Liu *et al. Chem. Biol. Technol. Agric.* (2022) 9:37
https://doi.org/10.1186/s40538-022-00301-7

Chemical and Biological Technologies in Agriculture

RESEARCH Open Access

Genome-level analysis of BpR2R3-MYB family genes transcribed in seedlings of *Betula platyphylla* and *BpR2R3-MYB15* enhanced flavonoid production

Huimin Liu, Zhongyang Yu, Guizhi Fan* and Baojiang Zheng*

Abstract

Background: Flavonoids have a wide range of biological activities in plant development, stress resistance and human health, etc. R2R3-MYBs are one of the key elements in regulation of flavonoid production, but their functional importance in *Betula platyphylla* remains elusive.

Methods: The full-length transcriptome data of 30-day-old seedlings of *Betula platyphylla* were used to identify BpR2R3-MYB family genes, and their gene structure, chromosome distribution and syntenic relationships were predicted by bioinformatics methods. *Agrobacterium*-mediated transient transformation was used to verify the function of *BpR2R3-pMYB15* in flavonoid production.

Results: 44 BpR2R3-MYB family genes expressed in seedlings of *Betula platyphylla* were identified and found to be unevenly distributed in 11 chromosomes. Among them, 90.90% of the *BpR2R3-MYBs* had introns, and only four genes had no introns. Five gene pairs with segment duplication were found, and their Ka/Ks ratios were less than 1. Thirty orthologs between *Betula platyphylla* and *Arabidopsis thaliana* and 68 orthologs between *Betula platyphylla* and *Populus trichocarpa* were detected. Five *BpR2R3-MYBs* were clustered with R2R3-MYB genes related to flavonoid synthesis, and *BpR2R3-pMYB15* had the highest correlation coefficients between the value of gene expression and flavonoid content. *BpR2R3-pMYB15* was cloned, and its transient overexpression obtained using *Agrobacterium*-mediated transformation positively regulated flavonoid production.

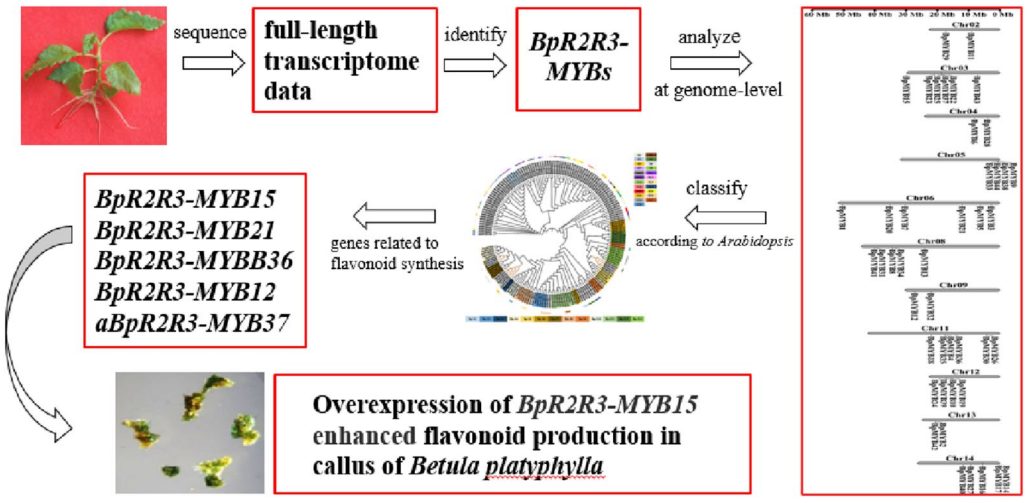
Conclusion: This work enriches the collection of R2R3-MYBs related to flavonoid production in seedlings of *Betula platyphylla*.

Keywords: R2R3-MYB, Gene family, Flavonoid, Seedlings of *Betula platyphylla*, Full-length transcriptome

*Correspondence: gzf325@126.com; shengmign2020@126.com

Key Laboratory of Saline-Alkali Vegetation Ecology Restoration, Ministey of Education, Northeast Forestry University, Harbin 150040, China

Graphical Abstract



Background

Flavonoids are a class of polyphenol phytochemicals that are abundant in many leaves, stems, flowers, fruits and other tissues or organs of higher plants [1]. About 8000 flavonoids have been discovered to date, ant they can be divided into six main subclasses, namely, the flavonols, flavanones, flavones, flavanols, isoflavones, and anthocyanidins [2, 3]. Flavonoids have a variety of biological activities in plant growth, development, and stress resistance to harsh environments, including (I) regulation of axillary bud or pollen tube growth [4, 5]; (II) provision of pigmentation for leaves, flowers, and fruits [3, 6]; (III) protection against biological and non-biological stresses [7–9]; (IV) acting as signal molecules between plant and microbe interactions [3]. Flavonoids also have medicinal properties including antitussive, expectorant, antibacterial, antifungal, anti-free radical, and anti-oxidation [10]. Concerns have been raised about their potential functions in plant and human health, and thus, understanding the molecular basis of flavonoid biosynthesis is crucial.

The biosynthesis of flavonoids proceeds via the phenylpropanoid pathway, and most of the key enzymes and transcription factors (TFs) involved in this pathway have been identified [9, 11]. The key (structural) enzymes in flavonoid biosynthesis include phenylalanine ammonia lyase (PAL), chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H), flavonol synthase (FLS), etc. [11]. These structural genes in flavonoid biosynthesis are transcriptionally regulated by MYB, bHLH, WD40, WRKY, zinc finger, and MADS box proteins, etc. [12]. R2R3-MYB, a type of MYB protein, is one of the key elements in the positive and negative regulation of the biosynthetic enzymes of flavonoids. For instance, *AtMYB11*, *AtMYB12*, *AtMYB14*, *AtMYB15*, and *AtMYB111* in *Arabidopsis* positively regulated flavonoid biosynthesis [13, 14]. *AtMYB60* in *A. thaliana*, *PtoMYB156* in *Populus tomentosa,* and *CmMYB012* in *Chrysanthemum x morifolium* negatively regulated flavonoid biosynthesis [15–17].

Betula platyphylla is a pioneer hardwood tree species with ecological, economic, and pharmacological activities, and it thrives in northeastern China, Russian Far East, Siberia, Mongolia, Northern Korea, and Japan [18]. Flavonoids are one of the main secondary metabolites in leaf extracts of *B. platyphylla*, which exhibits antifungal, anti-free radical, and antioxidant activities [19]. The overexpression of *BpCHS3* promotes flavonoid production and enhances the salt tolerance of *B. platyphylla* [20]. Given the importance of R2R3-MYB proteins in flavonoid biosynthesis in plants, the functional characterization of R2R3-MYB family members in flavonoid biosynthesis of *B. platyphylla*, especially in seedlings, has not been systematically investigated.

In this study, the latest *B. platyphylla* reference genome was used to characterize R2R3-MYB family members [18], which were screened from full-length transcriptome data of 30-day-old *B. platyphylla* seedlings. *BpR2R3-MYB15*, a R2R3-MYB gene predicted to be involved in flavonoid synthesis, was cloned and verified via transient transformation in *B. platyphylla*. The results of this

study can contribute to the functional characterization of R2R3-MYB transcription factors in seedlings of *B. platyphylla*.

Materials and methods

Identification of BpR2R3-MYB genes

The genomic sequence of *B. platyphylla* (accession code PRJNA285437) was published by Chen et al. [18], and the full-length transcriptome data of 30-day-old seedlings of *B. platyphylla* were obtained using Pacific Bioscience (PacBio) single-molecule real-time sequencing technology (accession code PRJNA816665). The hidden Markov model (HMM) of the R2R3-MYB DNA-binding domain (PF00249) downloaded from the Pfam database (http://pfam.xfam.org/) was used to search for the R2R3-MYB protein of *B. platyphylla* through the HMM search program (http://www.hmmer.org/). All putative proteins were subjected to conserved structural domain identification using SMART software (http://smart.embl.de/smart/batch.pl) and the NCBI-CDD database (https://www-ncbi-nlm-nih-gov-443.webvpn.nefu.edu.cn/cdd/) [21]. Forty-four *BpR2R3-MYB* family genes were identified and numbered according to the order in which they were found. We used ExPASy-ProtParam (http://web.expasy.org/protparam/) to analyze the physicochemical properties, protein molecular weights (MW), and theoretical isoelectric points (pI) of the 44 identified R2R3-MYB proteins of *B. platyphylla*. The subcellular localization was predicted using Plant-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/).

Chromosomal distribution analysis

The *B. platyphylla* genomic sequence was inputted into the function module of Genome Length Filter in TBtools software to obtain chromosome information [22]. Then, the chromosome and location information of the 44 *R2R3-MYBs* were entered into the function module of Gen Location Visualize (Advanced) in TBtools software to visualize the chromosomal distribution of the 44 *R2R3-MYBs*.

Construction of the phylogenetic trees

The amino acid sequences of the 44 BpR2R3-MYB proteins, 126 *A. thaliana* R2R3-MYB proteins, and 4 R2R3-MYB proteins related to flavonoid synthesis were aligned using Clustal W in MEGA X software [23–25]. PHYLOGENY in MEGA X software was used to construct a neighbor-joining tree through 1000 bootstrap replications.

Conserved motifs and gene structures

Multiple expectation maximization for motif elicitation (MEME) (https://meme-suite.org/meme/) was used to analyze the conserved motifs of the 44 BpR2R3-MYB proteins. The maximum number of motifs was set to 10 (width range of motif = 6–300 residues). PHYLOGENY in MEGA X software was adopted to construct a maximum-parsimony tree through 1000 bootstrap replications. Gene Structure View (Advanced) in TBtools software was used to visualize the exon/intron structures [22].

Collinearity analysis

The gene synteny, tandem, and segmental duplications of *BpR2R3-MYBs* among *B. platyphylla*, *A. thaliana*, and *P. trichocarpa* were analyzed using the One-Step MCScanX function in TBtools software [22]. The Advanced Circos and Multiple Synteny Plot in TBtools software were utilized to visualize intra-genomic and inter-genomic collinearity. Five gene pairs with segmental duplication were selected for the calculation of Ka (non-synonymous substitution sate) and Ks (synonymous substitution rate). The values of Ka and Ks were calculated with the Simple Ka/Ks Calculator (NG) in TBtools software. Generally, Ka/Ks < 1.0 indicates purifying or negative selection, Ka/Ks = 1.0 denotes neutral selection, and Ka/Ks > 1.0 means positive selection [26].

Plant materials

Leaves of 18 ten-year-old *B. platyphylla* trees planted in Northeast Forestry University were collected (3-h intervals on August 3–4, 2020, 45° 72′ 66″ N, 126° 64′ 47″ E) for daily cycle analysis, which provided a basis for sampling time in seedling. Thirty-day-old seedlings of *B. platyphylla* obtained from sterile seeds were treated with 90 μmol L^{-1} Cd treatment for 1 and 4 days, and 15-day-old calli of *B. platyphylla* (easily transformed at this stage) obtained from the stem of tissue-cultured seedlings were used for *Agrobacterium*-mediated transient transformation. The seedlings were planted in a woody plant medium supplemented with 20 g L^{-1} of sucrose. The calli were cultured in Gamborg's B_5 medium supplemented with 0.3 mg L^{-1} of 6-benzyladenine, 0.6 mg L^{-1} of thidiazuron, and 20 g L^{-1} of sucrose. The pH of the medium was adjusted to 5.6 ± 0.2 prior to autoclaving. Fresh samples frozen with liquid nitrogen were used for gene expression, and samples dried through the oven-drying method were used for the analysis of flavonoid or procyanidin content.

Cloning of full-length *BpR2R3-MYB15* and *BpR2R3-MYB21*

The full-length sequences of *BpR2R3-MYB15* and *BpR2R3-MYB21* were amplified by the following PCR primers:

Table 1 The sequence characteristics of 44 *R2R3-MYB* genes identified in *Betula platyphyll*a

Name	Amino acids	Molecular weight (kDa)	Theoretical pI	Subcellular localization
BpMYB1	423	46.36	6.04	Nuclear
BpMYB2	330	35.74	8.52	Chloroplast
BpMYB3	310	33.40	6.12	Nuclear
BpMYB4	322	36.43	5.52	Nuclear
BpMYB5	300	33.56	8.66	Nuclear
BpMYB6	309	34.31	5.75	Nuclear
BpMYB7	264	29.61	5.34	Nuclear
BpMYB8	218	24.37	9.55	Nuclear
BpMYB9	286	32.24	5.23	Nuclear
BpMYB10	218	24.76	9.57	Nuclear
BpMYB11	300	34.44	8.08	Nuclear
BpMYB12	217	24.32	9.02	Nuclear
BpMYB13	232	26.44	8.49	Nuclear
BpMYB14	262	29.28	7.7	Nuclear
BpMYB15	262	29.69	8.17	Nuclear
BpMYB16	357	40.85	9.47	Cytoplasmic
BpMYB17	305	33.21	7.69	Nuclear
BpMYB18	273	30.93	6.11	Nuclear
BpMYB19	293	33.33	6.25	Nuclear
BpMYB20	337	38.04	6.45	Nuclear
BpMYB21	288	32.76	6.73	Nuclear
BpMYB22	315	35.32	6.11	Nuclear
BpMYB23	251	29.27	5.57	Mitochondrial
BpMYB24	186	20.97	8.97	Nuclear
BpMYB25	339	39.04	5.19	Nuclear
BpMYB26	299	33.57	6.61	Nuclear
BpMYB27	287	32.57	7.27	Chloroplast
BpMYB28	363	40.95	5.38	Nuclear
BpMYB29	320	36.08	5.59	Nuclear
BpMYB30	401	43.97	5.94	Nuclear
BpMYB31	352	39.10	8.27	Nuclear
BpMYB32	272	30.63	5.45	Nuclear
BpMYB33	400	44.76	6.51	Nuclear
BpMYB34	435	48.14	6.76	Nuclear
BpMYB35	322	36.46	5.45	Nuclear
BpMYB36	245	28.21	8.72	Nuclear
BpMYB37	300	33.41	8.28	Nuclear
BpMYB38	287	32.36	5.12	Nuclear
BpMYB39	207	23.60	9.57	Nuclear
BpMYB40	366	40.76	6.8	Nuclear
BpMYB41	484	54.60	6.13	Nuclear
BpMYB42	324	34.92	7.62	Chloroplast
BpMYB43	331	36.83	6.09	Nuclear
BpMYB44	244	28.91	8.51	Nuclear

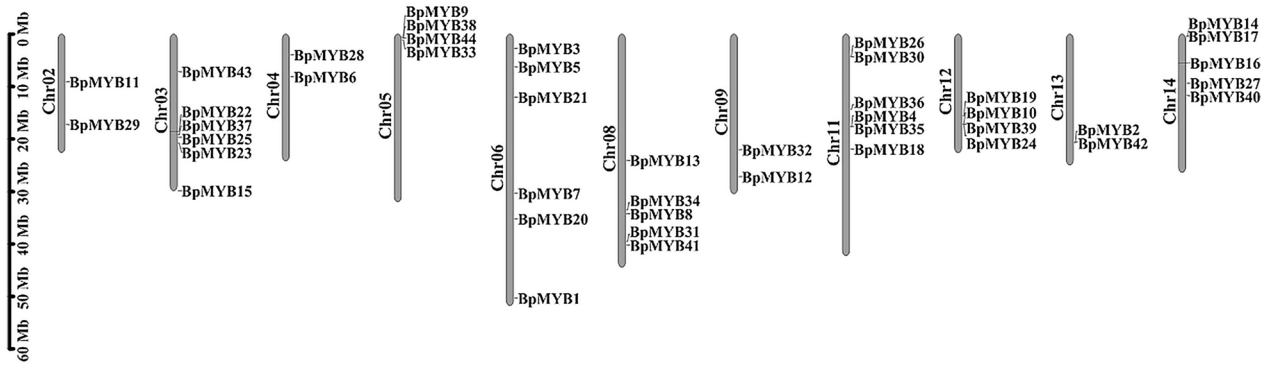


Fig. 1 Chromosomal locations of *BpR2R3-MYB* genes in *Betula platyphylla*. The chromosomal position of each *BpR2R3-MYB* gene was mapped according to the *Betula platyphylla* genome. The left scale with Mb indicates the length of the chromosome

BpR2R3-MYB15-F: GAGTCGCAATTACAACCACAG ATAT,

BpR2R3-MYB15-R: TCAATTATTATTCAACCTGCC ATTC,

BpR2R3-MYB21-F: AGAAAGAGATTCTGACGTAGA TGGG,

BpR2R3-MYB21-R: GCGTGCCACTACTAGGTTTAG ACTA.

PCR amplification was performed as follows: 94 C for 5 min; 35 cycles of 98 C for 10 s, 50 C for 45 s, and 72 C for 1 min; and 72 C for 10 min. Positive colonies (purified PCR amplification fragment ligated with pMD^{TM}18-T vector) were sequenced at Rui Biotech (Beijing).

Agrobacterium-mediated transient transformation

Agrobacterium tumefaciens strain LBA4404 harboring Cam1304-SubC-BpMYB15 (overexpression vector) or RhRNA-pTRV2-BpMYB15 (RNAi vector) was used to infect 15-day-old *B. platyphylla* calli (soaked in 25% sucrose for 5 min) for 1 h [27]. The infection solution consisted of 2 mM L^{-1} of MES-KOH (pH = 5.4), 10 mM L^{-1} of $CaCl_2$, 120 μM L^{-1} of acetosyringone (AS), 2% sucrose, 270 mM L^{-1} of mannitol, and 200 mg L^{-1} of dithiothreitol + 0.02% Tween. The infected calli were cultured in B_5 liquid medium containing 100 μmol L^{-1} of AS for 2 days in the dark at 28 °C. Then, the infected calli were washed with distilled water for analysis of gene expression and total flavonoid content. The transient expression of GUS was also histochemically assayed by staining the infected calli with X-GLUC solution in dark at 37 ℃ for 1 h.

Determination of the total flavonoid and procyanidin content

The collected fresh samples were dried at 105 ℃ for 15 min, and then dried to constant weight at 60 ℃. Dried samples (0.05 g) were accurately weighed and soaked in 5 mL of 65% ethanol for 24 h. After centrifugation at 2504*g* for 10 min, 1 mL of the supernatant solution was obtained for content analysis. The total flavonoid content was determined using the $AlCl_3$ colorimetric method with quercetin as the standard [28], and the linear equation was $y = 0.1151x + 0.0504$ ($R^2 = 0.996$), where x indicates the absorbance of the solution at 510 nm. Procyanidin content was analyzed via vanillin-hydrochloric acid spectrophotometric quantification with procyanidin B1 as the standard, and the linear equation was $y = 0.0018x + 0.0027$ ($R^2 = 0.991$), where x indicates the absorbance of the solution at 500 nm [29].

Gene expression analysis

The CTAB-based method was used to isolate the total RNA. The Taqman probes and primers are presented in Additional file 1: Table S1. PCR amplification was performed on an ABI Prism7500 real-time PCR system as follows: 95 °C for 30 s, followed by 40 cycles at 95 °C for 5 s and 60 °C for 34 s. Each RT-qPCR analysis was performed with three technical replicates. Gene expression data were calculated relative to the reference gene (*α-tubulin, TU*) following the $2^{-\Delta\Delta Ct}$ method [30]. The data of gene expression were $\log_{10}$ transformed and used for heatmap construction with TBtool software.

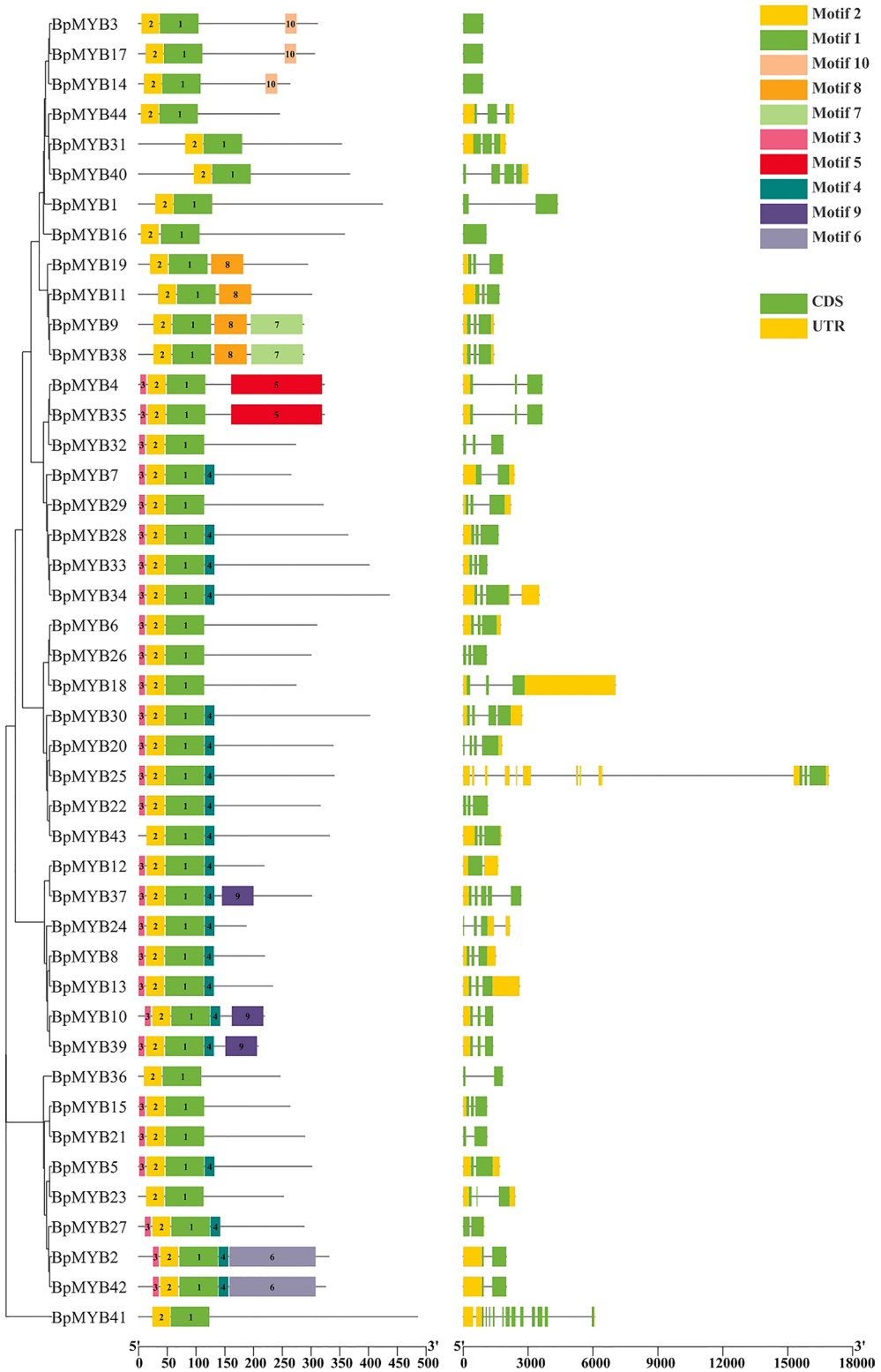


Fig. 2 Gene structures of the *BpR2R3-MYB* genes. **A** The phylogenetic tree of 44 BpR2R3-MYB proteins. **B** The conserved motifs in the BpR2R3-MYBs are represented by different colored boxes. **C** Exon/intron structures of BpR2R3-MYB genes. Green boxes, yellow boxes, and gray line indicate exons, untranslated regions (UTR), and introns at each BpR2R3-MYB gene, respectively

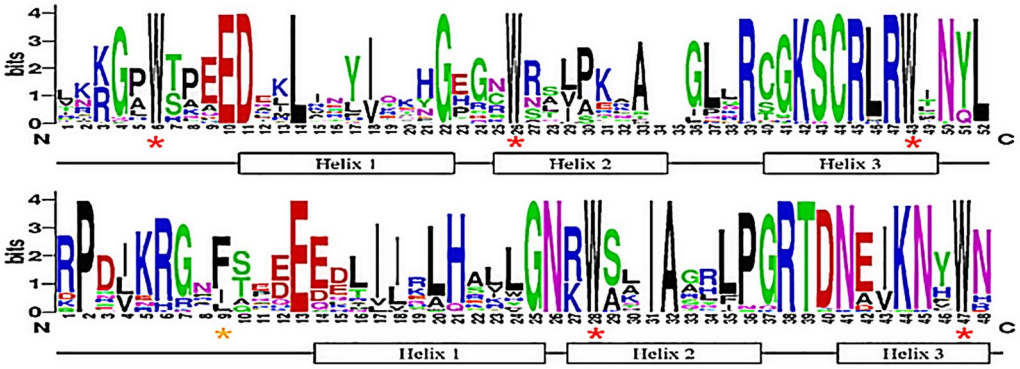


Fig. 3 Sequence conservation of the R2R3-MYB domain in *Betula platyphylla*. The bits indicate the conservation information content of each amino acid at specific positions in the sequence. Helix 1 to 3 are the positions of α-helice1 to 3, respectively. The red asterisks (*) represent the conserved tryptophan residues (W) in the R2R3-MYB repeats and the yellow asterisk, the replaced residues

Statistical analysis

The experiments were repeated three times. The data presented in the figures are means ± standard error. The data were analyzed through one-way ANOVA by using SPSS version 21.0. The different letters show significant differences among treatment means ($P < 0.05$, Tukey's test) [7].

Results

Identification of *BpR2R3-MYBs*

On the basis of the sequencing data of the full-length transcriptome of *B. platyphylla*, the genes with R2R3-MYB conserved structure domains were screened using the HMMer database, and the screened BpR2R3-MYB family genes were further verified using Pfam and CDD databases. Forty-four *BpR2R3-MYB* family genes were identified and numbered according to the order in which they were found (Table 1). The 44 BpR2R3-MYB proteins contained 186 (BpMYB24) to 484 (BpMYB41) amino acids, with molecular weights of 20.97 kDa (BpMYB24) and 54.60 kDa (BpMYB41) and an isoelectric point ranging from 5.12 (BpMYB38) to 9.57 (BpMYB10). The subcellular location predicted that most of the proteins were nuclear proteins. Only BpMYB2, BpMYB27, and BpMYB27 were distributed in the chloroplast. BpMYB16 and BpMYB23 were distributed in the cytoplasm and mitochondria, respectively.

Chromosome distribution of *BpR2R3-MYBs*

The 44 *BpR2R3-MYB* family genes were unevenly distributed in 11 chromosomes of *B. platyphylla*, and no distribution of *BpR2R3-MYB* was observed in chromosomes 1, 7, and 10 (Fig. 1). The maximum number of *BpR2R3-MYB* genes in one chromosome was six and located in chromosomes 3, 6, and 11. Two chromosomes (8 and 14) carried five *BpR2R3-MYB* genes, two chromosomes (5 and 12) had four *BpR2R3-MYB* genes, and four chromosomes (2, 4, 9, and 13) harbored two *BpR2R3-MYB* genes.

Analysis of the gene structure and conserved BpR2R3-MYB domains

The gene structure and domains of the BpR2R3-MYB proteins were analyzed using the online software MEME and TBtools. As shown in Fig. 2A, motifs with similar structures and domains were clustered into one clade, indicating that they had an analogous function. A total of 10 conserved amino acid motifs were identified in the BpR2R3-MYB proteins (Fig. 2B). Among them, Motifs 5, 6, 7, 8, 9, and 10 had no more than 5 occurrences in BpR2R3-MYB proteins, 20 BpR2R3-MYB proteins had Motif 4, 28 BpR2R3-MYB proteins had Motif 3, and all 44 BpR2R3-MYB proteins had the highly conserved R2-R3 structural domain of Motifs 1 and 2. Motif 1 was the R2-MYB structural domain (-W-(X19)-W-(X19)-W-), and Motif 2 was the R3-MYB structural domain (-F-(X19)-W-(X19)-W-) (Fig. 3). In the R3-MYB structural domain, the first W residue (position 9) was frequently substituted by phenylalanine (F) and less frequently by isoleucine (I), leucine (L), methionine (M), or tyrosine (Y).

To understand the structural diversity of BpR2R3-MYB, an exon–intron analysis was performed on the 44 *BpR2R3-MYBs* (Fig. 2C). The results showed that 90.90% (40/44) of the *BpR2R3-MYBs* had introns varying from 1 (*BpR2R3-MYB1*, *2*, *5*, *7*, *12*, *21*, *27*, *36*, *42*) to 12 (*BpR2R3-MYB 41*), and the four intron-less genes were *BpR2R3-MYB 3*, *14*, *16*, and *17*. In addition, 75% (33/44) of the *BpR2R3-MYBs* had untranslated regions (UTRs) varying

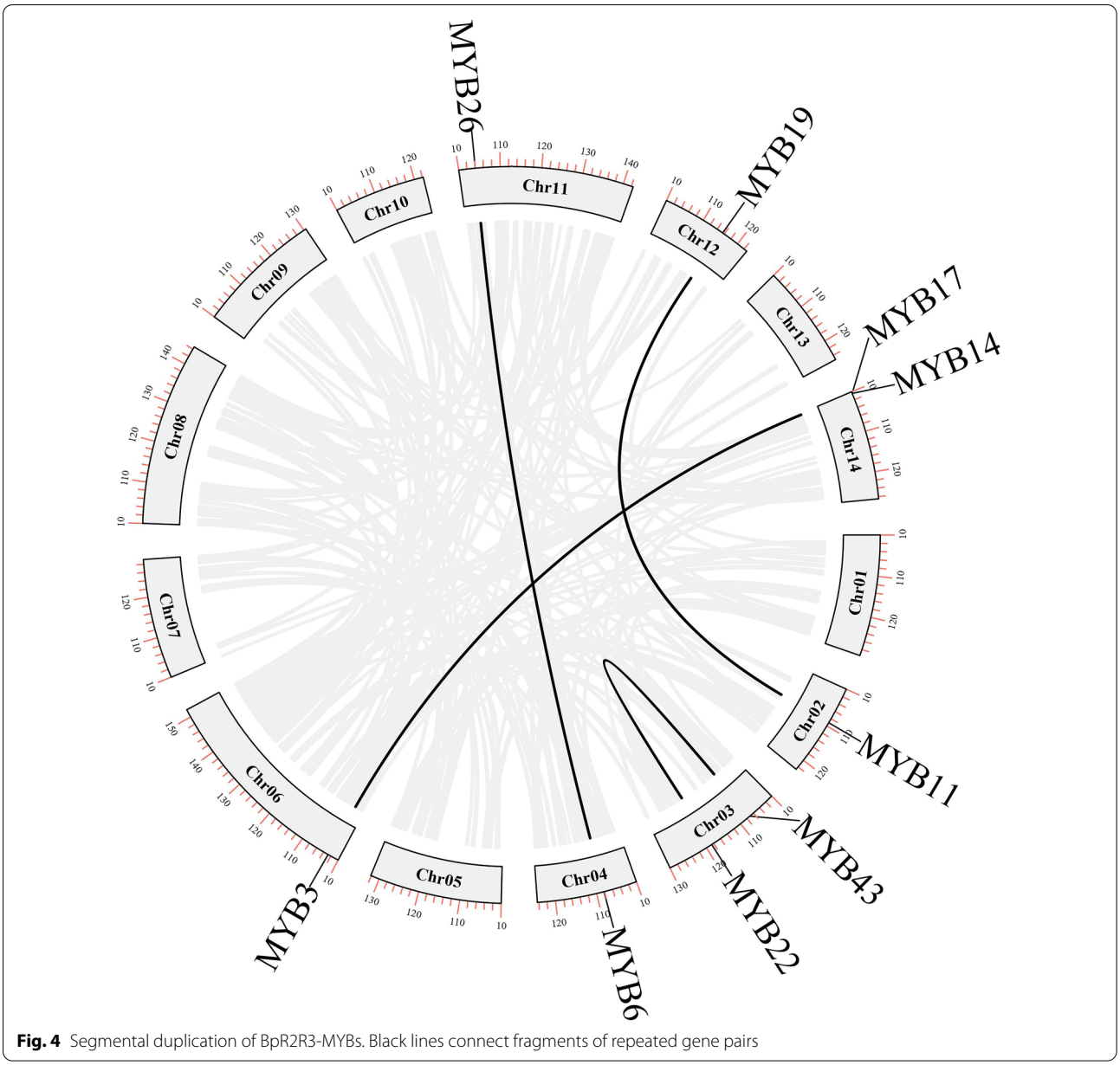


Fig. 4 Segmental duplication of BpR2R3-MYBs. Black lines connect fragments of repeated gene pairs

from 1 (*BpR2R3-MYB1*, *2*, *5*, *7*, *12*, *21*, *27*, *36*, *42*) to 3 (*BpR2R3-MYB 41*), and the 11 genes (*BpR2R3-MYB1*, *3*, *14*, *16*, *17*, *21*, *22*, *26*, *27*, *31*, *36*) had no UTR.

Evolutionary analysis of BpR2R3-MYBs

The Ka/Ks ratio was calculated to explore the evolutionary constraints on the BpR2R3-MYB gene family. The results showed that 20% (9/44) of the *BpR2R3-MYB* genes exhibited fragment duplication, and they were

Table 2 Ka, Ks and Ka/Ks of replication gene pairs of *BpR2R3-MYBs*

Duplicated gene pairs	Ka	Ks	Ka/Ks
BpMYB11&BpMYB19	0.35	2.12	0.17
BpMYB43&BpMYB22	0.24	1.84	0.13
BpMYB6&BpMYB26	0.23	1.21	0.19
BpMYB3&BpMYB17	0.24	1.81	0.13
BpMYB3&BpMYB14	0.24	1.81	0.13

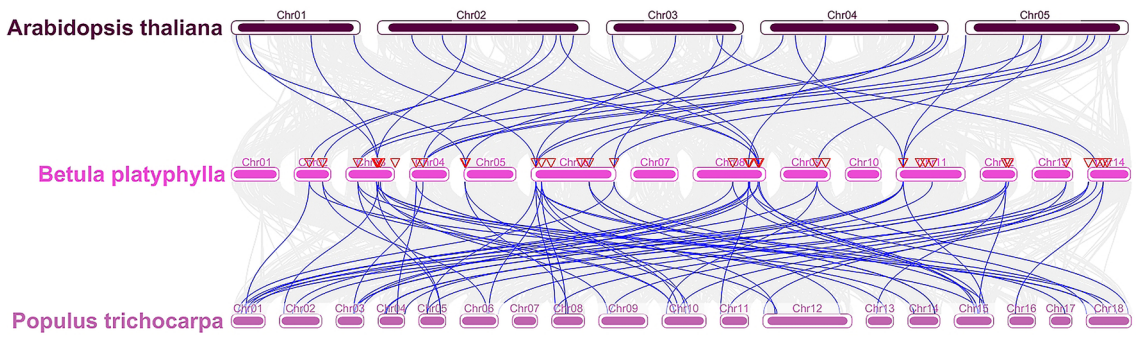


Fig. 5 Synteny analysis of BpR2R3-MYB genes in *Betula platyphylla, Arabidopsis thaliana*, and *Populus trichocarpa*. Gray lines in the background indicate the collinear blocks within *Betula platyphylla*, *Arabidopsis thaliana*, and *Populus trichocarpa*, the blue lines highlight the syntenic R2R3-MYB gene pairs

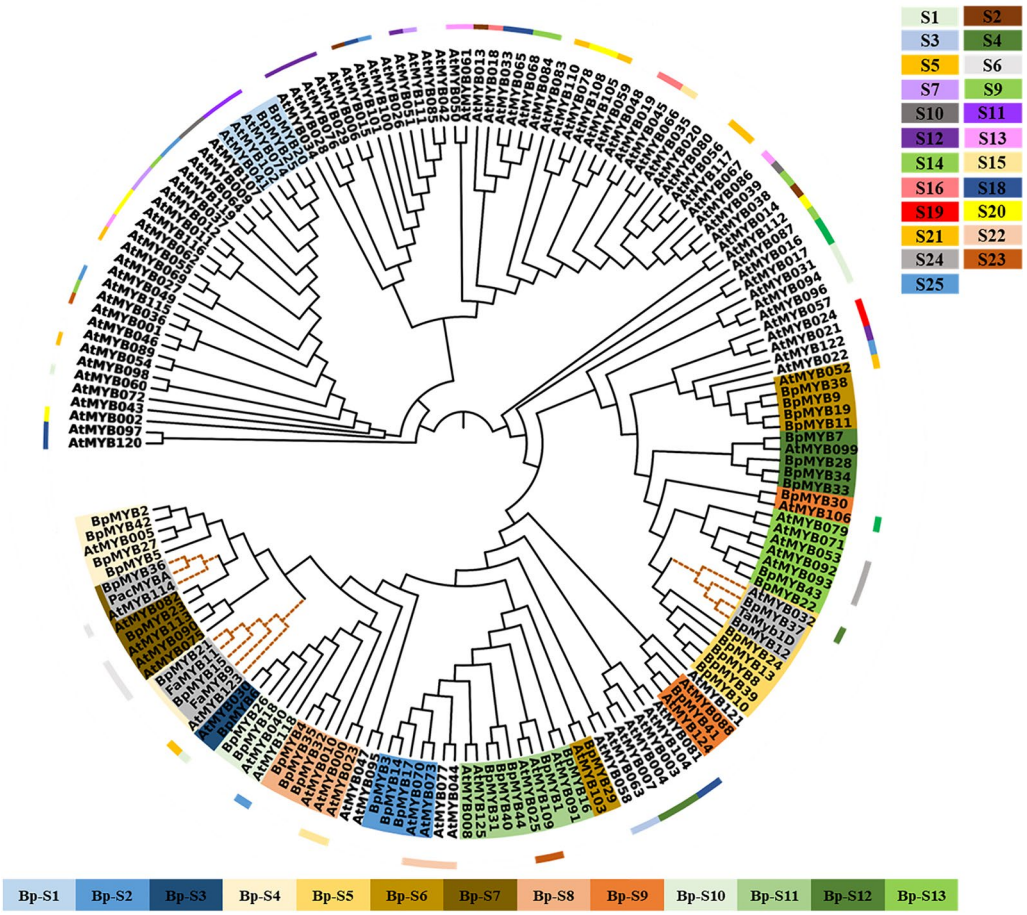


Fig. 6 Phylogenetic tree of *Betula platyphylla* and *Arabidopsis thaliana*. *Arabidopsis thaliana* subgroup classification markers are shown in the upper right corner, *Betula platyphylla* subgroup classification markers are shown in the bottom of figure, brown dashed lines indicate *BpR2R3-MYB* genes clustered with the reported flavonoid synthesis genes in one branch

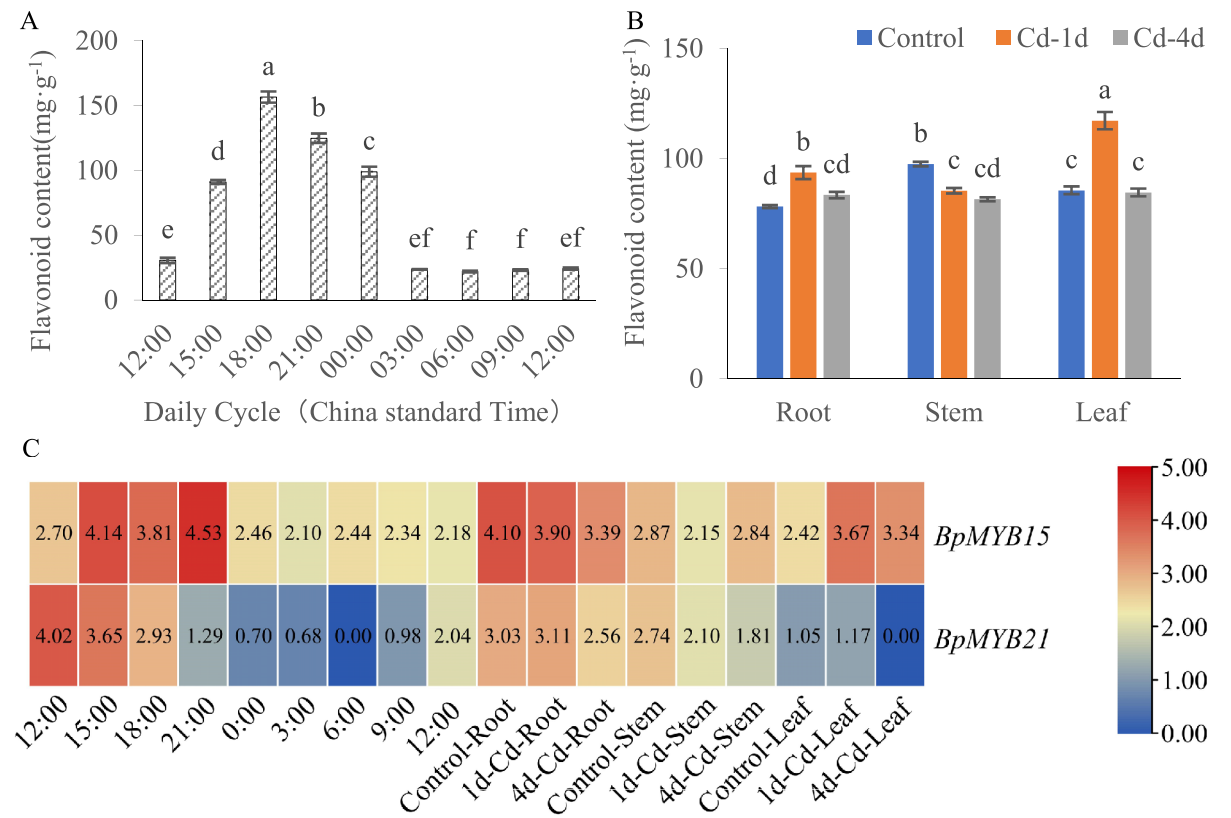


Fig. 7 Flavonoid content and gene expression of *BpR2R3-MYB15* and *BpR2R3-MYB21* under one daily cycle and 90 μmol L^{-1} Cd treatment. **A** Flavonoid content in one daily variation at 3-h intervals; **B** flavonoid content after 1d and 4d under 90 μmol L^{-1} Cd treatment; **C** gene expression of *BpR2R3-MYB15* and *BpR2R3-MYB21* under one daily cycle and 90 μmol L^{-1} Cd treatment. Different letters show significant differences among treatment means ($P < 0.05$, Tukey's test)

scattered in chromosomes 2, 3, 4, 6, 11, 12, and 14 (Fig. 4). Five gene pairs with segment duplication, namely, *BpMYB11*&*BpMYB19*, *BpMYB43*&*BpMYB22*, *BpMYB6*&*BpMYB26*, *BpMYB3*&*BpMYB17*, and *BpMYB3*&*BpMYB14* (Table 2), were found in the chromosomes. The Ka/Ks ratios of the five gene pairs were less than 1. In addition, the syntenic relationships of the R2R3-MYB genes showed that 30 orthologs existed between *B. platyphylla* and *A. thaliana,* and 68 orthologs existed between *B. platyphylla* and *Populus trichocarpa* (Fig. 5).

Phylogenetic analyses of BpR2R3-MYBs

In accordance with the classification of *Arabidopsis* R2R3-MYB proteins, we divided the BpR2R3-MYB proteins into 13 subgroups (Fig. 6). The average size of the subgroups was 3.38, and the size range was 1–7. Four R2R3-MYB proteins (*Fragaria x ananassa FaMYB9/FaMYB11*, *Prunus avium PacMYBA*, and *Triticum aestivum TaMyb1D*) related to flavonoid synthesis were also used to cluster with the 44 BpR2R3-MYB proteins. BpR2R3-MYB15 and BpR2R3-MYB21, BpR2R3-MYBB36, and BpR2R3-MYB12 and BpR2R3-MYB37 were clustered with *FaMYB9/FaMYB11*, *PacMYBA*, and *TaMyb1D*, respectively. We deduced that the five genes were related to flavonoid synthesis.

Correlation analysis of the flavonoid content and gene expression of *BpMYB15* and *BpMYB21*

The transcriptome sequencing data of nitrosoglutathione reductase (GSNOR) gene-silenced *B. platyphylla* plants (BpGSNOR-RNAi) and wild-type plants (WT) in our laboratory were used to analyze the correlation coefficients of the gene expression of the five *BpR2R3-MYBs* (*BpR2R3-MYB12*, *15*, *21*, *36*, and *37*) and the key enzyme genes of flavonoid synthesis (Additional file 1: Fig. S1 and Table S1). The results showed that the correlation coefficients of *BpR2R3-MYB15* and *BpR2R3-MYB21* were higher than those of the three other genes. Hence, we cloned *BpR2R3-MYB15* and *BpR2R3-MYB21* via PCR (Additional file 1: Fig. S2, 3).

Table 3 Correlation coefficients between flavonoid content and gene expression of *BpR2R3-MYB15* and *BpR2R3-MYB21*

Treatment	Time/tissue	*BpMYB15*	*BpMYB21*
Daily cycle	12:00	−0.12	−0.78
	15:00	−0.96	−0.23
	18:00	−0.69	−0.04
	21:00	0.83	0.96
	0:00	−0.88	−0.16
	3:00	−0.88	−0.69
	6:00	0.49	−0.72
	9:00	−0.99*	0.19
	12:00	0.07	−0.79
Control	Root	0.99*	−0.95
	Stem	−0.50	−0.54
	Leaf	0.98*	0.57
Cd-1d	Root	−0.93	0.96*
	Stem	0.25	0.99*
	Leaf	0.98*	−0.28
Cd-4d	Root	−0.49	−0.31
	Stem	−0.14	−0.72
	Leaf	−0.98*	0.93

Leaves of 18 decade-old *B. platyphylla* trees planted in Northeast Forestry University, which collected 3-h intervals on August 3–4 of 2020, were used for daily cycle analysis; leaves, stem, and root of 30-day-old seedlings of *B. platyphylla* were used for Cd treatment analysis

The flavonoid content and gene expression of *BpR2R3-MYB15* and *BpR2R3-MYB21* were further analyzed under one daily cycle and 90 μmol L^{-1} Cd treatment (Fig. 7). In one daily variation, the flavonoid content peaked at 18:00, and the time with high flavonoid content was from 15:00 to 0:00. The gene expression of *BpR2R3-pMYB15* and *BpR2R3-MYB21* peaked at 21:00 and 12:00, respectively. The flavonoid content and gene expression of *BpR2R3-pMYB15* and *BpR2R3-MYB21* in the leaves of the *B. platyphylla* plants reached the highest one day after Cd treatment, but the gene expression of *BpR2R3-pMYB15* and *BpR2R3-MYB21* in the stem and root of *B. platyphylla* mostly decreased after Cd treatment. The correlation coefficient of the gene expression of *BpR2R3-pMYB15* and flavonoid content was higher than that of of *BpR2R3-pMYB21* and flavonoid content (Table 3).

Overexpression of *BpR2R3-MYB15* enhanced flavonoid production

After 3 days of *Agrobacterium*-mediated transient transformation, the overexpression of *BpR2R3-MYB15* in *B. platyphylla* calli (5.72 times than that of untransformed calli) significantly enhanced the flavonoid and procyanidin contents and increased the gene expression of *BpCHI1*, *BpF3H*, and *BpDFR*, which are key enzyme genes for flavonoid synthesis. The silencing of *BpR2R3-MYB15* in *B. platyphylla* calli (0.68 times than that of untransformed calli) decreased the flavonoid and procyanidin contents and reduced the gene expression of *BpCHI1*, *BpF3H*, and *BpDFR* (Fig. 8).

Discussion

Transcriptome analysis provides insights into the spatiotemporal genes transcribed during plant growth and development processes or stress responses [31]. To clarify the function of R2R3-MYB family genes in the seedling development period of *B. platyphylla*, complete full-length transcriptome data of 30-day-old seedlings were generated using the PacBio Sequel System II, and 44 typical BpR2R3-MYB family genes with complete domains were identified. This study updated the collection of BpR2R3-MYB family genes in *B. platyphylla*.

The reported *B. platyphylla* genome provided an opportunity to investigate the gene structure and synteny of the identified BpR2R3-MYB family genes [18]. The 44 BpR2R3-MYB proteins had highly conserved R2 (-W-(X19)-W-(X19)-W-) and R3 (-F-(X19)-W-(X19)-W-) structural domains. The first W residue (position 9) in the R3-MYB structural domain is frequently substituted by phenylalanine (F) and less frequently by isoleucine (I), leucine (L), methionine (M), or tyrosine (Y). These substitutions in the R3 structural domain may result in the recognition of novel target genes and/or may significantly impair the DNA-binding activity [32]. Phylogenetic analysis of the BpR2R3-MYBs in this study showed that the genes in the same subgroups or subclades generally contained the same intron pattern, and most genes had no more than two introns. This result is in line with the results for other plants [33]. In addition, four genes had no intron, and *BpR2R3-MYB 41* had 12 introns. The *BpR2R3-MYB*s with different numbers of introns may be one of the reasons for the enlarged family members or functional diversity in *B. platyphylla*.

Plants experience gene duplication events, including tandem, fragment, and conversion duplication, in the process of evolution [34]. In our study, five gene pairs with segment duplication were found, and their Ka/Ks ratios were all less than 1. This result indicates that 22.73% of the genes of BpR2R3-MYBs (10/44) evolved under the effect of purifying selection. In addition, the number of orthologs between *B. platyphylla* and *P. trichocarpa* (68) was twice that between *B. platyphylla* and *A. thaliana* (30). We infer that *B. platyphylla* may be closer to *P. trichocarpa* than to *A. thaliana* in evolutionary branch, and similar results have been reported for other family genes of *B. platyphylla* [35, 36].

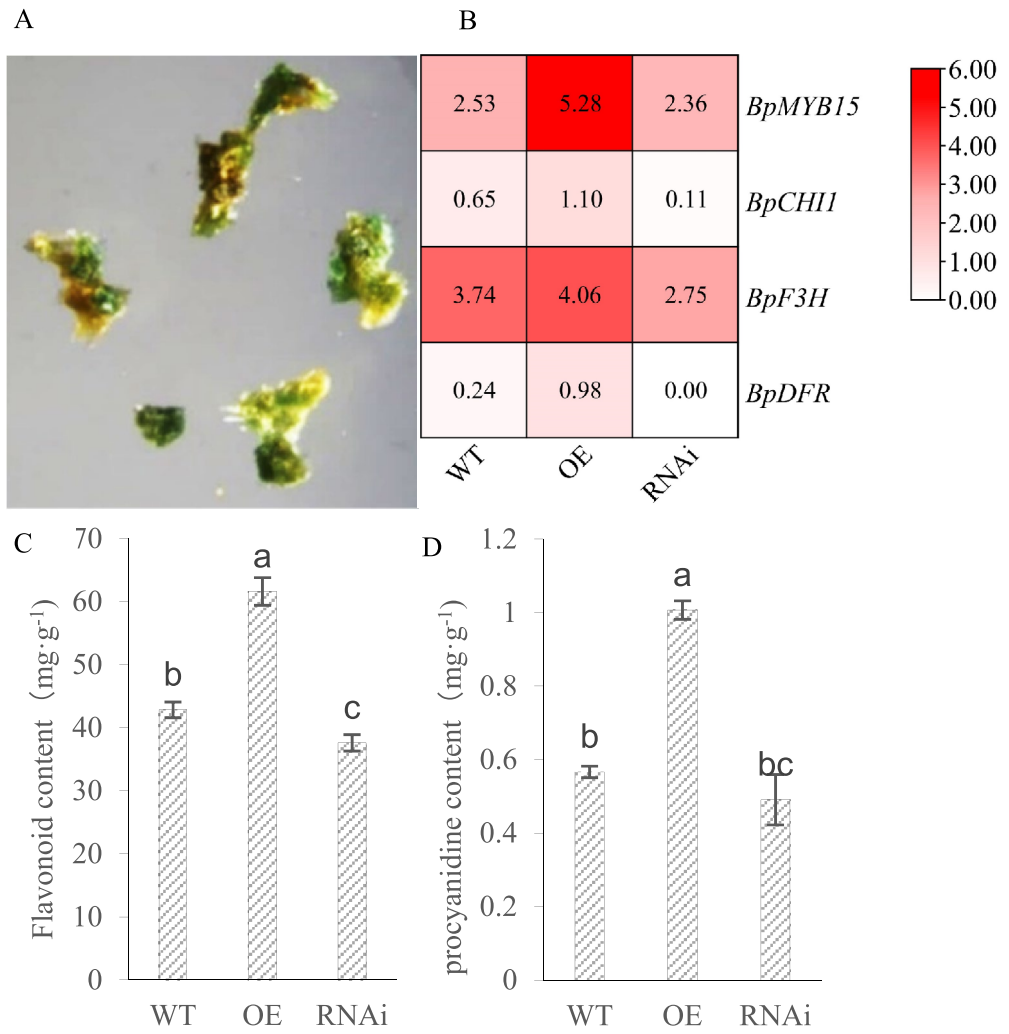


Fig.8 *BpR2R3-MYB15* regulated flavonoid production using *Agrobacterium*-mediated transient transformation. **A** GUS staining of transient transformed calli; **B** gene expression of *BpCHI1*, *BpF3H* and *BpDFR* in transient transformed calli; **C**, **D** flavonoid and procyanidins content in transient transformed calli, respectively. Different letters show significant differences among treatment means ($P < 0.05$, Tukey's test)

Forty-four BpR2R3-MYBs were unevenly distributed in 11 chromosomes of *B. platyphylla*, and similar results have been derived for *Arabidopsis*, *P. trichocarpa*, and six Rosaceae species [37, 38]. Combined with the results of the phylogenetic analyses, we found that 52.27% (23/44) of the BpR2R3-MYBs clustered into one subgroup were distributed in the same chromosome (Additional file 1: Table S3). In general, the members of the same subgroup had similar functions. Whether the above-mentioned BpR2R3-MYB genes in a subgroup have the same function and whether there is substitution or superposition of BpR2R3-MYB genes with the same function in the same chromosome will be verified experimentally in the future.

To clarify the identified BpR2R3-MYBs in flavonoid biosynthesis of *B. platyphylla*, *BpR2R3-MYB15* related to flavonoid biosynthesis was screened out based on phylogenetic analyses. The correlation coefficients between the gene expression of *BpR2R3-MYBs* and the key enzymes of flavonoid synthesis were determined using *GSNOR* transgenic seedlings via RNAi silencing or by using wild plants under one daily cycle and 90 μmol L^{-1} Cd treatment. On this basis, we further verified the function of *BpR2R3-MYB15* in flavonoid biosynthesis by using *Agrobacterium*-mediated transient transformation in the calli of *B. platyphylla*. Next, We will obtain *BpR2R3-MYB15*-transformed birch plants to further analyze its function

and mechanism in flavonoid synthesis. The results of this study lay a foundation for analyzing how *BpR2R3-MYBs* regulate flavonoid biosynthesis and the function of flavonoids in seedling development.

Conclusions

In this study, 44 BpR2R3-MYB family genes were identified based on the full-length transcriptome of 30-day-old seedlings of *Betula platyphylla*, and their gene structure, chromosome distribution, and syntenic relationships were analyzed at the genomic level. *BpR2R3-pMYB15*, one of the five *BpR2R3-MYBs* clustered with R2R3-MYB genes related to flavonoid synthesis, positively regulated flavonoid production via *Agrobacterium*-mediated transient transformation.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40538-022-00301-7.

Additional file 1: Table S1. Sequences of primer pairs for quantitative real-time RT-PCR assay. **Table S2.** Correlation coefficient between R2R3-MYB transcription factor and key enzyme genes of flavonoid synthesis pathway in *Betula platyphylla*. **Fig. S1.** Heat map of the gene expression of the R2R3-MYBs and key enzymes in the flavonoid synthesis pathway. **Fig. S2.** Electrophoresis chart of PCR products of BpMYB15 and BpMYB21. Note: A, BpMYB15; B, BpMYB21. **Fig. S3.** Amino acid sequences of BpMYB15 and BpMYB21. Note: A, BpMYB15; B, BpMYB21. **Table S3.** BpR2R3-MYBs clustered into one clade distributed on the same chromosome. **Fig. S4.** The plasmid map of Cam1304-SubC-BpMYB15 (A) and RhRNA-pTRV2-BpMYB15 (B)

Acknowledgements
The vector was donated by Dr. Yan P (Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences).

Author contributions
G-ZF conceived and designed the experiments. Z-YY and H-ML performed the research. G-ZF and B-JZ analyzed the data and wrote the paper. All authors have read and approved final manuscript.

Funding
This work was supported by the National Natural Science Foundation of China (32171829), Heilongjiang Touyan Innovation Team Program (Tree Genetics and Breeding Innovation Team).

Availability of data and materials
The data that support the finding of this study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
This manuscript is an original paper and has not been published in other journals. The authors agreed to keep the copyright rule.

Consent for publication
The authors agreed to the publication of the manuscript in this journal.

Competing interests
The authors declare no conflict of interests.

Received: 28 January 2022 Accepted: 3 May 2022
Published online: 03 June 2022

References

1. Taylor LP, Grotewold E. Flavonoids as developmental regulators. Curr Opin Plant Biol. 2005;8:317–23.
2. Alseekh S, Souza LP, Benina M, Fernie AR. The style and substance of plant flavonoid decoration; towards defining both structure and function. Phytochemistry. 2020;174: 112347.
3. Schijlen EGWM, Vos CHRD, Tunen AJV, Bovy AG. Modification of flavonoid biosynthesis in crop plants. Phytochemistry. 2004;65:2631–48.
4. Shi JB, Zhou H, Liu XH, Wang N, Xu QH, Yan GT. Correlation analysis of the transcriptome and metabolome reveals the role of the flavonoid biosynthesis pathway in regulating axillary buds in upland cotton (*Gossypium hirsutum* L.). Planta. 2021;254:7.
5. Chen WF, Xiao ZC, Wang YL, Wang JX, Zhai R, Wang KL, Espley R, Ma FW, Li PM. Competition between anthocyanin and kaempferol glycosides biosynthesis affects pollen tube growth and seed set of *Malus*. Hortic Res. 2021;8:173.
6. Ono E, Ruike M, Iwashita T, Nomoto K, Fukui Y. Co-pigmentation and flavonoid glycosyltransferases in blue *Veronica persica* flowers. Phytochemistry. 2010;71:726–235.
7. Yang JL, Zhang LX, Jiang L, Zhan YG, Fan GZ. Quercetin alleviates seed germination and growth inhibition in *Apocynum venetum* and *Apocynum pictum* under mannitol-induced osmotic stress. Plant Physiol Bioch. 2021;159:268–76.
8. Schenke D, Utami HP, Zhou Z, Gallegos MT, Cai DG. Suppression of UV-B stress induced flavonoids by biotic stress: Is there reciprocal crosstalk? Plant Physiol Bioch. 2019;134:53–63.
9. Ferreyra MLF, Rius SP, Casati P. Flavonoids: biosynthesis, biological functions, and biotechnological applications. Front Plant Sci. 2012;3:222.
10. Seth R, Kushwaha S, Luqman S, Meena A. Flavonoids as prospective aromatase inhibitors in breast cancer prevention/therapy. Curr Mol Pharmacol. 2021;14:1112–24.
11. Liu WX, Feng Y, Yu SH, Fan ZQ, Li XL, Li JY, Yin HF. The Flavonoid Biosynthesis Network in Plants. Int J Mol Sci. 2021;22:12824.
12. Xu WJ, Dubos C, Lepiniec L. Transcriptional control of flavonoid biosynthesis by MYB–bHLH–WDR complexes. Trends Plant Sci. 2015;20:176–85.
13. Stracke R, Ishihara H, Barsch GHA, Mehrtens F, Niehaus K, Weisshaar B. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. Plant J. 2007;50:660–77.
14. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. MYB transcription factors in *Arabidopsis*. Trends Plant Sci. 2010;15:573–81.
15. Park JS, Kim JB, Cho KJ, Cheon CI, Sung MK, Choung MG, Roth KH. *Arabidopsis* R2R3-MYB transcription factor AtMYB60 functions as a transcriptional repressor of anthocyanin biosynthesis in lettuce (*Lactuca sativa*). Plant Cell Rep. 2008;27:985–94.
16. Yang L, Zhao X, Ran LY, Li CF, Fan D, Luo KM. PtoMYB156 is involved in negative regulation of phenylpropanoid metabolism and secondary cell wall biosynthesis during wood formation in poplar. Sci Rep. 2017;7:41209.
17. Zhou LJ, Geng Z, Wang Y, Wang Y, Liu S, Chen C, Song A, Jiang J, Chen S, Chen F. A novel transcription factor CmMYB012 inhibits flavone and anthocyanin biosynthesis in response to high temperatures in chrysanthemum. Hortic Res. 2021;8:248.
18. Chen S, Wang YC, Yu LL, Zheng T, Wang S, Yue Z, Jiang J, Kumari S, Zheng CF, Tang HB, Li J, Li YQ, Chen JJ, Zhang WB, Kuang HH, Robertson JS, Zhao PX, Li HY, Shu SQ, Yordanov YS, Huang HJ, Goodstein DM, Gai Y, Qi Q, Min JM, Xu CY, Wang SB, Qu GZ, Paterson AH, Sankoff D, Wei HR, Liu GF, Yang CP. Genome sequence and evolution of *Betula platyphylla*. Hortic Res. 2021;8:37.
19. Liu Y, Miao XY, Wen LK. Study on the influence on flavonoid extracts from birch leaves on antioxidant activity of human umbilical vein endothelial cells. Sci Tech food Ind. 2017;38:82–6.

20. Jiang J, Li XY, Wang C, Wang F, Jiang J. Evaluation of salt tolerant performance of *BpCHS3* transgenic plants in *Betula platyphylla*. J Beijing Forestry Univ. 2019;41:1–6.
21. Schultz J, Milpetz F, Bork P, Ponting CP. SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci. 1998;95(11):5857–64.
22. Chen CJ, Chen H, Zhang Y, Thomas HR, Frank MH, He YH, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020;13:1194–202.
23. Schaart JG, Dubos C, De La Fuente IR, van Houwelingen AMML, de Vos RCH, Jonker HH, Xu WJ, Routaboul JM, Lepiniec L, Bovy AG. Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (*Fragaria x ananassa*) fruits. New Phytol. 2013;197:454–67.
24. Wei QH, Zhang F, Sun FS, Luo QC, Wang RB, Hu R, Chen MJ, Chang JL, Yang GX, He GY. A wheat MYB transcriptional repressor *TaMyb1D* regulates phenylpropanoid metabolism and enhances tolerance to drought and oxidative stresses in transgenic tobacco plants. Plant Sci. 2017;265:112–23.
25. Shen XJ, Guo XW, Guo X, Zhao D, Zhao W, Chen JS, Li TH. *PacMYBA*, a sweet cherry R2R3-MYB transcription factor, is a positive regulator of salt stress tolerance and pathogen resistance. Plant Physiol Bioch. 2017;112:302–11.
26. Zhu K, Wang X, Liu J, Tang J, Cheng Q, Chen JG, Cheng ZM. The grapevine kinome: annotation, classification and expression patterns in developmental processes and stress responses. Hortic Res. 2018;5:19.
27. Yan P, Zeng YJ, Shen WT, Tuo DC, Li XY, Zhou P. Nimble cloning: a simple, versatile, and efficient system for standardized molecular cloning. Front Bioeng Biotechnol. 2020;7:460.
28. Zhou YP, Lu YH, Wei DZ. Antioxidant activity of a Flavonoid-Rich extract of *Hypericum perfuratum* L. in vitro. J Agr Food Chem. 2004;52:5032–9.
29. Zhang Y, Ye JL, Liu CY, Xu Q, Long LC, Deng XX. The citrus CsPH4-Noemi regulatory complex is involved in proanthocyanidin biosynthesis via a positive feedback loop. J Exp Bot. 2020;71:1306–21.
30. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. Methods. 2001;25:402–8.
31. Shinozaki Y, Nicolas P, Fernandez-Pozo N, Ma QY, Evanich DJ, Shi YN, Xu YM, Zheng Y, Snyder SI, Martin LBB, Ruiz-May E, Thannhauser TW, Chen KS, Domozych DS, Catalá C, Fei ZJ, Mueller LA, Giovannoni JJ, Rose JKC. High-resolution spatiotemporal transcriptome mapping of tomato fruit development and ripening. Nat Commun. 2018;9:364.
32. Du H, Feng BR, Yang SS, Huang YB, Tang YX. The R2R3-MYB transcription factor gene family in Maize. PLoS ONE. 2012;7: e37463.
33. Guan YL, Liu SQ, Wu WH, Hong KY, Li RX, Zhu LM, Liu Y, Lu Y, Chen JH, Yang LM, Shi JS. Genome-wide identification and cold stress-induced expression analysis of the CBF gene family in Liriodendron chinense. J Forestry Res. 2021;32:2531–43.
34. Qiao X, Li QH, Yin H, Qi KJ, Li LT, Wang RZ, Zhang SL, Paterson AH. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. Genome Biol. 2019;20:38.
35. Chen BW, Ali S, Zhang X, Zhang YL, Wang M, Zhang QZ, Xie LN. Genome-wide identification, classification, and expression analysis of the JmjC domain-containing histone demethylase gene family in birch. BMC Genomics. 2021;22:772.
36. Chen S, Lin X, Zhang DW, Li Q, Zhao XY, Chen S. Genome-wide analysis of NAC gene family in *Betula pendula*. Forests. 2019;10:741.
37. Yang XY, Li J, Guo T, Guo B, Chen Z, An XM. Comprehensive analysis of the R2R3-MYB transcription factor gene family in *Populus trichocarpa*. Ind Crops Prod. 2021;168: 113614.
38. Liu H, Xiong JS, Jiang YT, Wang L, Cheng ZM. Evolution of the R2R3-MYB gene family in six Rosaceae species and expression in woodland strawberry. J Interg Arg. 2019;18:2753–70.

Publisher's Note