

RESEARCH

Open Access



Genome-wide identification and bioinformatics analysis of the WD40 transcription factor family and candidate gene screening for anthocyanin biosynthesis in *Rhododendron simsii*

Cheng Wang¹, Yafang Tang¹, Yan Li², Chao Hu¹, Jingyi Li¹ and Ang Lyu^{3*}

Abstract

WD40 transcription factors (TFs) constitute a large gene family in eukaryotes, playing diverse roles in cellular processes. However, their functions in the major ornamental plant, *Rhododendron simsii*, remain poorly understood. In this study, we identified 258 WD40 proteins in the *R. simsii* genome, which exhibited an uneven distribution across chromosomes. Based on domain compositions and phylogenetic analysis, we classified these 258 RsWD40 proteins into 42 subfamilies and 47 clusters. Comparative genomic analysis suggested that the expansion of the WD40 gene family predates the divergence of green algae and higher plants, indicating an ancient origin. Furthermore, by analyzing the duplication patterns of RsWD40 genes, we found that transposed duplication played a major role in their expansion. Notably, the majority of RsWD40 gene duplication pairs underwent purifying selection during evolution. Synteny analysis identified significant orthologous gene pairs between *R. simsii* and *Arabidopsis thaliana*, *Oryza sativa*, *Vitis vinifera*, and *Malus domestica*. We also investigated potential candidate genes involved in anthocyanin biosynthesis during different flower development stages in *R. simsii* using RNA-seq data. Specifically, we identified 10 candidate genes during the bud stage and 7 candidate genes during the full bloom stage. GO enrichment analysis of these candidate genes revealed the potential involvement of the ubiquitination process in anthocyanin biosynthesis. Overall, our findings provide a valuable foundation for further investigation and functional analysis of WD40 genes, as well as research on the molecular mechanisms underlying anthocyanin biosynthesis in *Rhododendron* species.

Keywords *Rhododendron simsii*, WD40, Duplication analysis, Anthocyanin biosynthesis, Candidate genes

*Correspondence:

Ang Lyu

lvang@hbaas.com

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



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

The WD40 transcription factors (TFs), also known as WD-repeat (WDR) proteins, constitute one of the largest TF families in eukaryotic organisms [1]. The WD40 TF is distinguished by the existence of multiple WD40 repeats, each consisting of 44–60 amino acids. These repeats adopt a seven-bladed β -propeller structure formed through repetitive folding. The N-terminus of each blade contains a glycine-histidine (GH) dipeptide with 11 to 24 residues, while the C-terminus is composed of the WD dipeptide [2, 3]. This arrangement allows for the formation of a stable repeat fold with a robust hydrogen bond network [4]. Besides the WD40 repeats, WD40 proteins often contain additional protein domains, resulting in lower sequence conservation among family members [5]. Different researchers have classified WD40 TFs in various ways, either based on phylogenetic analysis or domain compositions. For instance, in *Arabidopsis thaliana*, 237 AtWD40 TFs with a minimum of 4 WD40 repeats were grouped into 33 clusters, while in cucumber (*Cucumis sativus*), 191 CsWD40 TFs were categorized into 21 clusters [6]. Phylogenetic analysis divided 167 GbWD40, 204 FcWD40, 743 TaWD40, 225 SiWD40 and 187 RcWD40 TFs from *Ginkgo biloba*, *Ficus carica*, wheat (*Triticum aestivum*), *Setaria italica*, and *Rosa chinensis*, respectively, into 5 clusters, while domain composition analysis categorized them into 16, 12, 11, 12, and 15 subfamilies [7–11]. In apple (*Malus domestica*), 346 MdWD40 TFs were classified into 14 clusters based on phylogenetic analysis and 12 subfamilies based on domain composition analysis [12]. Similarly, in Potato (*Solanum Tuberosum*), 178 StWD40 TFs were grouped into 14 clusters based on the phylogenetic tree and 10 subfamilies based on domain compositions [13]. Additionally, phylogenetic analysis of 315 MiWD40 TFs from mango (*Mangifera indica*) and 231 CmWD40 TFs from *Cucurbita maxima* resulted in their classification into 11 and 5 clusters, respectively [14, 15].

WD40 TFs are typically recognized as adaptors that recruit other transcription factors to form protein or protein-DNA complexes [16, 17]. They play versatile roles in various physiological and biochemical processes, including cell cycle and division [18–20], DNA damage repair [21, 22], signal transduction [23, 24], flower development and flowering [25–27], hormones responses [28, 29], and stress responses [30–32]. One prominent function of WD40 is its involvement in the regulation of anthocyanin biosynthesis through the MYB-bHLH-WD40 (MBW) complex. The WD40 gene *TTG1* was initially identified in *Arabidopsis*, and mutations in this gene (*ttg1-1*) result in yellow seeds due to the absence of anthocyanin accumulation [33–35]. TTG1 can form MBW complexes with R2R3 MYB

transcription factors PAP1, PAP2, MYB113, MYB114 or TT2, and bHLH transcription factors GL3, EGL3 or TT8 [35–41]. Different MBW complexes regulate the biosynthesis of different flavonoids, and it has been demonstrated that the TTG1-TT8/GL3-PAP1/PAP2/MYB113/MYB114 complexes activate the expression of late biosynthetic genes, such as dihydroflavonol-4-reductase (DFR), anthocyanidin synthase (ANS), and UDP-glucose: flavonoid-3-O-glucosyltransferase (UFGT) in the anthocyanin biosynthetic pathway [38, 42, 43]. The TTG1-GL3/TT8-TT2 complexes activate the expression of DFR, ANS, BANYULS (BAN), TT19, and TT12 to regulate anthocyanin accumulation [37, 40, 43]. While *TTG1* orthologs in various plant species, such as *Camellia sinensis* [44], *Salvia miltiorrhiza* [45], *Vitis vinifera* [46], rice [47], and *Dendrobium candidum* [48], have been shown to regulate biosynthesis and accumulation of anthocyanin, no other WD40 genes have been identified in anthocyanin biosynthesis pathway. Consequently, there is a lack of literature investigating the role of WD40 genes in flower coloration, particularly in plants with a diverse range of flower colors. Therefore, further research is necessary to explore the potential involvement of WD40 genes in anthocyanin synthesis.

The *Rhododendron* genus (Ericaceae) is renowned for its extensive variety of corollas, encompassing more than 1000 species and 30,000 cultivars [49, 50]. *Rhododendron* exhibits a wide range of flower colors, including red, light red, pink, white, yellow, and blue. The aesthetic and economic value of *Rhododendron* as an ornamental plant is greatly influenced by the diversity of its flower colors. In *Rhododendron* species, flower pigmentation predominantly relies on the presence of anthocyanins and flavanols. Specifically, the composition and quantity of anthocyanin constituents determine the flower coloration [51, 52].

To date, there has been no comprehensive investigation of the WD40 gene family in *R. simsii*. In this study, we performed a genome-wide analysis of *R. simsii* WD40 (*RsWD40*) genes based on the published *R. simsii* genome [49]. We identified a total of 57 *RsWD40* genes. Subsequently, we examined their chromosome locations, categorized them into different groups, established their phylogenetic relationships, and investigated gene duplication events among the *RsWD40s*. Furthermore, we identified candidate *RsWD40* genes associated with anthocyanin biosynthesis using RNA-seq data from *Rhododendron* flower samples from three varieties with distinct flower colors. The findings of this research enhance our understanding of the *RsWD40* gene family and will facilitate genetic improvements in flower coloration in *Rhododendron* species.

Materials and methods

Investigation of WD40 TF family in *R. simsii* and other plant species

The complete protein sequence data of *R. simsii* was obtained from the Rhododendron Plant Genome Database (RPGD), accessible at <http://bioinfo.kib.ac.cn/RPGD/>. To identify potential RsWD40 proteins, we used the Hidden Markov Model (HMM) profile for WD40 (PF00400) as a query and performed a search using the HMMER 3.0 software [53]. The putative RsWD40 proteins were then validated by submitting them to the SMART [54] and Pfam [55] databases. Only sequences that exhibited at least one characteristic amino acid sequence of the WD40 repeat were retained.

In addition to *R. simsii*, we selected 21 other plant species to investigate the evolutionary dynamics of the WD40 family across the plant kingdom, representing a range of lower to higher plants. The acquisition of WD40 protein sequences from several species followed the protocol described in "Investigation of WD40 TF Family in *R. simsii* and Other Plant Species" section. The selected species include *Medicago truncatula*, *Populus trichocarpa*, *Brassica rapa*, *Theobroma cacao*, *Vitis vinifera*, *Solanum lycopersicum*, *Solanum melongena*, *Camellia sinensis*, *Sorghum bicolor*, *Azolla filiculoides*, *Salvinia cucullata*, *Selaginella moellendorffii*, *Physcomitrium patens*, *Chlamydomonas reinhardtii*, *Ostreococcus lucimarinus*, and *Cyanidioschyzon merolae*. Additionally, we retrieved WD40 protein sequences of the following species from previously published studies: *Cucumis sativus* [6], *Malus domestica* [12], *Arabidopsis thaliana* [6], *Oryza sativa* [5], and *Ginkgo biloba* [7]. The evolutionary relationships among these species were determined based on the PGDD database [56] and a previous study [57].

Characterization and genome distribution of RsWD40s

The coding sequence (CDS) and protein sequence of the RsWD40 TFs were obtained from RPGD database. To analyze their physical characteristics, all RsWD40 sequences were submitted to EXPASY (<https://web.expasy.org/protparam/>) for calculating the number of amino acids, molecular weight (MW), theoretical isoelectric point (pI), and grand average of hydropathicity (GRAVY) [58]. The position annotation of the RsWD40 genes was obtained from the RPGD database and visualized using TBtools software [59].

Classification and phylogenetic analysis of RsWD40s

To categorize the WD40 TFs family, two methods were employed: domain composition analysis and phylogenetic analysis. For the domain composition analysis, the amino acid sequences of the RsWD40 proteins were

searched against the SMART database [54] to identify their domain compositions. Based on their domain compositions, the RsWD40s were categorized into distinct subfamilies.

To construct the phylogenetic trees, the full-length sequences of RsWD40 proteins were used. First, the identified RsWD40 proteins were aligned using the T-Coffee software [60] with default parameters to ensure accurate sequence alignment. Then, an unrooted phylogenetic tree was constructed using the neighbor-joining method implemented by MEGA 7 [61]. The phylogenetic analysis employed the poisson model, pairwise deletion to handle gaps in the alignment, and 1,000 bootstrap replicates to assess the support for branching patterns. Subsequently, the resulting phylogenetic tree was visualized and enhanced using iTOL [62]. To assign RsWD40 proteins to specific clusters within the phylogenetic tree, a threshold of bootstrap values greater than 0.5 was applied, following the method described by Li et al. [6]. RsWD40 proteins exhibiting bootstrap values above this threshold were grouped together in the respective clusters.

Gene duplication analysis

To analyze gene duplication events among RsWD40 genes, we employed DupGen_finder [63]. This tool allows differentiation between various types of gene duplicates, including whole-genome duplication (WGD), tandem duplication, transposed duplication, proximal duplication, and dispersed duplication, using the default settings. The duplicated RsWD40 gene pairs were visualized using TBtools software [59]. To assess the evolutionary dynamics of the duplicated gene pairs, *Ka* (non-synonymous substitution rate), *Ks* (synonymous substitution rate), and *Ka/Ks* values were calculated using KaKs_Calculator (version 2.0) [64]. These values provide insights into the selective pressures acting on duplicated genes. Furthermore, we examined the synteny of WD40 genes between *R. simsii* and other plant species using MCS-canX [65]. The results of the synteny analysis were presented visually using TBtools software [59], allowing for a comprehensive understanding of the conservation and evolutionary relationships of WD40 genes across different plant species.

Expression analysis of RsWD40 genes

To investigate the expression patterns of RsWD40 genes, RNA-seq data from *Rhododendron* flower samples of three varieties (*Rhododendron wardii* var. *puralbum*, *Rhododendron simsii* Planch, and *Rhododendron hybridum* Ker Gawl, Figure S1), with white, pink, and red colored flowers, respectively, were analyzed. The plant materials used in this study were provided by Professor Xiaojing Wang from Guizhou University and were grown

under controlled conditions in a growth chamber. Flower samples were collected at the bud and full bloom stages. The transcript abundance of *RsWD40* genes was quantified using the Fragments per kilobase of exon model per million mapped reads (FPKM) method. The expression data for each gene were normalized by log₂ (FPKM) and presented as heat maps using TBtools software [59]. Differentially expressed genes (DEGs) were identified by comparing the expression levels between the different *Rhododendron* varieties. Genes with an absolute |fold change| ≥ 2 (log₂ |fold change| ≥ 1) and a *p*-value < 0.05 were considered as DEGs. To identify candidate genes associated with anthocyanin biosynthesis, a comparative strategy was employed. Specifically, the gene expression data from the same stages of floral development were used to compare colored and white flowers (WF versus RE, WF versus PF), resulting in two sets of DEGs that included both up- and down-regulated genes. The final set of DEGs was obtained by intersecting these two sets. The comparative strategy was applied separately for the bud and full bloom stages.

Confirmation of RNA seq analysis by qRT-PCR

To validate the RNA-seq results, qRT-PCR was conducted to validate the expression levels of candidate *RsWD40* genes related to anthocyanin biosynthesis at the bud and full bloom stages. The primers for qRT-PCR were designed using Beacon Designer 8, and their specific primer sequences can be found in Table S1. To ensure the accuracy and reliability of our experimental results, we adopted a strategy involving the use of at least two reference genes for data normalization [66–68]. The selection of *RsGAPDH* (Rhsim12G0106200) and *RsEF1α* (Rhsim02G0008200) as internal reference genes was based on their consistent expression levels in the RNA-seq data from the flowers of three different *Rhododendron* varieties at both developmental stages. Additionally, these genes have been utilized in previous studies investigating gene expression in the *Rhododendron* species [69, 70]. The qRT-PCR was carried out using the StepOnePlus™ System (Applied Biosystems, Foster City, CA, USA). The cycling parameters for qRT-PCR were as follows: 95 °C for 30 s, 95 °C for 5 s (40 cycles), and 60 °C for 30 s. Each qRT-PCR test included three biological and three technical replicates. The qRT-PCR reaction mixture was prepared using the TB Green® Premix Ex Taq™ kit (TaKaRa, Dalian, China). The data analysis was performed using the 2^{−ΔΔCt} method [71].

GO enrichment analysis of candidate *RsWD40* genes

GOATOOLS (<http://github.com/tanghaibao/GOatools>) was used to perform gene annotation, and Fisher's exact

test was employed to analyze the biological function enrichment of the candidate *RsWD40* genes. To reduce the likelihood of false positives, we applied the Bonferroni multiple testing correction, with a significance threshold set at a corrected *p*-value of less than 0.05. The results were visualized using a bubble dot diagram, which was generated on the website <http://www.bioinformatics.com.cn>.

Result

Identification and chromosomal locations of the *RsWD40*s

A total of 258 *RsWD40* genes were identified and renamed as *RsWD40-1* to *RsWD40-258* based on their chromosomal positions on Chr1-13 and the order of scaffolds (Fig. 1, Table S2). These genes were analyzed for their protein characteristics, including length, isoelectric point (PI), molecular weight (MW), and grand average of hydropathicity (GRAVY). The amino acid number of the *RsWD40* proteins ranged from 94 (*RsWD40-151*) to 3671 (*RsWD40-226*). The putative PIs of *RsWD40* proteins ranged from 4.40 (*RsD40-70*) to 9.76 (*RsWD40-63*). The MW of *RsWD40* proteins varied from 10.95 kDa (*RsWD40-151*) to 408.65 kDa (*RsWD40-226*). The vast majority of *RsWD40* proteins (248 out of 258) had a GRAVY (grand average of hydropathicity) value lower than 0, indicating they were hydrophilic. As shown in Fig. 1, the distribution of 240 *RsWD40*s (with the remaining 18 not being located) across the 13 chromosomes varies, with the maximum 33 *RsWD40*s found on chromosome 3, and only 9 *RsWD40*s found on chromosome 9.

Classification of *RsWD40* TFs

The *RsWD40* transcription factors (TFs) exhibit a diverse range of protein domains in addition to the WD40 repeats, as summarized in Table S1. Considering the presence of these additional domains, we classified the 258 *RsWD40* TFs into 42 distinct subfamilies, as shown in Table S3. Subfamily 1 (S1) consisted of 162 *RsWD40* TFs that exclusively contained WD40 repeats, while the remaining 96 *RsWD40* TFs sharing common domains were assigned to Subfamilies 2–42 (Table S3).

Due to variations in the number of WD40 repeats and the spacing of amino acids between them, aligning the WD40 repeats directly posed challenges. To overcome this, we employed the complete amino acid sequences of *RsWD40* proteins to construct an unrooted phylogenetic tree using MEGA 7 software. Figure 2 depicts the identification of 47 clusters (C1–C47) among the 258 *RsWD40* proteins, with a minimum bootstrap support of 0.5. Notably, out of the 258 *RsWD40* proteins, 112 could not be definitively assigned to any specific cluster through the phylogenetic analysis.

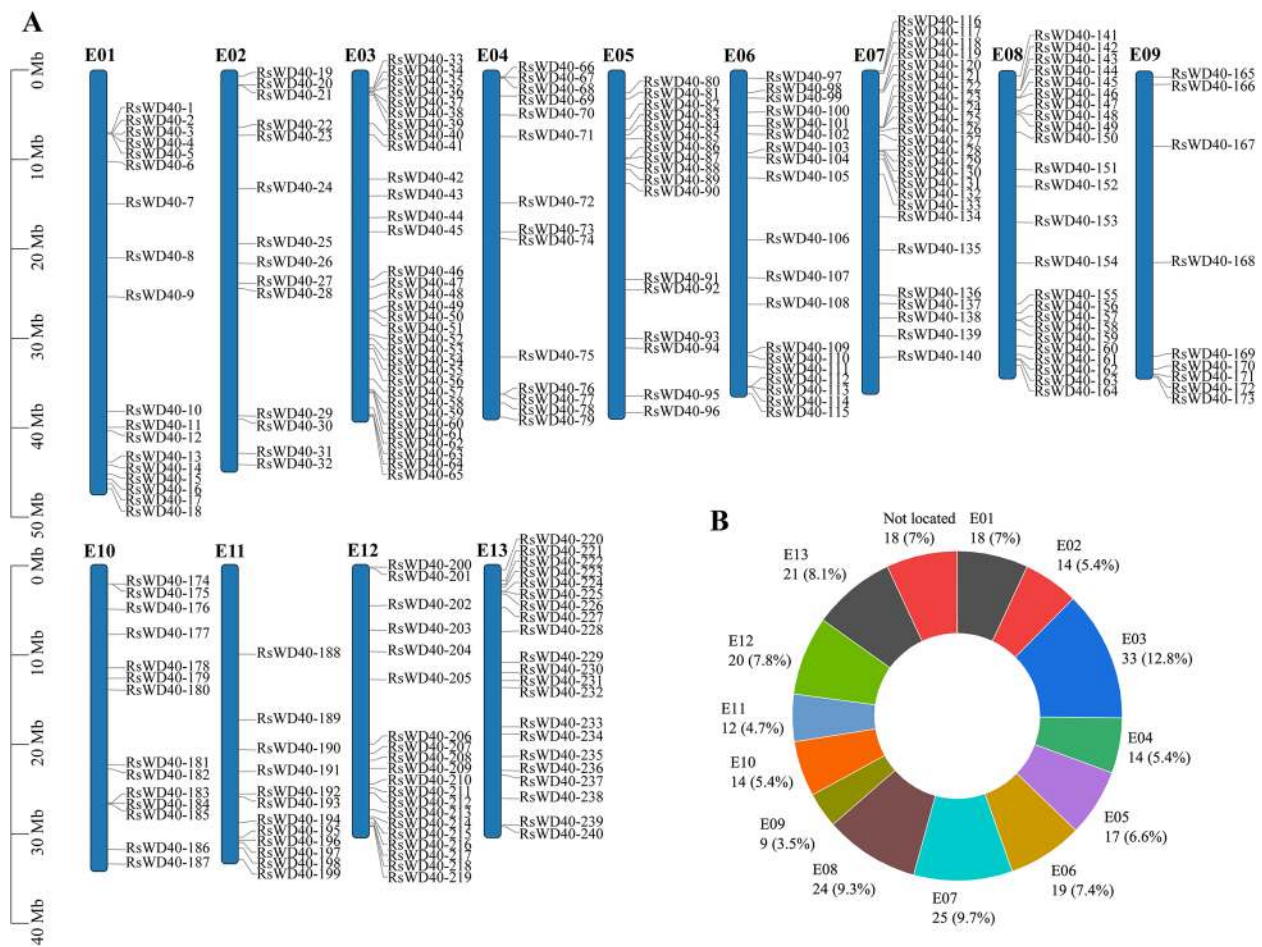


Fig. 1 Chromosomal locations of the *RsWD40s*. The relative length of chromosomes serves as a measure of its size. **A** The distribution *RsWD40* gene on 13 chromosomes. **B** Numbers of *RsWD40* genes on each *R. simsii* chromosome

Comparative genomic analysis of *WD40* genes

According to our investigation, we have identified a total of 258, 286, 332, 348, 243, 255, 262, 262, 323, 234, 227, 205, 204, 305, 223, 125, and 78 *WD40* genes from *R. simsii*, *M. truncatula*, *P. trichocarpa*, *B. rapa*, *T. cacao*, *V. vinifera*, *S. lycopersicum*, *S. melongena*, *C. sinensis*, *S. bicolor*, *A. filiculoides*, *S. cucullate*, *S. moellendorffii*, *P. patens*, *C. reinhardtii*, *O. lucimarinus*, and *C. merolae*, respectively (Table S4). Furthermore, previous studies have suggested that *C. sativus*, *M. domestica*, *A. thaliana*, *O. sativa*, and *G. biloba* species possess 191, 346, 230, 200, and 167 *WD40* genes, respectively [5–7, 12].

To explore the evolutionary patterns of the *WD40* gene family, we conducted a comparative genomic study of *WD40* genes in these 22 plant species. Figure 3 illustrates the evolutionary relationships among these species and the corresponding number of *WD40* genes in each genome. Our findings indicated that *WD40* genes are abundant in both lower and higher plants. Additionally, we calculated the gene density of *WD40* and

observed that it was higher in lower plants compared to higher plants. In higher plants, the gene density of *WD40* decreased starting from *P. patens* and then remained relatively stable. Interestingly, *G. biloba* had the lowest *WD40* gene density among all species due to its enormous genome size.

Gene duplication and *Ka/Ks* analysis

Gene duplication events were classified into five categories, namely WGD, tandem, proximal, transposed, and dispersed duplications, using DupGen_finder. Among these duplication events, transposed duplication accounted for the highest proportion at 33.7%, followed by dispersed duplication (23.6%), WGD (19%), proximal duplication (6.2%), and tandem duplication (5%). Additionally, 32 *RsWD40* genes (12.4%) were identified as singletons, indicating that they did not arise from duplication (Figure S2, Table S2). A total of 152 duplicate gene pairs were identified among the *RsWD40s* (Table S5, Fig. 4). We further analyzed the synonymous

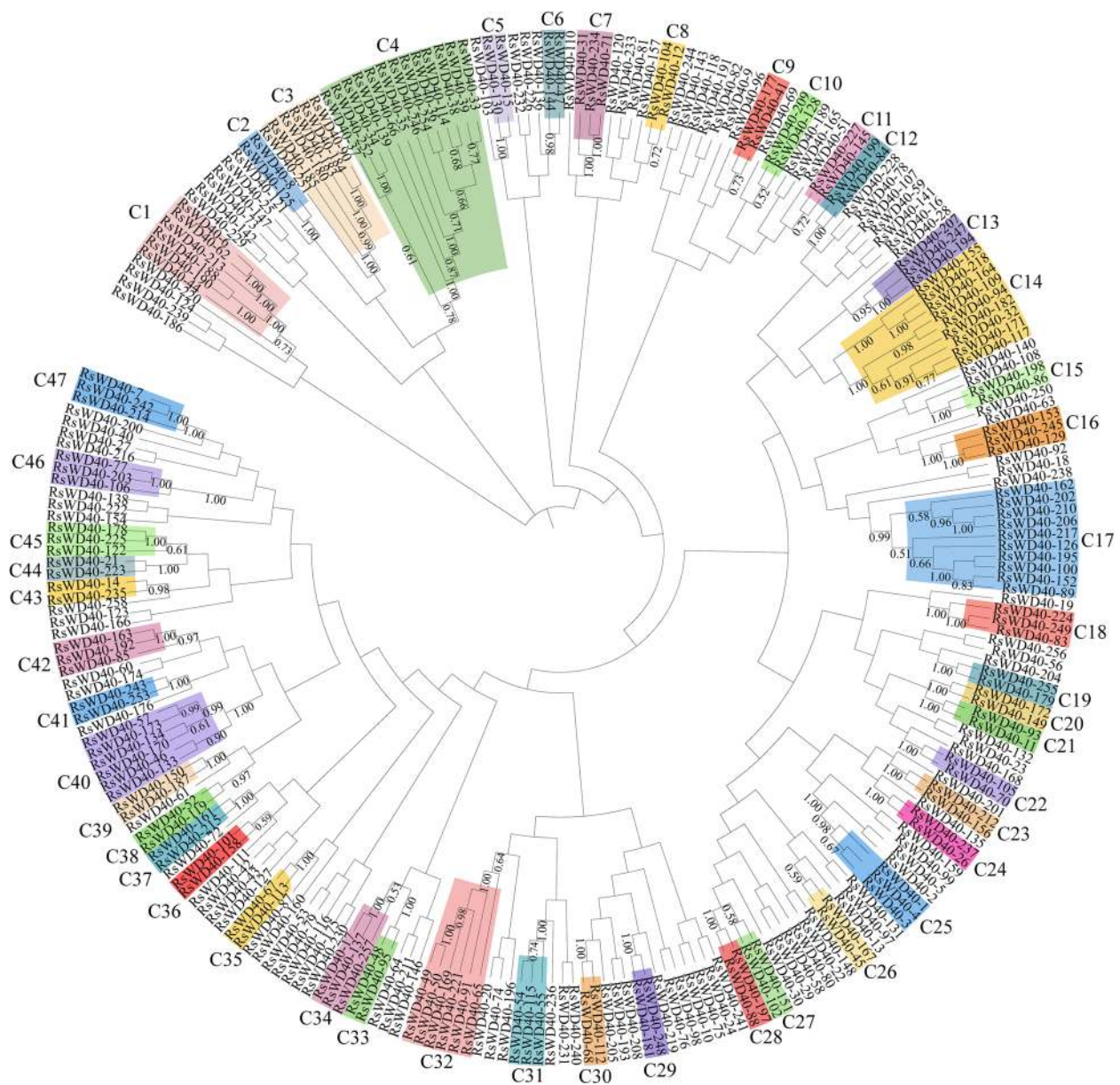


Fig. 2 Phylogenetic analysis of WD40 proteins in *R. simsii*. The phylogenetic tree was constructed using the neighbor-joining method implemented by MEGA 7, based on the complete amino acid sequences of RsWD40 proteins. The tree shows 47 phylogenetic clusters (C1-C47) with high bootstrap values. Bootstrap values lower than 0.5 are not displayed in the phylogenetic tree

(K_s) and non-synonymous (K_a) mutations in these gene pairs (Table S5). The K_a/K_s ratio was used as an indicator of selection pressure during evolution. The analysis revealed that the majority of *RsWD40* gene pairs exhibited a K_a/K_s ratio less than 1, suggesting the influence of purifying selection. This indicates that these genes have undergone functional constraints and selective pressure to maintain their essential functions. Only one gene pair (*RsWD40-246/RsWD40-254*) had a K_a/K_s ratio greater than 1, suggesting positive selection for advantageous mutations in these genes.

To investigate the potential evolutionary processes of the WD40 gene family, we conducted a synteny analysis to examine orthologous relationships of WD40 family genes between *R. simsii* and four other species (*A. thaliana*, *O. sativa*, *V. vinifera* and *M. domestica*) (Fig. 5). The analysis revealed a total of 164 pairs of orthologs between *R. simsii* and *A. thaliana*, 55 pairs of orthologs between *R. simsii* and *O. sativa*, 190 pairs of orthologs between *R. simsii* and *V. vinifera*, 306 pairs of orthologs between *R. simsii* and *M. domestica* (Table S6). These orthologous pairs of *RsWD40* genes and the corresponding *AtWD40*,

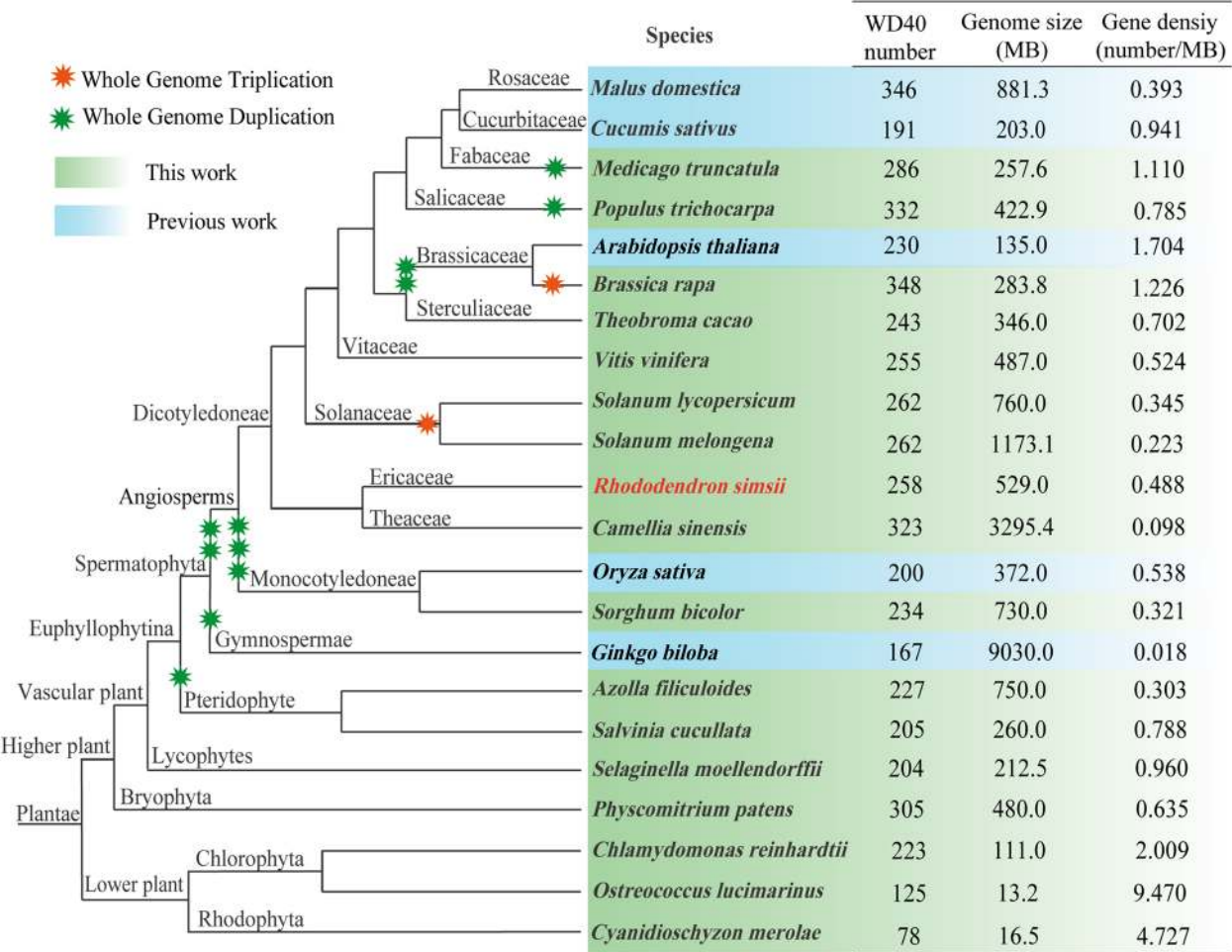


Fig. 3 The evolutionary relationships of the 22 plant species and the specific information of the *WD40* gene family in each genome. The left side of the figure shows the evolutionary relationships of the species, the right side shows the number of *WD40* gene family members in each species. Green data indicate information described in this work, while blue data indicate information previously published

OsWD40, *VvWD40* or *MdWD40* genes can be traced back to a common ancestor. The orthologous relationships varied, with some involving one *RsWD40* gene corresponding to one ortholog gene (e.g., *RsWD40-11*/AT3G45280 and *RsWD40-127*/LOC_Os01g72220), while others involved one *RsWD40* corresponding to multiple ortholog genes (e.g., *RsWD40-171*/AT5G24320/AT5G53500/AT5G54200 and *RsWD40-17*/LOC_Os02g50740/LOC_Os06g13140). These results suggest that duplication events have likely played a significant role in the evolution and functional diversification of the *WD40* family.

Expression profiling of *RsWD40* genes in three *Rhododendron* varieties

In this study, we utilized RNA-seq data to examine the expression profiles of *RsWD40* genes in three *Rhododendron* cultivars, aiming to identify *WD40* genes associated with anthocyanin biosynthesis in *R. simsii*. We identified

differentially expressed genes (DEGs) with absolute |fold change| ≥ 2 (log2 |fold change| ≥ 1) and p-value < 0.05 between the two *Rhododendron* varieties. Using this approach, we identified 24 DEGs between red flower (RF) and white flower (WF) varieties, and 24 DEGs between pink flower (PF) and white flower (WF) varieties, at the bud stage. By combining these two sets of DEGs, we aimed to identify candidate *RsWD40*s involved in anthocyanin biosynthesis. Our analysis revealed the up-regulation of 5 DEGs (*RsWD40-13*, 55, 76, 94, and 208) in colored-flower (PF and RF) varieties. This suggests that these genes may positively regulate anthocyanin production at the bud stage. On the other hand, the down-regulation of 5 DEGs (*RsWD40-64*, 101, 130, 180, and 216) in colored-flower (PF and RF) varieties indicates their potential involvement in negatively regulating anthocyanin biosynthesis at the bud stage (Fig. 6, Table S7 and S8). In addition, we conducted a similar comparative analysis at the full bloom stage. We

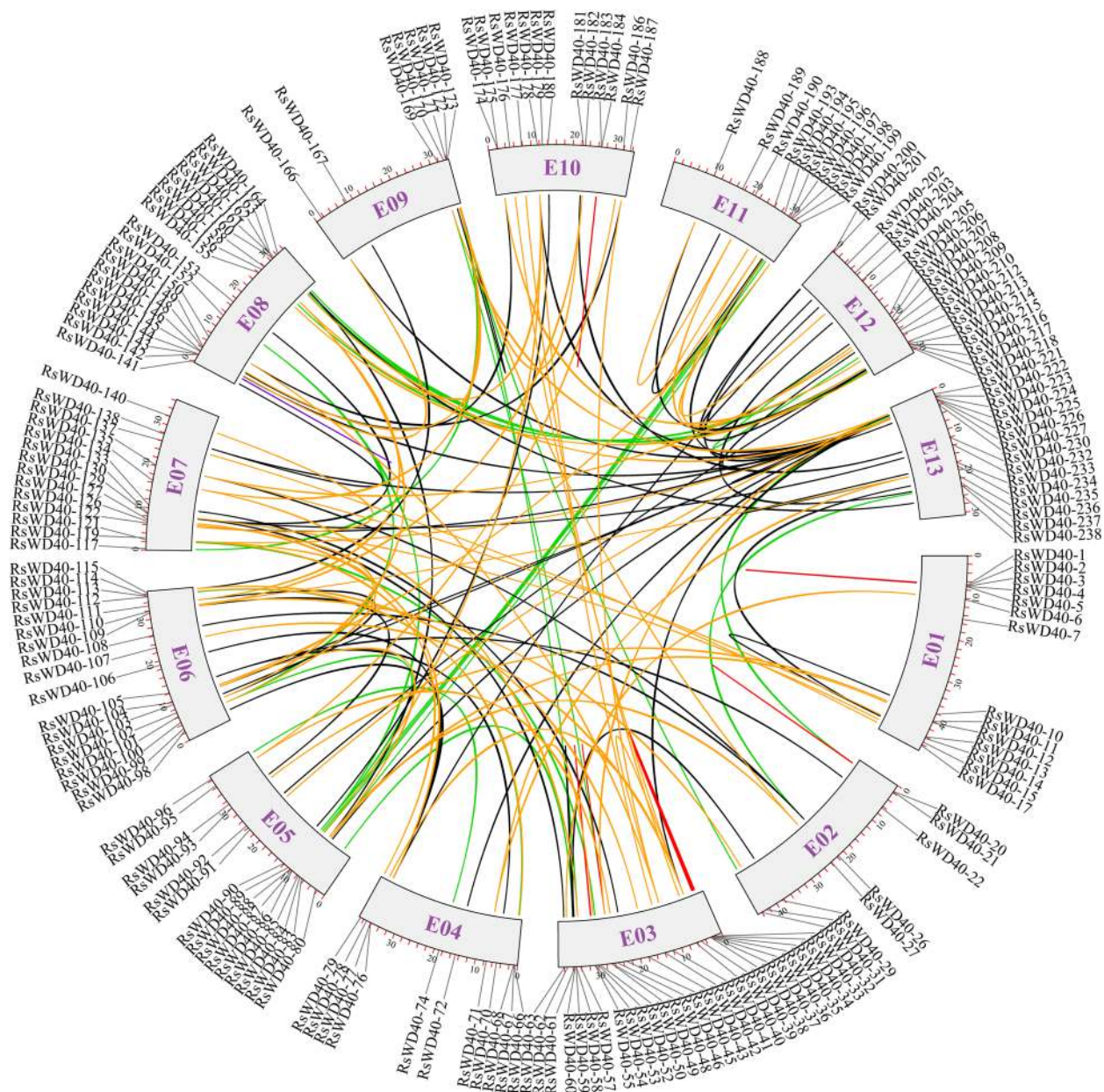


Fig. 4 Genomic locations of *RsWD40s* and all duplication gene pairs in the *R. simsii* genome. Duplicated *RsWD40* gene pairs were indicated by the colored lines: WGD duplication pairs (green), tandem duplication pairs (purple), proximal duplication pairs (red), transposed duplication pairs (black), dispersed duplication pairs (orange). Color boxes with a number inside it represents chromosomes

identified 3 candidate genes (*RsWD40-55*, *119*, and *233*) that may play a role in positively regulating anthocyanin production. Furthermore, we found 4 candidate genes (*RsWD40-64*, *139*, *211*, and *234*) that could potentially be involved in negatively regulating anthocyanin biosynthesis (Fig. 7, Tables S9 and S10).

Validation of RNA-Seq-based gene expression

To confirm the accuracy of the RNA-seq results, we selected and examined 15 candidate genes related to

anthocyanin biosynthesis at the bud and full bloom stages using qRT-PCR. It is important to note that the consistency between qRT-PCR and RNA-seq results may vary for different genes. Previous studies, such as Everaert et al. [72], have demonstrated that qRT-PCR and RNA-seq results are not always perfectly aligned. Approximately 85% of genes exhibited consistent expression patterns between the two methods. In our study, despite the disparities in fold change values between the RNA-seq and qRT-PCR data, the

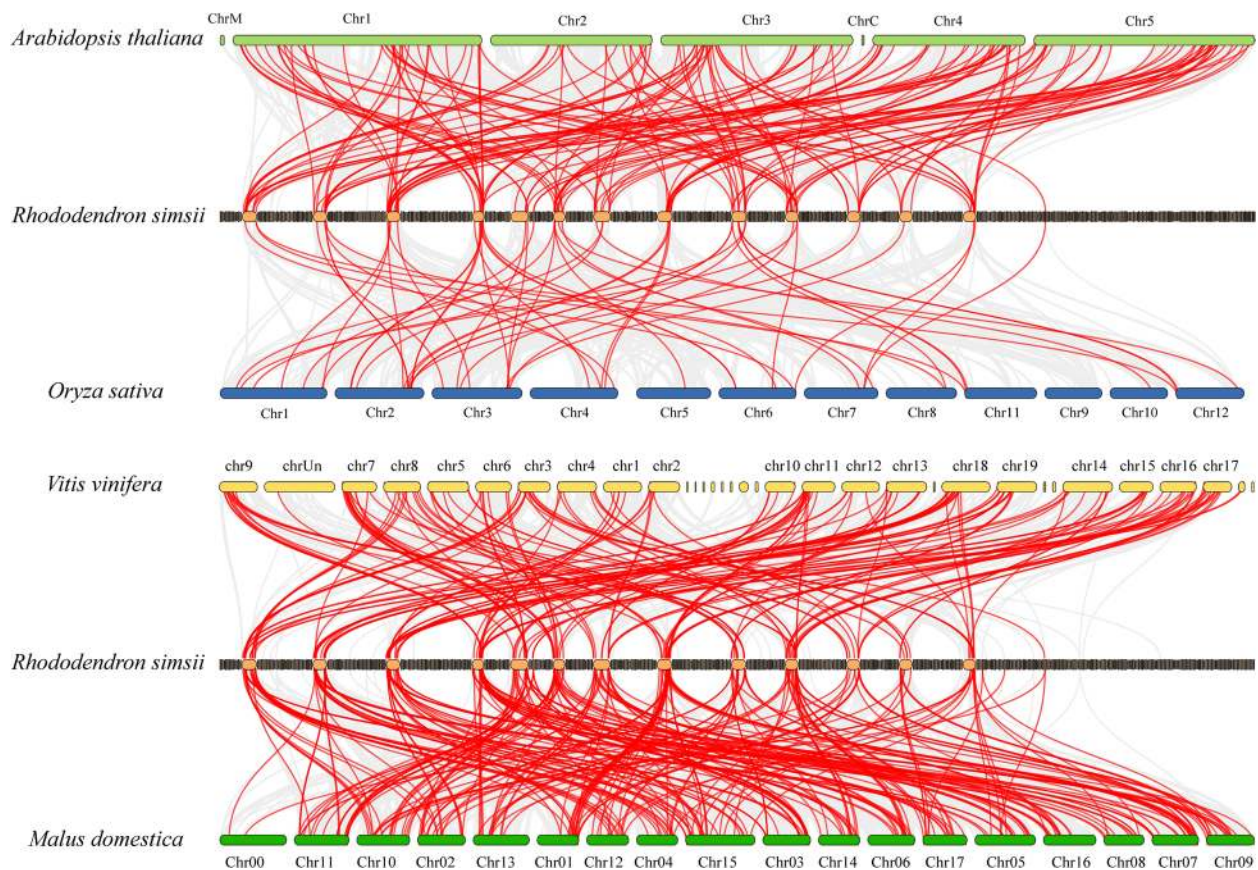


Fig. 5 Comparative physical mapping displaying the orthologous relationships of *WD40* genes among *R. simsii* and other four species (*A. thaliana*, *O. sativa*, *V. vinifera* and *M. domestica*)

overall expression trends remained consistent. For instance, based on the RNA-seq data, the expression of *RsWD40-76* was 8.0 and 5.9 times higher in red and pink flowers, respectively, compared to white flowers at the bud stage. However, in the qRT-PCR tests, the corresponding fold change values were 30.1 and 16.8. Additionally, some genes (*RsWD40-55*, *180*, *216*, and *101*) did not precisely follow the RNA-seq trend. Nevertheless, when comparing their expression levels between white flowers and colored flowers (including red and pink), both RNA-seq and qRT-PCR data demonstrated consistent results, despite potential differences in fold changes for red and pink flowers. In conclusion, the qRT-PCR results aligned well with the RNA-seq data in terms of the observed expression trends (Table S11, Fig. 8).

GO enrichment analysis of candidate *RsWD40* genes

In order to comprehensively understand the biological function of the candidate *RsWD40* genes, we conducted gene ontology (GO) annotation and enrichment analysis

on the 15 candidate genes (*RsWD40-13*, *RsWD40-55*, *RsWD40-64*, *RsWD40-76*, *RsWD40-94*, *RsWD40-101*, *RsWD40-119*, *RsWD40-130*, *RsWD40-139*, *RsWD40-180*, *RsWD40-208*, *RsWD40-211*, *RsWD40-216*, *RsWD40-233*, and *RsWD40-234*) investigated in this study. Our analysis revealed significant enrichment for 13 GO terms, comprising of 2 molecular function terms, 7 cellular component terms, and 4 biological process terms (Fig. 9, Table S12). The two enriched molecular function terms were "binding" and "protein binding", indicating the involvement of these candidate genes in binding interactions. The four enriched biological process terms were "regulation of biological process", "biological regulation", "regulation of cellular process", and "cellular component organization or biogenesis", suggesting their roles in regulating various cellular processes. Notably, in the cellular component category, the most enriched GO terms were "Cul4-RING E3 ubiquitin ligase complex", "cullin-RING ubiquitin ligase complex", and "ubiquitin ligase complex". This finding suggests that ubiquitination process may play an important role in anthocyanin biosynthesis in *R. simsii*.

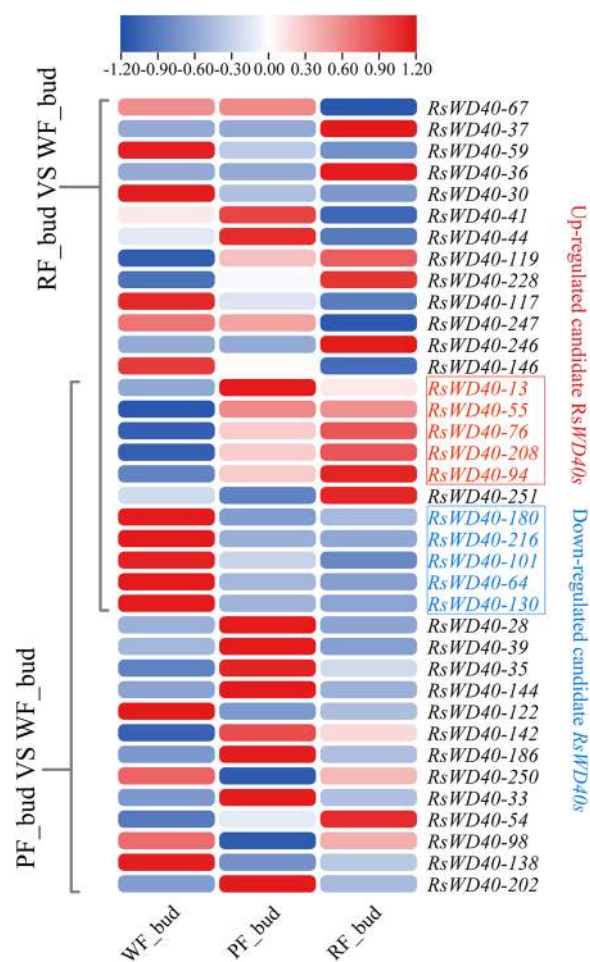


Fig. 6 Identification of candidate *RsWD40s* involved in anthocyanin biosynthesis across three different flower colors (white, red, and pink) in *Rhododendron* varieties at the bud stage. The heatmap shows the relative expression values of *RsWD40s*, with WF indicating white-flowering variety, RF indicating red-flowering variety, and PF indicating pink-flowering variety

Discussion

The grouping of gene families is crucial because genes within the same group often exhibit structural and functional similarities, making it easier to predict gene function. Previous studies have employed two approaches to group WD40 TFs families: by their domain composition or through phylogenetic analysis (see 'Introduction' section). The former approach involves grouping WD40 proteins into subfamilies based on their domain compositions. For instance, WD40 proteins from various species, including potato, wheat, apple, and *Ginkgo biloba*, have been classified into different subfamilies (ranging from 10 to 15) based on domain composition [7, 9, 12, 13]. However, some of these studies have classified WD40 proteins with different domain compositions into

the same group, referred to as "other subfamilies", suggesting the possibility of further subfamily divisions. In our study, we conducted a more comprehensive grouping of *RsWD40* proteins and identified 42 subfamilies. These results suggest that plant WD40 proteins contain a variety of domains in addition to the WD40 repeat. The second method involves grouping WD40 proteins based on phylogenetic analysis. Previous studies on WD40 proteins have reported low bootstrap values in phylogenetic trees for several species, including *Arabidopsis thaliana* [6], cucumber [6], *Setaria italica* [11], and potato [13], indicating poor support for many cluster assignments. Consequently, the grouping of WD40 proteins by phylogenetic analysis has exhibited significant variability among studies, even within the same species. For example, WD40 in *Arabidopsis* have been identified into 32, 14, 11, and 5 clusters in different studies [6, 13, 15, 73]. Here, we adopted a phylogenetic tree-based approach as previously described by Li et al. [6], which excluded genes with low support values to avoid incorrect assignments. As a result, we identified 47 distinct clusters for *RsWD40s*. It's important to note that the number of WD40 repeats in the WD40 protein can vary [74]. In our study, *RsWD40* proteins displayed a range of 1 to 12 WD40 repeats. The variable number of WD40 repeats and the diversity of other domains within the WD40 family members likely contribute to reduced sequence conservation, which could account for the low bootstrap values observed in the phylogenetic tree.

To explore the evolution of the WD40 gene family, we investigated the WD40 gene family in *R. simsii* and 21 other species, including 18 higher plants and 3 lower plants. Our comparative genomic analysis revealed that single-celled *C. merolae*, *O. lucimarinus* and *C. reinhardtii* have 78, 125 and 223 WD40 genes, respectively. The abundance of WD40 genes present in lower plants suggests their involvement in fundamental cellular processes essential for plant survival. This may have aided lower plants in adapting to early extreme environments. Interestingly, the number of WD40 genes in green algae (223) was found to be higher than that in higher plants such as rice (200), *C. sativus* (191), and *S. cucullata* (205). This suggests that no large-scale amplification of WD40 genes occurred in higher plants and that the expansion of WD40 gene family predates the divergence of green algae and higher plants. We also calculated the gene density of WD40 genes in different plant species and observed that lower plants had a high gene density. Starting from *P. patens*, the gene density of WD40 decreased and remained relatively stable in higher plants. *P. patens* is often used in evolutionary analyses as it allows for the reconstruction of genomic changes related to the conquest of land by comparing it with aquatic algae and vascular plants

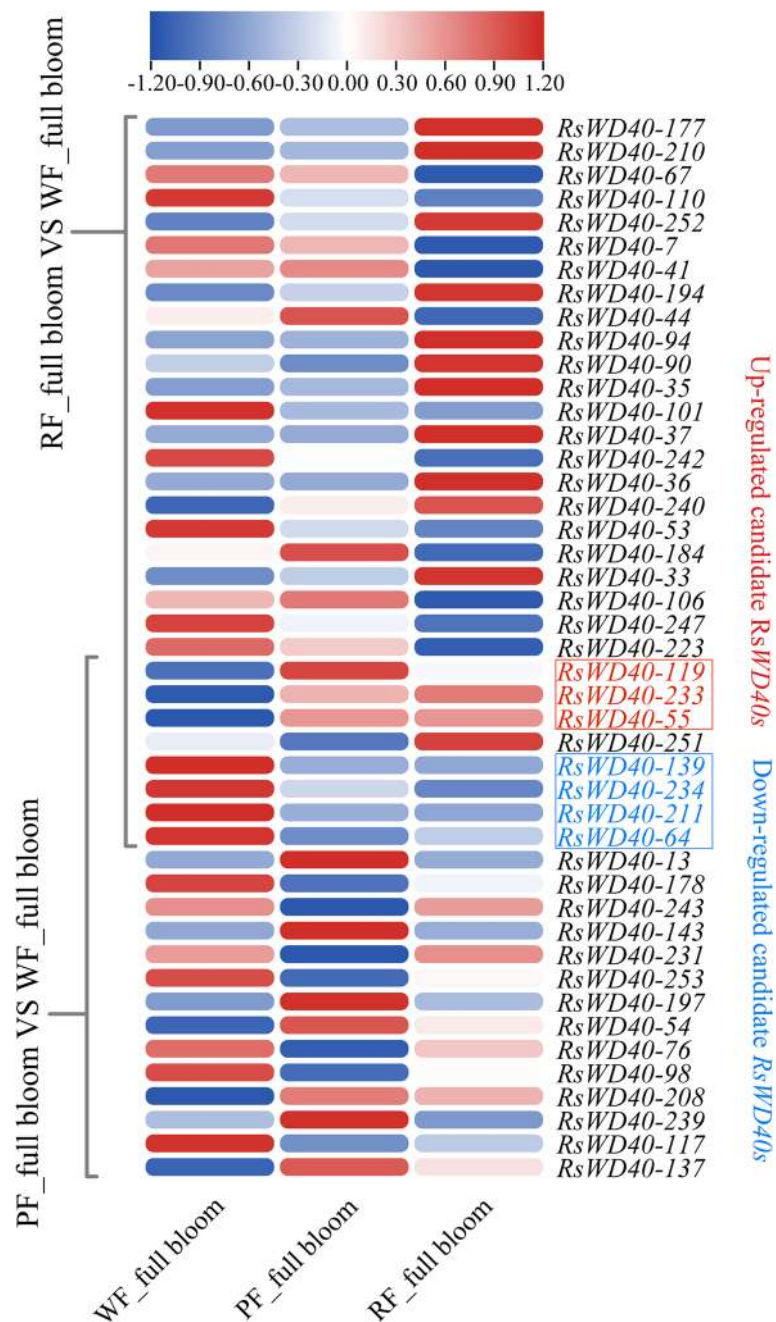


Fig. 7 Identification of candidate *RsWD40s* involved in anthocyanin biosynthesis across three different flower colors (white, red, and pink) in *Rhododendron* varieties at the full bloom stage. The heatmap shows the relative expression values of *RsWD40s*, with WF indicating white-flowering variety, RF indicating red-flowering variety, and PF indicating pink-flowering variety

[75]. The reduced gene density of *WD40* in higher plants suggests that *WD40* gene loss occurred during evolution of these species. This indicates that *WD40* genes have undergone diversification over time and have acquired new functions in different plant lineages. For example, some *WD40* genes have been found to regulate flowering and pollen tube growth in higher plants [27, 76]. These

functions may have emerged later in plant evolution, contributing to the diversification of plant lineages. Furthermore, we observed that *G. biloba* displayed an unusually low density of *WD40* genes compared to other species. This phenomenon can be attributed to the expansion of the *G. biloba* genome, which resulted in a significant increase in intron size primarily due to the insertion of

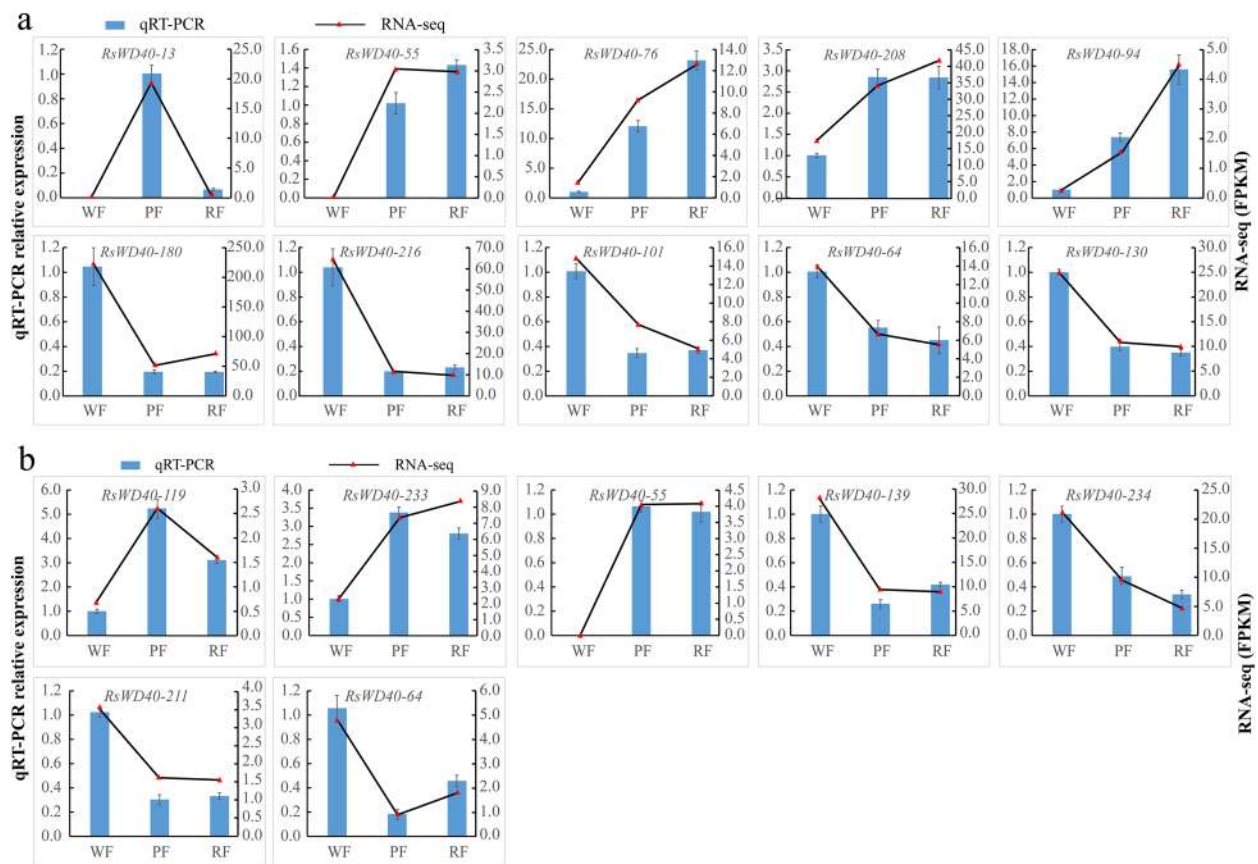


Fig. 8 qRT-PCR validation of 15 candidate *RsWD40* genes at the bud stage (a) and full bloom stage (b). WF, PF, and RF represent white-flowering variety, pink-flowering variety, the red-flowering variety, respectively. Data represent the mean of three biological replicates $\pm$ standard error of the mean. Standard errors are shown as bars above the columns

long terminal repeats rather than recent whole-genome duplication events [57].

Gene duplication plays a crucial role in the rapid expansion and evolution of gene families [77]. Such duplication events can be classified into five categories: WGD, tandem duplication, proximal duplication, transposed duplication, and dispersed duplication [63]. The different patterns of gene duplication contribute differentially to the expansion of specific gene families in plant genomes [78, 79]. For example, some gene families, including *WRKY*, *bHLH*, and *bZIP*, are more likely to expand through WGD and tandem duplications [80–82], while others such as *NBS-LRR* and *ERF* expand primarily through transposed duplication [83–85]. Our research on the *WD40* gene family in *R. simsii* suggested that transposed duplication was the primary cause of its expansion, with 33.7% of the *RsWD40* genes duplicated and retained through transposed duplication. Duplication of genes can lead to diverse results such as pseudogenization

(function loss), subfunctionalization, and neofunctionalization [86]. At both bud and full bloom stages, the gene pair *RsWD40-20/RsWD40-223*, resulting from WGD, exhibited different levels of expression. This suggests that duplicated genes may remain stable if they have different functional properties, which have been shaped by natural selection during the process of evolution. We also analyzed the rate of synonymous (K_s) to non-synonymous (K_a) mutation for all duplication gene pairs. Typically, genes are subject to three types of selection: purifying selection ($K_a/K_s < 1$), positive selection ($K_a/K_s > 1$), and neutral selection ($K_a/K_s = 1$) [87]. Our study observed that the K_a/K_s ratio of almost all *RsWD40* gene pairs was less than 1, indicating that these genes are mostly affected by purifying selection. Only one gene pair had a K_a/K_s ratio greater than 1, suggesting positive selection for advantageous mutations. Genomic comparison is an efficient means of transferring genomic knowledge of a well-studied taxonomic unit to a less studied one [88].

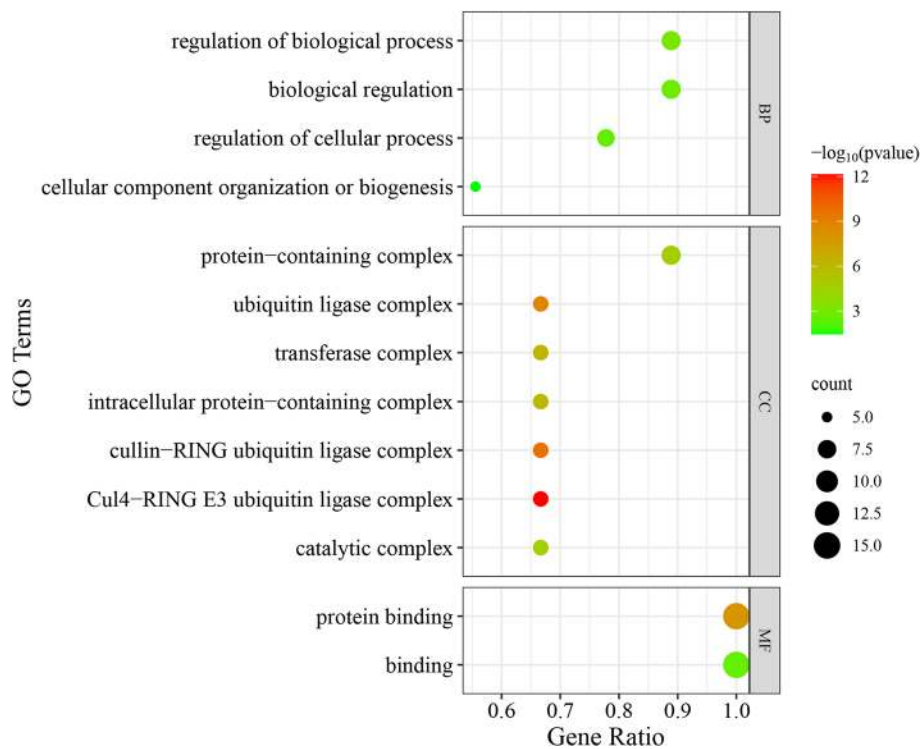


Fig. 9 GO enrichment of the candidate *RsWD40* Genes. GO was performed with three main categories: molecular function (MF), cellular component (CC) and biological process (BP). GO terms with P value < 0.05 were identified as significant

Our synteny analysis revealed 164 orthologous gene pairs between *R. simsii* and *A. thaliana*, 55 pairs with *O. sativa*, 190 pairs with *V. vinifera*, and 306 pairs with *M. domestica* (Table S6). This indicates that these gene pairs exhibit orthologous relationships, share common ancestors, have been conserved throughout evolution, and likely perform similar functions. The *WD40* genes identified in *A. thaliana*, *O. sativa*, *V. vinifera*, and *M. domestica* will serve as a reference for future studies of *RsWD40* genes.

Flower color is a crucial trait in plant breeding, and identifying genes associated with anthocyanin synthesis is essential for successful flower color breeding. The *WD40* gene *TTG1* in *A. thaliana* is well characterized for its involvement in anthocyanin synthesis, as discussed in the 'Introduction' section. However, *A. thaliana* lacks pigmentation, making it less suitable for studying anthocyanin synthesis. While some *TTG1* homologs have been studied in other species [44–48], a comprehensive genome-wide screening of *WD40* genes involved in anthocyanin synthesis in species with rich coloration has not been conducted. Therefore, in this study, we conducted a genome-wide screen for *WD40* genes involved in anthocyanin synthesis in *R. simsii* using RNA-seq data. The results of our study suggest that 10 candidate *RsWD40* genes may be involved in anthocyanin synthesis

during the bud stage, and 7 candidate *RsWD40* genes may be involved in anthocyanin synthesis during the bloom stage. Interestingly, there was little overlap between these two groups of genes, with only *RsWD40-55* and *RsWD40-64* shared between them. This finding indicates that different *RsWD40* genes may have distinct roles in anthocyanin biosynthesis during different flowering stages. Previous studies have reported the involvement of the E3 ubiquitin ligase-mediated ubiquitin pathway in anthocyanin biosynthesis in various plants [89–94]. Our GO enrichment analysis revealed that terms related to the "E3 ubiquitin ligase complex" were the most enriched GO terms, and a total of 6 candidate genes (*RsWD40-234*, *RsWD40-180*, *RsWD40-130*, *RsWD40-101*, *RsWD40-208*, and *RsWD40-233*) were associated with these terms. These results suggest that the ubiquitination process may also play an important role in anthocyanin biosynthesis in *R. simsii*. In *R. simsii*, the homologous gene for *AtTTG1* (AT5G24520) is *RsWD40-172* (Rhsim09G0209900) (they are orthologous gene pair, Table S6). However, its expression did not significantly differ across all three floral colors during the budding and blooming stages. This suggests that *RsTTG1* may not be responsible for anthocyanin biosynthesis in *Rhododendron*, and other *RsWD40* genes may have taken on its

function. Most of the up-regulated candidate *RsWD40* genes (7 of 8, Tables S9 and S11) identified in this study contained only WD40 repeat and had at least five WD40 repeats. These findings are consistent with the structural features of the AtTTG1 protein, implying a potential functional similarity between the candidate *RsWD40* genes and AtTTG1. However, further research is needed to confirm the potential involvement of the 15 candidate *RsWD40* genes related to anthocyanin biosynthesis.

Conclusion

In conclusion, this study provides a comprehensive analysis of the WD40 transcription factor (TF) family in the *R. simsii* genome. We classified 258 WD40 TFs into 42 subfamilies and 47 clusters based on their domain compositions and phylogenetic analysis. The distribution of *RsWD40* genes was found to be uneven across the 13 chromosomes. Comparative genomic analysis suggests that the expansion of the WD40 gene family occurred early in plant evolution, before the divergence of green algae and higher plants. Transposed duplication was identified as a major mechanism driving the expansion of the *RsWD40* gene family. Our analysis of selective pressure indicated that most duplication gene pairs underwent purifying selection. Through synteny analysis, we identified numerous orthologous gene pairs between *R. simsii* and other plant species, including *A. thaliana*, *O. sativa*, *V. vinifera*, and *M. domestica*. These findings highlight the conservation of certain genes and their potential functional similarities across different plant lineages. Furthermore, by utilizing RNA-seq data, we investigated the expression patterns of *RsWD40* genes and identified candidate genes potentially involved in anthocyanin biosynthesis during the bud and full bloom stages. GO enrichment analysis of these candidate genes revealed the potential involvement of the ubiquitination process in anthocyanin biosynthesis in *R. simsii*. Overall, this study enhances our understanding of the *RsWD40* family and provides valuable insights into the molecular mechanisms underlying anthocyanin biosynthesis in *Rhododendron* species.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09604-x>.

Additional file 1: Figure S1. The plant materials employed in this study, depicted from left to right, are *Rhododendron wardii* var. Puralbum, *Rhododendron simsii* Planch, and *Rhododendron hybridum* Ker Gawl.

Additional file 2: Figure S2. Proportion of genes originating from different replication events.

Additional file 3: Table S1. The primer sequences used for qRT-PCR. **Table S2.** The basic information of WD40 genes identified in *Rhododendron simsii*. **Table S3.** Grouping of *RsWD40* proteins based on domain compositions. **Table S4.** The WD40 genes in typical species. **Table S5.**

Gene duplications of WD40 genes in *R. simsii* with outlier Ka/Ks values.

Table S6. The orthologous relationships of the WD40 genes among *Rhododendron simsii* and other four species (*Arabidopsis thaliana*, *Oryza sativa*, *Vitis vinifera* and *Malus domestica*). **Table S7.** The expression of the *RsWD40* at the bud stage. The gene expression levels were calculated by using reads per kilobase per million (RPKM) measure. **Table S8.** The differentially expressed genes at the bud stage. **Table S9.** The expression of the *RsWD40* at the full bloom stage. The gene expression levels were calculated by using reads per kilobase per million (FPKM) measure. **Table S10.** The differentially expressed genes at the full bloom stage. **Table S11.** Relative expression of candidate *RsWD40* genes at the bud and full bloom stages, as determined by qRT-PCR. **Table S12.** The gene ontology (GO) analysis of candidate *RsWD40* genes.

Acknowledgements

The authors would like to acknowledge the contributions and support of Professor Xiaojing Wang from Guizhou University, who provided the plant materials necessary for this study. The authors would also like to express their gratitude to the journal editors and reviewers for their valuable comments and suggestions during the review process. Their feedback greatly contributed to improving the quality and clarity of this manuscript.

Authors' contributions

C.W. and A.L. designed the experiment and wrote the manuscript. Y.T. and J.L. performed the experiment. Y.L. and C.H. analyzed the data. C.W. and A.L. proofread the manuscript. All authors reviewed and approved the final version of the manuscript for publication.

Funding

This research was funded by Scientific Research Project of Education Department of Hubei Province (Q20222703); Scientific Research Project of Education Department of Hubei Province (B2022147); Natural Science Foundation of Hubei Province in China (2020CFB855).

Availability of data and materials

The RNA-seq data used in this study have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database under the identifier PRJNA952839. The data are publicly available at <http://www.ncbi.nlm.nih.gov/sra>. The materials used in this research, including plant samples, are available upon request from the corresponding author.

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 for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Key Laboratory for Quality Control of Characteristic Fruits and Vegetables of Hubei Province, College of Life Science and Technology, Hubei Engineering University, Xiaogan 432000, China. ²Department of Biology and Chemical Engineering, Weihai Vocational College, Weihai 264200, China. ³Institute of Quality Standard and Testing Technology for Agro-Products, Hubei Academy of Agricultural Science, Wuhan 430064, China.

Received: 26 April 2023 Accepted: 19 August 2023
Published online: 26 August 2023

References

- Stirnemann CU, Petsalaki E, Russell RB, Muller CW. WD40 proteins propel cellular networks. *Trends Biochem Sci.* 2010;35(10):565–74.
- Neer EJ, Schmidt CJ, Nambudripad R, Smith TF. The ancient regulatory-protein family of WD-repeat proteins. *Nature.* 1994;371(6495):297–300.
- Smith TF, Gaitatzes C, Saxena K, Neer EJ. The WD repeat: a common architecture for diverse functions. *Trends Biochem Sci.* 1999;24(5):181–5.
- Wu XH, Chen RC, Gao Y, Wu YD. The effect of Asp-His-Ser/Thr-Trp tetrad on the thermostability of WD40-repeat proteins. *Biochemistry.* 2010;49(47):10237–45.
- Ouyang Y, Huang X, Lu Z, Yao J. Genomic survey, expression profile and co-expression network analysis of *OsWD40* family in rice. *BMC Genomics.* 2012;13(1):100.
- Li Q, Zhao P, Li J, Zhang C, Wang L, Ren Z. Genome-wide analysis of the WD-repeat protein family in cucumber and *Arabidopsis*. *Mol Genet Genomics.* 2014;289(1):103–24.
- Zheng J, Yongling L, Feng X, Xian Z, Jiabao Y, Mingyue F, Xiaomeng L, Zhengyan C, Zhang W. Genome-wide identification of WD40 superfamily genes and prediction of WD40 gene of flavonoid-related genes in *Ginkgo biloba*. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca.* 2021;49(2):12086–12086.
- Fan Z, Zhai Y, Wang Y, Zhang L, Song M, Flaishman MA, Ma H. Genome-Wide Analysis of Anthocyanin Biosynthesis Regulatory WD40 Gene *FcTTG1* and Related Family in *Ficus carica* L. *Front Plant Sci.* 2022;13: 948084.
- Hu R, Xiao J, Gu T, Yu X, Zhang Y, Chang J, Yang G, He G. Genome-wide identification and analysis of WD40 proteins in wheat (*Triticum aestivum* L.). *BMC Genomics.* 2018;19:1–13.
- Sun YB, Zhang XJ, Zhong MC, Dong X, Yu DM, Jiang XD, Wang D, Cui WH, Chen JH, Hu JY. Genome-wide identification of WD40 genes reveals a functional diversification of COP1-like genes in Rosaceae. *Plant Mol Biol.* 2020;104(1–2):81–95.
- Mishra AK, Muthamilarasan M, Khan Y, Parida SK, Prasad M. Genome-wide investigation and expression analyses of WD40 protein family in the model plant foxtail millet (*Setaria italica* L.). *PLoS One.* 2014;9(1):e86852.
- Zhang B, Qu D, Yang H, Long X, Zhu Z, Yang Y, Zhao Z. Genome wide identification and characterization of apple WD40 proteins and expression analysis in response to ABA, drought, and low temperature. *Horticulturae.* 2022;8(2):141.
- Liu Z, Liu Y, Coulter JA, Shen B, Li Y, Li C, Cao Z, Zhang J. The WD40 gene family in potato (*Solanum Tuberosum* L.): genome-wide analysis and identification of anthocyanin and drought-related WD40s. *Agronomy.* 2020;10(3):401.
- Tan L, Salih H, Htet NNW, Azeem F, Zhan R. Genomic analysis of WD40 protein family in the mango reveals a TTG1 protein enhances root growth and abiotic tolerance in *Arabidopsis*. *Sci Rep.* 2021;11(1):2266.
- Chen C, Yang Y, Pan L, Xia W, Xu L, Hua B, Zhang Z, Miao M. Genome-Wide Identification of WD40 Proteins in *Cucurbita maxima* Reveals Its Potential Functions in Fruit Development. *Genes (Basel).* 2023;14(1):220.
- Jain BP, Pandey S. WD40 Repeat Proteins: Signalling Scaffold with Diverse Functions. *Protein J.* 2018;37(5):391–406.
- Dayebgadh G, Sardi ME, Florens L, Washburn MP. Biochemical Reduction of the Topology of the Diverse WDR76 Protein Interactome. *J Proteome Res.* 2019;18(9):3479–91.
- Kevei Z, Balaban M, Da Ines O, Tiricz H, Kroll A, Regulski K, Mergaert P, Kondorosi E. Conserved CDC20 cell cycle functions are carried out by two of the five isoforms in *Arabidopsis thaliana*. *PLoS ONE.* 2011;6(6): e20618.
- Caillaud MC, Paganelli L, Lecomte P, Deslandes L, Quentin M, Pecrix Y, Le Bris M, Marfaing N, Abad P, Favory B. Spindle assembly checkpoint protein dynamics reveal conserved and unsuspected roles in plant cell division. *PLoS ONE.* 2009;4(8): e6757.
- Saedler R, Jakoby M, Marin B, Galiana-Jaime E, Hulskamp M. The cell morphogenesis gene *SPRRIG* in *Arabidopsis* encodes a WD/BEACH domain protein. *Plant J.* 2009;59(4):612–21.
- Higa LA, Zhang H. Stealing the spotlight: CUL4-DDB1 ubiquitin ligase docks WD40-repeat proteins to destroy. *Cell Div.* 2007;2:5.
- Wakasugi M, Kawashima A, Morioka H, Linn S, Sancar A, Mori T, Nikaido O, Matsunaga T. DDB accumulates at DNA damage sites immediately after UV irradiation and directly stimulates nucleotide excision repair. *J Biol Chem.* 2002;277(3):1637–40.
- Chen R-H, Miettinen PJ, Maruoka EM, Choy L, Derynck R. A WD-domain protein that is associated with and phosphorylated by the type II TGF- β receptor. *Nature.* 1995;377(6549):548–52.
- Datta PK, Moses HL. STRAP and Smad7 synergize in the inhibition of transforming growth factor β signaling. *Molecular cellular biology.* 2000;20(9):3157–67.
- Jiang D, Kong NC, Gu X, Li Z, He Y. *Arabidopsis* COMPASS-like complexes mediate histone H3 lysine-4 trimethylation to control floral transition and plant development. *PLoS Genet.* 2011;7(3): e1001330.
- Pazhouhandeh M, Molinier J, Berr A, Genschik P. MSI4/FVE interacts with CUL4-DDB1 and a PRC2-like complex to control epigenetic regulation of flowering time in *Arabidopsis*. *Proc Natl Acad Sci.* 2011;108(8):3430–5.
- Yu D, Dong X, Zou K, Jiang XD, Sun YB, Min Z, Zhang LP, Cui H, Hu JY. A hidden mutation in the seventh WD40-repeat of COP1 determines the early flowering trait in a set of *Arabidopsis myc* mutants. *Plant Cell.* 2023;35(1):345–50.
- Liu WC, Zheng SQ, Yu ZD, Gao X, Shen R, Lu YT. WD 40-REPEAT 5a represses root meristem growth by suppressing auxin synthesis through changes of nitric oxide accumulation in *Arabidopsis*. *Plant J.* 2018;93(5):883–93.
- Qi T, Huang H, Wu D, Yan J, Qi Y, Song S, Xie D. *Arabidopsis* DELLA and JAZ proteins bind the WD-repeat/bHLH/MYB complex to modulate gibberellin and jasmonate signaling synergy. *Plant Cell.* 2014;26(3):1118–33.
- Zhu J, Jeong JC, Zhu Y, Sokolchik I, Miyazaki S, Zhu JK, Hasegawa PM, Bohnert HJ, Shi H, Yun DJ, et al. Involvement of *Arabidopsis* HOS15 in histone deacetylation and cold tolerance. *Proc Natl Acad Sci U S A.* 2008;105(12):4945–50.
- Biedermann S, Hellmann H. The DDB1a interacting proteins ATCSA-1 and DDB2 are critical factors for UV-B tolerance and genomic integrity in *Arabidopsis thaliana*. *Plant J.* 2010;62(3):404–15.
- Zhang D, Wang Y, Shen J, Yin J, Li D, Gao Y, Xu W, Liang J. OsRACK1A, encodes a circadian clock-regulated WD40 protein, negatively affect salt tolerance in rice. *Rice (N Y).* 2018;11(1):45.
- Walker AR, Davison PA, Bolognesi-Winfield AC, James CM, Srinivasan N, Blundell TL, Esch JJ, Marks MD, Gray JC. The *TRANSPARENT TESTA GLABRA1* locus, which regulates trichome differentiation and anthocyanin biosynthesis in *Arabidopsis*, encodes a WD40 repeat protein. *Plant Cell.* 1999;11(7):1337–49.
- Shirley BW, Kubasek WL, Storz G, Bruggemann E, Koornneef M, Ausubel FM, Goodman HM. Analysis of *Arabidopsis* mutants deficient in flavonoid biosynthesis. *Plant J.* 1995;8(5):659–71.
- Xu W, Dubos C, Lepiniec L. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci.* 2015;20(3):176–85.
- Qing L, Aoyama T. Pathways for epidermal cell differentiation via the homeobox gene *GLABRA2*: update on the roles of the classic regulator F. *J Integr Plant Biol.* 2012;54(10):729–37.
- Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant J.* 2004;39(3):366–80.
- Gonzalez A, Zhao M, Leavitt JM, Lloyd AM. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant J.* 2008;53(5):814–27.
- Petroni K, Tonelli C. Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Sci.* 2011;181(3):219–29.
- Xu W, Grain D, Bobet S, Le Gourrierec J, Thévenin J, Kelemen Z, Lepiniec L, Dubos C. Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in *Arabidopsis* seed. *New Phytol.* 2014;202(1):132–44.
- Zhou L-L, Shi M-Z, Xie D-Y. Regulation of anthocyanin biosynthesis by nitrogen in TTG1-GL3/TT8-PAP1-programmed red cells of *Arabidopsis thaliana*. *Planta.* 2012;236:825–37.
- Shi MZ, Xie DY. Biosynthesis and metabolic engineering of anthocyanins in *Arabidopsis thaliana*. *Recent Pat Biotechnol.* 2014;8(1):47–60.
- Deng Y, Lu S. Biosynthesis and regulation of phenylpropanoids in plants. *Crit Rev Plant Sci.* 2017;36(4):257–90.
- Liu Y, Hou H, Jiang X, Wang P, Dai X, Chen W, Gao L, Xia T. 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(6):1686.
- Li S, Wu Y, Kuang J, Wang H, Du T, Huang Y, Zhang Y, Cao X, Wang Z. SmMYB111 is a key factor to phenolic acid biosynthesis and interacts

- with both SmTTG1 and SmbHLH51 in *Salvia miltiorrhiza*. J Agric Food Chem. 2018;66(30):8069–78.
46. Jiu S, Guan L, Leng X, Zhang K, Haider MS, Yu X, Zhu X, Zheng T, Ge M, Wang C, et al. The role of VvMYBA2r and VvMYBA2w alleles of the MYBA2 locus in the regulation of anthocyanin biosynthesis for molecular breeding of grape (*Vitis* spp.) skin coloration. Plant Biotechnol J. 2021;19(6):1216–1239.
 47. Yang X, Wang J, Xia X, Zhang Z, He J, Nong B, Luo T, Feng R, Wu Y, Pan Y, et al. OsTTG1, a WD40 repeat gene, regulates anthocyanin biosynthesis in rice. Plant J. 2021;107(1):198–214.
 48. Jia N, Wang J, Wang Y, Ye W, Liu J, Jiang J, Sun J, Yan P, Wang P, Wang F, et al. The Light-Induced WD40-Repeat Transcription Factor DcTTG1 Regulates Anthocyanin Biosynthesis in *Dendrobium candidum*. Front Plant Sci. 2021;12: 633333.
 49. Yang FS, Nie S, Liu H, Shi TL, Tian XC, Zhou SS, Bao YT, Jia KH, Guo JF, Zhao W, et al. Chromosome-level genome assembly of a parent species of widely cultivated azaleas. Nat Commun. 2020;11(1):5269.
 50. Yan LJ, Liu J, Möller M, Zhang L, Zhang XM, Li DZ, Gao LM. DNA barcoding of *Rhododendron* (Ericaceae), the largest Chinese plant genus in biodiversity hotspots of the Himalaya-Hengduan Mountains. Mol Ecol Resour. 2015;15(4):932–44.
 51. Du H, Lai L, Wang F, Sun W, Zhang L, Li X, Wang L, Jiang L, Zheng Y. Characterisation of flower colouration in 30 *Rhododendron* species via anthocyanin and flavonol identification and quantitative traits. Plant Biol (Stuttgart). 2018;20(1):121–9.
 52. Wang S, Huang S, Yang J, Li Z, Zhang M, Fang Y, Yang Q, Jin W. Metabolite profiling of violet, white and pink flowers revealing flavonoids composition patterns in *Rhododendron pulchrum* Sweet. J Biosci. 2021;46:1–10.
 53. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011;7(10): e1002195.
 54. Letunic I, Khedkar S, Bork P. SMART: recent updates, new developments and status in 2020. Nucleic Acids Res. 2021;49(D1):D458–60.
 55. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49(D1):D412–9.
 56. Lee TH, Tang H, Wang X, Paterson AH. PGDD: a database of gene and genome duplication in plants. Nucleic Acids Res 2013;41(Database issue):D1152–1158.
 57. Liu H, Wang X, Wang G, Cui P, Wu S, Ai C, Hu N, Li A, He B, Shao X. The nearly complete genome of *Ginkgo biloba* illuminates gymnosperm evolution. Nature Plants. 2021;7(6):748–56.
 58. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. In: The Proteomics Protocols Handbook. Totowa: Humana Press; Springer; 2005. p. 571–607.
 59. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant. 2020;13(8):1194–202.
 60. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol. 2000;302(1):205–17.
 61. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35(6):1547–9.
 62. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293–6.
 63. Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. Genome Biol. 2019;20(1):1–23.
 64. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics. 2010;8(1):77–80.
 65. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7): e49.
 66. Derveaux S, Vandesompele J, Hellemans J. How to do successful gene expression analysis using real-time PCR. Methods. 2010;50(4):227–30.
 67. Kozera B, Rapacz M. Reference genes in real-time PCR. J Appl Genet. 2013;54(4):391–406.
 68. Nicot N, Hausman JF, Hoffmann L, Evers D. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. J Exp Bot. 2005;56(421):2907–14.
 69. De Keyser E, Desmet L, Van Bockstaele E, De Riek J. How to perform RT-qPCR accurately in plant species? A case study on flower colour gene expression in an azalea (*Rhododendron simsii* hybrids) mapping population. BMC Mol Biol. 2013;14:1–15.
 70. Ramage E, Soza VL, Yi J, Deal H, Chudgar V, Hall BD, Di Stilio VS. Gene Duplication and Differential Expression of Flower Symmetry Genes in *Rhododendron* (Ericaceae). Plants (Basel). 2021;10(10):1994.
 71. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods. 2001;25(4):402–8.
 72. Everaert C, Luyckaert M, Maag JLV, Cheng QX, Dinger ME, Hellemans J, Mestdagh P. Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. Sci Rep. 2017;7(1):1559.
 73. Salih H, Gong W, Mkulama M, Du X. Genome-wide characterization, identification, and expression analysis of the WD40 protein family in cotton. Genome. 2018;61(7):539–47.
 74. Wang Y, Hu XJ, Zou XD, Wu XH, Ye ZQ, Wu YD. WDSdb: a database for WD40-repeat proteins. Nucleic Acids Res. 2015;43(Database issue):D339–344.
 75. Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud P-F, Lindquist EA, Kamisugi Y. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science. 2008;319(5859):64–9.
 76. Kim YJ, Kim MH, Hong WJ, Moon S, Kim EJ, Silva J, Lee J, Lee S, Kim ST, Park SK. GORI, encoding the WD40 domain protein, is required for pollen tube germination and elongation in rice. Plant J. 2021;105(6):1645–64.
 77. Maher C, Stein L, Ware D. Evolution of Arabidopsis microRNA families through duplication events. Genome Res. 2006;16(4):510–9.
 78. 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.
 79. 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.
 80. Guo C, Guo R, Xu X, Gao M, Li X, Song J, Zheng Y, Wang X. Evolution and expression analysis of the grape (*Vitis vinifera* L.) WRKY gene family. J Exp Bot. 2014;65(6):1513–1528.
 81. Wang XJ, Peng XQ, Shu XC, Li YH, Wang Z, Zhuang WB. Genome-wide identification and characterization of PdbHLH transcription factors related to anthocyanin biosynthesis in colored-leaf poplar (*Populus deltoids*). BMC Genomics. 2022;23(1):244.
 82. Liu J, Chen N, Chen F, Cai B, Dal Santo S, Tornielli GB, Pezzotti M, Cheng ZM. Genome-wide analysis and expression profile of the bZIP transcription factor gene family in grapevine (*Vitis vinifera*). BMC Genomics. 2014;15:281.
 83. Liu Z, Xie J, Wang H, Zhong X, Li H, Yu J, Kang J. Identification and expression profiling analysis of NBS-LRR genes involved in *Fusarium oxysporum* f. sp. conglutinans resistance in cabbage. 3 Biotech. 2019;9:1–12.
 84. Li P, Chai Z, Lin P, Huang C, Huang G, Xu L, Deng Z, Zhang M, Zhang Y, Zhao X. Genome-wide identification and expression analysis of AP2/ERF transcription factors in sugarcane (*Saccharum spontaneum* L.). BMC Genomics. 2020;21(1):685.
 85. Zhong Y, Chen Z, Cheng Z. Different scales of gene duplications occurring at different times have jointly shaped the NBS-LRR genes in *Prunus* species. Mol Genet Genom. 2022;297:263–76.
 86. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000;290(5494):1151–5.
 87. Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet. 2002;18(9):486.
 88. Lyons E, Pedersen B, Kane J, Alam M, Ming R, Tang H, Wang X, Bowers J, Paterson A, Lisch D. Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. Plant Physiol. 2008;148(4):1772–81.
 89. Li YY, Mao K, Zhao C, Zhao XY, Zhang HL, Shu HR, Hao YJ. MdCOP1 ubiquitin E3 ligases interact with MdMYB1 to regulate light-induced anthocyanin biosynthesis and red fruit coloration in apple. Plant Physiol. 2012;160(2):1011–22.

90. Wu B-H, Cao Y-G, Guan L, Xin H-P, Li J-H, Li S-H. Genome-wide transcriptional profiles of the berry skin of two red grape cultivars (*Vitis vinifera*) in which anthocyanin synthesis is sunlight-dependent or-independent. *PLoS ONE*. 2014;9(8): e105959.
91. Zhou LJ, Li YY, Zhang RF, Zhang CL, Xie XB, Zhao C, Hao YJ. The small ubiquitin-like modifier E3 ligase MdSIZ1 promotes anthocyanin accumulation by sumoylating MdMYB1 under low-temperature conditions in apple. *Plant Cell Environ*. 2017;40(10):2068–80.
92. Maier A, Hoecker U. COP1/SPA ubiquitin ligase complexes repress anthocyanin accumulation under low light and high light conditions. *Plant Signal Behav*. 2015;10(1): e970440.
93. An JP, Liu X, Li HH, You CX, Wang XF, Hao YJ. Apple RING E3 ligase MdMIEL1 inhibits anthocyanin accumulation by ubiquitinating and degrading MdMYB1 protein. *Plant Cell Physiol*. 2017;58(11):1953–62.
94. Liang D, Zhu T, Deng Q, Lin L, Tang Y, Wang J, Wang X, Luo X, Zhang H, Lv X et al. PacCOP1 negatively regulates anthocyanin biosynthesis in sweet cherry (*Prunus avium* L.). *J Photochem Photobiol B*. 2020;203:111779.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

