

RESEARCH

Open Access



Genome-wide identification and characterization of PdbHLH transcription factors related to anthocyanin biosynthesis in colored-leaf poplar (*Populus deltoids*)

Xiao-jing Wang¹, Xu-qian Peng², Xiao-chun Shu³, Yu-hang Li³, Zhong Wang³ and Wei-bing Zhuang^{3*}

Abstract

Basic helix-loop-helix (bHLH) proteins are transcription factors (TFs) that have been shown to regulate anthocyanin biosynthesis in many plant species. However, the *bHLH* gene family in *Populus deltoids* has not yet been reported. In this study, 185 *PdbHLH* genes were identified in the *Populus deltoids* genome and were classified into 15 groups based on their sequence similarity and phylogenetic relationships. Analysis of the gene structure, chromosome location and conserved motif of the *PdbHLH* genes were performed by bioinformatic methods. Gene duplication analyses revealed that 114 *PdbHLH* were expanded and retained after WGD/segmental and proximal duplication. Investigation of cis-regulatory elements of *PdbHLH* genes indicated that many *PdbHLH* genes are involved in the regulation of anthocyanin biosynthesis. The expression patterns of *PdbHLHs* were obtained from previous data in two colored-leaf poplar (QHP and JHP) and green leaf poplar (L2025). Further analysis revealed that 12 candidate genes, including 3 genes (*PdbHLH57*, *PdbHLH143*, and *PdbHLH173*) from the subgroup III(f) and 9 gene from other groups, were positively associated with anthocyanin biosynthesis. In addition, 4 genes (*PdbHLH4*, *PdbHLH1*, *PdbHLH18*, and *PdbHLH164*) may be involved in negatively regulating the anthocyanin biosynthesis. These results provide a basis for the functional characterization of *bHLH* genes and investigations on the molecular mechanisms of anthocyanin biosynthesis in colored-leaf poplar.

Keywords: Colored-leaf poplar, Anthocyanin biosynthesis, *PdbHLH* genes, Phylogenetic analysis, Structure analysis, Expression pattern

Introduction

Colored-leaf plants played important roles in landscaping and urban beautification, which can form a nice scenery [1, 2]. Poplar is widely planted around the world due to its fast growth and better resistance to adversity, which can be used for producing timber, pulp, and paper [3].

Recently, many kinds of colored-leaf poplars have been bred, such as 'Zhonghong polar'(ZHP), 'Quanhong poplar' (QHP), 'Jinhong poplar' (JHP) and 'Caihong poplar' (CHP), which also bring great economic, social and ecological benefits [4, 5]. However, the molecular mechanisms of pigment formation in colored-leaf poplar are still unclear, and needed to be exploring.

The molecular mechanisms of anthocyanins biosynthesis in many species have been well characterized, such as *Arabidopsis* and rice. The anthocyanin biosynthesis was regulated mainly by two kinds of genes, enzyme-coding structural genes and transcription

*Correspondence: weibingzhuangnj@sina.com

³ The Jiangsu Provincial Platform for Conservation and Utilization of Agricultural Germplasm, Institute of Botany, Jiangsu Province and Chinese Academy of Sciences (Nanjing Botanical Garden Mem. Sun Yat-Sen), Nanjing 210014, China

Full list of author information is available at the end of the article



factor genes [6, 7]. The structural genes are conserved in many kinds of species, mainly including phenylalanine ammonia-lyase (*PAL*), chalcone synthase (*CHS*), chalcone isomerase (*CHI*), flavanone 3-hydroxylase (*F3H*), dihydroflavonol 4-reductase (*DFR*), anthocyanidin synthase (*ANS*), and flavonoid 3-O-glucosyltransferase (*UF3GT*), and the stable anthocyanins can be synthesized by various modifications, such as methylation, glycosylation and acylation [8, 9]. Several transcription factors have been reported to be involved in the anthocyanin biosynthesis, mainly including MYB, bHLH and WD40. The MYB transcription factors can regulate the anthocyanin biosynthesis independently or form MYB-bHLH-WD40 (MBW) complexes to regulate the anthocyanin biosynthesis [6, 10]. Up to now, many MYB TFs associated with anthocyanins biosynthesis has been studied in poplar, such as PtrMYB116, PtrMYB117, PtrMYB118, PtrMYB119, PtrMYB120 [11–13]. However, the functions of bHLH TFs in poplar have been investigated fewer compared with these of MYB TFs.

The bHLH TFs are the second largest family of plant TFs, and play a central role in anthocyanins biosynthesis [14, 15]. The conserved bHLH domain contains 50 to 60 amino acids and processes two functional regions: basic region and HLH region [16]. The basic region contained approximately 17 amino acids, which located at the N-terminus of bHLH domain and binded to a consensus hexanucleotide E-box (CANNTG) [17]. The HLH region with 50 amino acids was used to form the homodimers or heterodimers, which includes two alpha helixes separated by a variable loop [18]. bHLH TFs are involved in several plant metabolic pathways including flavonoids and anthocyanin biosynthesis [19]. The Lc protein was the first reported bHLH TFs in maize, which can be involved in the anthocyanin biosynthesis through regulating at least 2 structural genes [20]. Moreover, several bHLH TFs related to anthocyanin biosynthesis were identified and further characterized in other plants, such as *Arabidopsis* (*AtEGL3*, *AtGL3*, and *AtTT8*) [21], *V. vinifera* (*VvMYCA1*) [22], *N. tabacum* (*NtAn1* and *NtAn2*) [23], *M. domestica* (*MdbHLH3* and *MdMYC2*) [24, 25] and poplar (*PdTT8*) [26]. In poplar, PdTT8 directly interacts with PdMYB118 TF to regulate wound-induced anthocyanin biosynthesis [26]. However, several questions are still needed to be explored. Is there other bHLH TFs involved in the anthocyanin biosynthesis in poplar? What's the difference among the different bHLH TFs associated with anthocyanin biosynthesis in different species? What's the difference among the different bHLH TFs involving the anthocyanin biosynthesis in poplar? The genome-wide identification and characterization of *bHLH* gene

family in poplar could give us the answer of these questions.

The genome-wide identification and characterization of *bHLH* gene family has been conducted in many plants [14, 27, 28]. There is 162 *bHLH* genes in *Arabidopsis*, 192 in tobacco, 159 in tomato, and 188 in apple, which can be divided into 15–26 subfamilies [28]. Among these subfamilies, members of the III subfamily have been proved to be involved in anthocyanin synthesis [29]. Therefore, bHLH TF in poplar falling into the III subfamily might also be involved in anthocyanin synthesis. In present study, the phylogenetic analysis, gene or protein structures, gene duplication, and chromosome distribution of PdbHLH transcription factors were systematically and comprehensively investigated. Moreover, the expression pattern of *PdbHLH* genes in green leaf poplar (L2025) and colored leaf poplar (QHP) was evaluated with the released RNA-seq data [5]. To better explore the functions of *PdbHLH* genes in poplar, the expression pattern of *PdbHLH* genes in green leaf poplar (L2025) and colored leaf poplar (JHP) was further analyzed with the released RNA-seq data [30]. Our findings should not only provide a characterization of the *PdbHLH* gene superfamily but also provide insight into the roles of *PdbHLH* genes in the regulation of anthocyanin biosynthesis in colored-leaf poplar.

Methods

Identification of *PdbHLH* gene family in *P. deltoids*

The genome sequence and corresponding annotations of *P. deltoids* was downloaded from the DOE Joint Genome Institute website [31] (<http://genome.jgi.doe.gov/>). The hidden Markov Model profile of HLH (PF00010) domain obtained from Pfam database [32] (<http://pfam.xfam.org/>) was used to search candidate *PdbHLH* genes from *P. deltoids* genome using the HMMER3 software package [33]. Moreover, the physical localizations of all candidate genes and redundant sequences with the same chromosome location and short proteins (length < 100 aa) were further checked. Next, the Pfam [32] and SMART [34] (<http://smart.embl-heidelberg.de/>) databases was used to further verify the presence of HLH domain for all candidate protein sequences. After the above three steps, the identified protein sequences that contained the core domain (HLH) of known PdbHLH were regarded as putative homologs in the study.

Sequence analysis and structural characterization of *PdbHLH* genes in *P. deltoids*

Based on the genome sequence and annotation information, the exon-intron organization of *PdbHLH* genes, including intron distribution, number, and phases, was graphically displayed by the Gene Structure

Display Server GSDS2.0 (<http://gsds.cbi.pku.edu.cn/>). The MEME suite (<http://meme-suite.org/tools/meme>) was used to identify the conserved motifs of PdbHLH proteins [35]. The optimized parameters were employed for the analysis as follows: maximum numbers of different motifs, 20; minimum motif width, 6bp; maximum motif width, 50bp.

Chromosome distribution, gene duplication, and synteny analysis of *PdbHLH* genes in *P. deltoids*

The chromosome distribution of *PdbHLH* genes was extracted from the genome annotation database, and the MapChart software was used to visualize the chromosomal locations of the *PdbHLH* gene [36]. Gene duplication analyses for *P. deltoids* was conducted using the Multiple Collinearity Scan Toolkit (MCScanX) [37]. To identify candidate homologous gene pairs ($E < 1e^{-5}$), BLASTp was performed to search for potential homologous gene pairs across the whole *P. deltoids* genome. The potential homologous gene pairs were inputted into the program MCScanX with the default parameters to identify syntenic chains. MCScanX was used to further distinguish among whole-genome duplication (WGD)/segmental, dispersed proximal, and tandem duplication events in *PdbHLH* gene family [37]. Gene pairs identified in the same synteny block were used to calculate Ka and Ks values using PAML package [38].

Cis-elements in the promoter regions of *PdbHLH* genes and GO enrichment analyses

Conserved cis-regulatory elements in the promoter region of *PdbHLH* genes were identified by analyzing the 2000-bp sequence upstream of the transcription start site (TSS) obtained from TBtools. Promoter sequence analysis was performed using PlantCARE [39, 40].

GOATOOLS (<http://github.com/tanghaibao/GOatools>) was used to perform GO annotations for *PdbHLH* genes. The biological function enrichment analysis for *PdbHLH* genes was conducted using Fisher's exact test. Moreover, the Bonferroni multiple testing correction was used to minimize false positives, and functions were considered to be significantly enriched when their Bonferroni-corrected *P*-values (Padjust) were < 0.05 .

Phylogenetic relationships analysis of *PdbHLH* gene family in *P. deltoids*

The bHLH protein sequences, including PdbHLH, OsbHLH, and AtbHLH proteins from poplar, rice and *Arabidopsis* were aligned using the Muscle algorithm with default parameters. The best-fit model of protein evolution was selected using the Model-Generator program [41]. An unrooted neighbor-joining phylogenetic tree was constructed through multiple sequence alignments

of these bHLH proteins using MEGA 10.0, and these bHLH proteins were further grouped into different clades based on the topology of the phylogenetic tree. The parameters were as follows: pairwise deletion, Poisson model, and 1000 bootstrap replications.

Expression analyses of *PdbHLH* genes in 'Jinghong poplar' (JHP), 'Quanhong poplar' (QHP) and *Populus sp.* Linn. 2025 (L2025) by RNA-seq

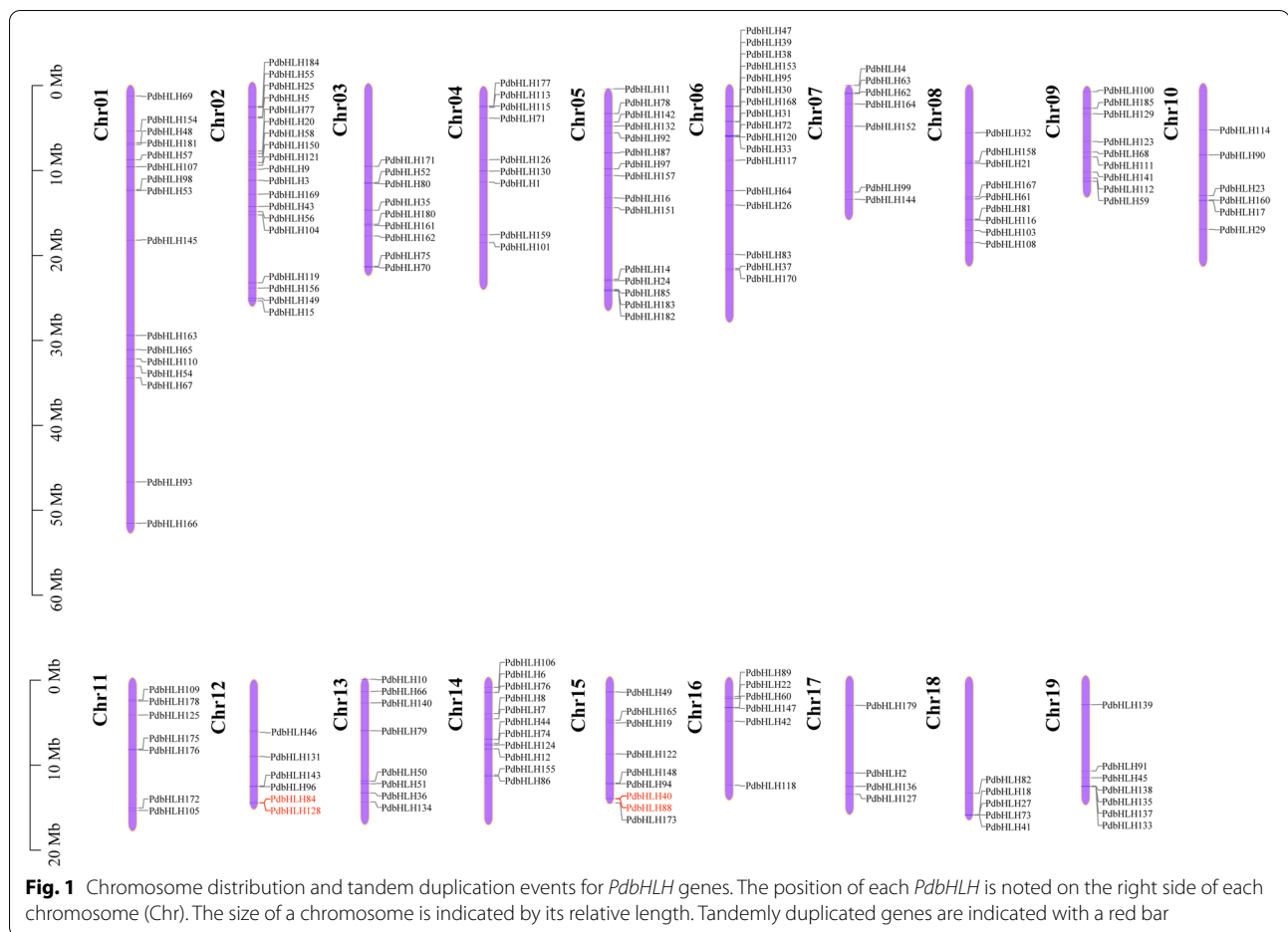
RNA-seq data for the *PdbHLH* genes were obtained from previous studies of differential gene expression in colored-leaf poplar (JHP and QHP) and green leaf poplar L2025 [5, 30] RNA-seq data for each *PdbHLH* were extracted, analyzed, normalized, and displayed in heat maps. The transcript abundance of *PdbHLH* genes was calculated as fragments per kilobase of exon model per million mapped reads (FPKM). The \log_2 (FPKM) from the RNA-seq data were subjected to hierarchical clustering with Cluster 3.0, and the results were graphically displayed using Java TreeView [35].

Validation of RNA-seq reliability in the leaves of JHP, QHP and L2025

To validate the reliability of RNA-seq results, the expression levels of 9 *PdbHLH* genes associated with anthocyanin biosynthesis in the leaves of JHP, QHP and L2025 were evaluated by qRT-PCR using an Applied Biosystems 7500 Real-Time PCR system (Applied Biosystems, Waltham, MA, USA). Gene-specific primers were designed according to the sequence of *PdbHLH* genes (Table S1), and the *ACTIN2* gene was used as a control gene [12]. The relative expression levels of 9 genes were analyzed using SPSS 17.0 with three biological replicates.

Protein interaction prediction

To investigate potential differential proteins that interact with PdbHLH proteins involved in anthocyanin biosynthesis, the putative PdbHLH protein sequences associated with anthocyanin biosynthesis were screened according to their phylogenetic analysis and expression pattern, and submitted to the online server STRING v10 (<https://string-db.org>). The PPI networks were constructed based on the active interaction sources with required confidence score > 0.4 in *Arabidopsis*, which included biological experiments, co-expression and databases. The PPI networks were visualized by Cytoscape software (v. 3.8.1). ClusterONE software was used to detect highly connected regions of the network, and the criteria are as follows: minimum density = 0.01, minimum size = 2 and edge weights = combined_score. The nodes and edges in the networks represent proteins and interactions, respectively. The interactions between bHLH proteins involved in anthocyanin biosynthesis



and potential differential proteins in JHP and QHP were screened by STRING.

Results

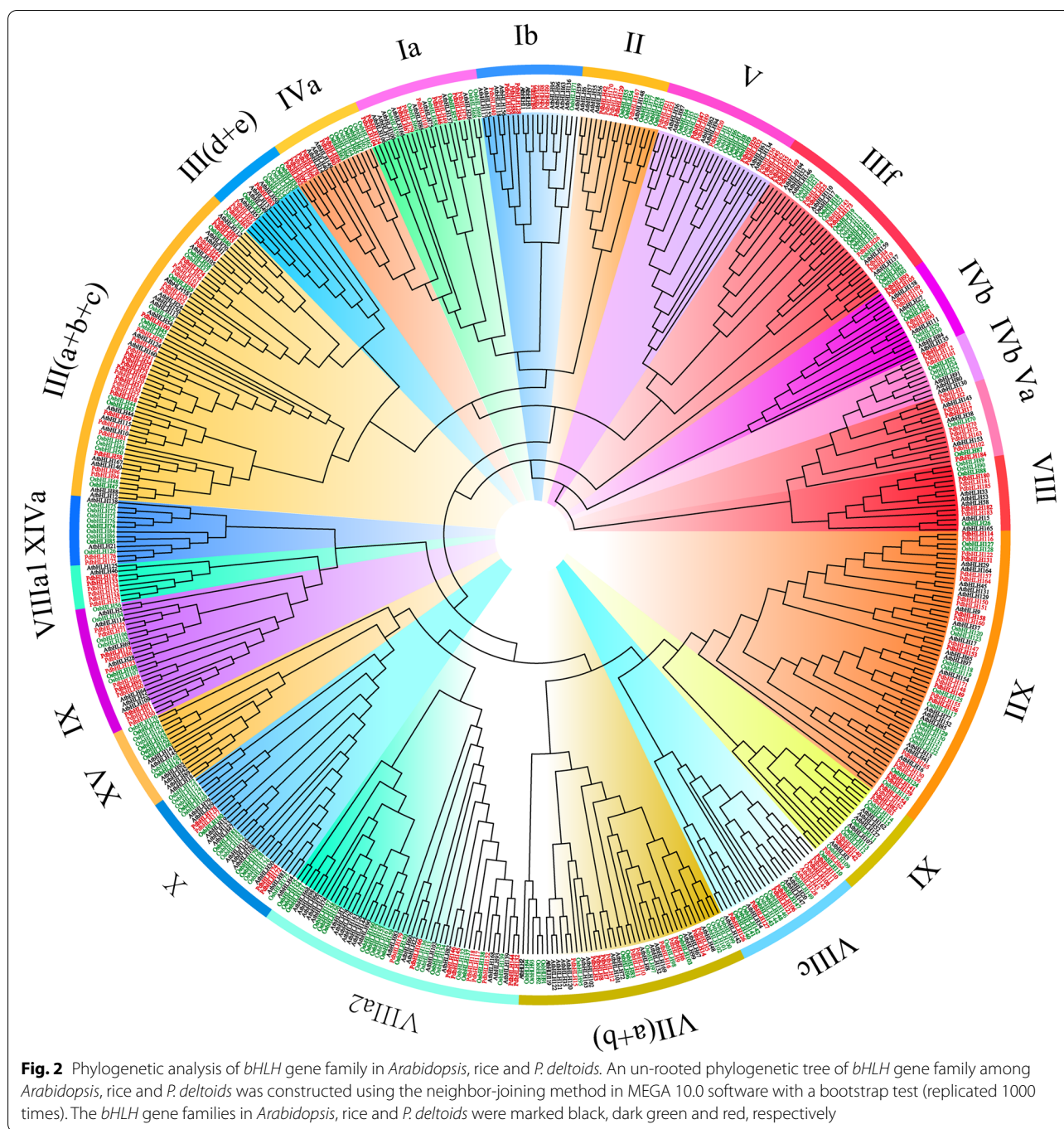
Identification and chromosome distribution of the *PdbHLH* gene family in *P. deltoids*

There are 185 *PdbHLH* family proteins identified in *P. deltoids* (Table S2). The *PdbHLH* proteins range in size from 129 (*PdbHLH72*) to 943 (*PdbHLH185*) amino acids, with an average length of approximately 373 amino acids. The molecular weight of the identified *PdbHLH* proteins ranged from 14.95 kDa (*PdbHLH72*) to 103.49 kDa (*PdbHLH185*), and the predicted isoelectric points of these proteins ranged from 4.62 (*PdbHLH58*) to 9.46 (*PdbHLH62*). In addition, the subcellular localization of the identified *PdbHLH* proteins was also predicted, and 181 of 185 (approximately 97.8%) *PdbHLH* proteins were localized in the nucleus (Table S2). According to the information of gene annotation, the predicted 185 *PdbHLH* genes were localized on poplar chromosomes. As shown in Fig. 1, the congregated region and the number of *PdbHLH*s is unevenly although the *PdbHLH* genes

distributed on all of the 19 chromosomes. Chromosome 2 possessed the largest number *PdbHLH* genes (19), while only four *PdbHLH* genes were present on chromosome 17. The percentage of *PdbHLH* genes per chromosome varied from 0.23% on chromosome 17 to 0.65% on chromosome 2 (Table S3).

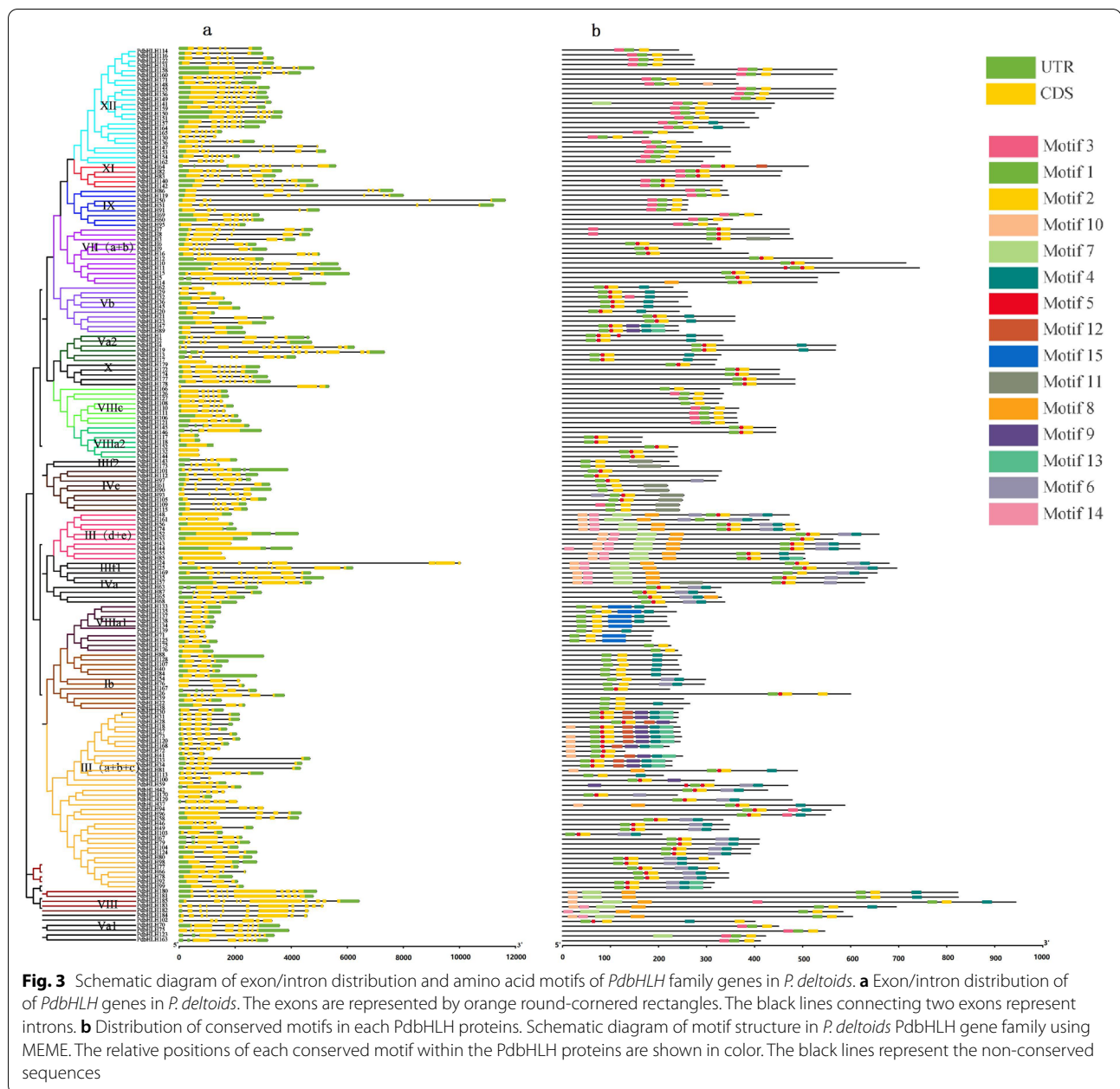
Phylogenetic analysis of the *PdbHLH* gene family among *Arabidopsis*, rice, and *P. deltoids*

An unrooted phylogenetic tree of the bHLH family genes in *P. deltoids*, *Arabidopsis*, and rice was constructed using the neighbor-joining method with the default parameters of MEGA 10.0 (Fig. 2). The *PdbHLH* family members of *P. deltoids* were divided into 15 groups (I to XV) according to the topology of the phylogenetic tree, as well as the classification and nomenclature of bHLH proteins in *Arabidopsis* and rice. The number of *PdbHLH* genes in different group was different. Group III, containing 37 *PdbHLH* genes, is the largest group in the 15 groups, which was further divided into 5 subgroups (subgroup a-e). Group XII and V contained 24 and 19 *PdbHLH* genes, respectively.



Group XV only contained the *bHLH* genes of *Arabidopsis* and rice, and didn't include the *bHLH* gene of *P. deltoideus*. Apart from the groups II, IV, V, and XII, the other groups contained much more *bHLH* genes in *P. deltoideus* than those in *Arabidopsis* and rice. In addition, some groups were further classified into two or more subgroups (Fig. 2). Subgroup VIII_{a1}, and III_{a-c} mainly contained the *bHLH* gene clusters in *P. deltoideus*,

indicating that the species-specific expansion of these genes occurred in *P. deltoideus* after the divergence of core eudicots. In addition, some subgroups just contained *AtbHLH* or *Os bHLH* genes without *PdbHLH* genes, suggesting that *bHLH* genes could have been either acquired and expanded in *Arabidopsis* or rice during evolution process or specially lost in *P. deltoideus*.



Gene structure and protein motif analysis of the *PdbHLH* gene family in *P. deltoidea*

To better analyze the structural diversity and motif composition of *PdbHLH* genes, the investigation of intron and exon distribution profile was conducted and visualized using the Gene Structure Display Server 2.0 (GSDS, Fig. 3). A total of 185 *PdbHLH* genes possessed exons varying from 1 to 12. Fourteen *PdbHLH* genes lacked introns and had only one exon, including *PdbHLH43*, *PdbHLH48*, *PdbHLH53*, *PdbHLH54*, *PdbHLH55*, *PdbHLH56*, *PdbHLH85*, *PdbHLH117*, *PdbHLH118*, *PdbHLH132*,

PdbHLH144, *PdbHLH152*, *PdbHLH176*, *PdbHLH179*. The majority (171 of 185) of the *PdbHLH* genes have 2 to 8 introns, and *PdbHLH19* contained 12 exons and 11 introns, which was the greatest number of exons in the total *PdbHLH* genes. As expected, gene structure analysis showed that most members in the same group had similar intron/exon compositions, including the numbers and length of exons. For example, all the members of subfamily XV have more than 5 exons, and all the members of group X have 3 exons. Thus, our results indicate that *PdbHLH* genes in the same group or subgroup had

similar gene structures, further verify the phylogenetic relationship of these *PdbHLH* genes (Fig. 2).

There were 15 putative conserved protein motifs identified in the *PdbHLH* proteins through MEME analysis (Motifs 1–15, Fig. 3, Fig. S1). Motif 1 the basic region and the first helix, and Motif 2 contained the second helix, both of which correspond to the bHLH domain. The bHLH domain has been identified in all *PdbHLH* proteins. Seven *PdbHLH* genes, including *PdbHLH38*, *PdbHLH39*, *PdbHLH175*, *PdbHLH176*, *PdbHLH166*, *PdbHLH127* and *PdbHLH108*, only contained the Motifs 1 and Motif 2. In each group, the components of the conserved motifs for most of the proteins were similar (Fig. 3b). For example, Motif 11, Motif 9, and Motif 15 were specifically distributed in the groups V, X, and XII, respectively. Motifs 6–8, 10–14 were commonly identified in all members of group IX.

Pivotal cis-elements in the promoter of *PdbHLH* genes in *P. deltoidea*

Cis-regulatory elements, which are usually restricted to 5' upstream areas of genes, are the binding sites of TFs, and are responsible for transcriptional regulation. Thus, the 2000 bp upstream of the transcription start site (TSS) of *PdbHLH* genes were used to explore gene regulation patterns with PlantCARE (Fig. S2). There were 23 functionally annotated cis-elements in the promoter of most *PdbHLH* genes, which were roughly divided into three categories: light-responsive elements (GT1-motif, I-box, AT1-motif, G-box, GATA-motif, GA-motif, Box II, Gap-box, L-box, CGTCA-motif, CAG-motif, AAAC-motif, and GTGGC-motif), stress-responsive elements [MYB-recognizing element (MRE), MYB-binding site (MBS), GC-motif, TC-rich repeats, LTR, and CCAAT-box]; and hormone-responsive elements (TCA-element, P-box, ABRE, TGACG-motif, TGA-element, GARE-motif, TATC-box, and SARE). The presence of MRE and MBS cis-elements in the *PdbHLH* gene promoter suggests that bHLH proteins might be transcriptionally regulated by MYB TFs in *P. deltoidea* to modulate the expression of downstream targets.

GO enrichment analysis of the *PdbHLH* genes in *P. deltoidea*

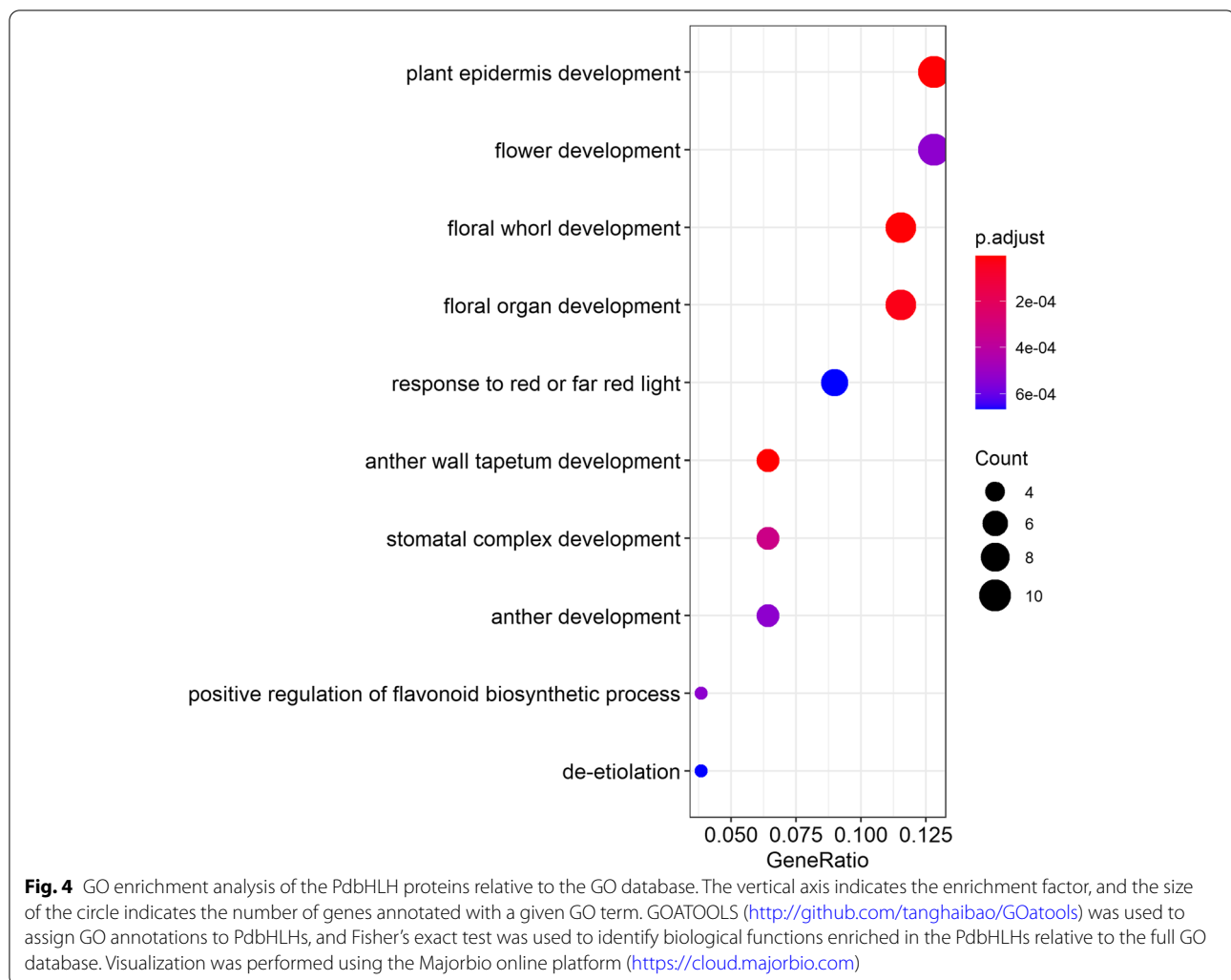
The bHLH proteins play a central role in a wide range of metabolic, physiological, and developmental processes in higher plants. To further investigate the biological functions of the *PdbHLH* genes in *P. deltoidea*, gene ontology (GO) annotation and enrichment analysis of the 185 *PdbHLH* genes were performed in present study (Fig. 4 and Table S4). Six molecular functions, two cellular components, and fifty-two biological processes in GO terms were enriched in the *PdbHLH* genes relative to the complete GO database. In the biological process category,

PdbHLH genes were mainly enriched in flower development ($n = 10$), plant epidermis development ($n = 10$), floral organ development (9), floral whorl development (9), plant organ morphogenesis (7), response to red or far red light (7) and so on. In the cellular component category, the genes were enriched in RNA polymerase II transcription (5) and transcription regulator complex (5). In the molecular function category, the genes were enriched in DNA-binding transcription activator activity (14), DNA-binding transcription activator activity and RNA polymerase II-specific (13), DNA-binding transcription factor activity and RNA polymerase II-specific (13), RNA polymerase II transcription regulatory region sequence-specific DNA binding (7), identical protein binding (7) and transcription factor binding (5). GO enrichment results suggested that *PdbHLH* TFs mainly involved in nucleic acid-binding TF activity, catalytic activity, developmental processes of cellular and multi-organism.

Gene duplication events of the *PdbHLH* gene family in *P. deltoidea*

Several gene duplication events, including WGD or segmental duplication, tandem duplication, dispersed gene duplication and others, promote the evolution of protein-coding gene families [42]. In the present study, the origins of duplicate genes for the *PdbHLH* gene family in *P. deltoidea* genome was detected with MCScanX package. Each member of *PdbHLH* gene family was assigned to one of five different categories: singleton, WGD/segment duplication, tandem, proximal and dispersed. Remarkably, 31 (16.7%), 57 (30.8%) and 57 (30.8%) of the *PdbHLH* family genes in *P. deltoidea* were duplicated and retained from singleton, proximal, and whole genome duplication (WGD)/segmental duplication, respectively (Fig. S3). Only five *PdbHLH* genes originated from tandem duplication. However, there was no *PdbHLH* genes originated from proximal in our results (Fig. S3).

To further explore the potential evolutionary mechanisms of *PdbHLH* gene family, the collinearity analysis of the *PdbHLH* gene family in *P. deltoidea* genome was performed using the all-vs.-all local BLASTP and MCS-can methods (Fig. 5). A total of 33 segmental duplication events with 52 *PdbHLH* genes were identified in the *P. deltoidea* genome, which accounted for 91.2% of WGD-type *PdbHLH* genes (Fig. 5 and Table S5). *PdbHLH* genes were located within synteny blocks on all chromosomes. Subsequently, non-synonymous to synonymous substitution ratio (Ka/Ks) was calculated using the pairwise model by maximum likelihood (PAML v8.0, Table S5). $Ka/Ks < 1$ indicates purifying selection, whereas $Ka/Ks > 1$ is the signature of positive selection (Hurst, 2002). In the present study, the Ka/Ks ratios of 32 *PdbHLH* gene pairs were less than one, implying that these genes are under

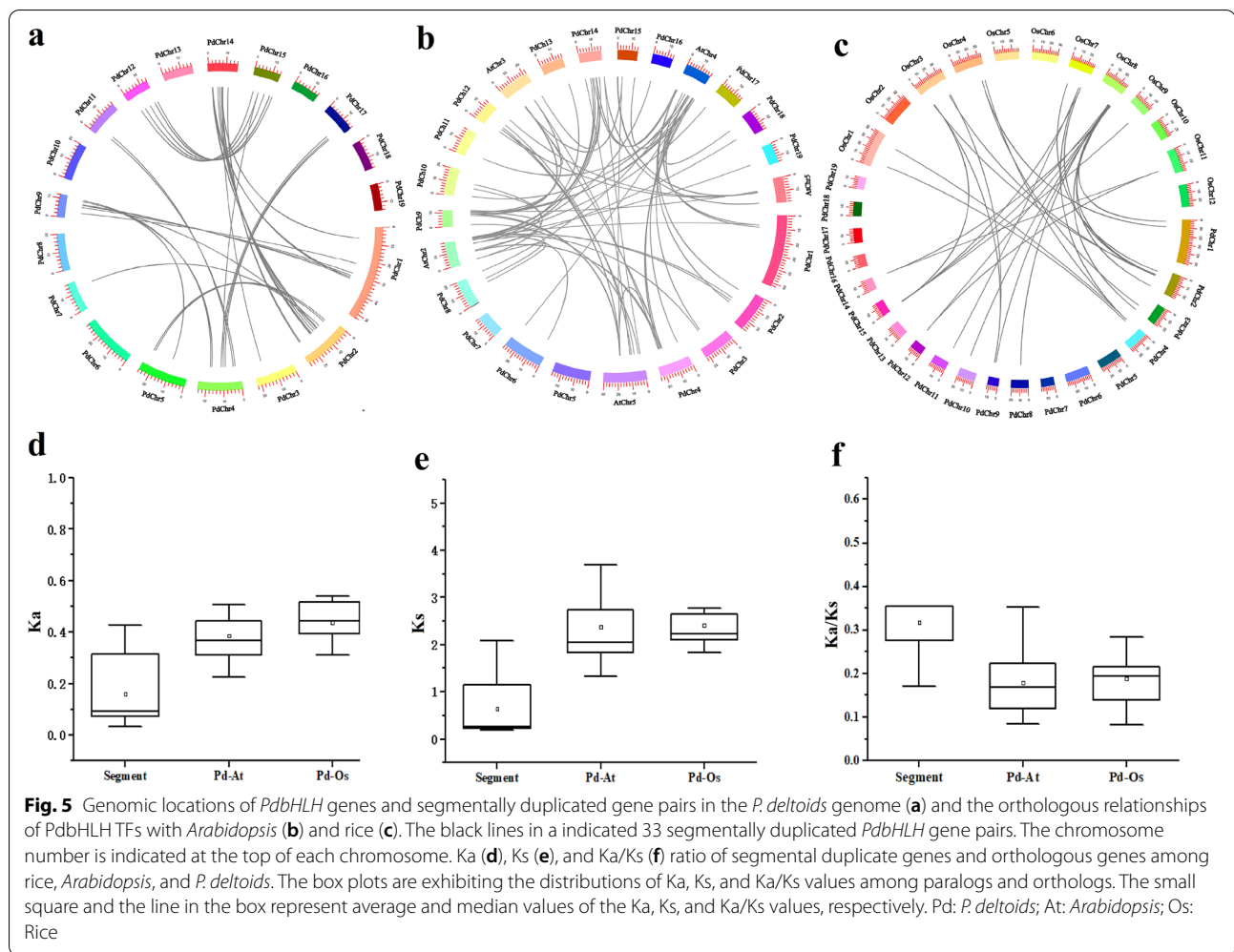


negative purifying selection, which maintained the functions of the *PdbHLH* gene family in *P. deltoids*. Moreover, K_s was usually used to estimate the evolutionary dates of genome or gene duplication events. The WGD/segmental duplicated events in *P. deltoids* occurred from 6.42 ($K_s=0.1927$) to 90.14 mya ($K_s=2.8244$).

The orthologous relationships of the *PdbHLH* family genes among *Arabidopsis*, Rice and *P. deltoids* were also investigated by collinearity analysis (Fig. 5, Table S6 and Table S7). There were 56 orthologous bHLH gene pairs between *P. deltoids* and *Arabidopsis* and 20 orthologs between *P. deltoids* and rice (Fig. 5). The number of orthologous events of PdbHLH-AtbHLH was much greater than that of PdbHLH-OsbHLH. The details of the collinear bHLH gene pairs were provided in Table S6 and S7.

Expression profile of the *PdbHLH* genes in colored-leaf poplar JHP and green-leaf poplar L2025

To evaluate the expression pattern of *PdbHLH* genes in colored-leaf poplar, the expression profiles of *PdbHLH* genes in the leaves of JHP and L2025 was evaluated using the previous RNA-seq data (Fig. 6, S4 and Table S8). The expression level of candidate *PdbHLH* genes in the leaves of JHP was more than 10 times that these in L2025, including 1 *PdbHLH* gene from Group I, 10 *PdbHLH* genes from Group III, 4 *PdbHLH* genes from Group IV, 4 *PdbHLH* genes from Group V, 5 *PdbHLH* genes from Group VII, 3 *PdbHLH* genes from Group IX, 2 *PdbHLH* genes from Group XI, 9 *PdbHLH* genes from Group XII, and 1 *PdbHLH* gene from Group XIII. These candidate *PdbHLH* genes are shown in Table S8. Among them, the expression levels of 12 *PdbHLH* genes (*PdbHLH141*, *PdbHLH95*, *PdbHLH140*, *PdbHLH57*, *PdbHLH136*, *PdbHLH91*, *PdbHLH94*, *PdbHLH156*, *PdbHLH1*, *PdbHLH173*, *PdbHLH148*, and *PdbHLH143*) were more than



100 times than that in L2025. In addition, the expression level of 23 *PdbHLH* genes in the leaves of L2025 was more than 10 times that these in JHP (Table S8). Among them, the expression level of 2 genes (*PdbHLH4* and *PdbHLH36*) in the leaves of L2025 was more than 40 times that these in JHP. These results suggested that these genes might be involved in regulating the anthocyanin biosynthesis.

Expression profile of the *PdbHLH* genes in colored-leaf poplar QHP and green-leaf poplar L2025

To further verify the expression pattern of *PdbHLH* genes, the expression level of *PdbHLH* genes in the leaves and buds of QHP and L2025 was evaluated using the previous RNA-seq data. A total of 102 *PdbHLH* genes were detected based on the RNA-seq analysis in leaves and buds of QHP and L2025 (Fig. 7, S5 and Table S9). Among them, the expression levels of 21 *PdbHLH* genes in the leaves and buds of QHP were significantly higher than that of L2025, and 2 genes (*PdbHLH143* and *PdbHLH9*)

were specifically expressed in the leaves and buds of QHP, indicating that these 2 genes might be involved in the positive regulation of anthocyanin biosynthesis. *PdPdbHLH123* gene was specifically expressed in the buds of poplar, and the expression level of this gene in the buds of QHP was more than 16 times than that of L2025. Moreover, the expression level of 20 *PdbHLH* genes in buds and leaves of L2025 were higher than those in QHP (Table S9). Among them, the expression level of 3 *PdbHLH* genes (*PdbHLH183*, *PdbHLH73*, and *PdbHLH85*) in leaves of L2025 were more than 5 times than those in QHP, and 2 genes (*PdbHLH67* and *PdbHLH74*) were specifically expressed in the leaves of L2025.

Combined with the comparative transcriptome data between the green leaf poplar and colored-leaf poplar (JHP and QHP), the candidate genes associated with leaf coloration were further screened (Table S10). The expression level of 12 *PdbHLH* genes (*PdbHLH12*, *PdbHLH131*, *PdbHLH143*, *PdbHLH156*, *PdbHLH160*, *PdbHLH173*, *PdbHLH20*, *PdbHLH57*, *PdbHLH7*, *PdbHLH82*,

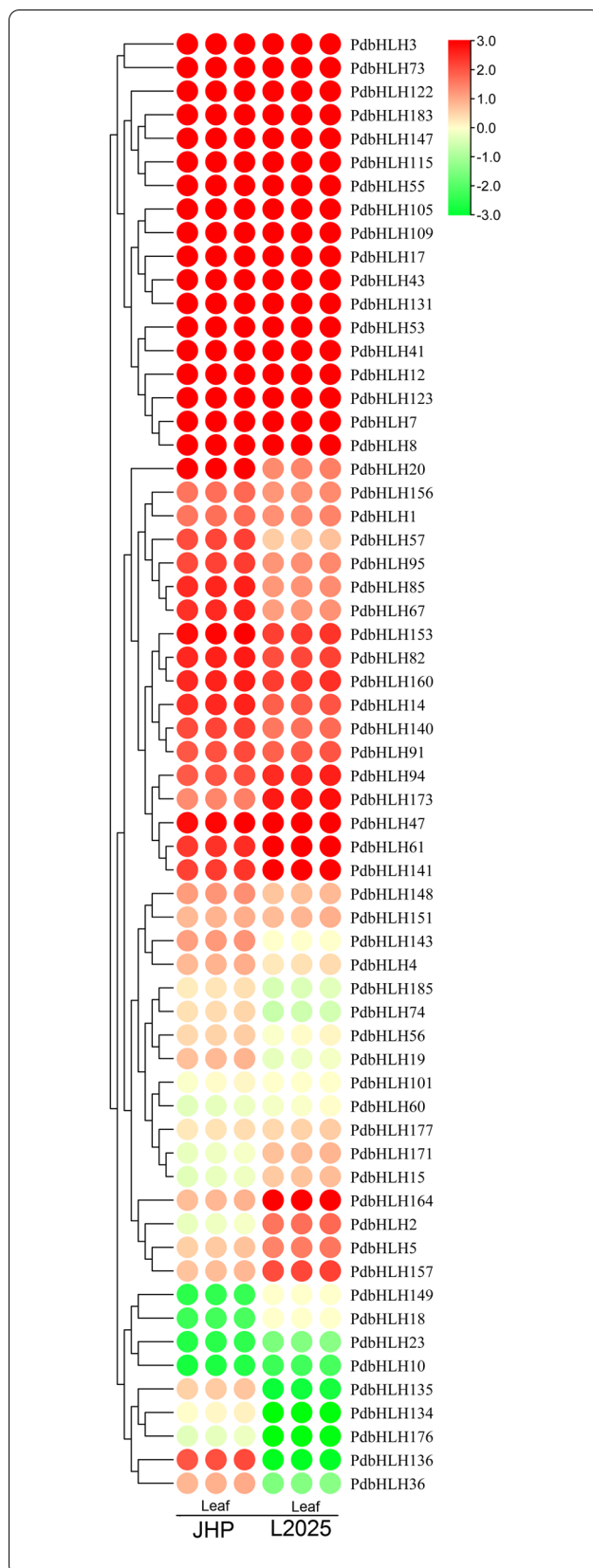


Fig. 6 Gene expression pattern of *PdbHLHs* in the leaves of JHP and L2025 by RNA-seq. The scale bars represent the log₂ transformations of the RPKM values

PdbHLH91, and *PdbHLH95*) in leaves of JHP was more than eight times than those in L2025, and their expression in leaves of QHP was different from those in L2025, suggesting that these genes may be involved in positively regulating the anthocyanin biosynthesis. In addition, the expression level 4 *PdbHLH* genes (*PdbHLH4*, *PdbHLH1*, *PdbHLH18*, and *PdbHLH164*) in leaves of L2025 was more than 10 times than those in JHP, and its expression in leaves of L2025 was slightly higher than those in QHP, suggesting that these genes may be involved in negatively regulating the anthocyanin biosynthesis. Therefore, these 16 *PdbHLH* genes might be involved in the anthocyanin biosynthesis in leaf coloration in poplar.

Validation of RNA-Seq-based gene expression

To validate the reliability of RNA-seq results, RT-PCR was performed on 9 genes associated with anthocyanin biosynthesis selected at random with high or low expression levels. Expression comparisons were performed in the leaves of L2025, QHP and JHP, and the expression trends in RT-PCR results were in agreement with the RNA-Seq data (Fig. S6).

Protein interaction prediction

Different bHLH proteins can form homodimers or heterodimers to bind DNA and regulate the transcription of downstream targets, which can interact with MYB and WD40 TFs to form a ternary complex (MBW) that regulates the expression of anthocyanin biosynthesis and structural genes [43, 44]. Thus, to further investigate the interaction networks of candidate *PdbHLH* TFs, protein interaction networks of sixteen *PdbHLH* proteins associated with anthocyanin biosynthesis were conducted. In our study, 16 candidate *PdbHLH* genes associated with leaf coloration were used to perform protein interaction networks, and 11 of them interacted with more than one protein (Fig. 8 and Table S11). In particular, two *PdbHLH* proteins (*PdbHLH131*, *PdbHLH156*) can interact with three kinds of PdMYBs, PdWDs and *PdbHLHs* (Fig. 8). *PdbHLH20*, *PdbHLH173* and *PdbHLH57* can interact with two kinds of PdMYBs, PdWDs and *PdbHLHs*, and the left ones can only interact with one kind of PdMYBs, PdWDs and *PdbHLHs*. However, most of them interact with PdMYBs. Interestingly, *PdbHLH57*, corresponding to PdTT8 associated with anthocyanin biosynthesis, can interact with many kinds of PdMYBs (such as PdMYB117, PdMYB112) and PdWDR6. It is

promoter of *CmDFR* gene [28]. VvMYC1, the first bHLH TF described in the grapevine, can control anthocyanin and proanthocyanidin biosynthesis through interacting with MYB5a, MYB5b, MYBA1/A2 and MYBPA1 [56]. In strawberry fruits, FvbHLH9 could form a HY5–bHLH9 complex with HY5 TF, which can positively regulate the anthocyanin biosynthesis [57]. PdTT8, one of bHLH transcription factors in poplar, physically interacted with PdMYB118 could form the PdMYB118–PdTT8 complex, which can regulate wound-induced the anthocyanin biosynthesis [26]. The *PdbHLH57* in our study corresponding to previous PdTT8 belonged to the III (f) subfamily, and *PdbHLH57* can interact with many kinds of PdMYBs, including PdMYB117 similar with PdMYB118, which is consistent with previous results. *PdbHLH57* can also interact with PdWDR6, which can form PdMYB–*PdbHLH57*–PdWDR6 complex to regulate the anthocyanin biosynthesis. However, the detailed functions are still needed to be explored. In addition, the expression level of *PdbHLH57* in the leaves of JHP is much higher than that in the leaves of L2025, and there is a slightly higher expression level than that in the leaves of L2025, which indicated that *PdbHLH57* might play important roles in the leaf coloration of JHP. Another two *PdbHLH* genes *PdbHLH143* and *PdbHLH173*, belonging to the III (f) subfamily, the expression level of which in colored-leaf of polar was higher than these in green-leaf of poplar, indicating that these two genes might also be involved in the anthocyanin biosynthesis in poplar. In addition, nine *PdbHLH* genes (*PdbHLH20*, *PdbHLH7*, *PdbHLH12*, *PdbHLH91*, *PdbHLH95*, *PdbHLH82*, *PdbHLH156*, *PdbHLH160*, and *PdbHLH13*) did not include into the III (f) subfamily, however, the expression level of which in colored-leaf of polar was also higher than these in green-leaf of poplar, suggesting that these genes may be used as the candidate genes to regulate the anthocyanin biosynthesis. *CpbHLH1* is a transcription factor from *Chimonanthus praecox*. Overexpression of *CpbHLH1* in *Arabidopsis* and tobacco resulted in a dramatic decrease in anthocyanin accumulation by repressing the expression of late biosynthesis genes in the flavonoid biosynthesis pathway [58]. In our study, four genes (*PdbHLH4*, *PdbHLH1*, *PdbHLH18*, and *PdbHLH164*) displayed higher expression in the leaves of L2025 than those in colored-leaf poplar (JHP and QHP), which indicated that these genes might be involved in reducing the anthocyanin accumulation.

Conclusions

In this study, 185 *PdbHLH* genes were identified in the *Populus deltoides* genome and were classified into 15 groups based on the sequence similarity and

phylogenetic relationships. Conserved domain, gene structure, and evolutionary relationships of *PdbHLH* genes were also established and analyzed. Investigation of cis-regulatory elements of *PdbHLH* genes indicated that many *PdbHLH* genes are involved in the regulation of anthocyanin biosynthesis. Comprehensive analysis revealed that 12 candidate genes, including 3 genes (*PdbHLH57*, *PdbHLH143*, and *PdbHLH173*) from the subgroup III(f) and 9 gene from other groups, were positively associated with anthocyanin biosynthesis. In addition, 4 genes (*PdbHLH4*, *PdbHLH1*, *PdbHLH18*, and *PdbHLH164*) may be involved in negatively regulating the anthocyanin biosynthesis. The above results could provide a basis for the functional characterization of *bHLH* genes, and also provide candidate genes for the future improvement of leaf colorization in *Populus deltoides*.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-08460-5>.

Additional file 1: Figure S1. Sequence logos of *PdbHLH* proteins.

Additional file 2: Figure S2. Cis-element analysis of *PdbHLH* genes from upstream 2000 bp sequence to the transcription start site.

Additional file 3: Figure S3. Proportion of genes originating from different replication events.

Additional file 4: Figure S4. Gene expression pattern of 185 *PdbHLH* genes in the leaves of JHP and L2025 by RNA-seq.

Additional file 5: Figure S5. Gene expression pattern of 185 *PdbHLH* genes in the buds and leaves of QHP and L2025 by RNA-seq.

Additional file 6: Figure S6. Relative expression levels of genes associate with anthocyanin biosynthesis in the leaves of L2025, JHP and QHP. Gene expression level was normalized with *ACTIN2*. All data represent the mean of three replicates with error bars indicating SD.

Additional file 7: Table S1. Specific primers used in relative quantitative real-time RT-PCR.

Additional file 8: Table S2. Information of *PdbHLH* genes identified in *Populus deltoides*.

Additional file 9: Table S3. The distribution ratio of *PdbHLH* genes on each chromosome in *P. deltoides*.

Additional file 10: Table S4. The gene ontology (GO) analysis of *PdbHLH* genes.

Additional file 11: Table S5. Segmentally duplicated *PdbHLH* gene pairs.

Additional file 12: Table S6. The orthologous relationships of the *bHLH* genes between *Arabidopsis* and *P. deltoides*.

Additional file 13: Table S7. The orthologous relationships of the *bHLH* genes between rice and *P. deltoides*.

Additional file 14: Table S8. The differentially expressed genes in JHP and L2025.

Additional file 15: Table S9. The differentially expressed genes in QHP and L2025.

Additional file 16: Table S10. The relative expression level of common genes in the leaves of L2025, JHP and QHP.

Additional file 17: Table S11. The protein-protein interaction used in this study.

Acknowledgements

The authors thank the lab members of the Bioinformatics Center of Guizhou University for their valuable advice on research design and paper discussion. Finally, we would like to thank the journal editors for their comments to improving our manuscript.

Authors' contributions

X.W. and W.Z. designed the experiment and wrote the manuscript. Y.L. and X.P. performed the experiment. X.S. and Z.W. analyzed the data. X.W. and W.Z. proofed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

This work was financially supported by the Jiangsu Agriculture Science and Technology Innovation Fund [CX (20)3039] and National Natural Science Foundation of China [31700585].

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files as follows:

- Zhang, F.; Zhao, J.; Wan, X.; Luo, X.; Li, W.; Sun, L.; Chen, Q. From green to red: large-scale transcriptome comparison of a bud sport in poplar (*Populus deltoides*). *Acta Physiol. Plant.* **2016**, *38*, 1-16, doi:<https://doi.org/10.1007/s11738-016-2259-7>.
- Tian, Y.; Rao S.; Li Q.; Xu M.; Wang A.; Zhang, H.; Chen, J. The coloring mechanism of a novel golden variety in *Populus deltoides* based on the RGB color mode. *Forestry Research* 2021, *1*: 5, doi:<https://doi.org/10.48130/FR-2021-0005>.

Declarations

Ethics approval and consent to participate

All plants materials involved in this research are used for scientific research, which are allowed to be used and provided free of charge in this study. These plant materials do not include any wild species at risk of extinction. No specific permits are required for sample collection in this study. We comply with relevant institutional, national, and international guidelines and legislation for plant study.

Consent to publication

Not applicable.

Competing interests

The authors declare no conflict of interest.

Author details

¹Key laboratory of Plant Resource Conservation and Germplasm Innovation in Mountainous Region (Ministry of Education), Guizhou University, Guiyang 550025, Guizhou, China. ²Institute of Horticulture, Guizhou Academy of Agricultural Sciences, Guiyang 550006, Guizhou, China. ³The Jiangsu Provincial Platform for Conservation and Utilization of Agricultural Germplasm, Institute of Botany, Jiangsu Province and Chinese Academy of Sciences (Nanjing Botanical Garden Mem. Sun Yat-Sen), Nanjing 210014, China.

Received: 17 December 2021 Accepted: 7 March 2022

Published online: 28 March 2022

References

- Yawei Q. Colored-leaf species resources and their landscaping application in Guang'an City. *J Landscape Res.* 2015;7:97 doi: [CNKI:SUN:JLDR.0.2015-06-027](https://doi.org/CNKI:SUN:JLDR.0.2015-06-027).
- Yanna H, Jinfang L. Plant landscape design language of small street parks in tropical coastal regions: a case study of Zhanjiang city. *J Landscape Res.* 2017;9:63 doi: [CNKI:SUN:JLDR.0.2017-02-017](https://doi.org/CNKI:SUN:JLDR.0.2017-02-017).
- Taylor. *Populus: arabidopsis for forestry. Do we need a model tree?* *Ann Bot-London.* 2002;90:681–9. <https://doi.org/10.1093/aob/mcf255>.
- Zhuang W, Liu T, Qu S, Cai B, Qin Y, Wang Z. Identification of candidate genes for leaf scorch in *Populus deltoides* by the whole genome resequencing analysis. *Sci Rep-UK.* 2018;8:1–11. <https://doi.org/10.1038/s41598-018-33739-7>.
- Zhang F, Zhao J, Wan X, Luo X, Li W, Sun L, et al. From green to red: large-scale transcriptome comparison of a bud sport in poplar (*Populus deltoides*). *Acta Physiol Plant.* 2016;38:1–16. <https://doi.org/10.1007/s11738-016-2259-7>.
- Liu H, Liu Z, Wu Y, Zheng L, Zhang G. Regulatory mechanisms of anthocyanin biosynthesis in apple and pear. *Int J Mol Sci.* 2021;22:8441. <https://doi.org/10.3390/ijms22168441>.
- Hu B, Lai B, Wang D, Li J, Chen L, Qin Y. Three LcABFs are involved in the regulation of chlorophyll degradation and anthocyanin biosynthesis during fruit ripening in *Litchi chinensis*. *Plant Cell Physiol.* 2019;60:448–61. <https://doi.org/10.1093/pcp/pcy219>.
- Montefiori M, Espley RV, Stevenson D, Cooney J, Allan AC. Identification and characterisation of F3GT1 and F3GGT1, two glycosyltransferases responsible for anthocyanin biosynthesis in red-fleshed kiwifruit (*Actinidia chinensis*). *Plant J.* 2011;65:106–18. <https://doi.org/10.1111/j.1365-3113.2010.04409.x>.
- Wu Q, Wu J, Li SS, Zhang HJ, Feng CY, Yin DD, et al. Transcriptome sequencing and metabolite analysis for revealing the blue flower formation in waterlily. *BMC Genomics.* 2016;17:897. <https://doi.org/10.1186/s12864-016-3226-9>.
- Muñoz C, Fanzone M, Lijavetzky D. Transcriptional regulation of the anthocyanin biosynthesis pathway in developing grapevine berries in cultivar 'Malbec' by putative R2R3 MYB negative regulators. *Sci Hortic-Asterdam.* 2019;257:108663. <https://doi.org/10.1016/j.scienta.2019.108663>.
- Cho JS, Nguyen VP, Jeon HW, Kim MH, Eom SH, Lim JY, et al. Overexpression of *PtMYB119*, a R2R3-MYB transcription factor from *Populus trichocarpa*, promotes anthocyanin production in hybrid poplar. *Tree Physiol.* 2016;36:1162–76. <https://doi.org/10.1093/treephys/tpw046>.
- Zhuang W, Shu X, Lu X, Wang T, Zhang F, Wang N, et al. Genome-wide analysis and expression profiles of PdeMYB transcription factors in colored-leaf poplar (*Populus deltoides*). *BMC Plant Biol.* 2021;21:1–16. <https://doi.org/10.1186/s12870-021-03212-1>.
- Wang H, Wang X, Song W, Bao Y, Jin Y, Jiang C, et al. PdMYB118, isolated from a red leaf mutant of *Populus deltoides*, is a new transcription factor regulating anthocyanin biosynthesis in poplar. *Plant Cell Rep.* 2019;38:927–36. <https://doi.org/10.1007/s00299-019-02413-1>.
- Zhao W, Liu Y, Li L, Meng H, Wu G. Genome-wide identification and characterization of bHLH transcription factors related to anthocyanin biosynthesis in red walnut (*Juglans regia* L.). *Front Genet.* 2021;12:632509. <https://doi.org/10.3389/fgene.2021.632509>.
- Goossens J, Mertens J, Goossens A. Role and functioning of bHLH transcription factors in jasmonate signalling. *J Exp Bot.* 2017;68:1333–47. <https://doi.org/10.1093/jxb/erw440>.
- Wang Y, Liu A. Genomic characterization and expression analysis of basic Helix-loop-Helix (bHLH) family genes in traditional Chinese herb *Dendrobium officinale*. *Plants.* 2020;9:1044. <https://doi.org/10.3390/plants9081044>.
- Zhang T, Lv W, Zhang H, Ma L, Li P, Ge L, et al. Genome-wide analysis of the basic Helix-loop-Helix (bHLH) transcription factor family in maize. *BMC Plant Biol.* 2018;18:1–14. <https://doi.org/10.1186/s12870-018-1441-z>.
- Niu X, Guan Y, Chen S, Li H. Genome-wide analysis of basic helix-loop-helix (bHLH) transcription factors in *Brachypodium distachyon*. *BMC Genomics.* 2017;18:1–20. <https://doi.org/10.1186/s12864-017-4044-4>.
- Arlotta C, Puglia GD, Genovese C, Toscano V, Karlova R, Beekwilder J, et al. MYB5-like and bHLH influence flavonoid composition in pomegranate. *Plant Sci.* 2020;298:110563. <https://doi.org/10.1016/j.plantsci.2020.110563>.
- Chandler VL, Radicella JP, Robbins TP, Chen J, Turks D. Two regulatory genes of the maize anthocyanin pathway are homologous: isolation of B utilizing R genomic sequences. *Plant Cell.* 1989;1:1175–83. <https://doi.org/10.1105/tpc.1.12.1175>.
- Bailey PC, Martin C, Toledo-Ortiz G, Quail PH, Huq E, Heim MA, et al. Update on the basic helix-loop-helix transcription factor gene family in *Arabidopsis thaliana*. *Plant Cell.* 2003;15:2497–502. <https://doi.org/10.1504/IJNET.2009.029397>.
- Matus JT, Poupin MJ, Cañón P, Bordeu E, Alcalde JA, Arce-Johnson P. Isolation of WDR and bHLH genes related to flavonoid synthesis in grapevine

- (*Vitis vinifera* L.). *Plant Mol Biol.* 2010;72:607–20. <https://doi.org/10.1007/s11103-010-9597-4>.
23. Bai Y, Pattanaik S, Patra B, Werkman JR, Xie CH, Yuan L. Flavonoid-related basic helix-loop-helix regulators, NtAn1a and NtAn1b, of tobacco have originated from two ancestors and are functionally active. *Planta.* 2011;234:363–75. <https://doi.org/10.1007/s00425-011-1407-y>.
 24. Espley RV, Hellens RP, Putterill J, Stevenson DE, Kutty-Amma S, Allan AC. Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant J.* 2007;49:414–27. <https://doi.org/10.1111/j.1365-3113X.2006.02964.x>.
 25. An JP, Li HH, Song LQ, Su L, Liu X, You CX, et al. The molecular cloning and functional characterization of MdMYC2, a bHLH transcription factor in apple. *Plant Physiol Biochem.* 2016;108:24–31. <https://doi.org/10.1016/j.plaphy.2016.06.032>.
 26. Wang H, Wang X, Yu C, Wang C, Jin Y, Zhang H. MYB transcription factor PdMYB118 directly interacts with bHLH transcription factor PdTT8 to regulate wound-induced anthocyanin biosynthesis in poplar. *BMC Plant Biol.* 2020;20:1–14. <https://doi.org/10.1186/s12870-020-02389-1>.
 27. Zhou X, Liao Y, Kim SU, Chen Z, Nie G, Cheng S, et al. Genome-wide identification and characterization of bHLH family genes from *Ginkgo biloba*. *Sci Rep-UK.* 2020;10:1–15. <https://doi.org/10.1038/s41598-020-69305-3>.
 28. Li M, Sun L, Gu H, Cheng D, Guo X, Chen R, et al. Genome-wide characterization and analysis of bHLH transcription factors related to anthocyanin biosynthesis in spine grapes (*Vitis davidii*). *Sci Rep UK.* 2021;11:1–13. <https://doi.org/10.1038/s41598-021-85754-w>.
 29. Shi Q, Li X, Du J, Li X. Anthocyanin synthesis and the expression patterns of bHLH transcription factor family during development of the Chinese jujube fruit (*Ziziphus jujuba* mill.). *Forests.* 2019;10:346. <https://doi.org/10.3390/f10040346>.
 30. Tian Y, Rao S, Li Q, Xu M, Wang A, Zhang H, et al. The coloring mechanism of a novel golden variety in *Populus deltoides* based on the RGB color mode. *Forestry Res.* 2021;1:5. <https://doi.org/10.48130/FR-2021-0005>.
 31. Nordberg H, Cantor M, Dushyko S, Hua S, Poliakov A, Shabalov I, et al. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Res.* 2014;42:D26–31. <https://doi.org/10.1093/nar/gkt1069>.
 32. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 2016;44:D279–85. <https://doi.org/10.1093/nar/gkv1344>.
 33. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol.* 2011;7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>.
 34. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* 2000;28:231–4. <https://doi.org/10.1093/nar/28.1.231>.
 35. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37:W202–8. <https://doi.org/10.1093/nar/gkp335>.
 36. Vooris RE. MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered.* 2002;93:77–8. <https://doi.org/10.1093/jhered/93.1.77>.
 37. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, et al. MScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012;40:e49–9. <https://doi.org/10.1093/nar/gkr1293>.
 38. Zhang Z, Li J, Zhao XQ, Wang J, Wong GKS, Yu J. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genom Proteom Bioinf.* 2006;4:259–63. [https://doi.org/10.1016/S1672-0229\(07\)60007-2](https://doi.org/10.1016/S1672-0229(07)60007-2).
 39. Lescot M. Plantcare, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 2002;30:325–7. <https://doi.org/10.1093/nar/30.1.325>.
 40. Sazegari S, Niazi A, Ahmadi FS. A study on the regulatory network with promoter analysis for *Arabidopsis* DREB-genes. *Bioinformatics.* 2015;11:101. <https://doi.org/10.6026/97320630011101>.
 41. Guindon S, Lethiec F, Duroux P, Gascuel O. PHYML online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* 2005;33:W557–9. <https://doi.org/10.1093/nar/gki352>.
 42. Li A, Mao L. Evolution of plant microRNA gene families. *Cell Res.* 2007;17:212–8. <https://doi.org/10.1038/sj.cr.7310113>.
 43. Liu Y, Hou H, Jiang X, Wang P, Dai X, Chen W, et al. A WD40 repeat protein from *Camellia sinensis* regulates anthocyanin and proanthocyanidin accumulation through the formation of MYB–bHLH–WD40 ternary complexes. *Int J Mol Sci.* 2018;19:1686. <https://doi.org/10.3390/ijms19061686>.
 44. Shin DH, Choi MG, Kang CS, Park CS, Choi SB, Park YI. A wheat R2R3-MYB protein PURPLE PLANT1 (TaPL1) functions as a positive regulator of anthocyanin biosynthesis. *Biochem Bioph Res Co.* 2016;469:686–91. <https://doi.org/10.1016/j.bbrc.2015.12.001>.
 45. Wang L, Xiang L, Hong J, Xie Z, Li B. Genome-wide analysis of bHLH transcription factor family reveals their involvement in biotic and abiotic stress responses in wheat (*Triticum aestivum* L.). *3. Biotech.* 2019;9:1–12. <https://doi.org/10.1007/s13205-019-1742-4>.
 46. Song XM, Huang ZN, Duan WK, Ren J, Liu TK, Li Y, et al. Genome-wide analysis of the bHLH transcription factor family in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). *Mol Gen Genomics.* 2014;289:77–91. <https://doi.org/10.1007/s00438-013-0791-3>.
 47. Freeling M. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu Rev Plant Biol.* 2009;60:433–53. <https://doi.org/10.1146/annurev.arplant.043008.092122>.
 48. Wang N, Xiang Y, Fang L, Wang Y, Xin H, Li S. Patterns of gene duplication and their contribution to expansion of gene families in grapevine. *Plant Mol Biol Report.* 2013;31:852–61. <https://doi.org/10.1007/s11105-013-0556-5>.
 49. Qiao X, Yin H, Li L, Wang R, Wu J, Wu J, et al. Different modes of gene duplication show divergent evolutionary patterns and contribute differently to the expansion of gene families involved in important fruit traits in pear (*Pyrus bretschneideri*). *Front Plant Sci.* 2018;9:161. <https://doi.org/10.3389/fpls.2018.00161>.
 50. Guo L, Chen Y, Ye N, Dai X, Yang W, Yin T. Differential retention and expansion of the ancestral genes associated with the paleopolyploidies in modern rosid plants, as revealed by analysis of the extensins supergene family. *BMC Genomics.* 2014;15:1–13. <https://doi.org/10.1186/1471-2164-15-612>.
 51. Du H, Liang Z, Zhao S, Nan MG, Tran LSP, Lu K, et al. The evolutionary history of R2R3-MYB proteins across 50 eukaryotes: new insights into subfamily classification and expansion. *Sci Rep-UK.* 2015;5:1–16. <https://doi.org/10.1038/srep11037>.
 52. Hao F, Yang G, Zhou H, Yao J, Liu D, Zhao P, et al. Genome-wide identification and transcriptional expression profiles of transcription factor WRKY in common walnut (*Juglans regia* L.). *Genes.* 2021;12:1444. <https://doi.org/10.3390/genes12091444>.
 53. Ma YF, Huang CP, Lu FR, Li JX, Han XM, Adeola AC, et al. OrthReg: a tool to predict cis-regulatory elements based on cross-species orthologous sequence conservation. *Zool Res.* 2020;41:471. <https://doi.org/10.2472/zj.issn.2095-8137.2020.099>.
 54. Zhang C, Feng R, Ma R, Shen Z, Cai Z, Song Z, et al. Genome-wide analysis of basic helix-loop-helix superfamily members in peach. *PLoS One.* 2018;13:e0195974. <https://doi.org/10.1371/journal.pone.0195974>.
 55. Li Y, Shan X, Gao R, Yang S, Wang S, Gao X, et al. Two IIIf clade-bHLHs from *Freesia hybrida* play divergent roles in flavonoid biosynthesis and trichome formation when ectopically expressed in *Arabidopsis*. *Sci Rep UK.* 2016;6:1–13. <https://doi.org/10.1038/srep30514>.
 56. Hichri I, Heppel SC, Pillet J, Léon C, Czemplak S, Delrot S, et al. The basic helix-loop-helix transcription factor MYC1 is involved in the regulation of the flavonoid biosynthesis pathway in grapevine. *Mol Plant.* 2010;3:509–23. <https://doi.org/10.1093/mp/ssp118>.
 57. Xiang LL, Liu XF, Li X, Yin XRG, D, Li F, Chen, K.S. A novel bHLH transcription factor involved in regulating anthocyanin biosynthesis in chrysanthemums (*Chrysanthemum morifolium* Ramat.). *PLoS One.* 2015;10:e0143892. <https://doi.org/10.1371/journal.pone.0143892>.
 58. Zhao R, Song X, Yang N, Chen L, Xiang L, Liu XQ, et al. Expression of the subgroup IIIf bHLH transcription factor CpbHLH1 from *Chimonanthus praecox* (L.) in transgenic model plants inhibits anthocyanin accumulation. *Plant Cell Rep.* 2020;39:891–907. <https://doi.org/10.1007/s00299-020-02537-9>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.