



OPEN ACCESS

EDITED BY
Jihong Hu,
Northwest A&F University, China

REVIEWED BY
Zhihua Wu,
Zhejiang Normal University, China
Zhichao Xu,
Northeast Forestry University, China

*CORRESPONDENCE
Zhezhi Wang
zzwang@snnu.edu.cn

[†]These authors have contributed
equally to this work

SPECIALTY SECTION
This article was submitted to
Plant Bioinformatics,
a section of the journal
Frontiers in Plant Science

RECEIVED 25 July 2022
ACCEPTED 15 September 2022
PUBLISHED 17 October 2022

CITATION
Hu S, Wang D, Wang W, Zhang C, Li Y,
Wang Y, Zhou W, Niu J, Wang S,
Qiang Y, Cao X and Wang Z (2022)
Whole genome and transcriptome
reveal flavone accumulation in
Scutellaria baicalensis roots.
Front. Plant Sci. 13:1000469.
doi: 10.3389/fpls.2022.1000469

COPYRIGHT
© 2022 Hu, Wang, Wang, Zhang, Li,
Wang, Zhou, Niu, Wang, Qiang, Cao and
Wang. This is an open-access article
distributed under the terms of the
Creative Commons Attribution License
(CC BY). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Whole genome and transcriptome reveal flavone accumulation in *Scutellaria baicalensis* roots

Suying Hu^{1†}, Donghao Wang^{1†}, Wentao Wang^{2,3},
Caijuan Zhang¹, Yunyun Li¹, Yueyue Wang¹, Wen Zhou¹,
Junfeng Niu¹, Shiqiang Wang¹, Yi Qiang¹,
Xiaoyan Cao¹ and Zhezhi Wang^{1*}

¹National Engineering Laboratory for Resource Development of Endangered Crude Drugs in Northwest China, Key Laboratory of the Ministry of Education for Medicinal Resources and Natural Pharmaceutical Chemistry, Shaanxi Normal University, Xi'an, China, ²Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China, ³University of Chinese Academy of Sciences, Beijing, China

Scutellaria baicalensis Georgi is a medicinal plant in the Lamiaceae family that contains high levels of 4'-deoxyflavone and other flavonoids in its roots. Therefore, it has strong potential as a plant resource for researching the biosynthesis of specific flavonoids. In this study, we report on a chromosome-level *S. baicalensis* genome assembled to nine chromosomes (376.81M) using PacBio, HiSeq XTen, and Hi-C assisted assembly. The assembly ratio was 99.22%, the contig N50 was 1.80 million bases, and the scaffold N50 was 40.57 million bases, with 31896 genes being annotated. Comparative genome analysis revealed that *S. baicalensis* and *Salvia miltiorrhiza* belonged to the same branch, and diverged 36.3 million years ago. Other typically correlated species were *Boea hygrometrica* and *Sesamum indicum*. We investigated the structural genes involved in flavonoid synthesis in combination with transcriptome sequencing analysis for different tissues (roots, stems, flowers, leaves) of purple, pink, and white flowers. The results revealed that *S.baiF6H* is involved in the accumulation of baicalein and was significantly increased in both purple roots vs. pink roots and white roots vs. pink roots. *S.baiMYB* gene family expression pattern analysis and co-expression network analysis revealed that *S.baiMYB* transcription factors primarily regulated the production of flavonoids in *S. baicalensis*. *S.baiMYB* serves as a major factor regulating flavonoid synthesis in the roots, where yeast one-hybrid assays revealed that these transcription factors could bind to the promoter regions of structural genes to control the accumulation of flavonoids. Genome and transcriptome sequencing, co-expression analysis, and yeast one-hybrid experiments provided valuable genetic resources for understanding flavonoid biosynthesis in *S. baicalensis*. These findings contribute to a better understanding of the accumulation of metabolites in Lamiaceae.

KEYWORDS

Scutellaria baicalensis, whole genome, transcriptomic, flavonoid biosynthesis, *S.baiMYB*

Introduction

Scutellaria baicalensis Georgi is a plant in the Lamiaceae family, and typically has purple flowers with bulbous roots that efficiently accumulate flavonoids. Its dried roots are called Huang-Qin (黄芩), which are primarily used to treat respiratory infections, diarrhea, dysentery, and liver disease in traditional Chinese medicine (Han et al., 2007). Its main active components consist of flavonoids and melatonin in the roots, of which baicalin comprises ~10–20%, and baicalin, baicalein, wogonin, and wogonin are the main active components (Bai et al., 2020). Its principal baicalin component has antibacterial, antiviral, anti-inflammatory, antioxidative, antitumor, neuroprotective, anticonvulsant, and cardiovascular protection properties that increase bone content, as well as anti-hyperglycemic properties (Zhao et al., 2016a).

In recent years, the structural genes of the flavonoid metabolic pathway have been investigated based on the *S. baicalensis* genome sequence (Zhao et al., 2019; Xu et al., 2020). Its flavonoid synthesis involved two pathways (classical flavonoid synthesis in the aerial components, and root-specific flavonoid 4'-deoxyflavonoid synthesis). The aerial flavonoid component is synthesized from an initial phenylalanine substrate via phenylalanine ammonia lyase (PAL), cinnamic acid 4-hydroxylase (C4H), 4-coumaric acid coenzyme A ligase (4CL), and 4-Coumaroyl coenzyme A (CoA). Subsequently, a 4-coumaroyl coenzyme A and three malonyl coenzyme A through the chalcone synthase (CHS) and chalcone isomerase (CHI) conversion of naringenin chalcone to naringenin. Flavonoid synthase (FNSII) synthesizes apigenin, after which flavonoid 6-hydroxylase is converted to scutellarin. Root-specific flavonoids are converted from phenylalanine through phenylalanine ammonia lyase (PAL) and cinnamate-CoA ligase (CLL-7) to cinnamoyl-CoA, whereas 4'-deoxyflavonoids are synthesized from cinnamoyl Coenzyme A, and malonyl coenzyme A are converted to pine chalcone by chalcone synthase (CHS) and chalcone isomerase (CHI). Next, chrysin is synthesized by flavonoid synthase (FNSII), and finally, flavonoid 6-Hydroxylase (F6H) and flavonoid 8-hydroxyl (F8H) enzymes are converted to scutellarin and wogonin (Pei et al., 2022). The final flavonoids undergo methylation and glycosylation modification to generate stable flavonoids. Among them, the function of F6H is well described. Contrasting with the accumulation patterns of baicalin and the expression patterns of *S. baiF6H*, which involved in the accumulation of baicalein in root of *S. baicalensis*. Zhao et al. reported the isolation and characterization of two *S. baiF6H* (SbCYP82D1.1 and SbCYP82D2) from *S. baicalensis*, which were able to convert chrysin to baicalein. When *S. baiF6H* was knocked down in *S. baicalensis* hairy roots, the content of baicalin was significantly reduced (Zhao et al., 2018). Liu et al. identified a F6H genes (CYP706X) from *Erigeron breviscapus* genome (Li et al., 2018). Xu et al. confirmed that seven F6H (CYP82D) genes could perform flavonoid 6-

hydroxylase (F6H) catalytic activity in *S. baicalensis* and *S. barbata* (Gao et al., 2022).

The release of plant genome sequences has gradually expanded the genome of the Lamiaceae family of plants. Although the flavonoid metabolic pathway has been analyzed, the comparative genomes of relative flavonoid synthesis pathways in *S. baicalensis* have not been investigated and the regulatory genes involved in high-content flavonoid synthesis pathways have not yet been explored. The flavonoid biosynthetic pathway is regulated by the conserved MYB-bHLH-WD40 (MBW) complex in plants (Xu et al., 2015). *PAP1* (*AtMYB75*) and *PAP2* (*AtMYB90*) are positive regulators of flavonoid synthesis in *Arabidopsis* (Maier et al., 2013), while *AtMYBL2* and *AtMYB4* are negative regulators (Jin et al., 2000; Matsui et al., 2008). Previous research has demonstrated that *AtMYB113* can regulate flavonoid synthesis in *S. baicalensis* via regulatory pathways that determine their efficient accumulation (Yuan et al., 2013; Qi et al., 2015). For this study, whole-genome sequencing and transcriptome data, gene co-expression network analysis, and a yeast one-hybrid test were combined to reveal that *S. baiMYB* spatiotemporally regulated the flavonoid synthesis process, which provides evidence for the efficient accumulation of flavonoids in *S. baicalensis*.

Materials and methods

Plant materials and growth conditions

Purple-, pink-, and white-flowered *S. baicalensis* were planted in the resource garden of Shaanxi Normal University. Genome sequencing and transcriptome materials were extracted from the same plants at the same time. The purple-, pink-, and white-flowered *S. baicalensis* were grown during their second year, and transcriptome sequencing was performed at the flowering stage.

PacBio genome sequencing and illumina sequencing

The genomic DNA of *S. baicalensis* was extracted using the CTAB method (Tel-zur et al., 1999); agarose electrophoresis was used to detect the DNA integrity; Nanodrop was used to detect the DNA purity and concentration, and Qubit was accurately quantified. A Covaris-TUBE was used for fragmentation and magnetic beads were enriched for large DNA fragments. All samples passed quality control, fault repair, end repair, stem-loop junction, and PacBio sequencing (Hackl et al., 2014). Following the removal of low-quality reads and adapter sequences, the clear data volume, quality, and length were qualified for subsequent assembly analysis.

Estimation of genome size

A survey was used to estimate the genome size, heterozygosity, and repeat sequences (Li et al., 2010). Third-generation assembly was conducted using 41 Gb subreads. All PacBio reads were first assembled using MECAT and Canu software and the resulting contigs were optimized using Quickmerge software (Boetzer and Pirovano, 2014); BUSCO was used to evaluate the assembly results data (Simao et al., 2015).

De Novo assembly

The cross-linking of young tissues was performed using formaldehyde. Once the chromosomes were extracted and qualified, the chromatin was digested *via* restriction endonuclease (HindIII/MboI), labeled with biotin, and subjected to blunt end ligation, Hi-C sample preparation, and DNA quality detection. After the Hi-C library was qualified, Illumina HiSeq X Ten was sequenced, the data was filtered, and high-quality reads were obtained, which were aligned with the genome. The restriction fragment interactions were analyzed and assembled in the chromosomes (Belton et al., 2012).

Genome annotation

Genome annotation consists primarily of repetitively annotated sequences, as well as gene structure and gene function prediction. Sequences akin to known repeat sequences were identified in the RepBase repeat sequence database using RepeatMasker and RepeatProteinMask software (Price et al., 2005). A *de novo* repeat sequence library was constructed using the RepeatModeler software and predicted with the RepeatMasker software. Genes were aligned using BLAST, Augustus, GlimmerHMM, SNAP, and GeneMark software, as well as the statistical characteristics of genome sequence data (Chaisson and Tesler, 2012). The gene structure was predicted using the EVidenceModeler method and was non-redundantly integrated. Functional gene annotation was derived from gene structural annotation and compared using various functional databases, including SwissProt, not, nr, Pfam, eggNOG, GO, and KEGG databases. Two strategies were used for the annotation of ncRNAs: the Rfam database to compare rRNAs and snRNAs, and miRNAs with known non-coding RNA libraries. The tRNA genomic sequences were predicted using tRNAscan-SE.

Evolutionary analysis

Gene family cluster analysis: Gene candidate species were filtered, which included several variants spliced transcripts of a gene, and only transcripts with the longest coding regions were

retained for further research (Parra et al., 2007). To ensure the efficacy of protein encoding, genes that encoded proteins with fewer than 50 amino acids were excluded. BLASTp was used to detect similarities between protein sequences from all species, where the e-value was set to 1e-5 by default. OrthoMCL software was used to cluster the results, with an expansion ratio of 1:5. The phylogenetic analysis of 567 single-copy orthologous gene families, MUSCLE1 (<http://www.drive5.com/muscle/>) alignment for each family, and alignment results were merged to generate a super alignment matrix. A 10-species (ML TREE) was constructed *via* 4d locus and PhyML software using the maximum likelihood method. Phylogenetic trees are typically developed using BRMC.via cmtree in the PAML software package (<http://abacus.gene.ucl.ac.uk/software/paml.html>). Gene families with abnormal gene numbers in individual species were filtered using the PGM (probabilistic graphical models) model in CAFE software (<http://sourceforge.net/projects/cafehahnlab/>). Gene gains and losses were simulated in the defined evolutionary tree, and gene family growth and contraction were analyzed using hypothesis testing. The MCscan program used Musclesoftware (<http://chibba.agtec.uga.edu/duplication/mcscan/>) to achieve multiple sequence alignment of the sequences within the block. The 4dTV value was then calculated and the overall species occurrence was estimated, depending on the abundance of the 4dTV value. The the genome and interspecies differentiation were duplicated.

RNA sequencing

Six tissues were harvested from *S. baicalensis*, namely flower, leaves, purple flower *S. baicalensis* of roots, stems, pink flower *S. baicalensis* of roots, white flower *S. baicalensis* of roots. Three biological replicates for each tissue were collected. The total RNA was extracted using phenol/chloroform. After qualifying the total RNA samples, magnetic beads bearing Oligo (dT) were used for enrichment. A polyA tail RNA was found at the 3' end of the mRNA. Polymerase I was used to create the second strand of cDNA, which was then purified using the QIAQuick PCR kit and eluted with an EB buffer. To complete the library preparation, the purified double-stranded cDNA was exposed to end repair, base A, and sequencing adapters, after which the target size fragments were recovered by agarose gel electrophoresis and PCR amplification. Following the construction of the library, Qubit3.0 was used for preliminary quantification, and Agilent 2100 was used to identify the insert size of the library. Quantitative PCR was performed using Bio-rad CFX 96. The Illumina X Ten platform was used to sequence the qualified libraries. The raw reads obtained from Illumina sequencing were processed to obtain high-quality sequences (Clean Reads) by eliminating low-quality sequences and decontaminating the adapters. Clean reads were aligned to our *de novo* genome of *S. baicalensis* using TopHat (Trapnell et al., 2012).

Reads provided the foundation for all future assessments and compared the differentially expressed genes at various thresholds. We selected $|\log_2 \text{fold-change (FC)}| \geq 1$, where $p\text{-value} < 0.05$ was considered significantly different. The DEGs were annotated using blast NCBI, Uniprot, GO, and KEGG databases, which were performed by hierarchical and *K*-means clustering. A hypergeometric test was used to identify the highly enriched pathways in differentially expressed genes to determine whether the pathways were considerably enriched (Mortazavi et al., 2008). Annotation of transcript datasets in different tissues (roots, stems, flowers, leaves) of purple, pink roots and white roots with FPKM are listed in Supplementary Tables S2, S3.

Phylogenetic analysis

Multiple sequence alignments were performed using ClustalW alignment (Larkin et al., 2007). The phylogenetic tree was designed using MEGA10 (NJ, 1,000 boots) (Tamura et al., 2013). 1.5 kb promoter fragments of structural genes were extracted and queried against PlantCARE. TF binding sites were illustrated using TBtools. *S. baiMYBs* sequences are given in Supplementary Table S4.

Coexpression analysis

Correlation analyses between structural genes and *S. baiMYB* involved in flavonoid synthesis with correlation coefficients > 0.8 were performed separately using advanced correlation network (pearson, $p < 0.05$) (Metware Cloud). Detailed samples are included in Supplementary Table S5.

Yeast one-hybrid

The promoter sequences of structural genes were searched for the genome. The promoters were inserted into the pHis2-Leu-GW vector using the In-Fusion method, and the *S. baiMYBs* gene was digested and ligated into the pGADT7 vector. The vectors were mixed to transform *Saccharomyces cerevisiae* Y187. Transformants were cultured in SD/-Leu/-Trp/-His medium and SD/-Leu/-Trp/-His + 60 mM 3-AT medium for 3 days, respectively. p53-his2 was transformed as a negative control, while p53-his2 and pGADT7-p53 were used as positive controls. The primer sequences are displayed in Supplementary Table S6.

Statistics

All experiments were performed with three biological replicates unless otherwise specified. The data were the

average of three technical repetitions expressed as mean \pm standard error. ANOVA analysis was used for statistical analysis; the probability value $P < 0.05$ was considered statistically significant.

Results

Whole-genome sequencing data of *S. baicalensis*

The entire nuclear *S. baicalensis* genome was sequenced using HiSeq X Ten, PacBio, and Hi-C assisted genome assembly, and the sequencing data of two other researchers were summarized (Figure S1). We obtained a genome of 376.81M in size, with a 99.22% assembly rate assembled into nine chromosomes (Figure 1A). The ScaffoldN50 was 40.57 million bases and the ContigN50 was 1.80 million bases. Genome annotation revealed that the *S. baicalensis* genome was 57.73% repetitive, which presents a chromosome-scale genome assembly for *S. baicalensis* with longer reads and coverage (Table 1). *Salvia miltiorrhiza*, *Arabidopsis thaliana*, and *Sesamum indicum* were used as reference genes for structural prediction. There were 31,896 gene annotations, with an average gene length of 2605.09 bp in *S. baicalensis* (Supplementary Table S1). An analysis of gene families in 15 closely related species of *S. miltiorrhiza*, and *S. indicum* and external species included *Olea europaea*, *Catharanthus roseus*, *Boea hygrometrica*, *Arabidopsis thaliana*, *Daucus carota*, outer species *Oryza sativa*, *Fagopyrum tataricum*, and *Macleaya cordata*, comprising 567 single-copy gene families of 25,706 clustered genes (Figure 1B). Venn diagram analysis selected 10 species, of which six gene families from the studied species were functionally annotated via phylogenetic analysis (Figure S2). The results revealed that *S. baicalensis* and *Salvia miltiorrhiza* belonged to one branch, whereas *S. baicalensis* and *S. miltiorrhiza* were closely related to each other. The divergence time was 36.3 million years ago (Mya) for *S. baicalensis* and *S. miltiorrhiza*, while for *B. hygrometrica* it was 57 Mya (Figure S3). *S. baicalensis* had a 2,109 gene family expansion and 609 gene family contraction comparison reference genome (Figure 1B). There was a 373 gene family expansion and 40 gene family contraction with significant differences in the phylogenetic tree. 567 single-copy paralogous gene pairs were investigated whole-genome duplication (WGD) events during *S. baicalensis* evolution. The *Ks* values showed genome duplication events at 0.36 and 0.74 in *S. baicalensis*. The genome data of *S. miltiorrhiza*, *B. hygrometrica* and *C. roseus* were represented species specific or different plant species, which analysed recent WGD event. The results showed that the *Ks* values peaked at approximately 0.25, 0.38 and 0.51 in *S. miltiorrhiza*, *B. hygrometrica* and *C. roseus*. The distribution of the *Ks* values

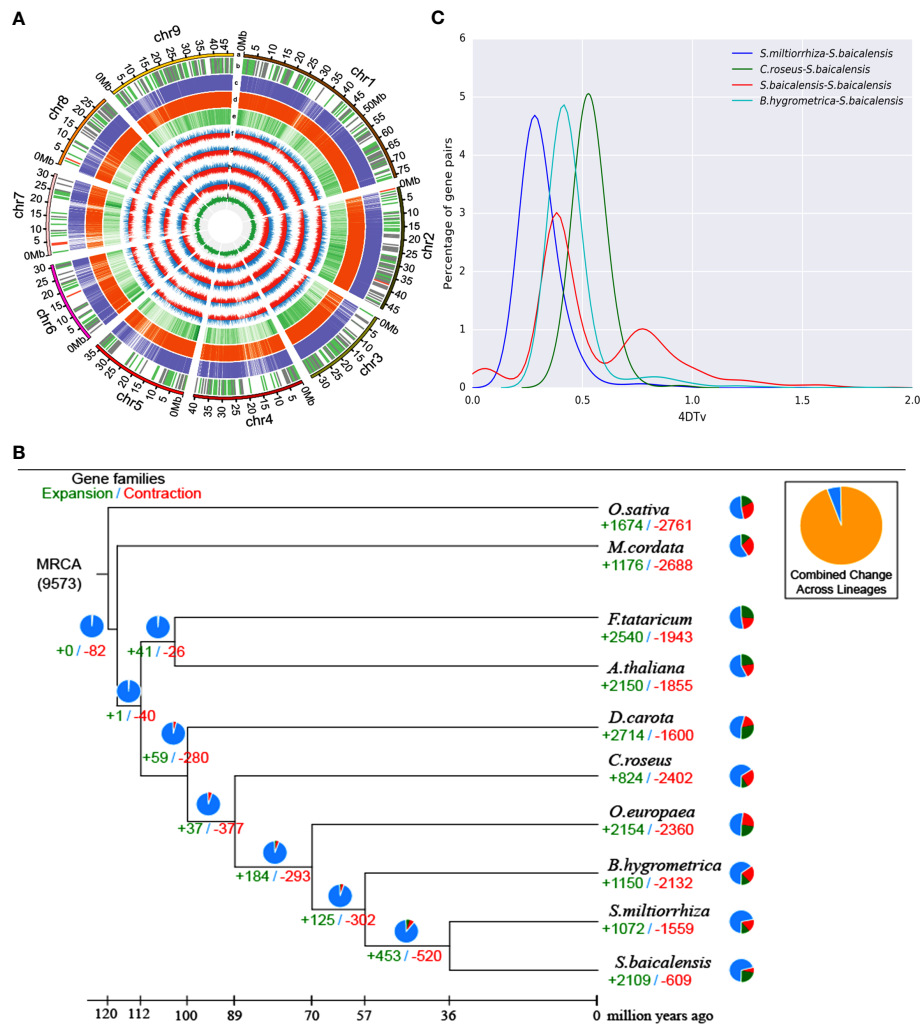


FIGURE 1

Genome and evolutionary analysis of *S. baicalensis*. (A) Overview of *S. baicalensis* genome assembly and annotation. (a) chromosome number and chromosome length; (b) non-coding RNA, rRNA, tRNA, and other types of non-coding RNA were distinguished on the genome (red, green, and grey, respectively); (c, d) Homologous gene data for *Salvia miltiorrhiza* and *Arabidopsis thaliana*; (e) Gene abundance (f-i) transcriptome of the expression data of flowers, leaves, stems and roots; (j) GC content. (B) Gene family expansion and contraction. Pie chart representing the proportion of gene expansion and contraction. (C) Whole-genome duplication analysis.

showed that a WGD events have occurred before the divergence of *B. hygrometrica* and *C. roseus*. There is a WGD events have occurred since the divergence of *S. miltiorrhiza* (Figure 1C).

Transcriptome analysis of *S. baicalensis*

A total of 56 million clean reads were obtained through Illumina X Ten sequencing, after which comparative and expression analyses were performed by hierarchical and K-means clustering (Figures 2A, B). The comparative expression patterns were significantly different between the roots, stems, flowers, and leaves. The DEGs were analyzed using KEGG

Enrichment, and the phenylpropane synthesis pathways and terpenoids were significantly different between the four tissues (Figure 2C).

Structural gene analyses of flavonoid biosynthesis pathways combined with transcriptome in *S. baicalensis*

The flavonoid biosynthesis expression patterns were summarized for *S. baicalensis*, and structural genes are displayed. (*S. baiPAL* (5), *S. baiCAH* (3), *S. bai4CL* (7), *S. baiCHS* (3), *S. baiFNSII* (2), *S. baiCHI* (3), *S. baiF6H* (2), *S. baiF8H* (1), and

TABLE 1 Statistics for genome sequencing of *S. baicalensis*.

Assembly feature	Qing Zhao etc.	Zhichao Xu etc.	Suying Hu etc.
Genome Size (Mbp)	408.14 Mb	441,861,814	365.92 Mb
Genome Length (bp)	386.63 Mb	376,971,275	376,808,447
ContigN50	880,642	2,102,880	1,803,970
Scaffold	386,627,338		376,862,547
ScaffoldN50	33.2 Mb	40,790,749	40,573,676
Max	87.96 Mb		76,118,278
Map rate (%)	96.50%	88.72%	99.22%
Complete BUSCos (C)	C: 64% [D: 22%], F: 1.7%, M: 3.4%, n:956	C:91.5%, F: 2.6%, M:5.9%, n: 1440	C: 90.2% [D: 11.7%], F: 2.6%, M: 2.8%, n: 1440
No. of predicted transcripts and proteins	28930	31896	33414
Average gene length (bp)	2,980.36	4553	2605
Average CDS length (bp)	1112.48	1130	1068
Hi-C assembly N50	33.20 Mb	40.8 Mb	40.57Mb

S.baiUGT (22)) (Figure 3A). Transcriptome data revealed that *S. baiF6H* was significantly more highly expressed in the roots of *S. baicalensis*, while the other structural genes were not specifically expressed in the four tissues (Figure 3B).

Transcriptome analysis of three germplasms of *S. baicalensis*

RNA-seq data identified DEGs in three germplasms of *S. baicalensis* (Figure 4B). The results showed that the structural genes of flavonoid biosynthesis significantly decreased in the white and pink roots (*S.baiCHS*, *S.baiCHI*, *S.baiFNSII*), while *S.baiF8H*, *S.baiF6H*, and *S.baiPFOMT* were upregulated in the white and purple roots (Figure 4C). The major bioactive chemical constituents (e.g., baicalin, wogonin, and scutellarin) were significantly higher in the *S. baicalensis* roots of the purple variety (Figure 4A). The DEGs were analyzed using KEGG enrichment, with the outcomes revealing that the phenylpropane synthesis pathways were significantly different between the three germplasms of *S. baicalensis* (Figure 4D).

Identification MYB transcriptional regulatory network in *S. baicalensis*

A total of 134 MYB candidate genes were identified through the Pfam database of the *S. baicalensis* genome. The results indicated that the phylogenetic tree separated the *S. baicalensis* MYB members into 20 groups, and a large number of *S.baiMYB* groups were not included in the *A. thaliana* branch (Figure 5A). A co-expression analysis of *S.baiMYBs* and flavonoid structural genes was performed according to the *S. baicalensis* root transcriptome data for white, pink, and purple flowers

(Figure 5B). The results showed that *S.baiMYBs* transcription factors were negatively correlated or positively correlated with *S.baiCHI*, *S.baiFNS*, and *S.baiUGT*. *S.baiMYBs* transcription factors were positively correlated with *S.baiF6H* and *S.baiF8H*. We selected 10 transcription factors related to flavonoid synthesis that were closely related to the *AtMYB* gene (e.g., *S.baiPAP1* (evm.model.contig94.43), *S.baiPAP2* (evm.model.contig34.129), *S.baiMYB111* (evm.model.contig357.324), *S.baiMYB116* (evm.model.contig76.33), *S.baiMYB70* (evm.model.contig57.4), *S.baiMYB60* (evm.TU.contig36.39), and *S.baiMYB5* (evm.TU.contig357.357) were positively correlated. *S.baiMYB3* (evm.model.contig535.20) and *S. baiMYB4* (evm.model.contig8.96) were negatively correlated. The prediction results showed that most MYBs could bind to the promoter regions of structural genes. However, *S.baiCLL-7*, *S.baiF6H*, and *S.baiF8H* had fewer binding sites (Figure 5C).

Yeast one-hybrid analysis MYB transcription factor

Transcription factor binding site prediction results verified that most MYBs bind with sites in the promoter regions of structural genes (Figure 6A). P53 and pGADT7 were used as positive and negative controls (Figure 6B), respectively, for the yeast one-hybrid analysis of candidate *S.baiMYB* transcription factors and structural genes (Figure 6C). The yeast one-hybrid results revealed that fewer MYBs could bind to the promoter regions of structural genes (*S.baiCLL-7*, *S.baiF6H*, and *S.baiF8H*). *S.baiMYB3*, *S.baiMYB4*, *S.baiPAP1*, and *S.baiPAP2* could bind to *S.baiPAL*, *CLL-7*, *S.baiFNSII*, and *S.baiF6H*, as well as *S.baiCHS* and *S.baiCHI* promoters. Furthermore, *S.baiMYB60*, *S.baiMYB70*, and *S.baiMYB111* could be combined with *S.baiPAL*, *CLL-7*, *S.baiFNSII*, and *S.baiF6H* promoters.

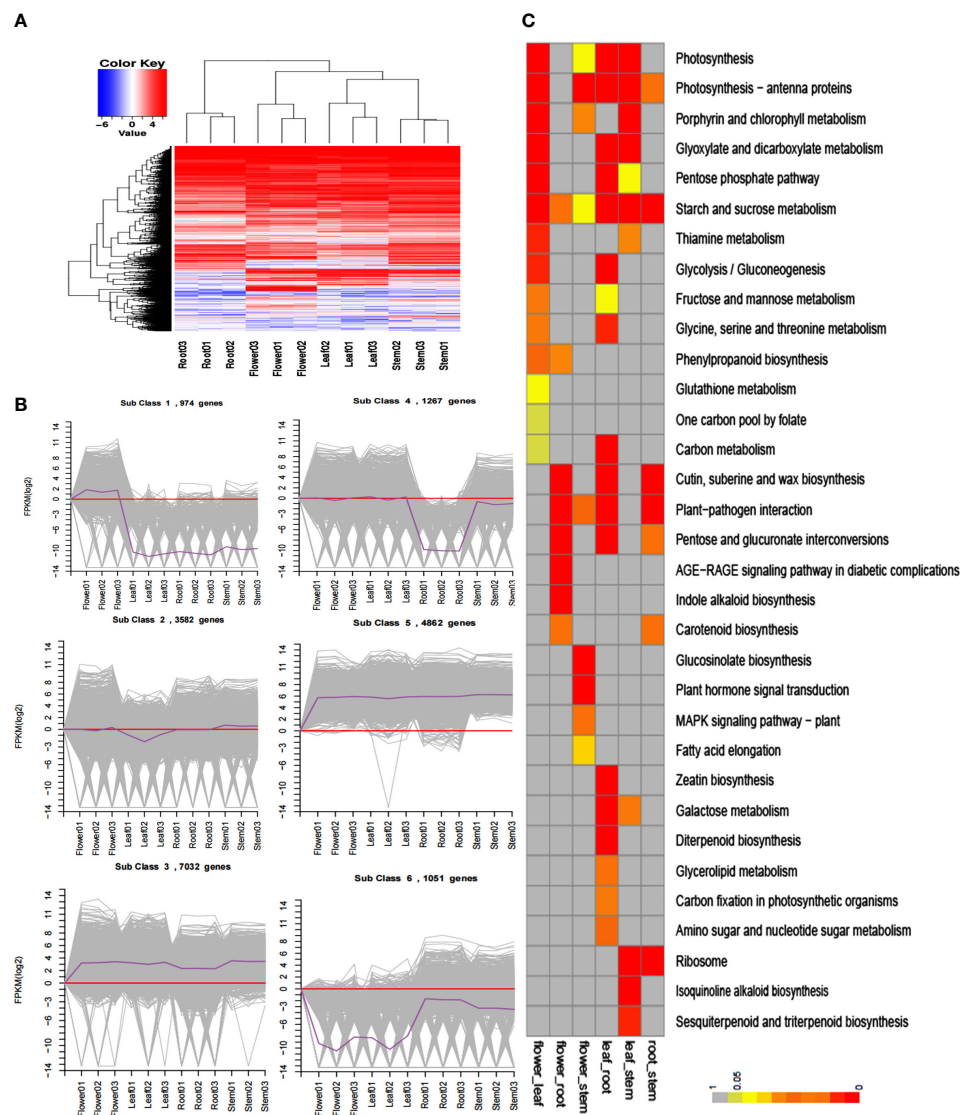


FIGURE 2
Transcriptome analysis of *S. baicalensis* roots, stems, flowers, and leaves. **(A)** Hierarchical cluster of differentially expressed genes. **(B)** K-means cluster. Grey lines show the expression patterns of genes in each cluster, and dark purple lines indicate that all genes in the cluster are in the sample. **(C)** KEGG enrichment. Different colors represent different levels of enrichment.

Discussion

As a popular Chinese herbal medicine, *S. baicalensis* has long been used in China (Zhao et al., 2016a). Traditional compound medicines such as Xiao Chai Hu Tang and Fei Fu Fang are extensively used in clinical practice. Xiao Chai Hu Tang has hepatoprotective effects, which were introduced to the alternative medicine community in the United States. These can increase chemotherapeutic effects for non-small cell lung cancers. Flavonoids (e.g., baicalin, baicalein, wogonin, and wogonin) are the primary active components in *S. baicalensis* (Bocho áková et al., 2003). Utilizing high-quality genome sequencing, we

investigated the molecular synthesis and efficient accumulation mechanisms of unique metabolites. Medicinal plants of the Lamiaceae family, such as *Mentha longifolia*, *S. miltiorrhiza*, *Salvia bowleyana*, *Scutellaria barbata*, *Salvia splendens*, *Ocimum basilicum*, *Origanum majoran a*, *Origanum vulgare*, *Pogostemon cablin*, and *Rosmarinus officinalis* have been sequenced (Meng et al., 2022). Among them, *S. miltiorrhiza*, *S. bowleyana*, *S. barbat*, and *P. cablin* were assembled chromosomes. *S. baicalensis* genome evolution analysis revealed that *S. barbata* and *S. baicalensis* diverged about 70,000 years ago (Ming et al., 2018). *S. splendens*, *S. miltiorrhiza*, *S. bowleyana*, and *S. barbara* are closely evolutionarily related.

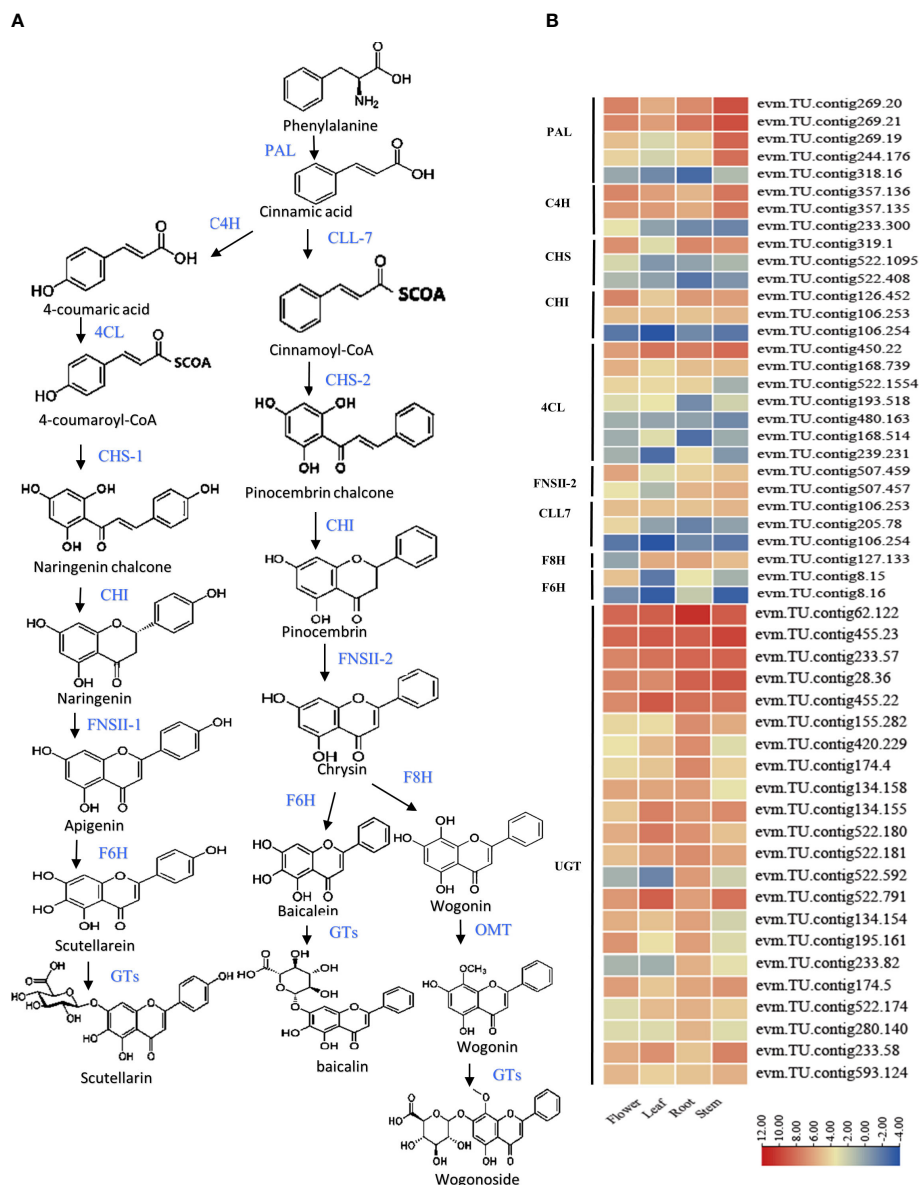


FIGURE 3
 Synthesis pathway of flavonoids in *S. baicalensis*. **(A)** Flavonoid biosynthesis pathway structural genes including phenylalanine ammonia lyase (*S.baiPAL*), cinnamate 4-hydroxylase (*S.baiC4H*), cinnamate-CoA ligase (*S.baiCLL-7*), 4-coumarate CoA ligase (*S.baiCLL-1*), chalcone synthase (*S.baiCHS-1*), pinocembrin- chalcone synthase (*S.baiCHS-2*), chalcone isomerase (*S.baiCHI*), flavone synthase II (*S.baiFNSII*), flavone 6-hydroxylase (*S.baiF6H*), flavone 8-hydroxylase (*S.baiF8H*), and 8-O-methyl transferase (*S.baiOMT*). **(B)** Heatmap of structural gene expression in stems, leaves, and flowers. Low to high expression is indicated by changes in color, from blue to red. Three biological replicates were used for each sample.

The medicinal components of *S. baicalensis* exist primarily in its roots (Wang et al., 2018). Zhao et al. found that the aboveground portion is mostly comprised of scutellarin, while the belowground portion is mostly made up of *S. baicalensis* flavonoids (Zhao et al., 2016b; Zhao et al., 2018). JA treatments have been shown to increase the expression of the structural genes of flavonoids in *S. baicalensis* (Yuan et al.,

2013). However, the high-efficiency accumulation of flavonoids in the roots of *S. baicalensis* remains unknown, as do the expression patterns of the genes involved. Transcriptome data revealed that the high expression of *S. baiF6H* was positively correlated with the synthesis of baicalein flavonoids and their efficient accumulation in the roots.

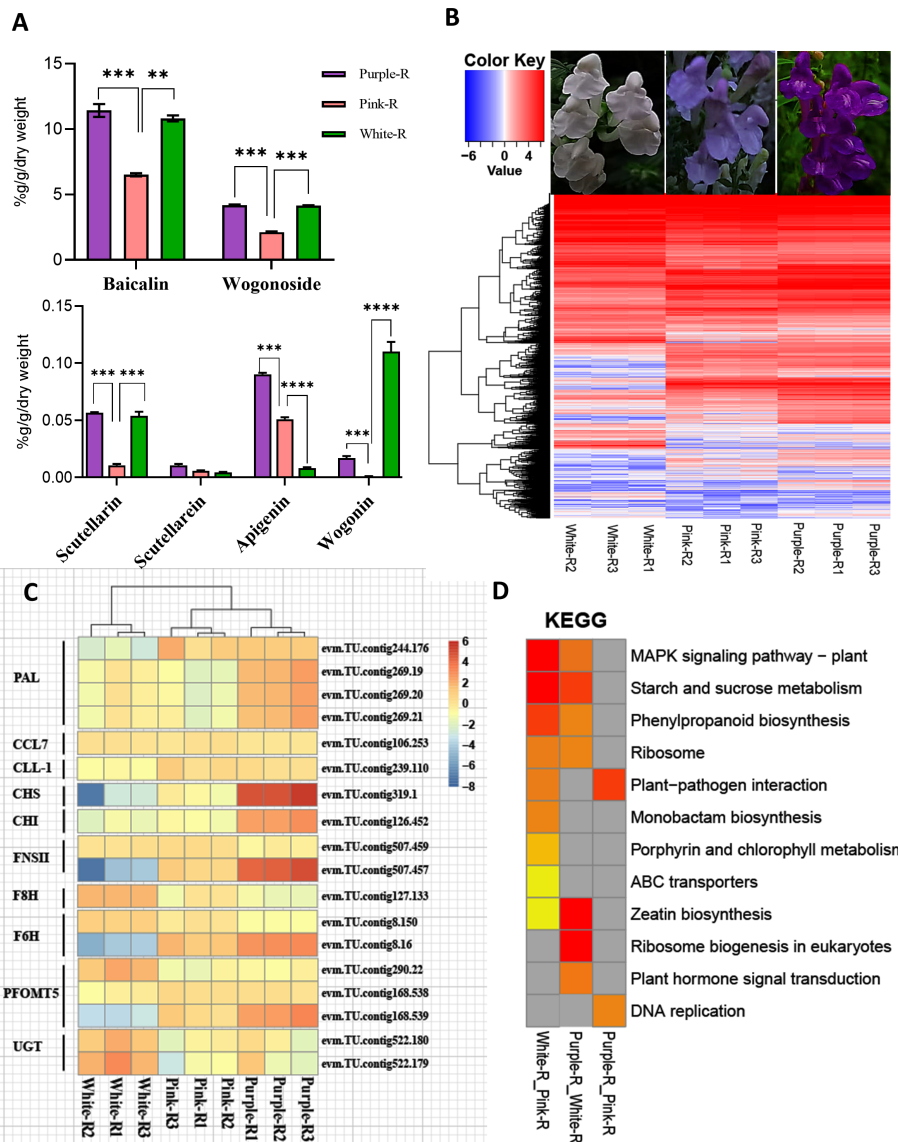


FIGURE 4

Transcriptome analysis of different germplasms of *S. baicalensis*. (A) Structural gene differential gene expression calorimetry. (B) KEGG clustering heatmap. (C) Histogram of baicalin, baicalein, wogonin, and wogonin contents. (D) Structural gene expression heatmap. Low to high expressions are indicated by changing colors, from blue to red. Data are the mean \pm SD from three biological replicates. Asterisks (*) indicate significant differences (** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$ based on Duncan's multiple range test).

The MYB-bHLH-WD40 (MBW) complex is a key factor in the regulation of flavonoids, with direct or indirect roles in regulating genes during the accumulation of plant flavonoids (Ramsay and Glover, 2005; Li, 2014). It is one of the largest gene families of MYB transcription factors in plants that directly regulate flavonoid synthesis. *AtMYB75/PAP1* (Rowan et al., 2009), *AtMYB90/PAP2* (Li et al., 2018), *AtMYB5* (Li et al., 2009), *AtMYB111* (Stracke et al., 2010), *AtMYB113* (Liu et al., 2016), and *AtMYB114* (Yao et al., 2017) (subfamily 6) positively

regulate the biosynthesis of anthocyanin in *A. thaliana* and can regulate the expression of genes such as chalcone synthase (*CHS*), chalcone isomerase (*CHI*), and flavonol 3-hydroxylase (*F3H*). Flavonoid synthesis inhibitors include *AtMYB34*, *AtMYB29*, *AtMYB76*, *AtMYBx*, *AtMYBL1*, *AtMYB3L*, *AtMYB27*, *AtMYB3*, *AtMYB4*, *AtMYB7*, and *AtMYB32* (Gates et al., 2018). The overexpression of *S.baiMYB* in tobacco can increase the accumulation of flavonoids (Qi et al., 2015). Most *S.baiMYB* in *S. baicalensis* and *AtMYB* are not in the same

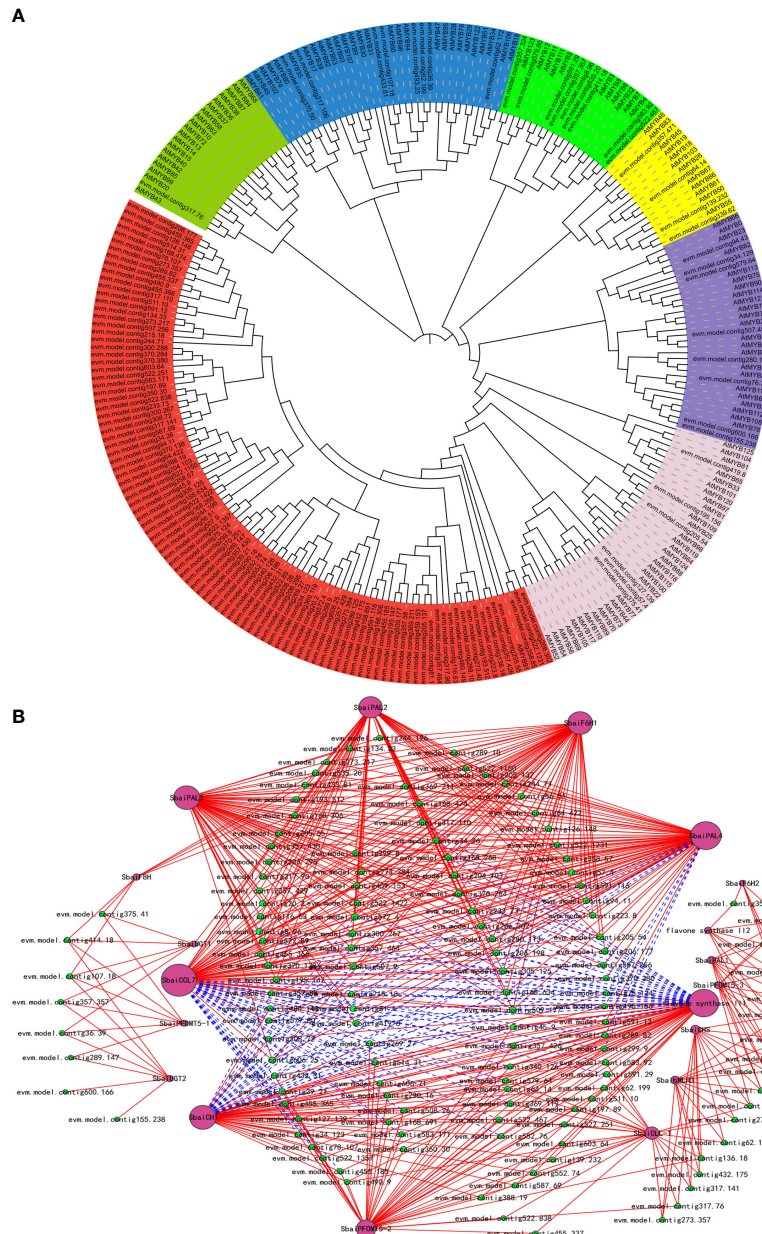


FIGURE 5
*S. bai*MYBs evolution and correlation analyses of structural genes. **(A)** *S. bai*MYBs and *At*MYBs were used to construct a phylogenetic tree (neighbor-joining method). **(B)** *S. bai*MYBs and structural gene co-expression network analysis. The data are mean \pm SD from three biological replicates.

evolutionary clade, and there are also fewer repressors clustered within the same evolutionary clade.

We used co-expression network analysis based on transcriptome sequencing results from rhizome mosaic leaves and *S. baicalensis* roots to describe the structural genes of flavonoid synthesis in *S. baicalensis*. We discovered fewer transcriptional repressors of the flavonoid contracting

process, while the majority of these MYBs served as positive regulators. *S. bai*MYB3 and *S. bai*MYB4, which act as repressors, were expressed at high levels in flowers but low levels in roots. The high expression of repressive *S. bai*MYBs in the aboveground portion of the plant decreased the flavonoid synthesis in *S. baicalensis*, which was highly accumulated in the roots. It was observed *via* transcriptional expression analysis

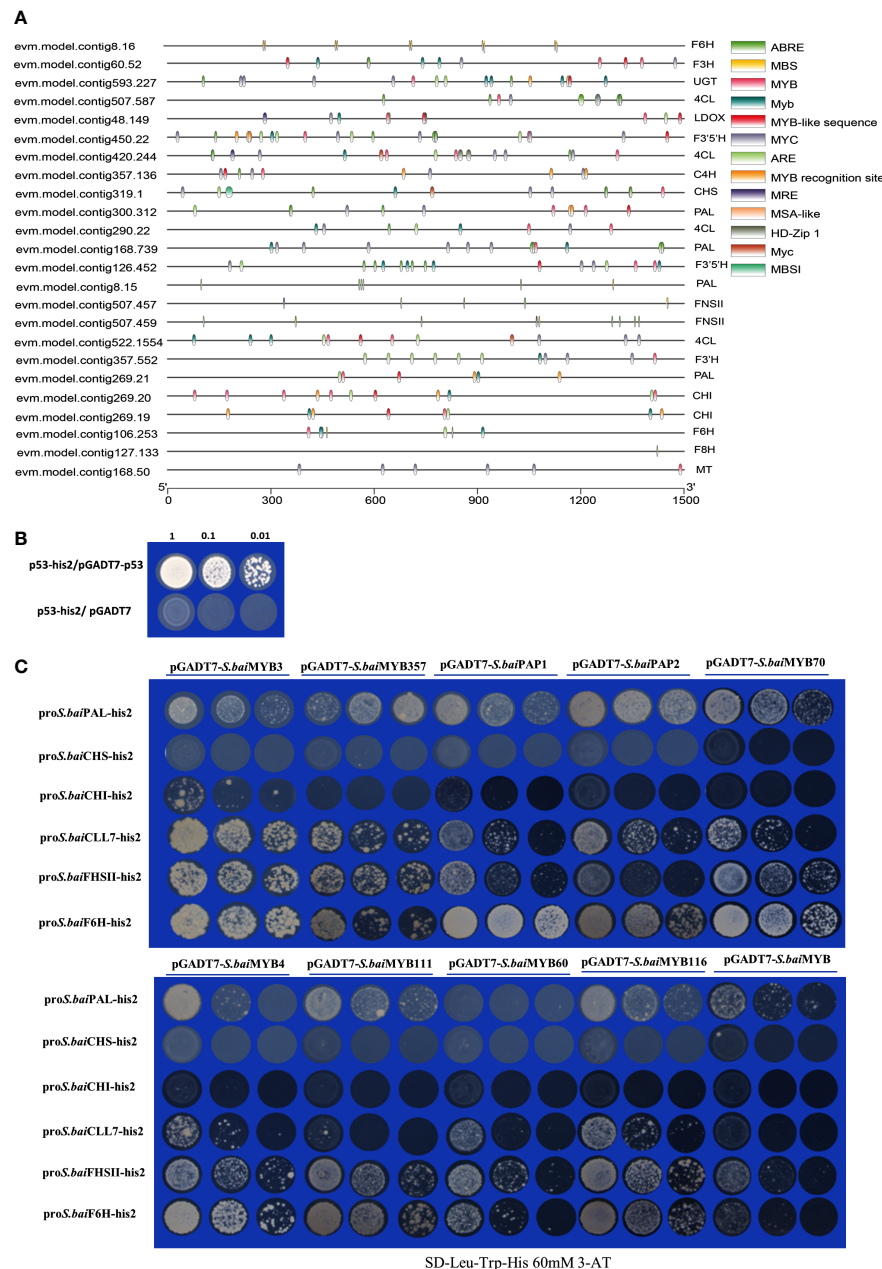


FIGURE 6

Yeast one-hybrid experiment. (A) Structural gene promoter region prediction. (B) pGADT7-p53 and p53-his2 are positive controls, and p53-his2 and pGADT7 are negative controls. (C) *S.baiMYBs* and structural gene promoter analysis.

that the specific spatiotemporal presence of *S. baiMYBs* and structural genes were highly significant, which was essential for flavonoid accumulation. According to yeast one-hybrid experiments, positive regulatory transcription factors regulated the expression of *S. baiCLL-7*, *S. baiFNSII*, and *S. baiF6H* genes. As such, this study elucidated the main reasons behind the efficient accumulation of flavonoids in *S. baicalensis*.

Conclusion

In this study, we reported the reference genome of *S. baicalensis*, while comparison with different tissues (roots, stems, flowers, leaves) of purple, pink, and white flowers revealed that *F6H S.baiF6H* is involved in the accumulation of baicalein in three germplasms of *S. baicalensis* roots. The *S.baiMYBs* gene family regulated the production of baicalein in *S. baicalensis* roots.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: <https://ngdc.cncb.ac.cn/>, PRJCA009554 & PRJCA009556.

Author contributions

ZW conceived of and designed the project. SH, DW, and WW performed the experiments and analyzed the data. SH wrote the original draft of the paper. YL, YW, WZ, JN, SW, YQ, and XC revised the paper. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the National Key Technologies R&D Program for Modernization of Traditional Chinese Medicine (2017YFC1701300), the Key Industry Chain Project of Shaanxi province (2022ZDLSF05-01), the National Natural Science Foundation of China (3217020611), the Central University Project (GK202003053), and the Fundamental Research Funds for the Central Universities (GK202205002, GK202205003, GK202205004, GK202205006).

References

- Bai, C., Yang, J., Cao, B., Xue, Y., Gao, P., Liang, H., et al. (2020). Growth years and post-harvest processing methods have critical roles on the contents of medicinal active ingredients of *scutellaria baicalensis*. *Ind. Crops Prod.* 158, 112985. doi: 10.1016/j.indcrop.2020.112985
- Belton, J. M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* 58, 268–276. doi: 10.1016/j.ymeth.2012.05.001
- Bocho áková, H., Paulová, H., Slanina, J., Musil, P., and Táborská, E. (2003). Main flavonoids in the root of *scutellaria baicalensis* cultivated in Europe and their comparative antiradical properties. *Phytother. Res.* 17, 640–644. doi: 10.1002/ptr.1216
- Boetzer, M., and Pirovano, W. (2014). SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinf.* 15, 211. doi: 10.1186/1471-2105-15-211
- Chaisson, M. J., and Tesler, G. (2012). Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinf.* 13, 238. doi: 10.1186/1471-2105-13-238
- Gao, R., Lou, Q., Hao, L., Qi, G., Tian, Y., Pu, X., et al. (2022). Comparative genomics reveal the convergent evolution of CYP82D and CYP706X members related to flavone biosynthesis in lamiaceae and asteraceae. *Plant J.* 109, 1305–1318. doi: 10.1111/tpj.15634
- Gates, D. J., Olson, B., Clemente, T. E., and Smith, S. D. (2018). A novel R3 MYB transcriptional repressor associated with the loss of floral pigmentation in *iochroma*. *New Phytol.* 217, 1346–1356. doi: 10.1111/nph.14830
- Hackl, T., Hedrich, R., Schultz, J., and Forster, F. (2014). Proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* 30, 3004–3011. doi: 10.1093/bioinformatics/btu392
- Han, J., Ye, M., Xu, M., Sun, J., Wang, B., and Guo, D. (2007). Characterization of flavonoids in the traditional Chinese herbal medicine-huangqin by liquid

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.1000469/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Interaction heat-map of chromosomal fragments based on Hi-C analysis.

SUPPLEMENTARY FIGURE 2

Common and unique gene families based on Venn diagram analysis.

SUPPLEMENTARY FIGURE 3

Estimation of divergence time analysis.

chromatography coupled with electrospray ionization mass spectrometry. *J. Chromatogr. B.* 848, 355–362. doi: 10.1016/j.jchromb.2006.10.061

Jin, H., Cominelli, E., Bailey, P., Parr, A., Mehrtens, F., Jones, J., et al. (2000). Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in arabidopsis. *EMBO J.* 19, 6150–6161. doi: 10.1093/emboj/19.22.6150

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Li, S. (2014). Transcriptional control of flavonoid biosynthesis: fine-tuning of the MYB-bHLH-WD40 (MBW) complex. *Plant Signal Behav.* 9, e27522. doi: 10.4161/psb.27522

Li, S. F., Milliken, O. N., Pham, H., Seyit, R., Napoli, R., Preston, J., et al. (2009). The arabidopsis MYB5 transcription factor regulates mucilage synthesis, seed coat development, and trichome morphogenesis. *Plant Cell* 21, 72–89. doi: 10.1105/tpc.108.063503

Liu, Y., Lin-Wang, K., Espley, R. V., Wang, L., Yang, H., Yu, B., et al. (2016). Functional diversification of the potato R2R3 MYB anthocyanin activators AN1, MYBA1, and MYB113 and their interaction with basic helix-loop-helix cofactors. *J. Exp. Bot.* 67, 2159–2176. doi: 10.1093/jxb/erw014

Li, N., Wu, H., Ding, Q., Li, H., Li, Z., Ding, J., et al. (2018). The heterologous expression of arabidopsis PAP2 induces anthocyanin accumulation and inhibits plant growth in tomato. *Funct. Integr. Genomics* 18, 341–353. doi: 10.1007/s10142-018-0590-3

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., et al. (2010). *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20, 265–272. doi: 10.1101/gr.097261.109

Maier, A., Schrader, A., Kokkelink, L., Falke, C., Welter, B., Iniesto, E., et al. (2013). Light and the E3 ubiquitin ligase COP1/SPA control the protein stability of

- the MYB transcription factors PAP1 and PAP2 involved in anthocyanin accumulation in arabidopsis. *Plant J.* 74, 638–651. doi: 10.1111/tj.12153
- Matsui, K., Umemura, Y., and Ohme-Takagi, M. (2008). AtMYBL2, a protein with a single MYB domain, acts as a negative regulator of anthocyanin biosynthesis in arabidopsis. *Plant J.* 55, 954–967. doi: 10.1111/j.1365-313X.2008.03565.x
- Meng, F., Tang, Q., Chu, T., Li, X., Lin, Y., Song, X., et al. (2022). TCMGP: an integrative database for traditional Chinese medicine plant genomes. *Hortic. Res.* 9, uhac060. doi: 10.1093/hr/uhac060
- Ming, J., Zhuoneng, L., and Guangxun, Z. (2018). Protective role of flavonoid baicalin from scutellaria baicalensis in periodontal disease pathogenesis: A literature review. *Complement. Ther. Med.* 38, 11–18. doi: 10.1016/j.ctim.2018.03.010
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226
- Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23, 1061–1067. doi: 10.1093/bioinformatics/btm071
- Pei, T., Yan, M., Huang, Y., Wei, Y., Martin, C., and Zhao, Q. (2022). Specific flavonoids and their biosynthetic pathway in scutellaria baicalensis. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.866282
- Price, A. L., Jones, N. C., and Pevzner, P. A. (2005). *De novo* identification of repeat families in large genomes. *Bioinformatics* 21 Suppl 1, i351–i358. doi: 10.1093/bioinformatics/bti1018
- Qi, L., Yang, J., Yuan, Y., Huang, L., and Chen, P. (2015). Overexpression of two R2R3-MYB genes from scutellaria baicalensis induces phenylpropanoid accumulation and enhances oxidative stress resistance in transgenic tobacco. *Plant Physiol. Biochem.* 94, 235–243. doi: 10.1016/j.plaphy.2015.06.007
- Ramsay, N. A., and Glover, B. J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci.* 10, 63–70. doi: 10.1016/j.tplants.2004.12.011
- Rowan, D. D., Cao, M., Lin-Wang, K., Cooney, J. M., Jensen, D. J., Austin, P. T., et al. (2009). Environmental regulation of leaf colour in red 35S:PAP1 arabidopsis thaliana. *New Phytol.* 182, 102–115. doi: 10.1111/j.1469-8137.2008.02737.x
- Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Stracke, R., Jahns, O., Keck, M., Tohge, T., Niehaus, K., Fernie, A. R., et al. (2010). Analysis of PRODUCTION OF FLAVONOL GLYCOSIDES-dependent flavonol glycoside accumulation in arabidopsis thaliana plants reveals MYB11-, MYB12- and MYB111-independent flavonol glycoside accumulation. *New Phytol.* 188, 985–1000. doi: 10.1111/j.1469-8137.2010.03421.x
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30, 2725–9. doi: 10.1093/molbev/mst197
- Tel-zur, S., Abbo, D., Myslabodski, D., and Mizrahi, Y. (1999). Modified CTAB procedure for DNA isolation from epiphytic cacti of the genera hylocereus and selenicereus (Cactaceae). *Plant Mol. Biol. Rep.* 17, 249–254. doi: 10.1023/A:1007656315275
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016
- Wang, Z. L., Wang, S., Kuang, Y., Hu, Z. M., Qiao, X., and Ye, M. (2018). A comprehensive review on phytochemistry, pharmacology, and flavonoid biosynthesis of scutellaria baicalensis. *Pharm. Biol.* 56, 465–484. doi: 10.1080/13880209.2018.1492620
- Xu, W., Dubos, C., and Lepiniec, L. (2015). Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci.* 20, 176–185. doi: 10.1016/j.tplants.2014.12.001
- Xu, Z., Gao, R., Pu, X., Xu, R., Wang, J., Zheng, S., et al. (2020). Comparative genome analysis of scutellaria baicalensis and scutellaria barbata reveals the evolution of active flavonoid biosynthesis. *Genomics Proteomics Bioinf.* 18, 230–240. doi: 10.1016/j.gpb.2020.06.002
- Yao, G., Ming, M., Allan, A. C., Gu, C., Li, L., Wu, X., et al. (2017). Map-based cloning of the pear gene MYB114 identifies an interaction with other transcription factors to coordinately regulate fruit anthocyanin biosynthesis. *Plant J.* 92, 437–451. doi: 10.1111/tj.13666
- Yuan, Y., Wu, C., Liu, Y., Yang, J., and Huang, L. (2013). The scutellaria baicalensis R2R3-MYB transcription factors modulates flavonoid biosynthesis by regulating GA metabolism in transgenic tobacco plants. *PLoS One* 8, e77275. doi: 10.1371/journal.pone.0077275
- Zhao, Q., Chen, X. Y., and Martin, C. (2016a). Scutellaria baicalensis, the golden herb from the garden of Chinese medicinal plants. *Sci. Bull. (Beijing)*. 61, 1391–1398. doi: 10.1007/s11434-016-1136-5
- Zhao, Q., Cui, M. Y., Levsh, O., Yang, D., Liu, J., Li, J., et al. (2018). Two CYP82D enzymes function as flavone hydroxylases in the biosynthesis of root-specific 4'-deoxyflavones in scutellaria baicalensis. *Mol. Plant* 11, 135–148. doi: 10.1016/j.molp.2017.08.009
- Zhao, Q., Yang, J., Cui, M. Y., Liu, J., Fang, Y., Yan, M., et al. (2019). The reference genome sequence of scutellaria baicalensis provides insights into the evolution of wogonin biosynthesis. *Mol. Plant* 12, 935–950. doi: 10.1016/j.molp.2019.04.002
- Zhao, Q., Zhang, Y., Wang, G., Hill, L., Weng, J. K., Chen, X. Y., et al. (2016b). A specialized flavone biosynthetic pathway has evolved in the medicinal plant, scutellaria baicalensis. *Sci. Adv.* 2, e1501780. doi: 10.1126/sciadv.1501780