Zheng J *et al.* (2021)
Notulae Botanicae Horti Agrobotanici Cluj-Napoca
Volume 49, Issue 2, Article number 12086
DOI:10.15835/nbha49212086
Research Article

Genome-wide identification of WD40 superfamily genes and prediction of WD40 genes involved in flavonoid biosynthesis in *Ginkgo biloba*

Jiarui ZHENG, Yongling LIAO*, Feng XU*, Xian ZHOU, Jiabao YE, Mingyue FU, Xiaomeng LIU, Zhengyan CAO, Weiwei ZHANG

*Yangtze University, College of Horticulture and Gardening, Jingzhou 434025, Hubei, China; Zhengjiarui1222@163.com; liaoyongling@yeah.net (*corresponding author); xufeng198@126.com (*corresponding author); chenxitianmu@126.com; yejiabao13@163.com; fumingyue1214@163.com; liuxm925@163.com; caozhengyan1998@163.com; wwzhangchn@163.com*

Abstract

The WD40 transcription factor family is a superfamily found in eukaryotes and implicated in regulating growth and development. In this study, 167 *WD40* family genes are identified in the *Ginkgo biloba* genome. They are divided into 5 clusters and 16 subfamilies based on the difference analysis of a phylogenetic tree and domain structures. The distribution of *WD40* genes in chromosomes, gene structures, and motifs is analyzed. Promoter analysis shows that five *Gb WD40* gene promoters contain the MYB binding site participating in the regulation of flavonoid metabolism, suggesting that these five genes may participate in the regulation of flavonoid synthesis in *G. biloba*. The correlation analysis is carried out based on FPKM value of *WD40* genes and flavonoid content in 8 tissues of *G. biloba*. Six *Gb WD40* genes that may participate in flavonoid metabolism are screened. The biological functions of the *WD40* family genes in *G. biloba* are systematically analyzed, providing a foundation for further elucidating their regulatory mechanisms. A number of *WD40* candidate genes involved in the biosynthesis and metabolism of *G. biloba* also predicted. This study presents an important basis and direction for conducting further research on the regulatory network of flavonoid synthesis and metabolism.

Keywords: flavonoid; genome; *Ginkgo biloba*; superfamily genes; WD40

Introduction

Ginkgo biloba L. is a typical gymnosperm left over from the Quaternary glaciation movement. G. biloba, which is also named white fruit tree, Gongsun tree, and "living fossil," is a deciduous tree that has experienced 270 million years of history, but its shape and structure have slightly changed (Gong *et al.*, 2008; Liu *et al.*, 2017). *G. biloba*, native to China, is an ancient dioecious plant with important medicinal properties (Lin *et al.*, 2011). *G. biloba* is rich in natural active ingredients, such as flavonoids and terpene trilactones (TTLs). Among them, flavonoids are commonly used as an herbal dietary supplement for the treatment of many diseases, and they can improve the psychological ability of patients with Alzheimer's disease (Albert *et al.*, 2018; Ni *et al.*, 2018). Flavonoids are a large class of secondary polyphenol metabolites in plants. They can be divided into six

Received: 19 Sep 2020. Received in revised form: 10 Jun 2021. Accepted: 10 Jun 2021. Published online: 18 Jun 2021.
From Volume 49, Issue 1, 2021, Notulae Botanicae Horti Agrobotanici Cluj-Napoca journal uses article numbers in place of the traditional method of continuous pagination through the volume. The journal will continue to appear quarterly, as before, with four annual numbers.

categories based on molecular structures: chalcones, flavones, flavonols, flavandiols, anthocyanins, and proanthocyanidins (PAs; also called condensed tannins). As antibacterial agents, flavonoids play an important role in the interaction and defense reaction between plants and microorganisms and have potential beneficial effects on human health (Winkel-Shirley *et al.*, 2001). Flavonoids in *G. biloba* extract are its active components, which play an important role in pharmacology (Ude *et al.*, 2013).

WD40 protein, also known as a WD40 repeat, widely exists in eukaryotes (Migliori *et al.*, 2012). It is characterized by a peptide motif with 40-60 amino acids, which are usually defined by the GH dipeptide (Gly-His) at the C terminal and the WD dipeptide (Trp-Asp) at the N terminal (van Nocker *et al.*, 2003). WD40 is widely involved in many functional processes, including signal transduction, cell division, vesicle formation, secondary metabolite synthesis, transcription regulation, cell cycle regulation, and chromatin histone modification (van Nocker *et al.*, 2003; Suganuma *et al.*, 2008; Xu *et al.*, 2011). WD40 protein acts as an adaptor in many protein complexes or protein-DNA complexes (Pesch *et al.*, 2015). It has been identified in plants, humans, and prokaryotes (Hu *et al.*, 2011; Zou *et al.*, 2016; Feng *et al.*, 2019). A total of 743 WD40 proteins are identified in the wheat genome, which is divided into 5 clusters and 11 subfamilies (Hu *et al.*, 2018). A total of 220 WD40 protein family genes are identified in the peach genome (Feng *et al.*, 2019). In *Oryza sativa*, a monocotyledonous model plant, 200 WD40 family members are divided into 5 clusters and 11 subfamilies based on their domain composition (Ouyang *et al.*, 2012). The WD40 protein family has been studied in many plants, but it is rarely reported in *G. biloba*. Many findings have shown that the WD40 protein is involved in flavonoid regulation (Xie *et al.*, 2016; Chen *et al.*, 2019). Some studies have demonstrated that flavonoid synthesis regulators interact with R2R3-MYBs or bHLHs, which either increase or decrease the flavonoid contents (Li *et al.*, 2014). The interaction between WD40, MYB, and bHLH transcription factor proteins has been widely studied. Flavonoid biosynthesis may be a good model for their interaction (Nakatsuka *et al.*, 2008).

Ye *et al.* (2019) prepared three generations of full-length transcriptome sequencing of eight different tissues of *G. biloba*, revealed 12 structural genes and transcription factor modules involved in flavonoid biosynthesis, and identified 7 hub genes participating in flavonoid biosynthesis and metabolism (Ye *et al.*, 2019). Wu *et al.* showed the key genes involved in flavonoid synthesis, transportation, and regulation through the transcriptome sequencing of *G. biloba* with different flavonoid contents (Wu *et al.*, 2018). However, few studies have been conducted on the WD40 gene involved in the flavonoid biosynthesis of *G. biloba*. In this study, the WD40 family genes in *G. biloba* is comprehensively and systematically identified and analyzed on the basis of the genomic data of *G. biloba* (http://gigadb.org/dataset/100613) (Guan *et al.*, 2019) and the transcriptome sequencing data of three generations of *G. biloba* obtained previously (Ye *et al.*, 2019). The *GbWD40* gene related to flavonoid biosynthesis is screened through the correlation analysis with flavonoid contents in eight different tissues of *G. biloba*. This work is performed to comprehensively identify the WD40 family gene in *Ginkgo* genome by (1) determining the exact number of IDs of the WD40 family gene in the *Ginkgo* genome; (2) phylogenetically analyzing *GbWD40* identified on the basis of the WD40 family gene in *Arabidopsis thaliana* and subclassifying the WD40 family members of *G. biloba* by combining with the domain structure; (3) analyzing the structure of the genes and proteins of each *GbWD40* member in *G. biloba*; (4) evaluating the chromosomal distribution; (5) carrying out a correlation analysis based on FPKM data and flavonoid content in eight tissues of *G. Biloba* and screening out *GbWD40* significantly correlated with flavonoid contents. Our study provides a basis for revealing the synthesis metabolism and content variation in flavonoids in *G. biloba*.

Materials and Methods

Plant materials and determination of flavonoids

The plant materials (31-years-old trees) of *G. biloba* was collected from *Ginkgo* Germplasm Repository of Yangtze University (N30.35, E112.14), China. The 8 independent sampled tissues included root (R), stem

(S), immature leaf (IL), mature leaf (ML), microstrobilus (M), ovulate strobilus (OS), immature fruit (IF), and mature fruit (MF). The tissue samples were immediately frozen in liquid nitrogen, and stored at ultra-low temperatures in a refrigerator for further analysis. The flavonoid contents in 8 tissues were determined according to the method of Ye *et al.* (2019). The flavonoid content determination was three biological replicates, with six technical replicates for each biological replicate.

Identification of WD40 superfamily genes in G. biloba

To order to identify WD40 proteins in *G. biloba*, the whole protein sequence of *GbWD40* was downloaded from the GIGADB (http://gigadb.org/dataset/100613, 2019), and the whole protein sequence of *A. thaliana* was downloaded from TAIR (https://www.arabidopsis.org/, TAIR 10). The hidden Markov model (HMM) profile of the domain (PF00400) was downloaded from Pfam (http://pfam.xfam.org/family/PF00400). The hmmsearch program (HMMER 3.0 package, https ://hmmer.org/) was employed against the whole protein sequence by using the HMM of the WD40 domain (PF00400) as the query file with E value ≤ 10^{-5}. To ensure the presence of the WD domain for each protein, the protein sequences were uploaded to the Batch CD search (https://www.ncbi.nlm.nih.gov/). After removing redundant sequences, all candidate proteins were evaluated via Pfam (https://pfam.xfam.org) (El-Gebali *et al.*, 2019) and Smart (https://smart.embl-heide lberg.de/) (Ponting *et al.*, 1999). Protein sequence, the coding sequence (CDS) and genomic sequence were extracted from ginkgo Genome Database by using TBtools software fasta extract tool (Chen *et al.*, 2020). The basic physical and chemical parameters (primary structure), including the number of amino acids, molecular weight (Mw), theoretical pI, aliphatic index, and grand average of hydropathicity (GRAVY), for each protein were collected from the ProtParam (*https://web.expasy.org/protparam/*) website.

Chromosome locations of GbWD40s

The General Feature Format (GFF) was downloaded from the GIGADB (http://gigadb.org/dataset/100613, 2019), and the physical location of all *GbWD40* genes on the chromosome was drawn using the TBtools tool (Chen *et al.*, 2020).

Phylogenetic analysis and classification of WD40 proteins in G. biloba

WD40 gene sequences of *A. thaliana* was downloaded from the TAIR database https://www.arabidopsis.org/index.jsp) and extracted using TBtools tool. Combining these 100 *WD40* superfamily proteins in *A. thaliana* with *WD40* proteins in *G. biloba*, one phylogenetic tree was constructed using MEGAX software (https://www.megasoftware.net). The all proteins sequence was aligned by using the MUSCLE algorithm and the tree was constructed with the neighbor-joining (NJ) method with a bootstrap test (1000 replicates). The phylogenetic tree was classified according to its topology and evolutionary relationship.

The gene structure prediction and motif distribution of the WD40 superfamily members in G. biloba

The exon–intron structures of *GbWD40*s was predicted by GSDS (http://gsds.cbi.pku.edu.cn/index.php) (Hu *et al.*, 2015). The conserved motifs of the *GbWD40*s proteins were searched in MEME 5.1.0 (http://meme-suite.org/tools/meme) (Bailey *et al.*, 2009) with a maximum of 20 motifs and under default parameters and drawn by TBtools (Chen *et al.*, 2020).

Promoter analysis of GbWD40s genes

The upstream 2000 bp genomic DNA sequences of the *WD40* genes were downloaded and submitted to PlantCARE (Lescot *et al.*, 2002) to predict putative cis-elements.

Correlation analysis of flavonoids

The Fragments Per Kilobase of transcript per Million fragments mapped (FPKM) of *WD40* was downloaded from the full-length transcriptome of *G. biloba* (Ye *et al.*, 2019), and the correlation coefficient was determined with OmicShare tools, which is a free online platform (htpp://www.omicshare.com/tools).

Gene ontology (GO) annotation

According to the annotations of the full-length transcriptome of *G. biloba* (Ye *et al.*, 2019), the GO annotations of WD40 family genes were classified and counted.

Results

Identification of WD40 genes in G. biloba and chromosomal distribution of GbWD40 genes

A total of 167 *GbWD40* family genes are identified and obtained on the basis of the HMM results and verification of conservative motifs. For convenience, the positions of 167 WD40 genes on the *Ginkgo* chromosome are sorted, and these genes are named *GbWD40-001* to *GbWD40-167* (Figure 1, Table S1) in combination with *G. biloba*. The length of WD40 protein sequence is in the range of 99-3230 aa, and the average length is about 620 aa. The relative molecular mass of the WD40 family protein is in the range of 11,116.5 (*GbWD40-015*) Da to 361,146.48 (*GbWD40-052*) Da, and the average molecular mass is 68,670.55 Da. The theoretical pI is in the range of 4.51 to 9.62. A total of 68 WD40 proteins with an average pI of more than 6.93. The instability index is in the range of 18.86 to 59.78, and the average is 41.11. A total of 96 WD40 protein sequences are considered to be unstable, and 71 are considered to be stable (Table S1). The grand average of hydropathicity (GRAVY) is in the range of −0.822 to 0.331, and the average value is −0.28. The length of 167 WD40 genome sequences is from 425 bp to 448,032 bp, with an average of 96,329 bp.

Phylogenetic analysis and subfamily classification of WD40 proteins in G. biloba

A phylogenetic analysis is conducted using the NJ method to identify evolutionary relationships among the *GbWD40* protein members. The WD40 protein sequence of *G. biloba* and the WD40 protein sequence of *A. thaliana* are used to construct an evolutionary tree in MEGAX based on the protein sequence similarities and their subsequent phylogenetic tree. *GbWD40* genes are divided into five clusters (Cluster I to V) based on the evolutionary relationship between the WD40 protein of *G. biloba* and the WD40 protein of *A. thaliana*, and these five clusters include 40, 54, 20, 61, and 92 *GbWD40* members, respectively (Figure 2, Table S1). According to their domain structure, the *GbWD40* protein is divided into 16 subfamilies. Among them, 116 proteins only contain the WD40 domain, which is classified as A subcategory. The 51 other *GbWD40* genes include other domains excluding the WD40 domain, and they are classified as subfamilies B to P (Figure 3; Figure S3). The domain structures of the genes in other families are similar except the subfamily P. This result indicates that WD40 proteins with similar domains are clustered together.

The motif distribution and gene structure of WD40 superfamily members in G. biloba

A motif analysis is carried out for each protein sequence by using a SMART online tool to confirm that *GbWD40*s are WD40 superfamily genes (Figure S3). The results show that motifs 1, 2, 3, and 4 contain 21, 15, 11, and 15 amino acids, and they are highly conserved in the WD40 protein sequence of *G. biloba*. Furthermore, 166, 164, 164, and 153 *GbWD40* proteins contain the corresponding motifs (Figure 4). The result also demonstrates that motif 1 is the most conserved in 20 motifs, followed by motifs 2 and 3. The sequences in motifs 1-20 correspond to logos 1–20 (Figure S2). The results suggest that the proteins in the same subfamily have similar motifs, such as subfamilies B and E (Figure 4).

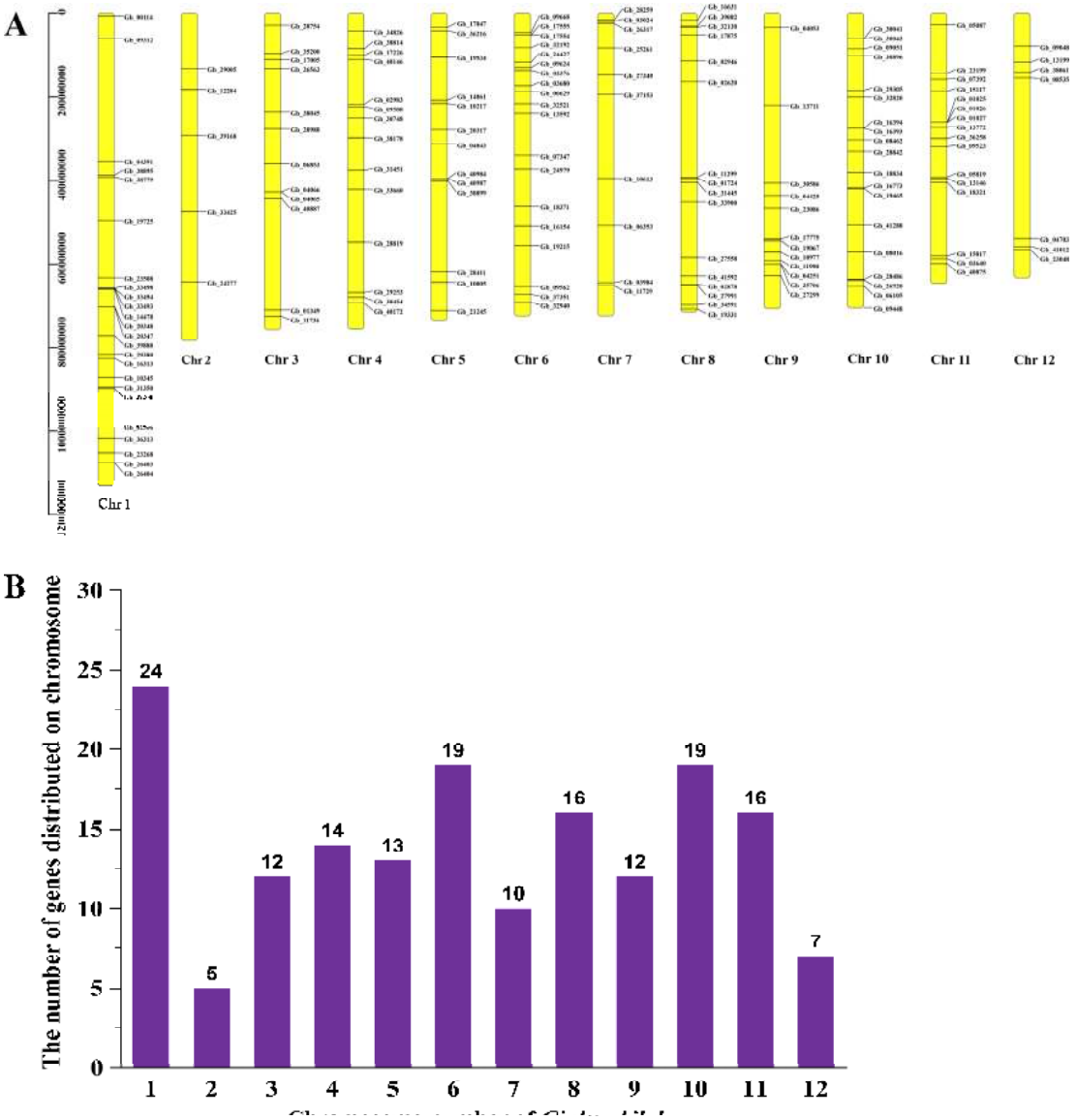

Figure 1. Uneven distributions of *GbWD40* genes on 12 chromosomes in *G. biloba*
(a) *GbWD40* gene distribution map on 12 *Ginkgo* chromosomes; (b) Statistics of *GbWD40*s on 12 chromosomes. Y-axis showed the number of *GbWD40*s on each chromosome; X-axis showed the chromosome number

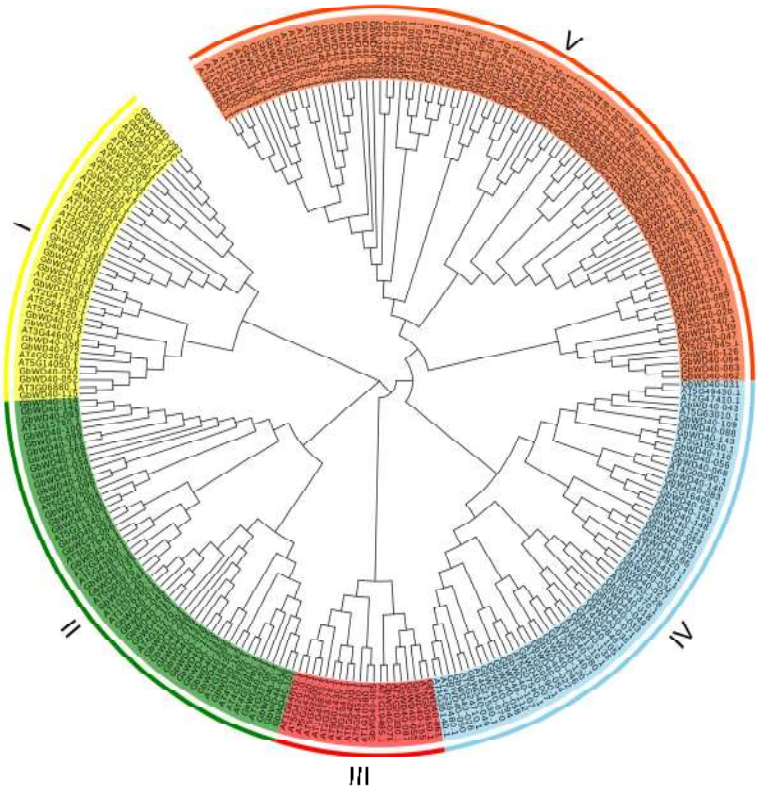

Figure 2. Phylogenetic relationship of WD40 superfamily proteins between *G. biloba* and *A. thaliana*
The tree was constructed by MEGAX. The 167 genes proteins are named GbWD40-001 to GbWD40-167 according to the position on the chromosomes. The Roman numerals (I-V) outside the large circle indicate the name of each cluster in the WD40 superfamily

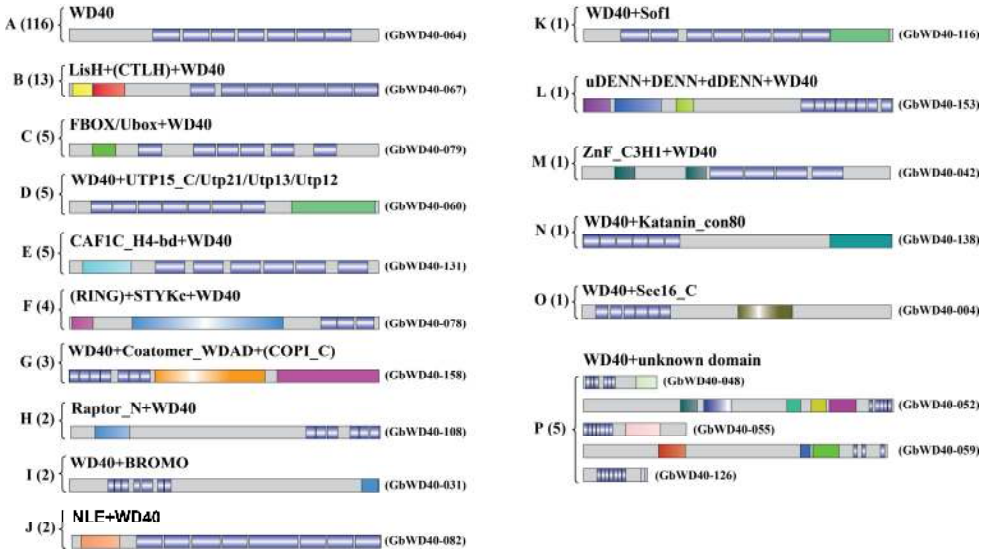

Figure 3. Diversified protein domain structure of *GbWD40*s
The WD40 repeats were shown in blue, while the other domains were marked in different colors. Texts in black above each cartoon indicated the types of domains. For each subfamily (from A to P), numbers in brackets gave the protein numbers for that subfamily, the structure diagram corresponds to the gene name behind. The figure shows a schematic diagram of the structure of *GbWD40*s. The structural boxes in different colors are the actual positions in the protein sequence

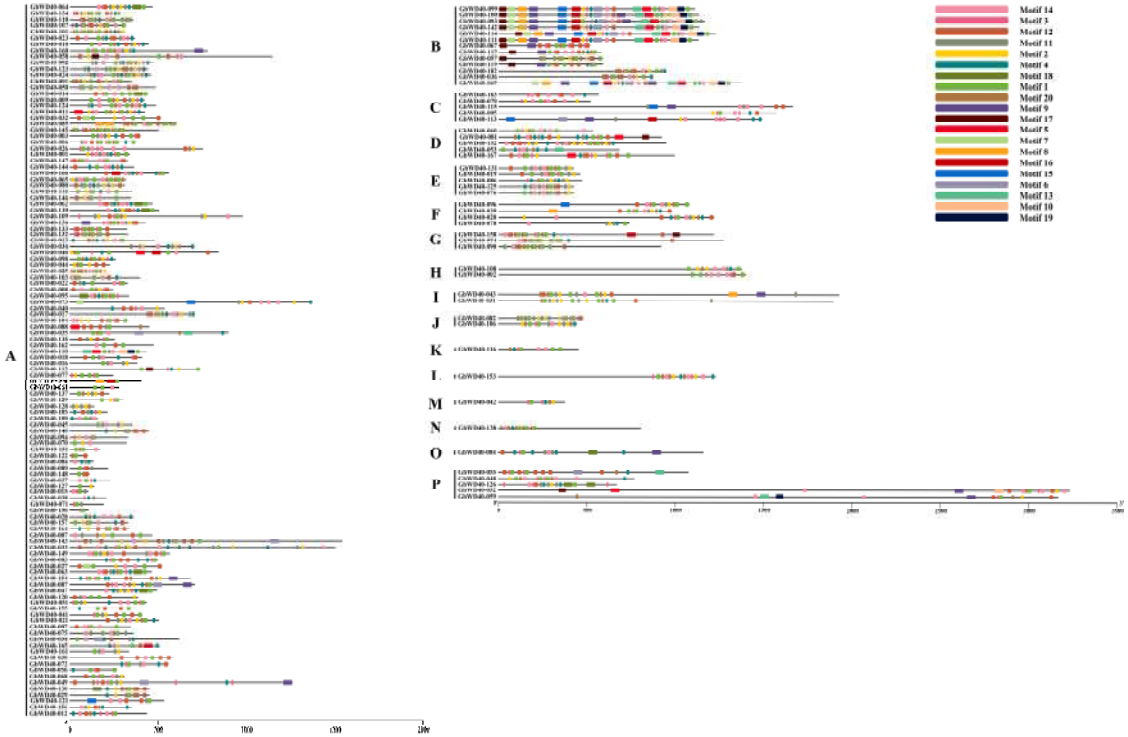

Figure 4. Distribution of conserved motifs in *GbWD40* superfamily proteins
Boxes of different colors represent 20 putative motifs. The Roman numerals (A to P) on the left indicate different subfamilies

The intron–exon schematic structure of *GbWD40* genes is constructed on the basis of the subfamily classification of 167 *GbWD40*s to study the distribution of introns and exons in each *GbWD40* gene (Figure 5). The analysis result shows that the difference between the numbers of introns and exons of *GbWD40* family genes is large even in the same subfamily. The gene structure, especially subfamily A containing 116 members, differs. However, subfamilies H, I, and J have only two members, but they slightly vary in gene structure. In 167 WD40 protein sequences, *GbWD40-52* has the most introns (total of 36), followed by *GbWD40-69*, which contains 31 introns. Next, *GbWD40-59* and *GbWD40-153* have 30 introns. A total of 21 (12.57%) *GbWD40*s have only one exon but have no introns. A total of 89 *GbWD40* genes contain 1-10 introns, the 57 remaining genes contain 11 or more introns (Figure 5, Table S1).

Expression profiles of GbWD40 genes in different tissues

The transcript abundance in different tissues, including root (R), stem (S), immature leaf (IL), mature leaf (ML), microstrobilus (M), ovulate strobilus (OS), immature fruit (IF), and mature fruit (MF) of *G. biloba*, are analysed on the basis of the transcriptome data (Ye *et al.*, 2019) to understand the tissue-specific expression patterns of the *GbWD40* genes (Figure 6). The results show that 73.7% (115/156) of *GbWD40* genes are highly expressed in eight tissues (FPKM > 5), whereas 8.33% (13/156) of *GbWD40* genes are lowly expressed in eight tissues (FPKM < 2).

Figure. 5. Gene structures of *GbWD40* genes

The exon/intron organization was drawn by the GSDS online tool. The letters (A to P) on the left indicate different subfamilies. The yellow box indicates the exon and the black line indicates the intron

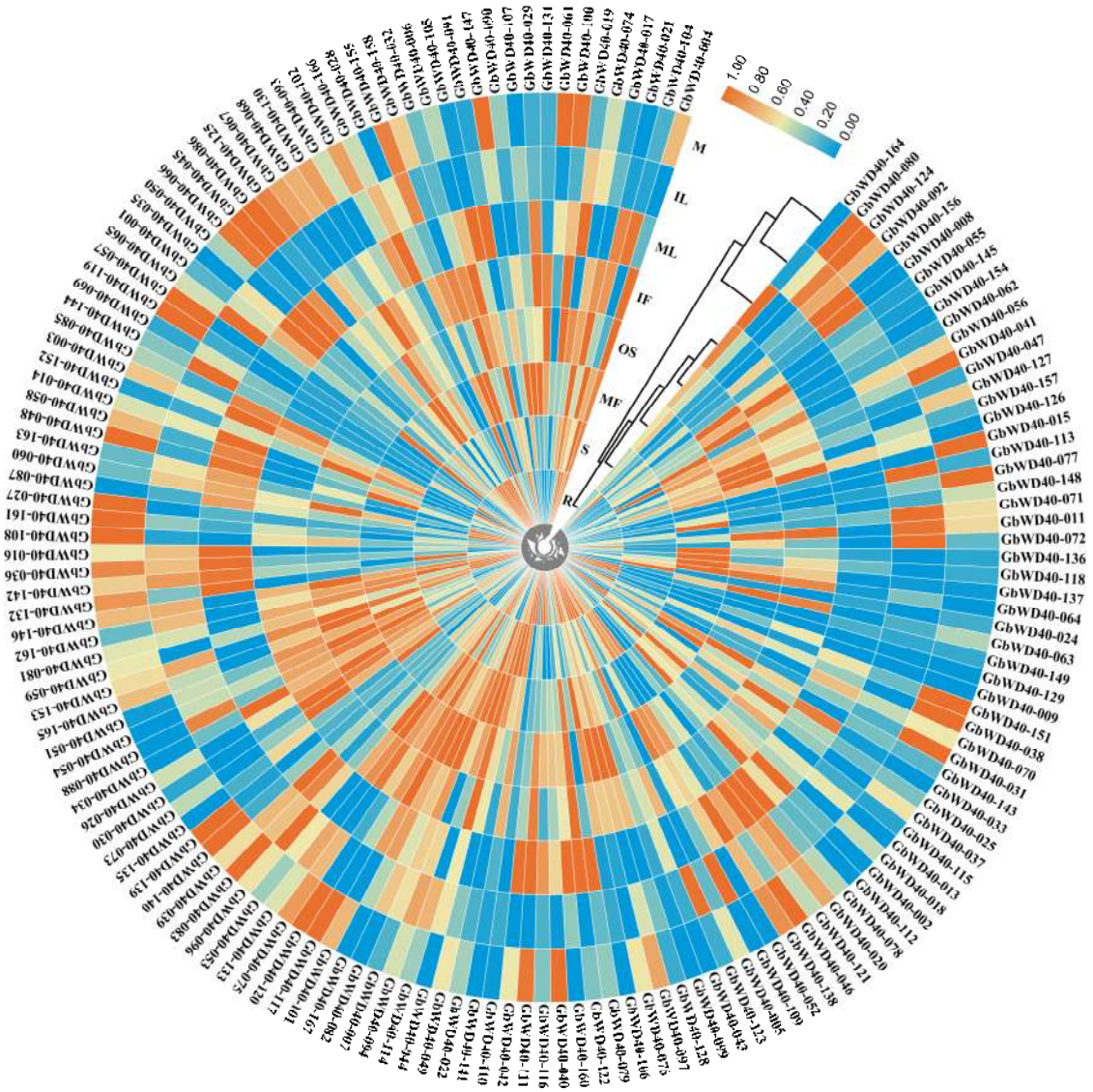

Figure 6. Expression profiles of *GbWD40* genes in different tissues
The R, S, IL, ML, M, OS, IF and MF represent root, stem, immature leaf, mature leaf, microstrobilus, ovulate strobilus, immature fruit and mature fruit, respectively

Candidate GbWD40 genes involved in flavonoid biosynthesis in G. biloba

The FPKM values of 156 *GbWD40* genes in eight different tissues of *G. biloba* are obtained from transcriptome data, that is, 11 *GbWD40* genes are not expressed in eight tissues. The correlation analysis performed on the FPKM values of 156 *GbWD40* genes and flavonoid content reveal that six *GbWD40* genes are significantly correlated with flavonoid contents (Table S2). In detail, the expression levels of *GbWD40-078* (R^2 = 0.91975, p < 0.01), *GbWD40-079* (R^2 = 0.89728, p < 0.01), *GbWD40-036* (R^2 = 0.84282, p < 0.01), *GbWD40-016* (R^2 = 0.83254, p < 0.05), *GbWD40-039* (R^2 = 0.82949, p < 0.05), and *GbWD40-166* (R^2 = 0.82408, p < 0.05) are significantly and positively correlated with flavonoid content. Therefore, these six genes may participate in the regulation of the flavonoid biosynthesis pathway in *G. biloba*.

Promoter analysis of GbWD40 genes

To further investigate the putative functions of *GbWD40* genes, we identified and analysed the potential cis-elements in the promoter regions of 2000-bp upstream of the start codon of *WD40* genes using PlantCARE software. As shown in Figure S1, the *GbWD40* genes are rich in cis-acting elements, including elements that respond to hormones (gibberellin-responsive element, cis-acting element involved in salicylic acid responsiveness, auxin-responsive element, cis-acting regulatory element involved in the MeJA-responsiveness, cis-acting element involved in the abscisic acid responsiveness), elements that respond to hormones (cis-acting element involved in low-temperature responsiveness, MYB binding site involved in drought-inducibility, cis-acting element involved in light responsiveness), MYB binding site involved in flavonoid biosynthetic genes regulation, MYBHv1 binding site, cis-regulatory element involved in endosperm expression, and cis-acting regulatory element essential for the anaerobic induction. Among those *GbWD40*s, *GbWD40-036(Gb_06853)*, *GbWD40-059(Gb_14861)*, *GbWD40-030(Gb_28754)*, *GbWD40-078(Gb_32521)*, and *GbWD40-163(Gb_38061)* have "MYB binding site involved in flavonoid biosynthetic genes regulation" (Figure 7). We speculate that these five *WD40* genes are involved in the synthesis of flavonoid in *G. biloba*.

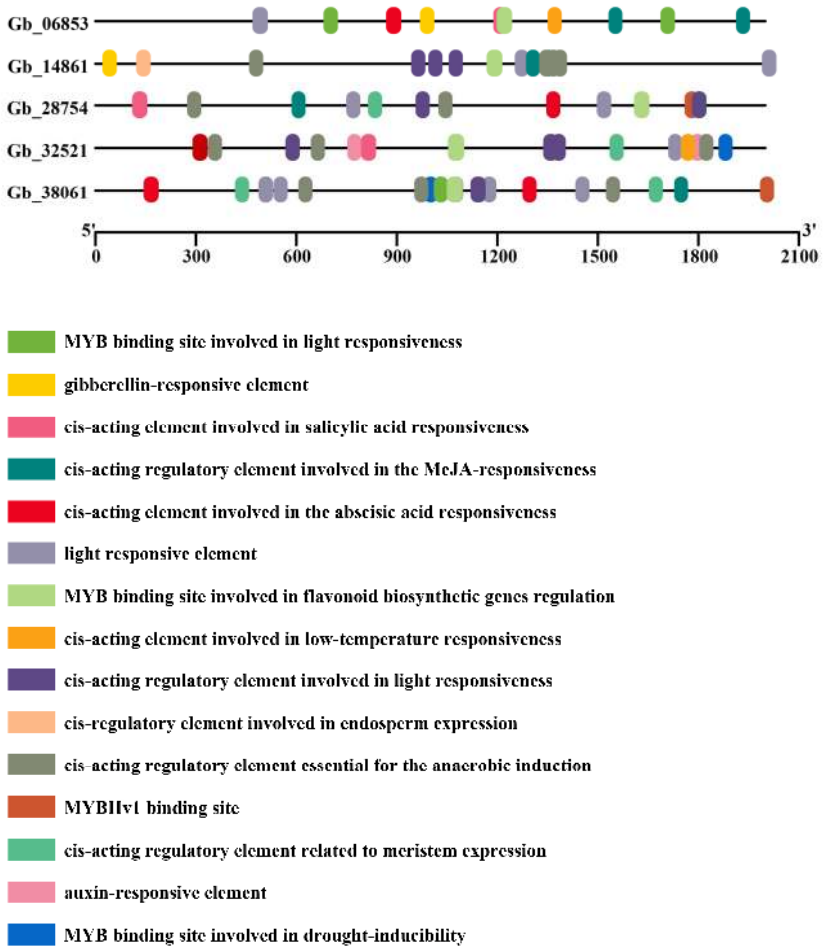

Figure 7. Promoter analysis of five *GbWD40* genes containing "MYB binding site involved in flavonoid biosynthetic genes regulation"

GO annotation analysis of GbWD40s genes

GO annotation analysis is performed on the target genes of *GbWD40*s. A total of 79 (47.31%) *GbWD40*s are divided into one or more functional GO categories, including molecular function, cellular component, and biological process (Figure 8). All the matched *WD40*s are annotated to 35 functional terms, including cellular component (20 subgroups), molecular function (10 subgroups), and biological process (5 subgroups; Figure 8). In the biological process, the following observations are found: 46 proteins with GO annotation enriched in cellular process, 40 proteins enriched in single-organism process, 33 proteins enriched in metabolic process, 22 proteins enriched in cellular component organization or biogenesis, 21 proteins enriched in biological regulation, 21 proteins enriched in biological process regulation. In the cell composition, the following findings are detected: 52 proteins with GO annotation enriched in cell and cell parts, 38 proteins enriched in macromolecular complex, 24 proteins enriched in organelles. In molecular functions, the following data are observed: 29 proteins with GO annotation enriched in binding and 21 proteins enriched in catalytic activities. Therefore, the members of this family likely play an important role in the regulation of protein–protein interactions and secondary metabolite synthesis.

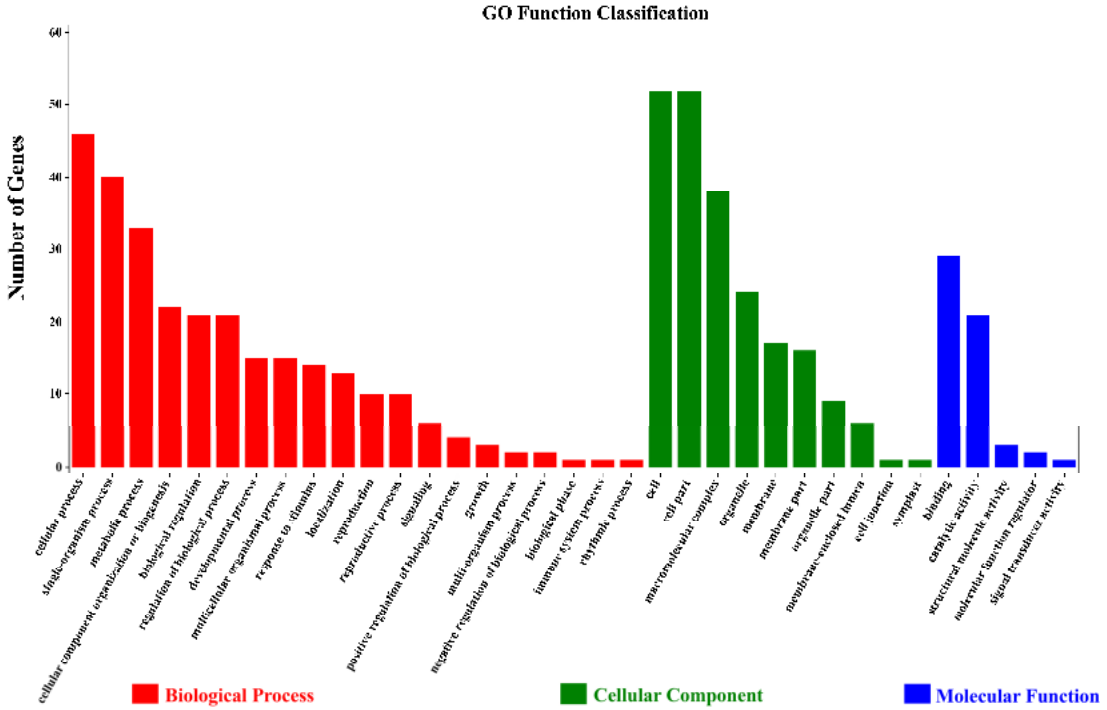

Figure 8. The Go annotation of *WD40* genes in *G. biloba*. All annotated GO terms including biological process, cellular component and molecular function of 79 *GbWD40*s

Discussion

The WD40 protein is a scaffold molecule for protein–protein interaction, which plays an important role in basic biological processes. They are highly conserved and abundant in eukaryotes, and they may play a key role in many biological processes, including signal transduction, protein trafficking, chromatin modification, and transcription (van Nocker *et al.*, 2003; Stirnimann *et al.*, 2010; Jain *et al.*, 2018). The WD40 family is identified and analysed in many species (Hu *et al.*, 2018; Salih *et al.*, 2018; Feng *et al.*, 2019; Liu *et al.*, 2020). These results show that WD40 family genes are amplified to different extents in different plant evolutionary stages. In this study, 167 WD40 proteins are identified in *G. biloba* genome. The analysis of the

positions of these *GbWD40* genes on the chromosome shows that WD40 genes in *G. biloba* are unevenly distributed among 12 chromosomes, which may be related to the evolution of *G. biloba*. The WD40 family genes in *G. biloba* are divided into 5 clusters according to the topological structure of the evolutionary tree and further divided into 16 subclasses according to the domain structure of *GbWD40* protein. A total of 116 proteins contain the WD40 domain, and they are classified as subclass A. The 51 remaining *GbWD40* genes consist of other domains, and they are categorized as subclasses B to P. In comparison with the classification of other plants, our results on the WD40 family genes in *G. biloba* show similarities and differences (Feng *et al.*, 2019; Liu *et al.*, 2020; Sun *et al.*, 2020). The *GbWD40* protein sequence is subjected to motif analysis. The result indicates that motif 1 is the most conserved in the WD40 family protein sequences in *G. biloba*.

The exon/intron diversification of gene family members plays an important role in the evolution of multiple gene families through the three main types of mechanisms, namely, exon/intron gain/loss, exonization/pseudoexonization, and insertion/deletion (Xu *et al.*, 2012). Of the 167 identified WD40 genes, 146 contain introns in the range of 1-36, and 21 genes have no introns. This result is consistent with Sun's identification result in *Rosa chinensis* (Sun et al., 2020). Introns are found in most WD40 genes, and no introns are detected in few genes. The WD40 family genes in *G. biloba* have experienced intron deletion or increase in evolution and development. The function of WD40 family members is affected by the size of exons and the length of introns. Studies on the gene structure of *GbWD40*s show that the size of exons and the length of introns are diversified, and the genome sequence length is in the range of 425-448 kb. This result indicates that the function of *GbWD40* genes may be diversified (Table S1).

Plant promoters are regulatory elements important for plant gene transcription and transcriptional level regulation (Danino *et al.*, 2015). Feyissa *et al.* (2019) observed that *SPL13*, as an inhibitory factor, can directly combine with the *DFR* promoter to affect its expression. Gou *et al.* (2011) indicated that *SPL9* can bind to the promoter region of *DFR*. Therefore, promoter analysis is essential for studies on gene function. The promoters of 167 *GbWD40* genes (2000 bp) are analysed, and *GbWD40-036* (II, B), GbWD40-059 (IV, P), GbWD40-030 (I, A), GbWD40-078 (III, F), and GbWD40-163 (I, C) contain the cis element of the "MYB binding site involved in the regulation of flavonoid biosynthesis genes." These five WD40 genes are distributed in Clusters I–IV, but do not belong to Cluster V, only containing subfamily A, B, C, F, and P. Therefore, genes with similar domain structures are not only the same in function but also different. Studies have shown that WD40 combines with bHLH and MYB to form a complex, which is involved in the regulation of flavonoid biosynthesis (Hichri *et al.*, 2011; Hong *et al.*, 2015). bHLH, WD40, and MYB are also involved as monomers in the regulation of flavonoid synthesis. In maize, *PAC1* (WD40), R (bHLH), and C1 (MYB) seem to be independently regulated (Carey *et al.*, 2004). In apples, *MdMYB10* does not seem to regulate the *MdbHLH3* and *MdbHLH33* expression (Espley *et al.*, 2007). The results demonstrate that the R2R3-MYB gene *GbMYBF2* plays a negative regulator role in flavonoid biosynthesis in *G. biloba* leaves (Xu *et al.*, 2014). Therefore, five *GbWD40* genes containing the cis-element of the "MYB binding site involved in the regulation of flavonoid biosynthesis genes" likely combine with MYB or combine with MYB and bHLH to form complexes. It plays a role in regulating the synthesis and metabolism of *G. biloba* and flavonoids.

TFs play essential roles in regulating flavonoid biosynthesis and transport. Interestingly, the MYB-bHLH-WD40 (MBW) ternary complex participates in flavonoid biosynthesis and transport processes because of its regulatory effect on many structural genes (Hichri *et al.*, 2011; Li, 2014). The correlation analysis of the FPKM value (156 *GbWD40* genes obtained from transcriptome data) and flavonoid content in eight *Ginkgo* tissues indicates that six WD40 genes significantly related to the flavonoid content of *G. biloba* are screened ($R^2 > 0.8$, $p < 0.05$ or $p < 0.01$), namely, *GbWD40-078* (III, F), *GbWD40-079* (I, C), *GbWD40-036* (II, B), *GbWD40-016* (II, A), *GbWD40-039* (II, F), and *GbWD40-166* (IV, A).

The synthesis of flavonoids in *A. thaliana* is relatively clear (Li, 2014), but the metabolic pathway of flavonoids in *G. biloba* is still unclear. A number of studies have shown that flavonoid synthesis is regulated by bHLH, MYB, WD40, and other transcription factors (Hichri *et al.*, 2011; Carey *et al.*, 2004; Li, 2014; Ye *et al.*, 2019). Five and six WD40 genes related to flavonoid synthesis are screened through promoter analysis and

correlation analysis, respectively. Two repeat genes are found in these 11 genes, which are also the key genes in the next step of functional verification. This result provides a direction for further studying WD40 in the synthesis and metabolism of flavonoids in *G. biloba*.

Conclusions

A total of 167 WD40 family genes in *G. biloba* genome are identified and analysed. A phylogenetic tree is constructed on the basis of 167 protein sequences and 100 WD40 family genes in *A. thaliana*. It is divided into 5 families based on the evolutionary relationship and further divided into 16 subfamilies from A to P based on the domain structure. The analysis on the promoters of *GbWD40* genes reveals that the promoters in *GbWD40-036* (*Gb_06853*), *GbWD40-059* (*Gb*_14861), *GbWD40-030* (*Gb_28754*), *GbWD40-078* (*Gb_32521*), and *GbWD40-163* (*Gb_38061*) genes contain the cis-acting element of the MYB binding site involved in the regulation of flavonoid biosynthesis genes related to flavonoid biosynthesis. The correlation analysis on the flavonoid content of eight different tissues in *G. biloba* and FPKM values of 156 *GbWD40*s in eight different tissues show that the expression of six *GbWD40*s is significantly related to flavonoid contents ($R^2 > 0.8$). In this study, WD40 family genes in the *G. biloba* genome are completely identified, the annotation of the *Ginkgo* genome is enriched, the prediction and development of WD40 family gene function in *G. biloba* are promoted. This study provides a clear direction for studying the synthesis and metabolism of flavonoids in *G. biloba*.

Authors' Contributions

Conceptualization: J.R.Z., Y.L.L. and F.X.; Methodology: J.R.Z., Y.L.L. and F.X.; Formal analysis: W.W.Z.; Data curation: J.R.Z., M.Y.F. and Z.Y.C.; Writing - original draft: J.R.Z.; Writing - review and editing: X.Z., X.M.L. and J.B.Y.; Funding acquisition: F.X. and Y.L.L. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by the National Natural Science Foundation of China, grant number 31901344 and 31270717.

Conflict of Interests

The authors declare that there are no conflicts of interest related to this article.

References

Albert NW, Thrimawithana AH, McGhie TK, Clayton WA, Deroles SC, Schwinn KE, ... Davies KM (2018). Genetic analysis of the liverwort *Marchantia polymorpha* reveals that R2R3-MYB activation of flavonoid production in response to abiotic stress is an ancient character in land plants. New Phytologist 218(2):554-566. *https://doi.org/10.1111/nph.15002*

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, ... Noble WS (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research 37:W202-W208. *https://doi.org/10.1093/nar/gkp335*

Carey CC, Strahle JT, Selinger DA, Chandler VL (2004). Mutations in the pale aleurone color1 regulatory gene of the *Zea mays* anthocyanin pathway have distinct phenotypes relative to the functionally similar TRANSPARENT TESTA *GLABRA1* gene in *Arabidopsis thaliana*. Plant Cell 16(2):450-464. *https://doi.org/10.1105/tpc.018796*

Chen C, Chen H, Zhang Y, Thomas Hannah R, Frank Margaret H, He Y, Xia R (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. Molecular Plant 13(8):1194-1202. *https://doi.org/10.1016/j.molp.2020.06.009*

Chen S, Wang S (2019). *GLABRA2*, a common regulator for epidermal cell fate determination and anthocyanin biosynthesis in *Arabidopsis*. International Journal of Molecular Sciences 20(20):4997. *https://doi.org/10.3390/ijms20204997*

Danino YM, Even D, Ideses D, Juven-Gershon T (2015). The core promoter: at the heart of gene expression. Biochimica et Biophysica Acta-Gene Regulatory Mechanisms 1849(8):1116-1131. *https://doi.org/10.1016/j.bbagrm.2015.04.003*

El-Gebali Sara, Mistry Jaina, Bateman Alex, Eddy Sean R, Luciani Aurelien, Potter Simon C, ... Finn Robert D (2019). The Pfam protein families database in 2019. Nucleic Acids Research 47(8):D427-D432. *https://doi.org/10.1093/nar/gky995*

Espley RV, Hellens RP, Putterill J, Stevenson DE, Kutty-Amma S, Allan AC (2007). Red colouration in apple fruit is due to the activity of the MYB transcription factor, *MdMYB10*. Plant Journal 49(3):414-427. *https://doi.org/10.1111/j.1365-313X.2006.02964.x*

Feng R, Zhang C, Ma R, Cai Z, Lin Y, Yu M (2019). Identification and characterization of WD40 superfamily genes in peach. Gene 710:291-306. *https://doi.org/10.1016/j.gene.2019.06.010*

Feyissa BA, Arshad M, Gruber MY, Kohalmi SE, Hannoufa A (2019). The interplay between miR156/SPL13 and DFR/WD40–1 regulate drought tolerance in alfalfa. BMC Plant Biology 19:434. *https://doi.org/10.1186/s12870-019-2059-5*

Gong W, Chen C, Dobes C, Fu CX, Koch MA (2008). Phylogeography of a living fossil: Pleistocene glaciations forced *Ginkgo biloba* L. (Ginkgoaceae) into two refuge areas in China with limited subsequent postglacial expansion. Molecular Phylogenetics and Evolution 48(3):1094-1105. *https://doi.org/10.1016/j.ympev.2008.05.003*

Gou JY, Felippes FF, Liu CJ, Weigel D, Wang JW (2011). Negative regulation of anthocyanin biosynthesis in *Arabidopsis* by a miR156-targeted SPL transcription factor. Plant Cell 23(4):1512-1522. *https://doi.org/10.1105/tpc.111.084525*

Guan R, Zhao Y, Zhang H, Fan G, Liu X, Zhou W, ... Chen W (2019). Updated genome assembly of *Ginkgo biloba* GigaScience database. Gigascience 2019. *http://dx.doi.org/10.5524/100613*

Hichri I, Barrieu F, Bogs J, Kappel C, Delrot S, Lauvergeat V (2011). Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. Journal of Experimental Botany 62(8):2465-2483. *https://doi.org/10.1093/jxb/erq442*

Hong Y, Tang X, Huang H, Zhang Y, Dai S (2015). Transcriptomic analyses reveal species-specific light-induced anthocyanin biosynthesis in chrysanthemum. BMC Genomics 16(1):202. *https://doi.org/10.1186/s12864-015-1428-1*

Hu XJ, Li T, Wang Y, Xiong Y, Wu XH, Zhang DL, ... Wu YD (2017). Prokaryotic and highly-repetitive WD40 proteins: a systematic study. Scientific Reports 7(1):10585. *https://doi.org/10.1038/s41598-017-11115-1*

Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G (2015). GSDS 2.0: an upgraded gene features visualization server. Bioinformatics 31(8): 1296-1297. *https://doi.org/10.1093/bioinformatics/btu817*

Hu R, Xiao J, Gu T, Yu XF, Zhang Y, Chang JL, ... He GY (2018). Genome-wide identification and analysis of WD40 proteins in wheat (*Triticum aestivum* L.). BMC Genomics 19(1): 803. *https://doi.org/10.1186/s12864-018-5157-0*

Jain BP, Pandey S (2018). WD40 repeat proteins: signalling scaffold with diverse functions. Protein Journal 37(5):391-406. *https://doi.org/10.1007/s10930-018-9785-7*

Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, ... Rombauts S (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Research 30(1):325-327. *https://doi.org/10.1093/nar/30.1.325*

Li S (2014). Transcriptional control of flavonoid biosynthesis: fine-tuning of the MYB-bHLH-WD40 (MBW) complex. Plant Signaling Behavior 9(1):e27522. *https://doi.org/10.4161/psb.27522*

Liu H, Cao F, Yin T, Chen Y (2017). A highly dense genetic map for *Ginkgo biloba* constructed using sequence-based markers. Frontiers in Plant Science 8: 1041. *https://doi.org/10.3389/fpls.2017.01041*

Lin XH, Zhang J, Li Y, Luo HM, Wu Q, Sun C, ... Chen S (2011). Functional genomics of a living fossil tree, *Ginkgo*, based on next-generation sequencing technology. Physiologia Plantarum 143(3):207-218. *https://doi.org/10.1111/j.1399-3054.2011.01500.x*

Liu Z, Liu YH, Coulter AJ, Shen BY, Li YM, Li CZ, ... Zhang JL (2020). The WD40 gene family in potato (*Solanum Tuberosum* L.): genome-wide analysis and identification of anthocyanin and drought-related WD40s. Agronomy 10(3):401. *https://doi.org/10.3390/agronomy10030401*

Migliori V, Mapelli M, Guccione E (2012). On WD40 proteins: propelling our knowledge of transcriptional control?. Epigenetics 7(8):815-822. *https://doi.org/10.4161/epi.21140*

Nakatsuka T, Haruta KS, Pitaksutheepong C, Abe Y, Kakizaki Y, Yamamoto K, ... Nishihara M (2008). Identification and characterization of R2R3-MYB and bHLH transcription factors regulating anthocyanin biosynthesis in gentian flowers. Plant and Cell Physiology 49(12):1818-1829. *https://doi.org/10.1093/pcp/pcn163*

Ni J, Dong LX, Jiang ZF, Yang XL, Chen ZY, Wu YH, Xu MJ (2018). Comprehensive transcriptome analysis and flavonoid profiling of *Ginkgo* leaves reveals flavonoid content alterations in day-night cycles. PLoS One 13(3):e0193897. *https://doi.org/10.1371/journal.pone.0193897*

Ouyang Y, Huang X, Lu Z, Yao J (2012). Genomic survey, expression profile and co-expression network analysis of *OsWD40* family in rice. BMC Genomics 13:100. *https://doi.org/10.1186/1471-2164-13-100*

Pesch M, Schultheiss I, Klopffleisch K, Uhrig JF, Koegl M, Clemen CS, ... Hulskamp M (2015). Transparent testa *GLABRA1* and *GLABRA1* compete for binding to *GLABRA3* in *Arabidopsis*. Plant Physiology 168(2):584-597. *https://doi.org/10.1104/pp.15.00328*

Ponting CP, Schultz J, Milpetz F, Bork P (1999). SMART: identification and annotation of domains from signalling and extracellular protein sequences. Nucleic Acids Research 27(1):229-232. *https://doi.org10.1093/nar/27.1.229*

Salih H, Gong W, Mkulama M, Du X (2018). Genome-wide characterization, identification, and expression analysis of the WD40 protein family in cotton. Genome 61(7):539-547. *https://doi.org/10.1139/gen-2017-0237*

Suganuma T, Pattenden SG, Workman JL (2008). Diverse functions of WD40 repeat proteins in histone recognition. Genes Development 22(10):1265-1268. *https://doi.org/10.1101/gad.1676208*

Stirnimann CU, Petsalaki E, Russell RB, Müller CW (2010). WD40 proteins propel cellular networks. Trends in Biochemical Sciences 35(10):565-574. *https://doi.org/10.1016/j.tibs.2010.04.003*

Sun YB, Zhang XJ, Zhong MC, Dong X, Yu DM, Jiang XD, ... Hu JY (2020). Genome-wide identification of WD40 genes reveals a functional diversification of COP1-like genes in *Rosaceae*. Plant Molecular Biology 104(1-2):81-95. *https://doi.org/10.1007/s11103-020-01026-7*

Ude C, Schubert-Zsilavecz M, Wurglics M (2013). *Ginkgo biloba* extracts: a review of the pharmacokinetics of the active ingredients. Clinical Pharmacokinetics 52(9):727-749. *https://doi.org/10.1007/s40262-013-0074-5*

van Nocker S, Ludwig P (2003). The WD-repeat protein superfamily in *Arabidopsis*: conservation and divergence in structure and function. BMC Genomics 4(1):50. *https://doi.org/10.1186/1471-2164-4-50*

Wu Y, Guo J, Zhou Q, Xin Y, Wang G, Xu LA (2018). De novo transcriptome analysis revealed genes involved in flavonoid biosynthesis, transport and regulation in *Ginkgo biloba*. Industrial Crops and Products 124:226-235. *https://doi.org/10.1016/j.indcrop.2018.07.060*

Winkel-Shirley B (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiology 126(2):485-493. *https://doi.org/10.1104/pp.126.2.485*

Xie Y, Tan H, Ma Z, Huang J (2016). Della proteins promote anthocyanin biosynthesis via sequestering *MYBL2* and JAZ suppressors of the MYB/bHLH/WD40 complex in *Arabidopsis thaliana*. Molecular Plant 9(5):711-721. *https://doi.org/10.1016/j.molp.2016.01.014*

Xu F, Ning YJ, Zhang WW, Liao YL, Li LL, Cheng H, Cheng SY (2014). An R2R3-MYB transcription factor as a negative regulator of the flavonoid biosynthesis pathway in *Ginkgo biloba*. Functional Integrative Genomics 14(1):177-189. *https://doi.org/10.1007/s10142-013-0352-1*

Xu C, Min J (2011). Structure and function of WD40 domain proteins. Protein Cell 2(3):202-214. *htpps://doi.org/10.1007/s13238-011-1018-1*

Xu G, Guo C, Shan H, Kong, H (2012). Divergence of duplicate genes in exon-intron structure. Proceedings of the National Academy of Sciences of the United States of America 109(4):1187-1192. *https://doi.org/10.1073/pnas.1109047109*

Ye JB, Cheng SY, Zhou X, Chen ZX, Kim SU, Tan JP, ... Zhu YX (2019). A global survey of full-length transcriptome of *Ginkgo biloba* reveals transcript variants involved in flavonoid biosynthesis. Industrial Crops and Products 139:111547. *https://doi.org/10.1016/j.indcrop.2019.111547*

Zou XD, Hu XJ, Ma J, Li T, Ye ZQ, Wu YD (2016). Genome-wide analysis of WD40 protein family in human. Scientific Reports 6:39262. *https://doi.org/10.1038/srep39262*