sequences from poplar, willow, *A. thaliana*, *P. patens*, *C. reinhardtii*, grapevine, maize, *B. napus*, apple, and ginkgo were clustered using OrthoFinder v2 with a cut-off *e*-value of 1×10^{-3} (Emms and Kelly, 2019).

¹https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Ptrichocarpa_er

²https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Spurpurea

³<http://pfam.xfam.org/>

⁴http://smart.embl.de/smart/set_mode.cgi?NORMAL=10

⁵https://phytozome.jgi.doe.gov/pz/portal.html#!search?show=KEYWORD&method=Org_Tcacao

⁶<https://www.arabidopsis.org/>

⁷<https://www.ncbi.nlm.nih.gov/>

The multiple sequence alignments and phylogenetic trees were constructed with MEGA X64 (Kumar et al., 2018) using the full-length sequences of identified R2R3 MYB proteins for each orthogroup. The phylogenetic tree was visualized using EVOLVIEW⁸ and a folding phylogenetic tree, comprised of 10 species was constructed (Subramanian et al., 2019). Certain biological functions of the R2R3 MYB proteins were predicted based on their homologs in the phylogenetic tree.

Gene Structure Analysis and Protein Motif Identification

The R2R3 MYB gene sequences and gene annotation files for poplar and willow can be found in the JGI genome portal. The gene structure data used were selected from a genome sequence annotation file. General Feature Format (GFF) files were used to extract the exact positions of all introns and exons in each predicted gene model. The number and length of the introns and exons were then calculated. Exon length (Exon 1–Exon 5) values were analyzed using Box plot in matplotlib for *Populus*, *Salix*, and *A. thaliana*. The exon-intron organization was determined by the online program Gene Structure Display Server (GSDS)⁹ comparing the predicted coding sequences to their corresponding full-length sequences (Hu et al., 2015).

To investigate conserved motifs in the R2R3 MYB protein sequences in more detail, we used the online program MEME¹⁰ to identify conserved motifs shared among the R2R3 MYB proteins (Bailey et al., 2009). The following parameter settings were used: any for the number of repetitions, the distribution of the motifs, zero or one per sequence, with up to 20 motifs in total. The minimum width of the motif was set at 2 amino acids and the maximum width at 250 amino acids (to identify long R2R3 domains). Other options used the default values. Only hits with e -values $< 1e^{-20}$ were retained for further analysis. The orders of the ID numbers of the R2R3 MYB genes in the exon-intron figures and motif distribution figures are the same as those in the phylogenetic trees. TBtools software graphically represented the motif distribution of R2R3 MYB genes (Chen et al., 2020).

Duplications and Syntenic Analysis of R2R3 MYB Genes

To determine the correspondence between poplar and willow R2R3 MYB genes, we analyzed the synteny and collinearity of these genes between the genomes by using MCScan software (Wang et al., 2012). The software is an algorithm that scans multiple genome sequences or sub-genome sequences to identify putative homologous chromosomal regions and then arranges these regions using genes as anchors (R2R3 MYB transcription factors in our case). We obtained information on the detailed chromosomal location of each R2R3 MYB gene from the genome annotation document. Based on this, the synteny and collinearity relationships of the R2R3 MYB genes of poplar and willow were visualized using Circos software (Krzywinski et al., 2009).

⁸<https://evolgenius.info/evolview-v2/#login>

⁹<http://gsds.cbi.pku.edu.cn/index.php>

¹⁰<http://alternate.meme-suite.org/tools/meme>

Plant Material

Root, stem, and leaf samples were collected from 6-month-old cultured poplar and willow plantlets grown under natural conditions in the greenhouse of Nanjing Forestry University (32°N, 118°W) under natural conditions. The male and female flower buds of the poplar were collected from the *Populus deltoides* planted on the campus of Nanjing Forestry University, and the male and female flower buds of the willow were collected from the *Salix suchowensis* planted in the Baima Base of Nanjing Forestry University. The flower buds were collected from September 2019 to March 2020. The male and female flower buds of both poplar and willow consist of six different developmental stages: T1 (flower bud differentiation stage), T2 (flower bud dormancy stage), and T3–T6 (pollen grain production from formation to maturity). All the collected samples were immediately frozen in liquid nitrogen and stored at -80°C for RNA extraction.

Gene Expression Analysis

Total RNA was extracted from the different poplar and willow tissues using the RNAprep Pure Polysaccharide Polyphenol Plant Total RNA Extraction Kit (TIANGEN, NanJing, China) following the manufacturer's instructions. The RNA quality was assessed by 1.0% agarose gel electrophoresis. The first-strand cDNA was synthesized using 1 μg total RNA and One-Step gDNA Removal and cDNA Synthesis SuperMix (TIANGEN, NanJing, China). The reverse-transcribed cDNA was diluted with ddH₂O at a ratio of 1:10 and used as a template for qRT-PCR. Primers were designed using Primer 5.0 based on the laboratory's existing poplar and willow CDS database, and primer specificity was determined by BLAST. *PtUBQ* (Potri.015G013600.1) and *DnaJ* (SapurV1A.0212s0110) were used as internal control genes for poplar and willow, respectively (Gutierrez et al., 2008; Zhang Y. X. et al., 2017). The total reaction mixture of each PCR experiment comprised of forward and reverse primers (4 μM each), 2 μl of cDNA (diluted 10 times), 10 μl of PowerUpTM SYBRTM Green Master Mix (Applied Biosystems, United States), and RNase-free water to 20 μl . The reaction was conducted in the 7,500 Fast Real-Time PCR System (Applied Biosystems, United States). The qRT-PCR reaction conditions were: 95°C for 3 min, followed by 40 cycles of 95°C for 15 s, 60°C for 15 s, and 72°C for 30 s. At the end of each experiment, the default parameters were used to conduct melting curve analysis for 60 s at 55–95°C in 0.3°C increments. The expression of the genes was analyzed according to the $2^{-\Delta\Delta\text{Ct}}$ method (Livak and Schmittgen, 2001).

Prediction of Cis-Acting Elements in the Promoter Region of GAMYB Genes in Poplar and Willow

The 2,000-bp sequence upstream of the transcription starting point of the poplar and willow GAMYB genes was identified from the databases and the PlantCARE online website¹¹ was used to

¹¹<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>

predict the sequences and *cis*-acting elements of the promoter regions (Lescot et al., 2002).

RESULTS

Identification of *Populus* and *Salix* R2R3 MYB Transcription Factors

According to the known sequence characteristics of the R2R3 MYB domain, 192 and 216 R2R3 MYB transcription factors were identified in the recently updated poplar genome database (version 3.1) and willow genome database (version 1.0), respectively (Tuskan et al., 2006; Zhou et al., 2018; Zhou R. et al., 2020). The identification of the poplar R2R3 MYB genes was consistent with the 192 R2R3-MYB members reported by Wilkins et al. (2009), and two fewer members than Chai et al. (2014). In the reference version (i.e., V3.1), the amino acid sequence of *PttMYB076* does not have the typical R2R3 MYB domain and belongs to 1R-MYB, therefore it was not defined as R2R3 MYB, while *PttMYB235* has no homolog gene in version 3.1. Thus, we only have 192 R2R3 MYB members. The R2R3 MYB domain sequence in poplar and willow was found to be analogous to that of other plant species, containing approximately 106 amino acids with only a few deletions or insertions. The domain contained characteristic amino acid residues with evenly distributed and highly conserved Trp (W) residues (Chen et al., 2021). Based on the HMM results, we manually verified the accuracy of the sequences and deleted incomplete R2R3 MYB domain sequences and R-R-type groups not listed as R2R3 MYB but classified as Atypical-MYB (Jiang and Rao, 2020). The accession numbers of the R2R3 MYB genes from poplar and willow are listed in **Supplementary Tables 1, 2**, respectively, together with information of their positions on the chromosomes. The amino acid sequences and CDS sequences of the poplar proteins are listed in **Supplementary Sequences 2, 3**, respectively, and the corresponding willow sequences in **Supplementary Sequences 4, 5**. The accession numbers and corresponding amino acid sequences of 126 R2R3 MYB genes in *A. thaliana* are listed in **Supplementary Sequence 5** (Dubos et al., 2010).

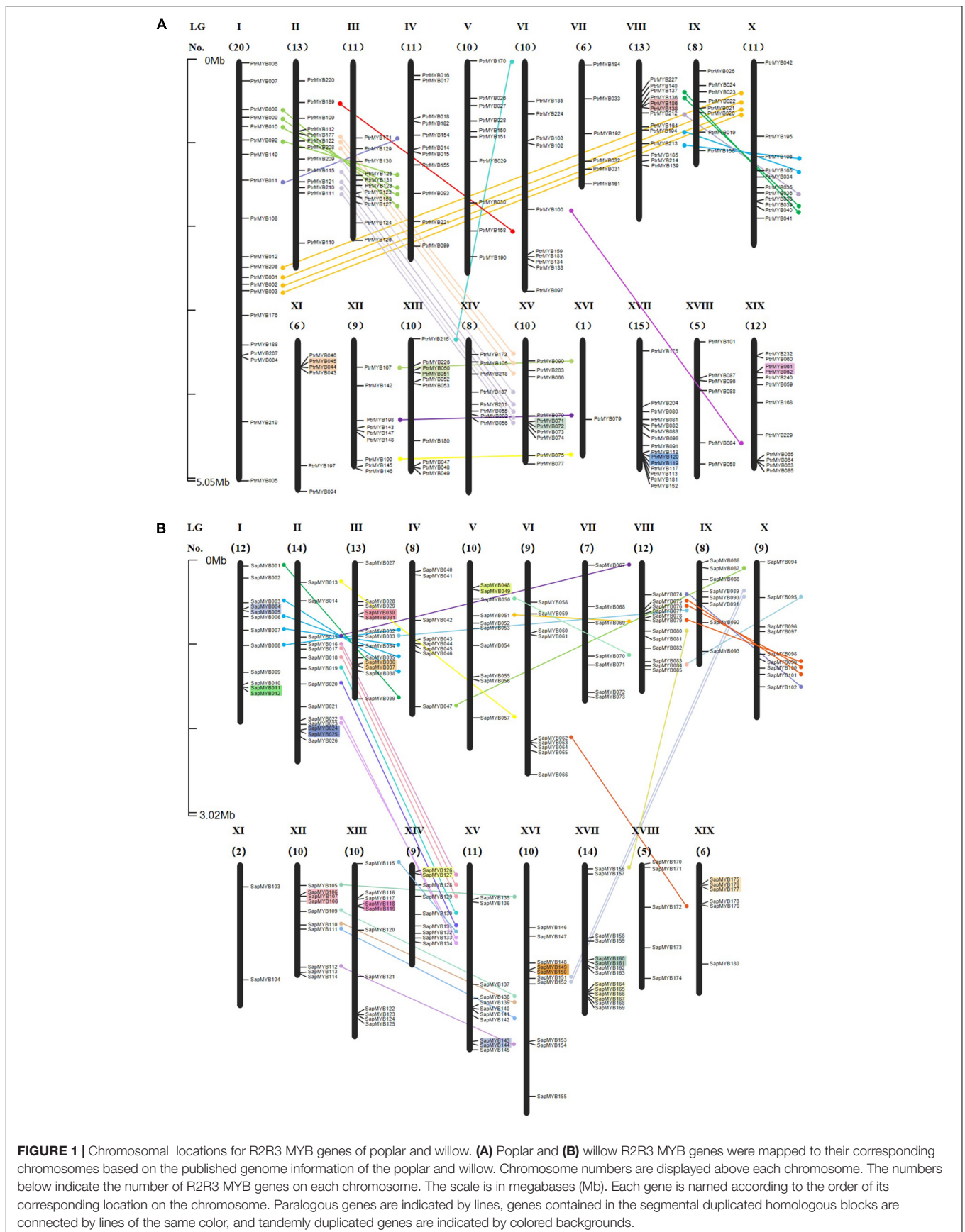
Chromosomal Distribution and Identification of Paralogous Gene Pairs in *Populus* and *Salix* R2R3 MYB Transcription Factors

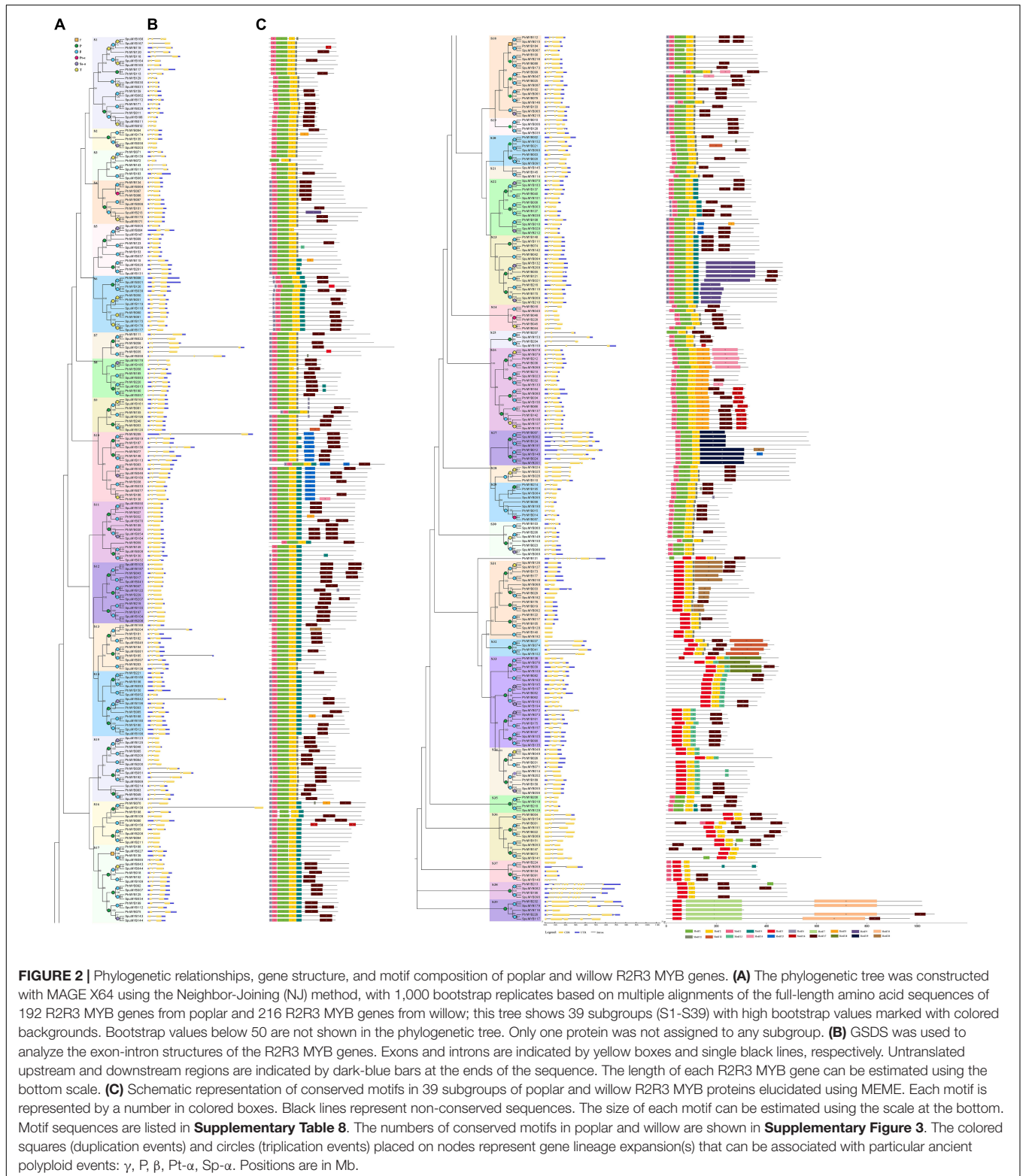
To date, although there have been reports on expansion events in the poplar R2R3 MYB family (Chai et al., 2014), the situation in the willow is still unclear. To determine the relationship between genetic divergence and gene duplication within the R2R3 MYB family in the Salicaceae and using the genome annotation information, 189 poplar R2R3 MYB genes (**Figure 1A**) and 180 willow R2R3 MYB genes (**Figure 1B**) were mapped to the corresponding 19 chromosomes, respectively. According to the order of their corresponding locations on the chromosome, the willow R2R3 MYB genes were named *SpuMYB001* to *SpuMYB180*. 36 R2R3 MYB genes in the willow that could not be mapped to a chromosome, remaining only on unassembled

scaffold fragments, were named *SpuMYB181* to *SpuMYB216*. The poplar has only three R2R3 MYB genes that are not mapped to chromosomes, and the naming of the poplar R2R3 MYB genes refers to published articles (Wilkins et al., 2009; Chai et al., 2014). The number of willow R2R3 MYB genes that were not mapped to chromosomes was significantly higher than that in the poplar, which may be the reason for the incomplete assembly of the willow genome. We found that the genes mapped on the chromosomes were unevenly distributed. The distribution numbers of the R2R3 MYB genes in the poplar and willow were compared, showing that there were significant differences in the number of genes on chromosomes I and XVI with the poplar having 20 R2R3 MYB genes on chromosome I while there were only 12 on chromosome I in the willow. In contrast, only one R2R3 MYB gene was identified on chromosome XVI in poplar while the willow had 10 genes on chromosome XVI. These observations suggest that chromosomes I and XVI undergo chromosomal rearrangement (Dai et al., 2014).

The presence of paralogous genes is the result of gene family expansion, which is mainly produced by segmental duplication and tandem duplication events. One or more genes located on the same chromosome within 100 kb and having high sequence similarities (> 70%) with their counterparts were considered to be tandem duplicates, denoted as the “T” event (Liu et al., 2014; Wang M. et al., 2018). A segmental duplication event is produced by a genome-wide replication event (Magwanga et al., 2018). The gene duplication seen in the poplar indicates a single genome-wide event. We refer to this duplication event as a “salicoid” duplication event, that is, the “P” event, which occurred around 58 Mya (Tuskan et al., 2006). About 6 Mya after the “salicoid” duplication event, two major inter-chromosomal rearrangements and several minor intra-chromosomal rearrangements occurred. After this chromosomal rearrangement event, called the “ β ” event, the ancestors of the modern willow appeared, indicating that willow and poplar were differentiated from a common paleotetraploid ancestor (Hou et al., 2016). The genome duplication in poplar and willow is very recent, occurring around 8–13 Mya, defined as the “Pt- α ” events and “Sp- α ” events, respectively (Sterck et al., 2005; Berlin et al., 2010). The Ks value reflects the generation time of the homologous gene to some extent and is often used as the basis for assessing the gene duplication events (Cui et al., 2006; Hou et al., 2015).

In this study, paralogous gene pairs were identified to determine the different types of duplication responsible for the R2R3 MYB family expansion in poplar and willow and the reasons for the differences in the numbers of R2R3 MYB genes. Using whole-genome analysis of gene duplications, we identified 81 pairs of paralogous genes in the poplar R2R3 MYB family (**Supplementary Table 3**), while 86 pairs of paralogous genes were identified in the willow (**Supplementary Table 4**). There were 70 pairs of homologous genes observed in the poplar, generated by the “salicoid” duplication event in the Salicaceae around 58 Mya. In addition, the recent poplar genome replication event i.e., Pt- α produced 6 pairs of tandem repeat genes and 3 pairs of paralogous genes (**Figure 2A**). In addition, two additional pairs of homologous genes were generated in





the γ triplication event shared by the poplar, willow, and *A. thaliana*. The rate of evolution of genes in the willow genome was about twice that of poplar genome (Hou, 2016). There were four pairs of paralogous genes in the willow that were

produced in the γ triplication event and 42 pairs generated in the “salicoid” duplication event. Surprisingly, 17 pairs of tandem repeat genes were identified in the willow, indicating that the willow R2R3 MYB family has been actively replicated. The

amino acid sequences of *SpuMYB126* and *SpuMYB127* and *SpuMYB160* and *SpuMYB161* are the same; their corresponding cDNA and DNA sequences are also very similar (100 and 99.62%, 100 and 99.9%, respectively). At the same time, three (*SpuMYB175*, *SpuMYB176*, *SpuMYB177*) tandem repeats and four (*SpuMYB164*, *SpuMYB165*, *SpuMYB166*, *SpuMYB167*) tandem repeats were identified. Their amino acid sequences are highly similar, indicating recent replication events and providing strong evidence that active gene replication has contributed significantly to the expansion of the willow R2R3 MYB gene family. After genome duplication, antagonism eliminates most of the duplicate genes, so there are some genes produced by dispersed duplication. Different replication events appear to have led to the complexity of the R2R3 MYB genes in the willow genome (Dai et al., 2014; Hou et al., 2015). It also provides a theoretical basis suggesting that the willow has undergone more environmental adaptation than the poplar and has evolved more varieties. poplar contains 5 groups with about 40 species. Compared with poplar, willow has a wide variety of species, divided into 33–38 groups, and comprising about 520 species (Skvortsov, 1999; Hou, 2016).

Synteny Analysis of R2R3 MYB Transcription Factors and Identification of Orthologous Gene Pairs

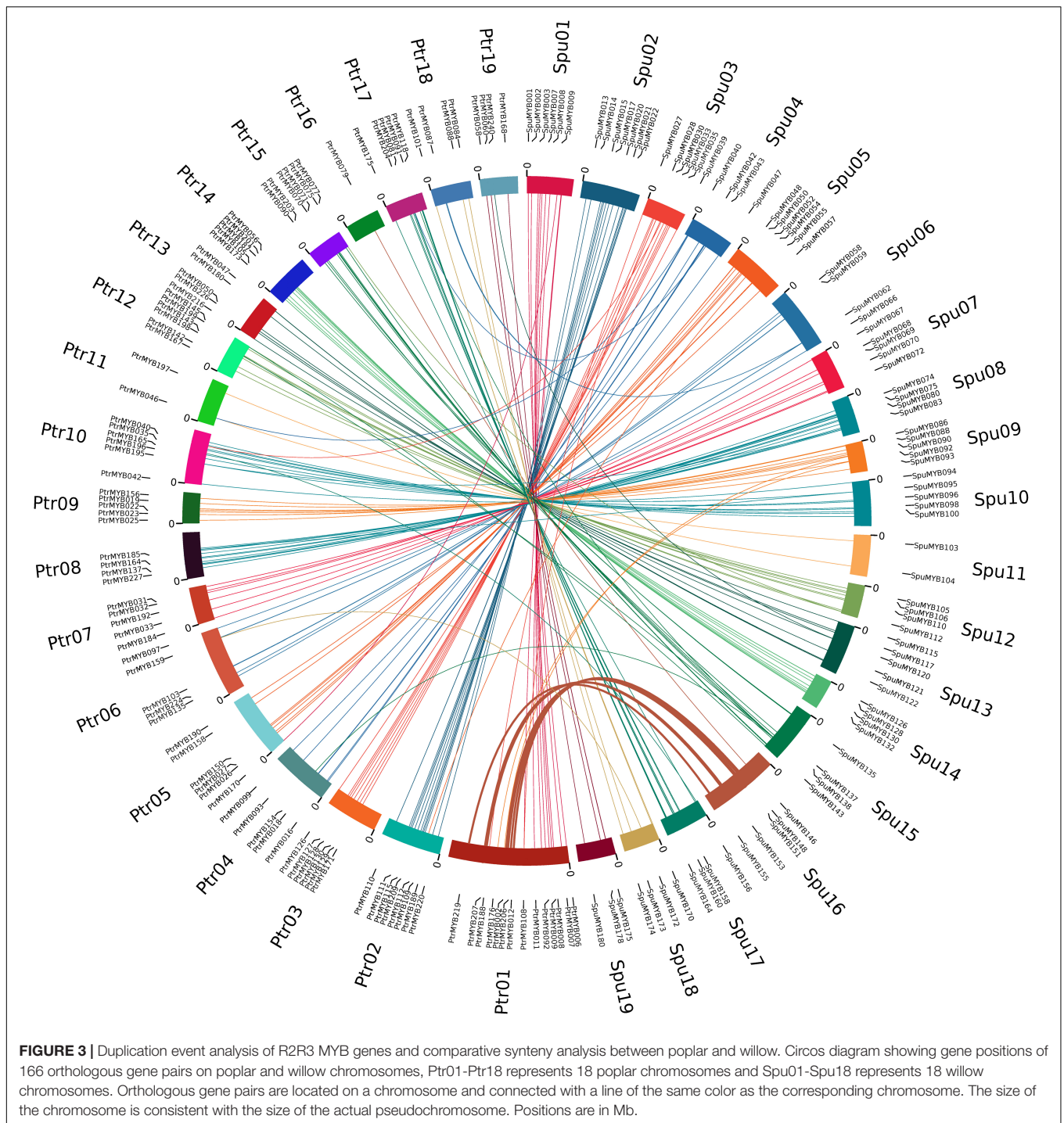
Genome-wide replication events in plants have been considered as a mechanism for diversification and adaptation to the environment. Although we found that the R2R3 MYB gene family of the willow had undergone more active replication than the poplar and that there were large phenotypic changes during their evolution, the two species also share many common characteristics, including the same chromosome number of $2n = 38$ and the common “Salicoid” genomic duplication that occurred around 58 Mya, resulting in high macroscopic homology, and they also retained similar characteristics in terms of growth, reproduction and other developmental processes (Tuskan et al., 2006; Dai et al., 2014). To further analyze the diversity and evolutionary conservation of members of the R2R3 MYB gene family between the poplar and willow, we used OrthoFinder V2 analysis, combined with phylogeny, gene structures, and sequence similarities to identify the orthologous genes between the two species. The results showed that most of the R2R3 MYB genes were preserved during the evolution of the poplar and willow. We identified 166 pairs of R2R3 MYB orthologous genes between the poplar and willow. In addition to 117 pairs of one-to-one orthologs, the genomes also contain 45 pairs of one-to-many and 4 pairs of many-to-many co-orthologous genes (Supplementary Table 5), indicating that these genes have expanded to varying degrees. Among them, 158 pairs of orthologous genes were generated during the β event of the fission or fusion of chromosomes in the ancestral genome 6 Mya (Figure 2A). There was also high bootstrap support ($\geq 90\%$) with short branch lengths at the ends of the branches, indicative of high homology and recent co-evolutionary origins (Figure 3). Comparison of the R2R3 MYB genes from poplar and willow

revealed that genes with orthologous relationships tended to be clustered together in the phylogenetic tree in contrast to the paralogous genes, indicating that amplifications in the R2R3 MYB gene family were more likely to have occurred in the ancestral species.

During evolution, large segmental duplications and small-scale tandem duplications have been the two main mechanisms for generating new genes. We found that the orthologous R2R3 MYB genes on the poplar chromosomes II–XV and XVII–XIX had similar distributions on the corresponding willow chromosomes, suggesting that they originated from the same tetraploid ancestor. However, no orthologous genes were found on the lower parts of chromosome 1 in both and willow although seven pairs of orthologous genes were detected on chromosome XVI in the willow. This chromosomal distribution of poplar and willow orthologous genes is consistent with previous reports that have described major inter-chromosomal rearrangements and chromosome fission and fusion involving chromosomes I and XVI between the two genomes (Dai et al., 2014; Hou et al., 2016). In addition, we have also identified orthologous genes on other corresponding chromosomes of the poplar and willow; although there are not as many genes as between chromosomes I and XVI, this confirms the conclusion reached by Hou et al. (2019) that, in addition to the large degree of recombination between the two important chromosomes in the poplar and willow, subtle changes between corresponding chromosomes may have also occurred (Hou et al., 2019). This provides a reference for the study of R2R3 MYB differentiation mechanisms after genome-wide replication events in the Salicaceae.

Phylogeny, Conserved Gene Structures, and Protein Motif Analysis of *Populus* and *Salix* R2R3 MYB Transcription Factors

Members of the R2R3 MYB gene family in the poplar and willow were subdivided into 39 subgroups based on clades with at least 50% bootstrap support and named S1–S39 (Figure 2A). In our phylogenetic analysis, only the poplar gene *PtrMYB130* did not fit into any subgroup. In addition, the tree (Supplementary Figure 1) topology using maximum likelihood (ML) analysis was essentially the same as the NJ tree (Figure 2A), indicating that the two phylogenetic trees are in good agreement. To analyze the similarity and diversity of the gene structures within the subgroups, we used the GSDS tool to draw the exon/intron schematic structures of 408 R2R3 MYB genes in poplar and willow (Figure 2B). Usually, there is at least one intron in the R2R3 MYB DNA-binding domain, but in our analysis, we found that 5.7% (11/192) and 4.2% (9/216) of poplar and willow R2R3 MYB genes, respectively, lacked introns. In addition, the most common R2R3 MYB intron-exon structure, seen in 73% (140/192) of poplar and 72% (156/216) of willow genes, was a typical splicing of 3 exons and 2 introns. Some R2R3 MYB genes were found to have 2 or 4 exons. However, a total of 5 R2R3 MYB sequences (*PtrMYB131*, *PtrMYB213*, *SpuMYB082*, *PtrMYB196*, and *SpuMYB096*) possessed more than 6 introns. Among them, *PtrMYB213* and *SpuMYB082*, *PtrMYB196*, and *SpuMYB096* were



identified as orthologous genes. They formed a separate subgroup (S38) containing a complex exon-intron structure. Although there are some differences in their intron-exon structure, it is worth noting that the results of the phylogenetic analysis of the poplar and willow R2R3 MYB family indicated that genes in the same subgroups usually contained very similar exon-intron patterns, a pattern that appeared to be conserved. For example, the R2R3 MYB genes in subgroup 4 (S4) had 3 exons and 2

introns, those in subgroup 34 (S34) had 2 exons and 1 intron, while the genes with the most exons were all found in subgroup 38 (S38). The analysis of the exon-intron structures of the R2R3 MYB genes supports our phylogenetic analysis. Our data indicate that there were some subgroups (i.e., S7, S10, S13, S14, S16, and S33) of R2R3 MYB with a specific exon-intron structure in which the introns were significantly longer. We speculate that the introns in the R2R3 MYB domain may play an important role in

the evolution of the R2R3 MYB gene family through unknown mechanisms. Simultaneously, we found that of the 166 orthologs identified, most gene pairs had the same exon-intron pattern and were highly conserved. The R2R3 MYB genes are not only highly conserved in their exon-intron structures but are also conserved in length. Exons 1 and 2 usually encode almost the entire R2R3 MYB DNA-binding domain, making the first two exons almost the same size in plant species. Generally, exons 1 and 2 are very similar in length (exon 1, 132 bp; exon 2, 129 bp) and are also highly conserved (exon 1, 34% in poplar, 38% in willow; exon 2, 61% in poplar, 62% in willow). Exon 3 has the greatest diversity in sequence length (poplar, 30–1,131 bp; willow, 33–1,131 bp) (**Supplementary Figure 2**). We found that the exon lengths of the R2R3 MYB family members in poplar and willow are consistent with the average exon lengths in *A. thaliana* and grapes (**Supplementary Figure 2**; Matus et al., 2008). The restricted lengths of exons 1 and 2 in R2R3 MYB genes explain the high degree of conservation in the R2R3 MYB domain seen during plant evolution, suggesting that these conservative features play key roles in specific subgroup functions. Exons 3–12 encode the last region of the R3 repeat and the C-terminal region of the R2R3 MYB protein. The length and sequence differences in the exons can lead to different functional motifs and domains, which can explain functional divergence between R2R3 MYB homologs within and between plant species.

The C-terminal region next to the MYB domain in R2R3 MYB proteins usually has large sequence variations, although the domains in each protein subgroup contain functionally important motifs that are less conserved than those in the MYB domain. To analyze the conserved motifs in R2R3 MYB proteins, a total of 20 conserved motifs in the poplar and willow R2R3 MYB proteins were identified using the online MEME tool (**Figure 2C**). Comparing the distributions of 20 motifs between the poplar and willow, it was found that the number of motifs was similar between the two species and the motifs did not differ significantly in composition. The lengths of these 20 motifs varied between 8 and 250 amino acids, and the number of motifs in each R2R3 MYB sequence varied between 3 and 8. Most R2R3 MYB proteins had motifs 1, 2, 3, 6, 11, and 17 (**Supplementary Figure 3**), 1, 2, and 3 constituting the conserved R2 and R3 domains of “W-(X19)-W-(X19)-W-...-F/I-(X18)-W-(X18)-W-“. Motif 6 was also identified as the MYB domain, and the other motif functions are unknown. However, while motifs 1, 2, and 3 directly followed the MYB domain in the sequence and were particularly conserved in most subgroups, most motifs were selectively distributed between specific clades of the phylogenetic tree, with each R2R3 MYB subgroup having a common motif. For example, in subgroup 15 (S15), all R2R3 MYB proteins contained motifs 1, 2, 3, 6, 11, and 17, and in subgroup 12 (S12), all the proteins contained motifs 1, 2, 3, 4, 6, and 17. However, certain motifs appear to be unique to certain subgroups, for example, motif 19 was found in subgroup 27 (S27), while motifs 7 and 10 were only identified in subgroup 39 (S39). Some of the genes present in the same group did encode proteins that differed in their shared motifs. For example, *PtrMYB072* in subgroup 3 (S3) lacked motif 3, *SpuMYB211* in subgroup 16 (S16) lacked motif 6, *PtrMYB012* in subgroup 27 (S27) had an additional

motif 20, and *SpuMYB148* and *PtrMYB168* in subgroup 14 (S14) had additional motifs 15 and 8, respectively. Generally, most R2R3 MYB genes clustered in the same subgroup showed common motif characteristics, implying that members of the same subgroup have similar functions. Subgroup 39 (S39) was located at the far end of the phylogenetic tree, and most of its members (4/5) had a very long C-terminal region. MEME search results found two unexpectedly large, conserved motifs (motifs 7 and 10) in the C-terminal region of this subgroup, and a large, conserved motif (motif 20) was found in subgroup 27 (S27). It is possible that this may indicate the presence of specific functions in this R2R3 MYB subgroup, however, the biological functions of these proteins require further verification.

Putative Functions of *Populus* and *Salix* R2R3 MYB Transcription Factors

Amino acid changes in motifs and domains, as well as sequence length diversity are sources of functional differences. To examine the evolutionary relationships among the identified poplar and willow R2R3 MYBs and R2R3 MYBs from other species, the R2R3 MYB genes of 10 species (poplar, willow, *A. thaliana*, *P. patens*, *C. reinhardtii*, grapevine, maize, *B. napus*, apple, and ginkgo) were analyzed. According to the orthogroups identified by Orthofinder V2 and with reference to previous research on the R2R3 MYB gene family in different species (Stracke et al., 2001), the R2R3 MYB genes were subdivided into 53 subgroups (**Figure 4**). Since R2R3 MYB proteins in the same subgroup usually control the same molecular regulatory pathway, our results will provide support for the functional prediction of these poplar and willow R2R3 MYB genes. To predict the biological functions of poplar and willow R2R3 MYB, we referred to gene functions that have been fully verified in *A. thaliana* as well as gene functions that have been reported in other species, resulting in the functional annotation of 30 subgroups. For example, members of subgroup 5 (S5) and subgroup 6 (S6) are related to anthocyanin biosynthesis (He et al., 2021), subgroup 1 (S1) members participate in responses to biotic and abiotic stress (Froidure et al., 2010), while subgroup 9a (S9a) members are associated with trichome development (Wu et al., 2018; Zhou et al., 2021), while members of subgroup 37 (S37) and subgroup 46 (S46) all play key roles in lignin biosynthesis (Tohge et al., 2005; Zhou et al., 2009).

Although the number of genes contained in each subgroup was similar between the poplar and willow, we nevertheless identified several evolutionarily different subgroups. For example, the number of willow genes in subgroup 34 (S34) was double that of the poplar, indicating that after willow differentiation, the willow genes in this subgroup replicated more actively in this subgroup than in poplar, including two pairs of tandem repeat genes (*SpuMYB011* and *SpuMYB012*, *SpuMYB030* and *SpuMYB031*). Remarkably, several clades (S13, S28, S32, S34, and S37) of R2R3 MYB genes have undergone different degrees of expansion in woody plants, especially in the poplar and willow. It may be that many genes evolved differently after differentiation from the common ancestor, including some lineage specific gene duplication and gene loss events

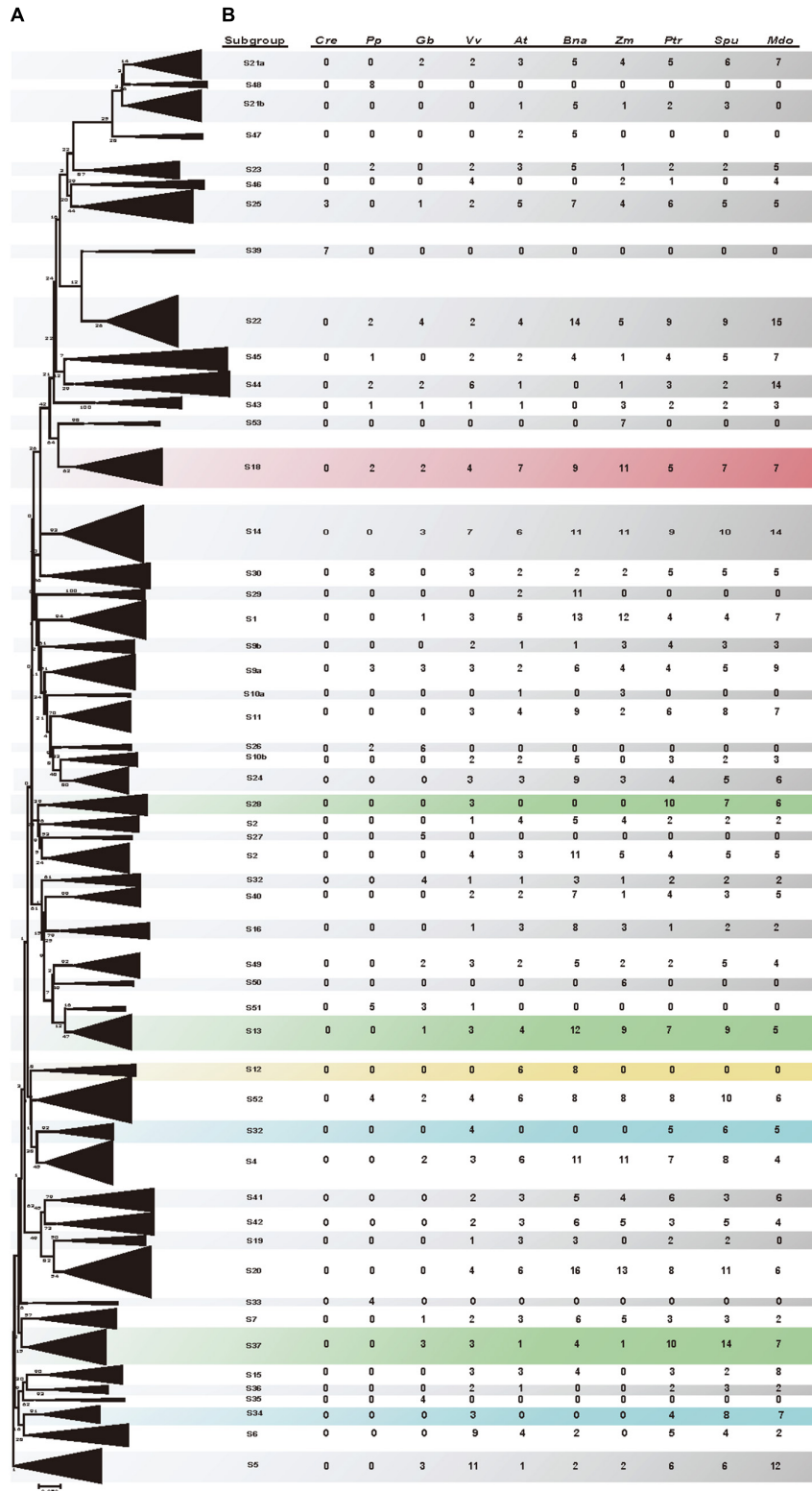


FIGURE 4 | Phylogenetic tree of poplar, willow, *A. thaliana*, *P. patens*, *C. reinhardtii*, grapevine, maize, *B. napus*, apple, and ginkgo R2R3 MYB proteins. **(A)** Neighbor-joining phylogenetic tree (Supporting sequences 4) and Maximum likelihood (ML) phylogenetic tree (Supporting sequences 5) constructed using 1,431 amino acid sequences (Supporting sequences 5) including all of the R2R3-MYB proteins. Folding triangles represent 53 subgroups of the R2R3 MYB protein cluster, consistent with OrthoFinder clustering orthogroup, and prediction of the function of the subgroups, each triangle represents a R2R3-MYB subgroup. *(Continued)*

FIGURE 4 | (Continued)

Supplementary Figure 4 shows an unfolded evolutionary tree with full classification of members. **(B)** Subgroup names are included next to each clade together with a short name to simplify nomenclature. The number of genes of each species for each subgroup is also included. Subgroups in general expanded in woody species and have been confirmed to be related to lignin synthesis are highlighted in green, and subgroups in general expanded in woody species and predicted to be related to wood development are highlighted in blue. The specific subgroups of Cruciferae are shown in yellow, and the subgroups related to stamen development are shown in red.

(Soler et al., 2015). Among them, subgroups 13, 28, and 37 (S13, S28, and S37), which are related to xylem development, have significantly expanded. It was found that the R2R3 MYB genes of subgroup 28 (S28) were highly expressed in vessels and fiber during wood formation, which could regulate the biosynthesis of lignin, cellulose, and xylose (McCarthy et al., 2010). Although the members of the subgroup 32 (S32) and subgroup 34 (S34) have not yet undergone functional verification, the grape, poplar, willow, and *A. thaliana* all evolved from the same ancient hexaploid ancestor. Both subgroups in the grape have R2R3 MYB members. In contrast, there were no R2R3 MYB members in *A. thaliana*, but more genes are replicated in the poplar and willow, so we speculate that the genes of subgroup 32 (S32) and subgroup 34 (S34) likely have similar functions and participate in wood formation. This further suggests that the genes of these subgroups may have specialized roles that were either acquired in poplar and willow or lost in *A. thaliana* after the divergence from the last common ancestor, and they are good candidates for further functional and phylogenetic analysis. In contrast, we also found that certain clades showed a bias toward *A. thaliana* species in the phylogenetic trees. For example, while subgroup 12 (S12) included six *A. thaliana* R2R3 MYB proteins and eight *B. napus* R2R3 MYB proteins, it did not include any poplar or willow genes, nor genes from other species. Recently, it has been reported that the members of this subgroup *AtMYB34*, *AtMYB51*, and *AtMYB122* are involved in the biosynthesis of glucosinolates (Frerigmann and Gigolashvili, 2014). This hypothesis was based on the understanding that glucosinolate compounds are Cruciferae plant-specific (Perez et al., 2021). Although the evolutionary links of all the subgroups are unclear, this study provides the groundwork for future functional studies of poplar and willow R2R3 MYB genes.

Identification of R2R3 MYB Candidate Genes Related to Stamen Development in Poplar and Willow

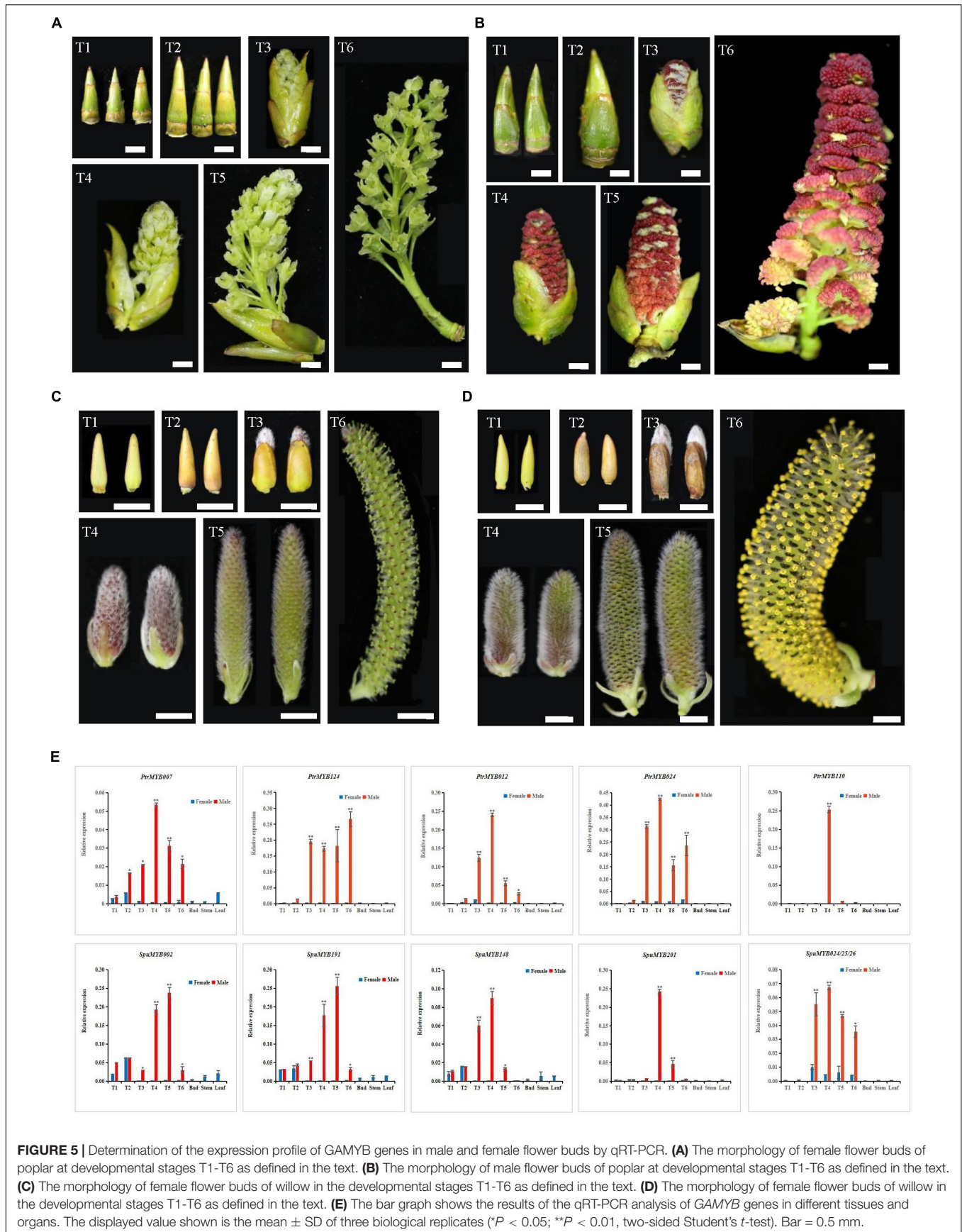
Both poplar and willow are dioecious. The male plants are known to produce pollen, which is an important allergen. Therefore, we chose the GAMYB gene to understand the genes' potential regulatory role in the development of male flowers. We used qRT-PCR technology to analyze the expression patterns of this subgroup of genes in six different developmental stages of male and female flower buds, leaf buds, stems and leaves of both poplar and willow.

The results showed that the expression levels of 12 poplar and willow R2R3 MYB genes the subgroup 18 (S18) differed in different organs and tissues (**Figure 5**). The vast majority of these R2R3 MYB genes had relatively low expression levels in different developmental stages of female flower buds, flower bud

differentiation (T1) and dormancy (T2) stages of male flower buds as well as in leaf bud, stem, and leaf. From pollen formation to maturity (T3-T6), the expression levels first increased and then decreased, with similar expression patterns in both poplar and willow. *PtrMYB124* had a high expression level during the entire pollen development period, with its highest expression during the pollen maturity stage. The sequence similarities between *SpuMYB024*, *SpuMYB025*, and *SpuMYB026* were as high as 98.91%, and it was difficult to design specific primers to distinguish them. Therefore, we designed universal primers for these three genes to analyze their expression patterns and found that, compared with other GAMYB genes, these three genes were not only expressed in male flowers but were also detected in female flowers, although their expression in the latter was very low. It is speculated that these genes may have acquired different functions during evolution and thus may have a wider range of functions. The tissue specificities and high similarities of GAMYB genes expression in poplar and willow indicate functional conservation in members of this subgroup. Combined with the functional analysis of GAMYB genes in other species. This indicates that the GAMYB genes play an important role in the development of poplar and willow pollen and is an ideal candidate gene for the functional analysis of male flower development in poplar and willow.

Cis-Element Identification in the Promoter Regions of GAMYB Genes

Cis-acting regulatory elements play an important role in the overall regulation of gene expression. Evidence indicates that genes with similar expression patterns may contain the same regulatory elements in their promoters. The GAMYB gene also plays an important role in the development of flowers, especially the development of anthers. To further understand the possible regulatory mechanism of GAMYB in the development of male poplar and willow flower buds, the PlantCARE database was used to identify the cis-acting regulatory elements (CREs) at 2,000 bp upstream of these genes, and a total of 1543 CREs were identified (**Supplementary Table 6**). It was found that all poplar and willow GAMYB promoter sequences contained TATA-box and CAAT-box core elements. Most genes also contain multiple types of MYB-related elements such as MBS, MYB, and MYC (**Supplementary Figure 6**). GAMYB is a gibberellin-dependent α -amylase transcription positive regulator, which can specifically bind to the GA response element in the promoter region of the α -amylase gene and other GA-regulated hydrolase genes (Murray et al., 2003). The TATC-box and P-box are cis-acting elements involved in the gibberellin reaction (Chen et al., 2021). We found that 9 GAMYB gene promoters contained one or more of these two elements, of which *SpuMYB002* contained 4 P-box



and 1 TATC-box elements, the *PtrMYB043* promoter contained 3 P-box, *PtrMYB096* and *SpuMYB191* both contained 2 P-box, while *PtrMYB002*, *PtrMYB010*, *SpuMYB024*, *SpuMYB025*, and *SpuMYB026* each contained one P-box element. This indicates that these genes are combined with GA response elements to participate in the GA signaling pathway and regulate the development of flower buds by regulating the expression of genes in this pathway. It has been found that other GAMYB gene promoters contain other hormone response-related elements. For example, *SpuMYB201* contains the auxin response element TGA-element, and *PtrMYB033* contains the cis-acting element ABRE that participates in the abscisic acid response. MYB transcription factors usually form MBW complexes with bHLH and WD40 (Liu Y. F. et al., 2021). The bHLH transcription factor mainly regulates downstream target genes by recognizing the E-box (CANNTG). Among them, *PtrMYB012*, *PtrMYB110*, *SpuMYB024*, *SpuMYB025*, *SpuMYB026*, *SpuMYB201*, and *SpuMYB002* all contain G-box cis-elements, indicating that these genes may be regulated by bHLH transcription factors.

DISCUSSION

The “salicoid” duplication event in the Salicaceae family occurred around 58 Mya (Tuskan et al., 2006; Dai et al., 2014). About 6 Mya after the whole-genome duplication, chromosome breakage and fusion occurred in the paleotetraploid genome, and the genome doubled again (Tuskan et al., 2006; Dai et al., 2014). During the subsequent evolution, chromosome 1 of the poplar tree had a centromere break, and the lower part of chromosome I fused with the telomere of chromosome XVI to form chromosome I of willow, the upper part of which evolved to form the willow chromosome XVI (Hou et al., 2015, 2016). This chromosomal rearrangement event marked the appearance of the ancestors of the modern willow, indicating that the willow resulted from the poplar chromosome breakage and fusion. During the subsequent evolution after this differentiation, both poplar and willow not only retained the common Salicaceae characteristics but also evolved their own unique characters to adapt to the environment. Therefore, the comparative analysis of the important gene families of the poplar and willow is helpful to identify the retention and loss of key gene family members after the differentiation of the two species. The MYB transcription factor superfamily is one of the largest gene families in plants, and R2R3 MYB transcription factors participate in plant cell differentiation, growth, metabolism, and various stress responses (Millard et al., 2019a; Chang et al., 2020). Through comparative analysis of the R2R3 MYB transcription factor family in the poplar and willow, it was found that they have both retained most of the gene members of the ancestral species. It was found the family expanded faster in the willow than in the poplar through tandem replication or chromosome segment duplication. The mean substitution rates of willow and poplar are 1.09×10^{-9} and 0.67×10^{-9} /site/year, respectively (Dai et al., 2014). The mean substitution rate in the willow was significantly higher than that in the poplar, which indicated that the evolutionary rate in the willow was faster than that in the poplar (Dai et al., 2014).

Therefore, we infer that the rapid expansion of the R2R3 MYB gene family during the rapid evolutionary process in the willow may have resulted in more sub-functionalization, allowing the newly generated gene family members to be preserved and promoting better adaptation to a complex environment.

We combined our analysis with the poplar, willow, *A. thaliana*, *Physcomitrella patens*, *Chlamydomonas reinhardtii*, grapevine, maize, *Brassica napus*, apple, and ginkgo R2R3 MYB gene families for a comprehensive classification and evolutionary analysis. Phylogenetic analysis divided the R2R3 MYB transcription factors in the 10 species into 53 subgroups. Due to the conservation of the R2R3 MYB genes, genes with similar or identical functions were classified into the same subgroup, providing a reliable basis for studying the functions of related genes in this gene family. Thus, the R2R3 MYB members of subgroups (S13, S28, S31) have been confirmed to be involved in lignin biosynthesis in other species, and these subgroups have expanded to different degrees in the poplar and willow. Therefore, it is speculated that the R2R3 MYB genes of the poplar and willow clustered in the same subgroup are also related to wood development (Zhou et al., 2009). The *A. thaliana* R2R3 MYB genes of subgroup 6 (S6), *AtMYB75* (*PAP1*), *AtMYB90* (*PAP2*), *AtMYB113*, and *AtMYB114*, are the main transcription factors regulating anthocyanin biosynthesis (Borevitz et al., 2000; Bhargava et al., 2010). As it is speculated that there are functional similarities between poplar and willow and their R2R3 MYB members on the same branch, these genes may play important roles in regulating anthocyanin biosynthesis in poplar and willow. In subgroup 1 (S1), *AtMYB30*, *AtMYB60*, and *AtMYB96* can activate transcription factors that respond to stress and play important roles in plant growth under adverse conditions (Raffaele and Rivas, 2013; Rusconi et al., 2013). It is speculated that poplar and willow members of this subgroup with close genetic distances have similar functions and can respond to stress conditions. In addition, *AtMYB33* and *AtMYB65* of subgroup 18 (S18) participate in the GA signal transduction pathway to control flowering, resulting in the name of GAMYB for these genes (Gao et al., 2020). This gene subgroup has been found to have similar regulatory roles in *H. vulgare*, *A. thaliana*, *O. sativa*, and *C. sativus*. GAMYB genes have been shown to play important roles in the development of flowers, especially the development of anthers (Liu et al., 2010; Zhang et al., 2014; Abraham et al., 2016). Therefore, the study of the flowering mechanism in the poplar and willow is of significant value.

This study analyzed the expression patterns of the GAMYB gene in subgroup 18 (S18) at different developmental stages of male and female flower buds of poplar and willow. It is speculated that they play a role in the process of male flower formation. We further predicted the cis-acting elements in their promoter regions and found that most of the GAMYB gene promoter regions contain cis-acting elements related to hormone response, with 75% of the GAMYB gene promoters containing the cis-acting elements TATC-box and P-box related to the GA signaling pathway. Similarly, the GAMYB gene promoter sequence also contained numerous other cis-acting elements related to light response, plant growth and development, regulation, and the stress response. This implies that while members of this subgroup

may have different biological functions and abilities to participate in signaling pathways, most are involved in GA signaling to regulate flowering pathways. We next intend to verify these gene functions through transgene studies and protein interactions, amongst other methods, which will lay a foundation for the understanding of the role of R2R3 MYB transcription factors in regulating male flower development and provide directions for breeding new non-pollen-producing varieties of male poplar and willow plants.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

TY and HW designed the research and revised the manuscript. FZ conducted the experiments and data analysis, and drafted the manuscript. YC participated in data analysis. All authors reviewed and approved the final manuscript.

FUNDING

This study was financially supported by the Natural Science Foundation of China (31800562), the Key Research and Development Project of Jiangsu Province, China (BE2021366), the National Key Research and Development Plan of China (2016YFD0600101), and the Youth Science and Technology Innovation Fund of Nanjing Forestry University (CX2018008).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.721558/full#supplementary-material>

REFERENCES

- Abraham, Z., Iglesias-Fernández, R., Martínez, M., Rubio-Somoza, I., Diaz, I., Carbonero, P., et al. (2016). A Developmental Switch of Gene Expression in the Barley Seed Mediated by HvVP1 (Viviparous-1) and HvGAMYB Interactions. *Plant Physiol.* 170, 2146–2158. doi: 10.1104/pp.16.00092
- Aharoni, A., De Vos, C. H. R., Wein, M., Sun, Z., Greco, R., Kroon, A., et al. (2001). The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. *Plant J.* 28, 319–332. doi: 10.1046/j.1365-313x.2001.01154.x
- Ambawat, S., Sharma, P., Yadav, N. R., and Yadav, R. C. (2013). MYB transcription factor genes as regulators for plant responses: an overview. *Physiol. Mol. Biol. Plants* 19, 307–321. doi: 10.1007/s12298-013-0179-1
- Aya, K., Hiwatashi, Y., Kojima, M., Sakakibara, H., Ueguchi-Tanaka, M., Hasebe, M., et al. (2011). The Gibberellin perception system evolved to regulate a pre-existing GAMYB-mediated system during land plant evolution. *Nat. Commun.* 2:544.
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208.
- Baldoni, E., Genga, A., and Cominelli, E. (2015). Plant MYB transcription factors: their role in drought response mechanisms. *Int. J. Mol. Sci.* 16, 15811–15851. doi: 10.3390/ijms160715811
- Baumann, K., Perez-Rodriguez, M., Bradley, D., Venail, J., Bailey, P., Jin, H. L., et al. (2007). Control of cell and petal morphogenesis by R2R3 MYB transcription factors. *Development* 134, 1691–1701. doi: 10.1242/dev.02836
- Berlin, S., Lagercrantz, U., Arnold, S. V., Ost, T., and Rönnerberg-Wästljung, A. C. (2010). High-density linkage mapping and evolution of paralogs and orthologs in *Salix* and *Populus*. *BMC Genomics* 11:129. doi: 10.1186/1471-2164-11-129

Supplementary Figure 1 | Maximum likelihood phylogenetic tree of poplar and willow R2R3 MYB genes.

Supplementary Figure 2 | Distribution of exon lengths in poplar and willow R2R3 MYB genes.

Supplementary Figure 3 | Distribution and numbers of different motifs in the poplar and willow R2R3 MYB genes.

Supplementary Figure 4 | Neighbor-joining phylogenetic tree of poplar, willow, *A. thaliana*, *P. patens*, *C. reinhardtii*, grapevine, maize, *B. napus*, apple, and ginkgo R2R3 MYB genes.

Supplementary Figure 5 | Maximum likelihood phylogenetic tree of poplar, willow, *A. thaliana*, *P. patens*, *C. reinhardtii*, grapevine, maize, *B. napus*, apple, and ginkgo R2R3 MYB genes.

Supplementary Figure 6 | Cis-element analysis of the promoter regions of GAMYB genes.

Supplementary Table 1 | List of 192 identified poplar R2R3 MYB transcription factors.

Supplementary Table 2 | List of 216 identified willow R2R3 MYB transcription factors.

Supplementary Table 3 | List of poplar R2R3 MYB paralogous gene pairs.

Supplementary Table 4 | List of willow R2R3 MYB paralogous gene pairs.

Supplementary Table 5 | List the poplar and willow R2R3 MYB orthologous gene pairs identified by OrthoFinder v2.

Supplementary Table 6 | List of different cis-acting elements.

Supplementary Table 7 | List of primer sequences used in qRT-PCR.

Supplementary Table 8 | List different Motif sequences.

Supplementary Sequence 1 | List of sequences of 192 poplar R2R3 MYB proteins.

Supplementary Sequence 2 | List of the CDS sequences of 192 poplar R2R3 MYB proteins.

Supplementary Sequence 3 | List of the sequences of 216 willow R2R3 MYB proteins.

Supplementary Sequence 4 | List of the CDS sequences of 216 willow R2R3 MYB proteins.

Supplementary Sequence 5 | List of the sequences of R2R3 MYB proteins from eight species: *A. thaliana* (126 members), *P. patens* (50 members), *C. reinhardtii* (10 members), *V. vinifera* (134 members), *Z. mays* (157 members), *B. napus* (249 members), *M. domestica* (228 members) and *G. biloba* (69 members).

- Bhargava, A., Mansfield, S. D., Hall, H. C., Douglas, C. J., and Ellis, B. E. (2010). MYB75 functions in regulation of secondary cell wall formation in the *Arabidopsis* inflorescence stem. *Plant Physiol.* 154, 1428–1438. doi: 10.1104/pp.110.162735
- Borevitz, J. O., Xia, Y., Blount, J., Dixon, R. A., and Lamb, C. (2000). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell* 12, 2383–2394. doi: 10.1105/tpc.12.12.2383
- Cao, Z. H., Zhang, S. Z., Wang, R. K., Zhang, R. F., and Hao, Y. J. (2013). Genome Wide Analysis of the Apple MYB Transcription Factor Family Allows the Identification of *MdoMYB121* Gene Confering Abiotic Stress Tolerance in Plants. *PLoS One* 8:e69955. doi: 10.1371/journal.pone.0069955
- Chai, G. H., Wang, Z. G., Tang, X. F., Yu, L., Qi, G., Wang, D., et al. (2014). R2R3-MYB gene pairs in *Populus*: evolution and contribution to secondary wall formation and flowering time. *J. Exp. Bot.* 65, 4255–4269. doi: 10.1093/jxb/eru196
- Chang, D., and Duda, T. F. (2012). Extensive and continuous duplication facilitates rapid evolution and diversification of gene families. *Mol. Biol. Evol.* 29, 2019–2029. doi: 10.1093/molbev/mss068
- Chang, X. J., Xie, S. P., Wei, L. L., Lu, Z. L., Chen, Z. H., Chen, F., et al. (2020). Origins and Stepwise Expansion of R2R3-MYB Transcription Factors for the Terrestrial Adaptation of Plants. *Front. Plant Sci.* 11:575360. doi: 10.3389/fpls.2020.575360
- Chen, C. J., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y. H., et al. (2020). TBtools: an Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009
- Chen, X. J., Wang, P. J., Gu, M. Y., Lin, X. Y., Hou, B. H., Zheng, Y. C., et al. (2021). R2R3-MYB transcription factor family in tea plant (*Camellia sinensis*): genome-wide characterization, phylogeny, chromosome location, structure and expression patterns. *Genomics* 113, 1565–1578. doi: 10.1016/j.ygeno.2021.03.033
- Chen, Y. H., Yang, X. Y., He, K., Liu, M. H., Li, J. G., Gao, Z. F., et al. (2006). The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Mol. Biol.* 60, 107–124. doi: 10.1007/s11103-005-2910-y
- Cheng, H. Q., Han, L. B., Yang, C. L., Wu, X. M., Zhong, N. Q., Wu, J. H., et al. (2016). The cotton MYB108 forms a positive feedback regulation loop with CML11 and participates in the defense response against *Verticillium dahliae* infection. *J. Exp. Bot.* 67, 1935–1950. doi: 10.1093/jxb/erw016
- Cominelli, E., Galbiati, M., Vavasseur, A., Conti, L., Sala, T., Vuylsteke, M., et al. (2015). A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance. *Curr. Biol.* 15, 1196–1200. doi: 10.1016/j.cub.2005.05.048
- Costache, A., Berghi, O. N., Cergan, R., Dumitru, M., Neagos, A., Popa, L. G., et al. (2021). Respiratory allergies: *Salicaceae* sensitization (Review). *Exp. Ther. Med.* 21:609.
- Cui, L., Wall, P. K., Leebens-Mack, J. H., Lindsay, B. G., Soltis, D. E., Doyle, J. J., et al. (2006). Widespread genome duplications throughout the history of flowering plants. *Genome Res.* 16, 738–749. doi: 10.1101/gr.4825606
- Dai, X. G., Hu, Q. J., Cai, Q. L., Feng, K., Ye, N., and Yin, T. M. (2014). The willow genome and divergent evolution from poplar after the common genome duplication. *Cell Res.* 24, 1274–1277. doi: 10.1038/cr.2014.83
- Du, H., Feng, B. R., Yang, S. S., Huang, Y. B., and Tang, Y. X. (2012a). The R2R3-MYB transcription factor gene family in maize. *PLoS One* 7:e37463. doi: 10.1371/journal.pone.0037463
- Du, H., Yang, S. S., Liang, Z., Feng, B. R., Liu, L., Huang, Y. B., et al. (2012b). Genome-wide analysis of the MYB transcription factor superfamily in soybean. *BMC Plant Biol.* 12:106. doi: 10.1186/1471-2229-12-106
- Du, H., Liang, Z., Zhao, S., Nan, M. G., Tran, L. S. P., Lu, K., et al. (2015). The evolutionary history of R2R3-MYB proteins across 50 eukaryotes: new insights into subfamily classification and expansion. *Sci. Rep.* 5:11037.
- Du, H., Wang, Y. B., Xie, Y., Liang, Z., Jiang, S. J., Zhang, S. S., et al. (2013). Genome-wide identification and evolutionary and expression analyses of MYB-related genes in land plants. *DNA Res.* 20, 437–448. doi: 10.1093/dnares/dst021
- Dubos, C., Stracke, R., Grotewold, E., Weissshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in *Arabidopsis*. *Trends. Plant Sci.* 15, 573–581.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432.
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238.
- Fang, Q., Wang, X. Q., Wang, H. Y., Tang, X. W., Liu, C., Yin, H., et al. (2020). The poplar R2R3 MYB transcription factor PtrMYB94 coordinates with abscisic acid signaling to improve drought tolerance in plants. *Tree Physiol.* 40, 46–59. doi: 10.1093/treephys/tpz113
- Frampton, J., Leutz, A., Gibson, T., and Graf, T. (1989). DNA-binding domain ancestry. *Nature* 342:134. doi: 10.1038/342134a0
- Frerigmann, H., and Gigolashvili, T. (2014). MYB34, MYB51, and MYB122 distinctly regulate indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. *Mol. Plant* 7, 814–828. doi: 10.1093/mp/ssu004
- Froidure, S., Canonne, J., Daniel, X., Jauneau, A., Brière, C., Roby, D., et al. (2010). AtsPLA2-alpha nuclear relocalization by the *Arabidopsis* transcription factor AtMYB30 leads to repression of the plant defense response. *Proc. Natl. Acad. Sci. U. S. A.* 107, 15281–15286. doi: 10.1073/pnas.1009056107
- Futamara, N., Mori, H., Kouchi, H., and Shinohara, K. (2000). Male flower-specific expression of genes for polygalacturonase, pectin methylesterase and beta-1,3-glucanase in a dioecious willow (*Salix gilgiana* Seemen). *Plant Cell Physiol.* 41, 16–26. doi: 10.1093/pcp/41.1.16
- Gao, X., Zhang, Q., Zhao, Y. Q., Yang, J., He, H. B., and Jia, G. X. (2020). The lre-miR159a-LrGAMYB pathway mediates resistance to grey mould infection in *Lilium regale*. *Mol. Plant Pathol.* 21, 749–760. doi: 10.1111/mpp.12923
- Gocal, G. F., Sheldon, C. C., Gubler, F., Moritz, T., Bagnall, D. J., MacMillan, C. P., et al. (2001). GAMYB-like genes, Flowering, and Gibberellin Signaling in *Arabidopsis*. *Plant Physiol.* 127, 1682–1693. doi: 10.1104/pp.127.4.1682
- Gutierrez, L., Mauriat, M., Guénin, S., Pelloux, J., Lefebvre, J. F., Louvet, R., et al. (2008). The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. *Plant Biotechnol. J.* 6, 609–618. doi: 10.1111/j.1467-7652.2008.00346.x
- Hajiebrahimi, A., Owji, H., and Hemmati, S. (2017). Genome-wide identification, functional prediction, and evolutionary analysis of the R2R3-MYB superfamily in *Brassica napus*. *Genome* 60, 797–814. doi: 10.1139/gen-2017-0059
- He, Q. L., Jones, D. C., Li, W., Xie, F. L., Ma, J., Sun, R., et al. (2016). Genome-Wide Identification of R2R3-MYB Genes and Expression Analyses During Abiotic Stress in *Gossypium raimondii*. *Sci. Rep.* 6:22980.
- He, Y. Q., Zhang, X. Y., Li, L. Y., Sun, Z. T., Li, J. M., Chen, X. Y., et al. (2021). SPX4 interacts with both PHR1 and PAP1 to regulate critical steps in phosphorus-status-dependent anthocyanin biosynthesis. *New Phytol.* 230, 205–217. doi: 10.1111/nph.17139
- Hou, J. (2016). *Genomics Basis Leading to the Divergence Between Willow and Poplar and the Differentiation in their Sex Determinationsystems*. China: Nanjing Forestry University.
- Hou, J., Wei, S. Y., Pan, H. X., Zhuge, Q., and Yin, T. M. (2019). Uneven selection pressure accelerating divergence of *Populus* and *Salix*. *Hortic Res.* 6:37. doi: 10.1038/s41438-019-0121-y
- Hou, J., Ye, N., Dong, Z. Y., Lu, M. Z., Li, L. G., and Yin, T. M. (2016). Major Chromosomal Rearrangements Distinguish Willow and Poplar After the Ancestral “Salicoid” Genome Duplication. *Genome Biol. Evol.* 8, 1868–1875. doi: 10.1093/gbe/evw127
- Hou, J., Ye, N., Zhang, D. F., Chen, Y. N., Fang, L. C., Dai, X. G., et al. (2015). Different autosomes evolved into sex chromosomes in the sister genera of *Salix* and *Populus*. *Sci. Rep.* 5:9076.
- Hu, B., Jin, J. P., Guo, A. Y., Zhang, H., Luo, J. C., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817
- Jakoby, M. J., Falkenhan, D., Mader, M. T., Brininstool, G., Wischnitzki, E., Platz, N., et al. (2008). Transcriptional profiling of mature *Arabidopsis*

- trichomes reveals that *NOECK* encodes the MIXTA-like transcriptional regulator MYB106. *Plant Physiol.* 148, 1583–1602.
- Jiang, C. K., and Rao, G. Y. (2020). Insights into the diversification and evolution of R2R3-MYB transcription factors in plants. *Plant Physiol.* 183, 637–655. doi: 10.1104/pp.19.01082
- Jiang, I., Tsai, C. K., Chen, S. C., Wang, S. H., Amiraslano, I., Chang, C. F., et al. (2011). Molecular basis of the recognition of the ap65-1 gene transcription promoter elements by a Myb protein from the protozoan parasite *Trichomonas vaginalis*. *Nucleic Acids Res.* 39, 8992–9008. doi: 10.1093/nar/gkr558
- Jin, H., and Martin, C. (1999). Multifunctionality and diversity within the plant MYB-gene family. *Plant Mol Biol.* 41, 577–585.
- Kamenova, I., Mukherjee, P., Conic, S., Mueller, F., El-Saafin, F., Bardot, P., et al. (2019). Co-translational assembly of mammalian nuclear multisubunit complexes. *Nat. Commun.* 10:1740.
- Ko, S. S., Li, M. J., Ho, Y. C., Yu, C. P., Yang, T. T., Lin, Y. J., et al. (2021). GAMYB TF modulates bHLH142 and is homeostatically regulated by TDR TF during rice anther tapetal and pollen development. *J. Exp. Bot.* 72:erab190.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Kumar, S., Stecher, G., Li, M., Nkayaz, C., and Tamura, K. (2018). MEGA X: molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Peer, Y. V. D., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. doi: 10.1093/nar/30.1.325
- Letunic, I., and Bork, P. (2018). 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* 46, D493–D496.
- Li, H., Wang, Y., Wang, Z., Guo, X., Wang, F., Xia, X. J., et al. (2016). Microarray and genetic analysis reveals that *gsa-miR159b* plays a critical role in abscisic acid-mediated heat tolerance in crafted cucumber plants. *Plant Cell Environ.* 39, 1790–1804. doi: 10.1111/pce.12745
- Li, X. X., Guo, C., Ahmad, S., Wang, Q., Yu, J., Liu, C., et al. (2019). Systematic Analysis of MYB Family Genes in Potato and Their Multiple Roles in Development and Stress Responses. *Biomolecules* 9:317. doi: 10.3390/biom9080317
- Linger, B. R., and Price, C. M. (2009). Conservation of telomere protein complexes: shuffling through evolution. *Crit. Rev. Biochem. Mol. Biol.* 44, 434–446. doi: 10.3109/10409230903307329
- Liu, C., Yu, H. S., Rao, X. L., Li, L. G., and Dixon, R. A. (2021). Abscisic acid regulates secondary cell-wall formation and lignin deposition in *Arabidopsis thaliana* through phosphorylation of NST1. *Proc. Natl. Acad. Sci. U. S. A.* 118:e2010911118. doi: 10.1073/pnas.2010911118
- Liu, Y. F., Ma, K. X., Qi, Y. W., Lv, G. W., Ren, X. L., Liu, Z. D., et al. (2021). Transcriptional Regulation of Anthocyanin Synthesis by MYB-bHLH-WDR Complexes in Kiwifruit (*Actinidia chinensis*). *J. Agric. Food Chem.* 69, 3677–3691. doi: 10.1021/acs.jafc.0c07037
- Liu, S. L., Pan, A. Q., and Adams, K. L. (2014). Protein Subcellular Relocalization of Duplicated Genes in *Arabidopsis*. *Genome Biol. Evol.* 6, 2501–2515. doi: 10.1093/gbe/evu191
- Liu, Z. H., Bao, W. J., Liang, W. Q., Yin, J. Y., and Zhang, D. B. (2010). Identification of *gamyb-4* and analysis of the regulatory role of GAMYB in rice anther development. *J. Integr. Plant Biol.* 52, 670–678. doi: 10.1111/j.1744-7909.2010.00959.x
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔC_T} Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Lv, Y., Yang, M., Hu, D., Yang, Z. Y., Ma, S. Q., Li, X. H., et al. (2017). The OsMYB30 transcription factor suppresses cold tolerance by interacting with a JAZ protein and suppressing β-amylase expression. *Plant Physiol.* 173, 1475–1491. doi: 10.1104/pp.16.01725
- Ma, D., and Constabel, C. P. (2019). MYB Repressors as Regulators of Phenylpropanoid Metabolism in Plants. *Trends Plant Sci.* 24, 275–289. doi: 10.1016/j.tplants.2018.12.003
- Magwanga, R. O., Lu, P., Kirungu, J. N., Cai, X. Y., Zhou, Z. L., Wang, X. X., et al. (2018). Whole Genome Analysis of Cyclin Dependent Kinase (CDK) Gene Family in Cotton and Functional Evaluation of the Role of CDKF4 Gene in Drought and Salt Stress Tolerance in Plant. *Int. J. Mol. Sci.* 19:2625. doi: 10.3390/ijms19092625
- Matus, J. T., Aquea, F., and Arce-Johnson, P. (2008). Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across *Vitis* and *Arabidopsis* genomes. *BMC Plant Biol.* 8:83. doi: 10.1186/1471-2229-8-83
- McCarthy, R. L., Zhong, R. Q., Fowler, S., Lyskowski, D., Piyasena, H., Carleton, K., et al. (2010). The poplar MYB transcription factors, PtrMYB3 and PtrMYB20, are involved in the regulation of secondary wall biosynthesis. *Plant Cell Physiol.* 51, 1084–1090. doi: 10.1093/pcp/pcq064
- Millard, P. S., Kragelund, B. B., and Burrow, M. (2019a). R2R3 MYB Transcription Factors - Functions outside the DNA-Binding Domain. *Trends Plant Sci.* 24, 934–946. doi: 10.1016/j.tplants.2019.07.003
- Millard, P. S., Weber, K., Kragelund, B. B., and Burrow, M. (2019b). Specificity of MYB interactions relies on motifs in ordered and disordered contexts. *Nucleic Acids Res.* 47, 9592–9608. doi: 10.1093/nar/gkz691
- Murray, F., Kalla, R., Jacobsen, J., and Gubler, F. (2003). A role for *HvGAMYB* in anther development. *Plant J.* 33, 481–491. doi: 10.1046/j.1365-313x.2003.01641.x
- Nesi, N., Jond, C., Debeaujon, I., Caboche, M., and Lepiniec, L. (2001). The *Arabidopsis* TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *Plant Cell* 13, 2099–2114. doi: 10.1105/tpc.13.9.2099
- Noda, S., Koshihata, T., Hattori, T., Yamaguchi, M., Suzuki, S., and Umezawa, T. (2015). The expression of a rice secondary wall-specific cellulose synthase gene, *OsCesA7*, is directly regulated by a rice transcription factor, *OsMYB58/63*. *Planta* 242, 589–600. doi: 10.1007/s00425-015-2343-z
- Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., et al. (1994). Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell* 79, 639–648. doi: 10.1016/0092-8674(94)90549-5
- Perez, V. C., Dai, R., Bai, B., Tomiczek, B., Askey, B. C., Zhang, Y., et al. (2021). Aldoximes are precursors of auxins in *Arabidopsis* and maize. *New Phytol.* 25, 1449–1461. doi: 10.1111/nph.17447
- Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Res.* 46, W200–W204.
- Raffaele, S., and Rivas, S. (2013). Regulate and be regulated: integration of defense and other signals by the AtMYB30 transcription factor. *Front. Plant Sci.* 4:98. doi: 10.3389/fpls.2013.00098
- Rusconi, F., Simeoni, F., Francia, P., Cominelli, E., Conti, L., Riboni, M., et al. (2013). The *Arabidopsis thaliana* MYB60 promoter provides a tool for the spatio-temporal control of gene expression in stomatal guard cells. *J. Exp. Bot.* 64, 3361–3371. doi: 10.1093/jxb/ert180
- Seo, P. J., and Park, C. M. (2010). MYB96-mediated abscisic acid signals induce pathogen resistance response by promoting salicylic acid biosynthesis in *Arabidopsis*. *New Phytol.* 186, 471–483. doi: 10.1111/j.1469-8137.2010.03183.x
- Shi, M., Zhang, H., and Wu, X. D. (2017). ALYREF mainly binds to the 5' and the 3' regions of the mRNA in vivo. *Nucleic Acids Res.* 45, 9640–9653. doi: 10.1093/nar/gkx597
- Skvortsov, A. K. (1999). *Willows of Russia and Adjacent Countries, Taxo-nomical and Geographical Revision*. Joensuu: University of Joensuu.
- Soler, M., Camargo, E. L. O., Carocha, V., Cassan-Wang, H., Clemente, H. S., Savelli, B., et al. (2015). The *Eucalyptus grandis* R2R3-MYB transcription factor family: evidence for woody growth-related evolution and function. *New Phytol.* 206, 1364–1377. doi: 10.1111/nph.13039
- Sterck, L., Rombauts, S., Jansson, S., Sterky, F., Rouzé, P., and Peer, Y. V. D. (2005). EST data suggest that poplar is an ancient polyploid. *New Phytol.* 167, 165–170. doi: 10.1111/j.1469-8137.2005.01378.x
- Stracke, R., Werber, M., and Weisshaar, B. (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* 4, 47–56.
- Subramanian, B., Gao, S. H., Lercher, M. J., Hu, S. N., and Chen, W. H. (2019). Evolvview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res.* 47, W270–W275.
- Tang, X. F., Wang, D., Liu, Y., Lu, M. Z., Zhuang, Y. M., Xie, Z., et al. (2020). Dual regulation of xylem formation by an auxin-mediated *PaC3H17-PaMYB199*

- module in *Populus*. *New Phytol.* 225, 1545–1561. doi: 10.1111/nph.16244
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673
- Tohge, T., Nishiyama, Y., Hirai, M. Y., Yano, M., Nakajima, J., Awazuhara, M., et al. (2005). Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. *Plant J.* 42, 218–235. doi: 10.1111/j.1365-313x.2005.02371.x
- Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604.
- Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77
- Wang, F. P., Wang, X. F., Zhang, J. C., Ma, F. W., and Hao, Y. J. (2018). MdMYB58 Modulates Fe Homeostasis by Directly Binding to the MdMATE43 Promoter in Plants. *Plant Cell Physiol.* 59, 2476–2489.
- Wang, M., Xu, Z. C., Ding, A., and Kong, Y. Z. (2018). Genome-Wide Identification and Expression Profiling Analysis of the Xyloglucan Endotransglucosylase/Hydrolase Gene Family in Tobacco (*Nicotiana tabacum* L.). *Genes (Basel)*. 9:273. doi: 10.3390/genes9060273
- Wang, Y. P., Tang, H. B., Debarry, J. D., Tan, X., Li, J. P., Wang, X. Y., et al. (2012). MCSscanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49.
- Wang, Z., Tang, J., Hu, R., Wu, P., Hou, X. L., Song, X. M., et al. (2015). Genome-wide analysis of the R2R3-MYB transcription factor genes in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*) reveals their stress and hormone responsive patterns. *BMC Genomic* 16:17. doi: 10.1186/s12864-015-1216-y
- Wang, Z. F., Mao, Y. L., Guo, Y. J., Gao, J. H., Liu, X. Y., Li, S., et al. (2020). MYB Transcription Factor161 Mediates Feedback Regulation of Secondary wall-associated NAC-Domain1 Family Genes for Wood Formation. *Plant Physiol.* 184, 1389–1406. doi: 10.1104/pp.20.01033
- Wei, H., Movahedi, A., Xu, C., Sun, W. B., Wang, P., Li, D. W., et al. (2020). Characterization, expression profiling, and functional analysis of *PtDef*, a defensin-encoding gene from *Populus trichocarpa*. *Front. Microbiol.* 11:106. doi: 10.3389/fmicb.2020.00106
- Wilkins, O., Nahal, H., Foong, J., Provart, N. J., and Campbell, M. M. (2009). Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. *Plant Physiol.* 149, 981–993. doi: 10.1104/pp.108.132795
- Wong, D. C. J., Schlechter, R., Vannozzi, A., Höll, J., Himmam, I., Bogs, J., et al. (2016). A systems-oriented analysis of the grapevine R2R3-MYB transcription factor family uncovers new insights into the regulation of stilbene accumulation. *DNA Res.* 23, 451–466.
- Wu, H. T., Tian, Y., Wan, Q., Fang, L., Guan, X. Y., Chen, J. D., et al. (2018). Genetics and evolution of MIXTA genes regulating cotton lint fiber development. *New Phytol.* 217, 883–895. doi: 10.1111/nph.14844
- Yang, C. Q., Fang, X., Wu, X. M., Mao, Y. B., Wang, L. J., and Chen, X. Y. (2012). Transcriptional regulation of plant secondary metabolism. *J. Integr. Plant Biol.* 54, 703–712.
- Yang, X. M., Zhou, T. T., Wang, M. K., Li, T. T., Wang, G. B., Fu, F. F., et al. (2021). Systematic investigation and expression profiles of the Gbr2R3-MYB transcription factor family in ginkgo (*Ginkgo biloba* L.). *Int J Biol Macromol.* 172, 250–262. doi: 10.1016/j.ijbiomac.2021.01.053
- Zhang, J., Yuan, H. W., Yang, Q. S., Li, M., Wang, Y., Li, Y. J., et al. (2017). The genetic architecture of growth traits in *Salix matsudana* under salt stress. *Hortic. Res.* 4:17024.
- Zhang, Y., Zhang, B., Yang, T. W., Zhang, J., Liu, B., Zhan, X. Q., et al. (2020). The GAMYB-like gene *SlMYB33* mediates flowering and pollen development in tomato. *Hortic. Res.* 7:133.
- Zhang, Y., Zhang, X. L., Liu, B., Wang, W. J., Liu, X. W., Chen, C. H., et al. (2014). A GAMYB homologue CsGAMYB1 regulates sex expression of cucumber via an ethylene-independent pathway. *J. Exp. Bot.* 65, 3201–3213. doi: 10.1093/jxb/eru176
- Zhang, Y. X., Han, X. J., Chen, S. S., Zheng, L., He, X. L., Liu, M. Y., et al. (2017). Selection of suitable reference genes for quantitative real-time PCR gene expression analysis in *Salix matsudana* under different abiotic stresses. *Sci. Rep.* 7:40290.
- Zhang, Z., Li, J., Zhao, X. Q., Wang, J., Wong, G. K. S., and Yu, J. (2006). KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genom. Prot. Bioinform.* 4, 259–263. doi: 10.1016/s1672-0229(07)60007-2
- Zhou, F. W., Wu, H. T., and Yin, T. M. (2021). Characteristics of MIXTA/MIXTA-like genes and their functions in regulating plant epidermal cell differentiation. *J. Nanjing Forest. Univ.* 45, 229–237.
- Zhou, J. L., Lee, C. H., Zhong, R. Q., and Ye, Z. H. (2009). MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in *Arabidopsis*. *Plant Cell* 21, 248–266. doi: 10.1105/tpc.108.063321
- Zhou, R., Macaya-Sanz, D., Carlson, C. H., Schmutz, J., Jenkins, J. W., Kudrna, D., et al. (2020). A willow sex chromosome reveals convergent evolution of complex palindromic repeats. *Genome Biol.* 21:38.
- Zhou, R., Macaya-Sanz, D., Rodgers-Melnick, E., Carlson, C. H., Gouker, F. E., Evans, L. M., et al. (2018). Characterization of a large sex determination region in *Salix purpurea* L. (*Salicaceae*). *Mol. Genet. Genomics* 293, 1437–1452. doi: 10.1007/s00438-018-1473-y
- Zhou, W., Zhang, Q., Sun, Y., Yang, L., and Wang, Z. (2020). Genome-wide identification and characterization of R2R3-MYB family in *Hypericum perforatum* under diverse abiotic stresses. *Int. J. Biol. Macromol.* 15:145.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zhou, Chen, Wu and Yin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.