Integrated transcriptomic and WGCNA analyses reveal candidate genes regulating mainly flavonoid biosynthesis in Litsea coreana var. sinensis

Na Xie
Guizhou University
Qiqaing Guo
hnguoqiqiang@126.com
Guizhou University
Huie Li
Guizhou University
Gangyi Yuan
Guizhou University
Qin Gui
Guizhou University
Yang Xiao
Guizhou University
Mengyun Liao
Guizhou University
Lan Yang
Guizhou University

Research Article

Keywords: Litsea coreana var. sinensis, Flavonoids, Full-length transcriptome sequencing, WGCNA, Candidate genes

Posted Date: September 28th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3376521/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Plant Biology on April 1st, 2024. See the published version at https://doi.org/10.1186/s12870-024-04949-1.

Abstract

Litsea coreana var. sinensis is a popular ethnic herb and beverage plant known for its high flavonoid content, which has been linked to a variety of pharmacological benefits and crucial health-promoting impacts in humans. The progress in understanding the molecular mechanisms of flavonoid accumulation in this plant has been hindered due to the deficiency of genomic and transcriptomic resources. We investigated the presence of four valuable flavonoids—quercetin-3-O-β-D-galactoside, quercetin-3-O-β-D-glucoside, kaempferol-3-O-β-D-galactoside, and kaempferol-3-O-β-D-glucoside in 98 samples, using high-performance liquid chromatography. In addition, we utilized Oxford Nanopore Technology (ONT) sequencing to derive full-length transcriptome sequences. Subsequent ONT sequencing yielded a total of 126,977 unigenes, of which 107,977 were successfully annotated in seven public databases. Within the annotated unigenes, 3,781 were categorized into 58 transcription factor families. A weighted gene co-expression network analysis identified two co-expression modules, MEpink and MEturquoise, that showed strong positive correlation with flavonoid content. Within these modules, four transcription factor genes (R2R3-MYB, NAC, WD40, and ARF) and four key enzyme-encoding genes (CHI, F3H, PAL, and C4H) emerged as potential hub genes. Among them, the R2R3-MYB (LcMYB123) as a homologous gene to AtMYB123/TT2, was speculated to play a significant role in flavonol biosynthesis based on phylogenetic construction. Our findings provided a theoretical foundation for further research into the molecular mechanisms of flavonoid biosynthesis, as well as a valuable resource for novel gene identification and genetic improvement in high-flavonoid L. coreana var. sinensis.

Introduction

Tea is a popular non-alcoholic beverage worldwide. *Litsea coreana* var. *sinensis* is a member of the Lauraceae, which is the main raw material plant of Eagle tea, and has been consumed for over a thousand years in southwest China [1]. In recent years, studies showed that *L. coreana* var. *sinensis* contained all kinds of compounds such as flavonoids, vitamins, polysaccharides, fatty acids and polyphenols while flavonoids are believed to be the major bioactive components in it [2]. As the main polyphenolic compounds in tea leaves, flavonoids can be divided into flavonols, anthocyanins, chalcones, dihydroflavonols, isoflavones, flavones, and flavanols depending on the carbon of the C-ring to which the B-ring is linked as well as the level of unsaturation and oxidation of the C-ring [3]. It contributes to the bitterness and astringency of tea as well as health benefits like anti-inflammatory, anticancer, immunomodulatory, antimicrobial and antioxidant bioactivities [4, 5]. It has been identified that *L. coreana* var. *sinensis* chiefly comprises four variants of flavonol glycosides [6], namely quercetin-3-O-*β*-D-galactoside (Q-3-gal), quercetin-3-O-*β*-D-glucoside (Q-3-glu), kaempferol-3-O-*β*-D-galactoside (K-3-gal) and kaempferol-3-O-*β*-D-glucoside (Q-3-glu) [7]. These constituents have demonstrated anti-obesity [8], anti-inflammatory and high antioxidant properties [9, 10]. Furthermore, the flavonoids within the plant have been associated with the unique aroma and flavor characteristic to the tea, with kaempferol and quercetin glycosides being the primary contributors to bitterness and astringency [11]. In light of its noteworthy medicinal and nutritional attributes, the synthesis of flavonoids and the corresponding molecular

mechanisms in *L. coreana* var. *sinensis* have become an area of keen interest among scholars. Nevertheless, the biosynthetic pathway and regulatory genes in *L. coreana* var*. sinensis* remains unclear.

Flavonoid biosynthesis is an intricate process involving both early and late biosynthetic genes [12]. Most of the genes encoding enzymes involved in flavonoid biosynthesis have been discovered and described in plants [13]. Flavonoids are synthesized from phenylalanine initially catalyzed by phenylalanine ammonia-lyase (PAL), followed by cinnamate 4-hydroxylase (C4H), 4-coumarate: CoA ligase (4CL), chalcone synthase (CHS), chalcone isomerase (CHI), flavone 3-hydroxylase (F3H), flavonoid 3-hydroxylase (F3'H), flavonol synthase (FLS), and UDP-glucosyltransferase (UGT) [14]. Transcription factors (TFs) such as bZIP, WRKY, MADS-box, ARF, NAC, and especially bHLH, MYB, and WD40 controlled flavonoid biosynthesis through regulating the biosynthetic genes involved in flavonoid biosynthesis pathway [15–17]. Among these, bHLH TFs can control anthocyanin, PA, and flavonol biosynthesis within the flavonoid pathway[18]. A key player in the regulatory networks that govern flavonoid biosynthesis in plants is the R2R3-MYB TFs. In *Arabidopsis*, there are 126 R2R3-MYB TFs, which can be divided into 25 subgroups, and the 5^{th}, 6^{th}, 7^{th}, 15^{th}, 19^{th} subgroups are directly implicated in the biosynthesis of flavonoids [9, 10]. For instance, the SG7 MYBs (*MYB11/PFG2, MYB12/PFG1 and MYB111/PFG3*) and SG19 MYBs *(MYB21, MYB24*, and *MYB57*) control flavonol biosynthesis [17, 19]. Furthermore, *MYB123/TT2*, a member of subgroup 5, is involved in the biosynthesis of proanthocyanidins (PAs) [20], while *AtMYB75/PAP1*, *AtMYB90/PAP2*, *AtMYB113* and *AtMYB114*, all of subgroup 6, are known to modulate anthocyanin biosynthesis[21]. In addition, MYB TFs can also interact with bHLH and WDR proteins to form a dynamic transcriptional activation complex (MBW complex) to regulate the anthocyanin and PA biosynthesis [22].

The assembly and annotation of a high-grade, chromosome-scale genome of *L. coreana* var. *sinensis* was recently accomplished, yet the resulting data remains unpublished [23]. RNA-seq based on high-throughput next-generation sequencing has been widely applied in plant studies, such as *Pinus massoniana* [24], and *Camellia sinensis* [25]. The majority of transcriptomes are currently created utilizing second-generation sequencing (SGS) technologies on the Illumina platform [26]. However, Illumina reads' short length limits their accurate identification of complex transcript isoforms [27]. The third-generation sequencing (TGS) technologies, such as Pacific Bioscience (PacBio) and Oxford Nanopore Technologies (ONT), have the capability to process substantial data quantities and read long sequences and full-length gene transcripts [28]. Compared to PacBio, ONT provides full-length, single-molecule transcriptome analysis, with exceptionally long reads, at a reduced cost and with a more straightforward operation [29]. The integration of second- and third-generation sequencing techniques can produce high-quality assemblies that facilitate the investigation of complex biological questions [30]. Weighted gene co-expression network analysis (WGCNA) is a valuable and reliable R tool that can identify key modules and hub genes related to target traits. For example, Hou et al. [31] combined WGCNA and transcriptomic analyses revealed a co-expression gene network associated with rutin synthesis. Through the application of WGCNA, MYB, ERF, WRKY, bZIP, and bHLH were identified as regulatory factors in the accumulation of flavonoids in mango [32]. A concurrent study by Ju et al. [33] identified 17 co-expression

modules and found MEcoral1 module was related to flavonoid metabolism. Flavonoids are the main active ingredients in *L. coreana* var. *sinensis*; however, the molecular mechanisms underlying the characteristic accumulation of flavonoids are largely unknown.

In the current study, the full-length transcriptome of *L. coreana* var. *sinensis* was sequenced utilizing ONT sequencing methods, establishing a vital and substantial public resource for genomic, gene expression, and functional genomic research within this particular species. Subsequently, WGCNA was conducted to investigate the relationships between major flavonoid content and gene expression and to identify potential hub genes for flavonoid production. The results would provide candidate genes for further industrial development, breeding and genetic improvement of *L. coreana* var. *sinensis*.

Materials And Methods

Plant materials

A bud with two young leaves of 98 trees were collected in May 2021, respectively. The trees are grown in a germplasm nursery in Xi Feng County, Guizhou Province, China (27°7’ N, 106°38’ E). These resources were collected previously from different regions: (1) Dao Zhen (DZ) (2) Xi Shui (XS) (3) Kai Yang (KY) (4) Mei Tan (MT) (5) Zhen An (ZA). A scrupulous selection process was applied to ensure the health and uniformity of the samples in terms of height and canopy width. All samples were immediately frozen in liquid nitrogen and stored at -80 °C before RNA extraction and metabolite analysis.

Flavonoid quantification

The extraction and determination of flavonoids were performed using a previously published method with slight modifications [34, 35]. K-3-glu, Q-3-glu, K-3-gal and Q-3-gal were purchased from Shanghai Yuanye Bio-Technology Co., Ltd (Shanghai, China). Flavonoids were detected by HPLC, equipped with Shim-pack GIST C18 (250 × 4.6 mm, 5 μm, Shimadzu, Japan). The water used was prepared with a Millipore water purification system. The freeze-dried samples were pulverized for 30 s at 30 Hz. Each sample, weighing 0.5 g (dry weight), was carefully weighed and subsequently mixed with 20 ml of 70% ethanol. This mixture was subjected to an extraction process for a duration of 60 minutes, and then sonicated at a frequency of 40 kHz for an additional 20 minutes. The samples were then centrifuged at 5000 rpm for 15 minutes, after which the supernatant was passed through a microporous filter membrane (0.22 μm pore size) and transferred into HPLC vials. The elution program was as follows: a linear gradient from 16 to 24% B for 0-26 min, and then 20% B for 26–40 min, followed by washing and equilibration. The injection volume was 10 μl with a flow rate of 1 ml/min and detected at a wavelength of 355 nm, while column temperature was kept at 35℃.

Transcriptome sequencing

Total RNA was extracted using RNA-prep Pure Plant Plus Kit (TIANGEN, Beijing, China). The integrity and purity of the isolated RNA were determined using a NanoPhotometer spectrophotometer (Thermo Scientific, Waltham, USA) and an Agilent 2100 Bioanalyzer (Agilent Technologies, California, USA), respectively. Following this, the mRNA fragments were utilized for the synthesis of the first cDNA strand, which subsequently led to the formation of a second strand, generating double-stranded cDNA. The synthesized cDNA was purified using magnetic beads, the ends were repaired, and a single nucleotide A was added. Sequencing adaptors were ligated onto each blunt-ended cDNA molecule. The prepared samples were subsequently sequenced using an Illumina NovaSeq 6000 platform, resulting in the generation of paired-end reads. Raw reads were filtered by removing the adapter sequences, null reads, low-quality reads, unknown nucleotides comprising more than 5% and ambiguous sequence “N”. Clean reads obtained from each library were assembled using Trinity software under de novo transcriptome assembly strategy [36]. Oxford Nanopore Technologies' standard cDNA library sequencing protocol was employed for the preparation of the Nanopore sequencing libraries from the captured RNA. In total, 15 samples were combined for the construction of the library. Full-length transcriptome sequencing was conducted on the Oxford Nanopore PromethION platform for long read sequencing. The TransDecoder tool (v3.0.0) (https://github.com/TransDecoder/TransDecoder/wiki) were utilized to predicted the coding sequences and their corresponding amino acid sequences. Expression levels were measured using the Transcripts per Million (TPM) approach. The sequencing results from Short-Read Sequencing (SGS) were used to correct the Third Generation Sequencing (TGS) data and to perform an analysis of gene expression levels, subsequent to the assembly of the transcriptome. The entire process of library construction and transcriptome sequencing were carried out by Kaitai-bio-Company (Hangzhou, China).

Gene functional annotation and transcription factor identification

The functional annotations of the unigenes were conducted through BLASTx, utilizing an E-value threshold of 10^{-5} across seven public databases: non-redundant protein sequences (NR), eukaryotic Orthologous Groups (KOG), the database of Homologous protein family (Pfam), non-redundant protein sequence database (Swiss-Prot), Translation of EMBL (TrEMBL), Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). NR database results were preferentially selected. The identification of transcription factor families was achieved by querying the unigenes against the transcription factor protein sequences contained in the Plant Transcription Factor Database (PlantTFDB; http://planttfdb.gao-lab.org//v5.0).

Weighted gene co-expression network analysis

The WGCNA was performed using the WGCNA R package (v4.1.3) [37]. The raw transcriptomic data underwent standardization before the computation of the median absolute deviation (MAD) for each gene. The top 20,000 genes ranked according to their MAD values were then selected, focusing on those

displaying large inter-sample variation, to serve as the foundation for the WGCNA construction. Genes demonstrating high degrees of clustering were allocated to distinct modules. The expression profile of module genes in each sample was displayed by module eigengene, which is the first principal component of gene expression. Gene modules were calculated based on the default settings with minmodulesize = 30, mergecutheight = 0.25, and power = 6. The correlations between modules and four flavonoids content were determined by estimating the module-trait relationships. The resulting networks were then visualized, and the degree value was calculated using Cytoscape (v3.10.0). In the co-expression module, hub genes were predominantly employed to represent closely related genes exhibiting substantial connectivity. The top 30 genes in each module based on degree were identified as hub genes. From the target modules, we pinpointed the transcription factor genes involved in the flavonoid biosynthesis pathways, in addition to the structural genes. TBtools (v1.120) was utilized to generate gene expression heatmaps for selected samples and candidate genes [38].

Phylogenetic analysis

The R2R3-MYB protein sequences related to flavonoid biosynthesis from different plants were obtained from the NCBI database (http://www.ncbi.nlm.nih.gov/). The phylogenetic analysis and multiple alignments were conducted using MEGA7 software. The phylogenetic tree was conducted by neighbor-joining (NJ) method with 1000 bootstrap replicates and visualized by the iTOL web tool (https://itol.embl.de/). Multiple alignments of protein sequences were visualized using GeneDoc.

qRT-PCR analysis.

To validate the accuracy and reliability of the RNA-seq data, 10 genes related to the flavonoid biosynthesis pathway from pink module were selected for validation using qRT-PCR. The RNA was extracted from sample XS13. Reverse transcription was performed using TUREscript RT Kit with gDNA Eraser (Aidlab, Beijing, China) and real-time fluorescence quantitative PCR was conducted using 2×Sybr Green Qpcr Mix (Aidlab, Beijing, China) on An IQ5 real-time PCR detection system (Bio-Rad) (Applied Biosystems, Waltham, USA). The 18S rRNA gene was utilized as the internal control. The qPCR was performed as follows: 94°C for 2 min, 40 cycles of 94 °C for 5 s, 60 °C for 30 s. The relative expression levels were calculated based on the values of three technical replicates and the $2^{-\Delta\Delta Ct}$ method was used for expression calculations [39]. The primers utilized for qRT-PCR procedure were designed by Primer 3 (v0.4.0) (http://primer3.ut.ee) and listed in Table S1.

Statistical analysis.

The conducted experiments were all performed in triplicate, with the resultant data being expressed as the mean value ± the standard error (SE). Significant differences were calculated using the least significant

difference (LSD) test (defined as $P < 0.05$). All statistical analysis were performed using SPSS software (version 19), and all the figures were plotted by Origin 2018 software.

Results

Determination of main flavonoid contents

The contents of four flavonoids, namely K-3-gal, K-3-glu, Q-3-glu and Q-3-gal were determined by HPLC and the results showed significant differences among these four kinds of flavonol glycosides. The most prevalent flavonoid within the samples was K-3-glu, which exhibited the the highest content, followed by Q-3-glu, Q-3-gal and K-3-gal. The mean contents of K-3-glu and Q-3-glu within the samples were determined to be 8.00 mg/g and 2.26 mg/g, respectively (Fig. 1A).

Transcriptome sequencing

Individual transcriptome assemblies were generated for each of the 98 samples. Approximately 50,267,369 raw reads were generated using an Illumina sequencing platform. The Q20 (sequencing error rate < 1 %), Q30 (sequencing error rate < 0.1 %), and GC content among all the libraries were about 98.05%, 95.05%, and 47.58% (Table 1). After raw reads filtered, 37.28-71.54 million clean reads were obtained, with 5.54-10.65 G clean base.

The transcriptome integrity was assessed using BUSCO v5.0.0, revealing that 22.1% of the sequences were complete and single-copy duplicated, 3.2% were fragmented, and only 0.5% were missing. The full-length transcriptome sequencing resulted in 126,977 unigenes for analysis, boasting an average length of 1199.9 bp and an N50 length of 1,345 bp. The distribution of the transcripts by size was as follows: 61.55% were between 201 and 1000 bp in length, 27.06% were between 1001 and 3000 bp, and and the remaining 11.39% exceeded 3000 bp in length (Fig. S1).

Table 1 Quality of reads obtained after RNA-sequencing

Classification	Maximum	Minimum	Average
Raw Reads	72,776,102	37,886,518	50,267,369
Clean Reads	71,541,570	37,282,590	49,247,699
Clean Bases (G)	10.65	5.54	7.33
Q20 (%)	98.50	96.84	98.05
Q30 (%)	95.88	93.43	95.05
GC Content (%)	49.36	46.37	47.58

Functional annotation and transcription factor analysis

All 107,977 unigenes were produced and successfully annotated to the following seven public databases: GO, KEGG, KOG, NR, Pfam, Swiss-Prot, and TrEMBL. Specifically, 54,892 (41.65%), 53,827 (42.39%),

61,557 (48.48%), 76,446 (60.20%), 96,438 (75.95%), 66,072 (52.03%), and 53,978 (42.51%) unigenes were annotated respectively (Table S2). Homology comparison of all predicted proteins found that majority of them are homologous to the species *Cinnamomum micranthum f. kanehirae* (73.54%), followed by *Artemisia annua* (3.22%), *Nelumbo nucifera* (2%), *Vitis vinifera* (1.72%), *Macleaya cordata* (1.12%), *Picea sitchensis* (0.96%), and *Camellia sinensis* (0.66%), among others (Fig. 2A).

Based on sequence homology, a total of 54,892 unigenes were annotated to one or more GO terms, and classified into three primary functional categories: biological process (BP), cellular component (CC), and molecular function (MF) (Fig. S2). The BP category was dominated by the terms 'cellular process' (35,412 unigenes) and 'metabolic process' (28,503 unigenes). The MF category featured 'catalytic activity' (24,745 unigenes), 'binding' (19,144 unigenes), and 'transcription regulator activity' (3,692 unigenes) as the top subcategories. For the CC category, the leading terms were 'cell part' (46,078 unigenes), 'cell' (46,078 unigenes), 'organelle' (35,713 unigenes), 'organelle part' (19,515 unigenes) and 'membrane' (18,020 unigenes). The KOG analysis classified 61,557 unigenes into 25 different categories (Fig. 2B). The most common subclasses were 'general function prediction only' (16,375 unigenes), 'signal transduction mechanisms' (6,684 unigenes) and 'posttranslational modification, protein turnover, chaperones' (6,624 unigenes). Totally, 53,827 unigenes annotated in the KEGG database, of which 'Protein families: genetic information processing' (25,267 unigenes), 'Protein families: signaling and cellular processes' (9,055 unigenes) and 'Protein families: metabolism' (8,122 unigenes) were the top three pathways in the *L. coreana* var. *sinensis* (Fig. 2C).

TFs are integral regulators of plant growth and metabolism, functioning as both activators and repressors. A total of 3,781 unigenes could be further classified into 58 TF families. Among these, bHLH transcription factors were the most abundant (314, 8.30%), succeeded by NAC (253, 6.69%), MYB-related (240, 6.35%), C2H2 (219, 5.72%), bZIP (192, 5.08%), ERF (188, 4.97%), and MYB (180, 4.76%) (Fig. S3). This comprehensive identification of TFs offers an abundant source of information for studying the roles of TFs in various biochemical pathways in *L. coreana* var*. sinensis.*

Weighted gene co-expression network analysis

To identify candidate genes related to flavonoid biosynthesis, we conducted a WGCNA based on gene expression data and the contents of four flavonoids. As a result, 23 modules were obtained, including the grey module, which contained all genes that did not belong to any other modules (Fig. S4). The size of modules ranged from 47 to 3103 unigenes (Fig. 3B). TFs present in each module varied in quantity from 2 to 145 (Fig. 3C). In addition, the relationship between modules and traits were examined and the pink module displayed a significantly strong positive correlation with four flavonoids, with the highest correlation observed with K-3-gal ($r = 0.74$, $p = 5.7e*10^{-4}$) (Fig. 3A).

As part of investigation into the potential functions of genes in each module, KEGG enrichment analysis was performed for genes within each module. The top 20 enriched pathways in each module were

revealed that the genes from pink and turquoise modules were enriched in pathways related to flavonoid biosynthesis, such as “flavonoid biosynthesis”, “flavone and flavonol biosynthesis”, “biosynthesis of secondary metabolites”, and “metabolic pathways” (Fig. 4). Additionally, module-trait relationships indicated a positive correlation between K-3-glu and the turquoise module ($r = 0.4$, $p = 5.1e*10^{-5}$). Notably, numerous genes encoding key enzymes were found within these two modules, including but not limited to: *FLS*, *PAL*, *4CH*, *CHS*, *CHI*, *F3H*, and *UGT*. Due to the positive correlation between modules and flavonoid contents, as well as the results of KEGG pathway analysis, pink and turquoise modules were selected for further analysis.

Identification of hub genes related to flavonoid biosynthesis

In gene co-expression network, the top 30 genes with the most degrees within the module were identified as hub genes. The enrichment analysis showed that genes within pink and turquoise modules were enriched in flavonoid biosynthesis, so the co-expression networks were constructed using Cytoscape. From the pink module, six genes *NODE64739g19765i1*, *NODE70694g17362i1*, *NODE12843g1147i21*, *NODE61449g421i59*, *NODE31353g4722i1*, and *NODE12636g299i40* were identified as transcription factor *MYB*, *NAC*, *CHI*, *C4H*, *PA*L, and *F3H*, respectively (Fig. 5A). In addition, *ARF* (*NODE42009g6374i3*) and *WD40* (*NODE25184g3897i2*) in turquoise module were also identified (Fig. 5B). Surprisingly, based on the gene expression data, structural genes and transcription factor genes (*MYB*, *bHLH, WD40*) related to flavonoid synthesis were analyzed within the pink and turquoise modules. Notably, high expression levels were found in samples with high flavonoid content (Fig. 6), which indicated that the genes within the pink and turquoise modules may have potential roles in the biosynthesis of flavonoids. Figure 7 visualizes the transcriptional expressions of key enzymes in flavonoid biosynthesis pathway.

The identified *NODE64739g19765i1* in *L. coreana* var. *sinensis* was named *LcMYB123*. The LcMYB123 protein contained two MYB DNA-binding domains, according to the conserved protein domain analysis. A multiple sequence alignment and phylogenetic tree of LcMYB123 with other flavonoid-related R2R3-MYBs were generated to further define its roles (Fig. 8). The phylogenetic analysis showed that LcMYB123 had a higher sequence similarity to AtMYB123/TT2, PbMYB9 and MdMYB9.

Validation of gene expression by qRT-PCR

To verify the reliability of transcriptome, 10 genes related to flavonoid biosynthesis were selected for verification by qRT-PCR in the pink module. The correlation coefficient between RNA-seq and qRT-PCR results reached 0.9331 ($P<0.01$) (Fig. 9B). It indicated that the transcriptome sequencing data was highly accurate.

Discussion

The main flavonoids in L. coreana var. sinensis

L. coreana var. *sinensis*, commonly known as the Eagle tea plant. Due to its high flavonoid content, Eagle tea has become one of the most popular traditional drinks in southwest of China, producing from buds or leaves of *L. coreana* var. *sinensis* [6]. Researches on flavonoids are beneficial to improve the development and utilization of this species. In contrast to traditional green tea, the major contituents in Eagle tea are flavonol glycosides, which have antioxidant, anti-inflammatory, and antineoplastic activities [41]. Yu et al. [42] first isolated and identified K-3-glu, and Q-3-glu in *L. coreana* var. *sinensis* and Chen et al. [43] subsequently characterized Q-3-gal and K-3-gal. Ma et al. [6] further corroborated that K-3-gal and Q-3-gal were the dominant components in it. However, K-3-glu and Q-3-glu were higher in content compared to the other two in this study. This could potentially be attributed to variations in environmental conditions at the sampling sites.

Sequencing and annotation

The secondary metabolism of *L. coreana* var. *sinensis* is poorly understood due to the lack of genomic data. It has been widely acknowledged that RNA-seq analysis is a valuable method for identifying genes involved in the metabolic processes in non-model species [44]. Furthermore, the advent of TGS technologies offering a cost-effective approach for annotating the transcriptomes. In this research, ONT was used for the first time to sequence the full-length transcriptome of *L. coreana* var. *sinensis*, and 107,977 unigenes were successfully annotated in seven public databases. Our findings would be a great resource for gene expression studies, selective breeding, and cultivation of *L. coreana* var. *sinensis*, even related species in the Lauraceae family.

Candidate structural genes involved in flavonoid biosynthesis

The process of flavonoid biosynthesis initiates with the phenylpropanoid pathway [45]. In this pathway, phenylalanine is converted to p-coumaroyl-CoA catalyzed by PAL, C4H, and 4CL. Following this, CHS, the inaugural enzyme in the flavonoid biosynthesis pathway, facilitates the conversion of three molecules of malonyl-CoA and one molecule of 4-coumaroyl-CoA into naringenin chalcone. This naringenin chalcone is subsequently transformed into naringenin catalyzed by CHI [32]. F3H catalyzes the conversion of naringenin to dihydroflavonols, which are ultimately modified into kaempferol, quercetin, and myricetin by FLS, from their respective precursors, dihydrokaempferol, dihydroquercetin, and dihydromyricetin [12]. There exists an interaction between FLS and dihydroflavonol 4-reductase (DFR) resulting in a competitive pathway leading either to flavonol or anthocyanin biosynthesis [46]. Glycosylated flavonoids, which are prevalent natural compounds in plants, are derived from the conversion of flavonols to flavonol glycosides, a reaction catalyzed by UGTs [47].

WGCNA is widely used in evaluating transcriptomics and metabolomics data in order to identify co-expression modules associated with specific traits [48]. Khan et al. [49] identified five key enzyme-encoding genes (C*HI*, *F3'H*, *DFR*, *LAR* and *UFGT*) that controlled flavonoid biosynthesis. Under cold stress conditions, the expression levels of several genes (*PAL*, *4CL*, *CHS*, *ANR*, *FLS*, and *LAR*) were significantly up-regulated, leading to an increase in flavonols in *Tetrastigma hemsleyanum* [50]. In this research, 23 gene co-expression modules were obtained using WGCNA, of which the pink and turquoise modules were found to be significantly enriched in the flavonoid biosynthesis pathway. As candidate hub genes, four structural genes (*CHI*, *C4H*, *PAL*, and *F3H*) that have been shown to play direct roles in the flavonoid pathway were screened [51–54]. This observation strongly suggested that *CHI*, *C4H*, *PA*L, and *F3H* may play crucial roles in the accumulation of flavonoids in *L. coreana* var. *sinensis.*

Candidate transcription factors related to flavonoid biosynthesis

TFs, bind to cis-acting elements in target genes via their special gene binding regions to activate the expression of genes [55]. The MYB proteins are classified into four main types that are called 1R-MYB (MYB-related), 2R-MYB (R2R3-MYB), 3R-MYB (R1R2R3-MYB), and 4R-MYB (R1R2R2R1/2-MYB) containing one to four MYB repeats, respectively [56]. Among these, R2R3-MYB is the most common and largest MYB TFs, playing a central role in regulating the transcriptional metabolism of flavonoid synthesis in plants [57]. Several members of the R2R3-MYB have been recognized; however, none have been reported in *L. coreana* var. *sinensis* previously. TRANSPARENT TESTA 2 (TT2)-type MYBs are an R2R3-MYBs that have been widely identified as regulators in the PA pathway [20]. TT2-type MYBs have also been found to regulate the biosynthesis of flavonols. It was worth noting that *FtMYB31*, a homologous gene of *AtMYB123/TT2*, was proved to enhance the accumulation of rutin and total flavonols by upregulating the *CHS*, *F3H*, and *FLS* genes in transgenic tobacoo [58]. The expression of *PpMYB123* was positively correlated with flavonols in nectarine, making it a promising candidate for flavonol regulation [59]. *PbMYB9* stimulated anthocyanin and flavonol production by binding to the *PbUFGT1* promoter [60]. Similarly, *LcMYB123 (NODE64739g19765i1)* was clustered with *AtMYB123/TT2*, *PbMYB9* and *MdMYB9*. It suggested that *LcMYB123* is able to activate the flavonol pathway; however, more research is needed to validate the functions of this gene.

Other types of TFs, such as NAC, ARF, and WD40, were also found to be central in regulating flavonoid synthesis. According to Yao et al. [61], *FtWD40* upregulated the expression of *NtDFR* and *NtANS* while downregulating *NtFLS*. In apple, *MdNAC9* may promote flavonol accumulation by activating *MdFLS* [62]. Overexpression of *PaNAC03* decreased flavonol biosynthesis by downregulating *CHS*, *F3'H*, and *PaLAR3* [15]. *ARF2* directly regulated the expression of *MYB12* and *FLS* genes in the flavonol pathway, and indirectly regulates *MYB11* and *MYB111* genes increased flavonol content in *Arabidopsis* [16]. Another study revealed that auxin impacts flavonol accumulation through ARF protein regulation of *CHS, CHI, FLS*, and *F3'H* [63]. Consequently, four TF genes (*R2R3-MYB*, *NAC*, *ARF*, *WD40*) screened in this study might influence flavonoid biosynthesis by modulating the expression of structural genes or TF genes. Thus, this

work is beneficial as a reference to reveal the complex regulatory mechanism of flavonoid biosynthesis in *L. coreana* var. *sinensis* in the future.

Conclusion

In this study, we sequenced full-length transcriptomes in *L. coreana* var. *sinensis* using the ONT sequencing technologies. A total of 126,977 unigenes were found from the generated data, with 107,977 successfully annotated in seven established public databases. Through a comprehensive transcriptome analysis combined with WGCNA, eight candidate genes influencing flavonoid biosynthesis were identified, including four structural genes and four TF genes. The transcription factor LcMYB123 was suggested to have a potential role in the biosynthesis of flavonols in *L. coreana* var. *sinensis.* This research presents valuable insights into candidate genes that might be instrumental in the future genetic manipulation and improvement of *L. coreana* var. *sinensis*.

Abbreviations

ONT, Oxford Nanopore Technology; Q-3-gal, Quercetin-3-O-*β*-D-galactoside; Q-3-glu, Quercetin-3-O-*β*-D-glucoside; K-3-gal, Kaempferol-3-O-*β*-D-galactoside; Q-3-glu, Kaempferol-3-O-*β*-D-glucoside; PAL, Phenylalanine ammonia-lyase; C4H, Cinnamate 4-hydroxylase; 4CL, 4-coumarate: CoA ligase; CHS, Chalcone synthase; CHI, Chalcone isomerase; F3H, Flavone 3-hydroxylase; F3'H, Flavonoid 3-hydroxylase; FLS, Flavonol synthase; UGT, UDP-glucosyltransferase; SGS, Second-generation sequencing; TGS, Third-generation sequencing; PacBio, Pacific Bioscience; WGCNA, Weighted gene co-expression network analysis; TF, Transcription factors; HPLC, High-performance liquid chromatography

Declarations

Supplementary information

Supplementary material associated with this article can be found in the online version.

Acknowledgements

Dr. Qiling Song and Fenfang Long are thanked for their valuable advice of the manuscript.

Authors' contributions

Q.Q. Guo and H.E Li designed the study, and revised the manuscript. N. Xie, G.Y. Yuan and Q. Gui performed the experiments. Y. Xiao, M.Y. Liao and L. Yang analyzed data. N. Xie wrote the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (32060349).

Data availability

The sequenced raw reads generated in this study have been submitted to the NCBI SRA database with BioProject: PRJNA992466.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Conpeting interest

The authors declare no conflict of interest.

References

1. Yuan GY, Guo QQ, Xie N, Zhang YQ, Luo SQ, Gui Q, Yao S. Population dynamics and protection strategies of *Litsea coreana* Levl. var. *sinensis* with Guizhou Province in China as an example. Journal of Mountain Science. 2023;20(4):1090–101.
2. Jia XJ, Li P, Wan JB, He CW. A review on phytochemical and pharmacological properties of *Litsea coreana*. Pharm Biol . 2017;55(1):1368–74.
3. Shi J, Yang GZ, You QS, Sun SL, Chen RH, Lin Z, Simal-Gandara J, Lv HP. Updates on the chemistry, processing characteristics, and utilization of tea flavonoids in last two decades (2001-2021). Critical Reviews in Food Science and Nutrition. 2023;63(20):4757–84.
4. Chen AY, Chen YC. A review of the dietary flavonoid, kaempferol on human health and cancer chemoprevention. Food Chemistry. 2013;138(4):2099–107.
5. Tao H, Li LY, He YQ, Zhang XY, Zhao Y, Wang QM, Hong GJ. Flavonoids in vegetables: improvement of dietary flavonoids by metabolic engineering to promote health. Critical Reviews in Food Science and Nutrition. 2022;:1–15.
6. Ma TT, Huang C, Meng XM, Zhang QL, Zhang L, Lv XW, Jin Y, Xie J, Li Jun. Fingerprint analysis of hawk-tea by high-performance liquid chromatography. Food Chemistry. 2011;129(1):551–6.
7. Chen ZL, Ma TT, Huang C, Zhang L, Zhong J, Han JW, Hu TT, Li J. Efficiency of transcellular transport and efflux of flavonoids with different glycosidic units from flavonoids of *Litsea coreana* L. in a MDCK epithelial cell monolayer model. European Journal of Pharmaceutical Sciences. 2014;53:69–76.
8. Li S, Pan JH, Hu X, Zhang Y, Gong DM, Zhang GW. Kaempferol inhibits the activity of pancreatic lipase and its synergistic effect with orlistat. Journal of Functional Foods. 2020;72:104041.

9. Deliorman-Orhan D, Şenol S, Kartal M, Orhan I. Assessment of antiradical potential of *Calluna vulgaris* (L.) Hull and its major flavonoid. J Sci Food Agric. 2009;89(5):809–14.
10. Lesjak M, Beara I, Simin N, Pintać D, Majkić T, Bekvalac K, Orčić D, Mimica-Dukić N. Antioxidant and anti-inflammatory activities of quercetin and its derivatives. Journal of Functional Foods. 2018;40(8):68–75.
11. Lin N, Liu XY, Zhu WF, Cheng X, Wang XH, Wan XC, Liu LL. Ambient ultraviolet B signal modulates tea flavor characteristics via shifting a metabolic flux in flavonoid biosynthesis. J Agric Food Chem. 2021;69(11):3401–14.
12. Xia X, Gong R, Zhang CY. Integrative analysis of transcriptome and metabolome reveals flavonoid biosynthesis regulation in *Rhododendron pulchrum* petals. BMC Plant Biol. 2022;22(1):401.
13. Saito K, Yonekura-Sakakibara K, Nakabayashi R, Higashi Y, Yamazaki M, Tohge T, Fernie AR. The flavonoid biosynthetic pathway in *Arabidopsis*: structural and genetic diversity. Plant Physiol Biochem. 2013;72:21–34.
14. Ishihara H, Tohge T, Viehöver P, Fernie AR, Weisshaar B, Stracke R. Natural variation in flavonol accumulation in *Arabidopsis* is determined by the flavonol glucosyltransferase BGLU6. EXBOTJ. 2016;67(5):1505–17.
15. Dalman K, Wind JJ, Nemesio-Gorriz M, Hammerbacher A, Lundén K, Ezcurra I, Elfstrand M. Overexpression of *PaNAC03*, a stress induced NAC gene family transcription factor in Norway spruce leads to reduced flavonol biosynthesis and aberrant embryo development. BMC Plant Biol. 2017;17(1):6.
16. Jiang WB, Xia Y, Su XJ, Pang Y. ARF2 positively regulates flavonols and proanthocyanidins biosynthesis in *Arabidopsis thaliana*. Planta. 2022;256(2):44.
17. Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F, Niehaus K, Weisshaar B. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. Plant J. 2007;50(4):660–77.
18. Li PH, Chen BB, Zhang GY, Chen LX, Dong Q, Wen JQ, Mysore KS, Zhao J. Regulation of anthocyanin and proanthocyanidin biosynthesis by *Medicago truncatula* bHLH transcription factor MtTT8. New Phytol. 2016;210(3):905–21.
19. Schilbert HM, Glover BJ. Analysis of flavonol regulator evolution in the Brassicaceae reveals *MYB12*, *MYB111* and *MYB21* duplications and *MYB11* and *MYB24* gene loss. BMC Genomics. 2022;23(1):604.
20. Nesi N, Jond C, Debeaujon I, Caboche M, Lepiniec L. The *Arabidopsis TT2* gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. Plant Cell. 2001;13(9):2099–114.
21. Muñoz-Gómez S, Suárez-Baron H, Alzate JF, González F, Pabón-Mora N. Evolution of the Subgroup 6 R2R3-MYB genes and their contribution to floral color in the perianth-bearing piperales. Front Plant Sci. 2021;12:633227.

22. Liu YJ, Hou H, Jiang XL, Wang PQ, Dai XL, Chen W, Gao LP, 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.
23. Zhang BH, Yao XZ, Chen HF, Lu. High-quality chromosome-level genome assembly of *Litsea coreana* L. provides insights into Magnoliids evolution and flavonoid biosynthesis. Genomics. 2022;114(4):110394.
24. Xiao F, Zhao Y, Wang XR, Liu Q, Ran J. Transcriptome analysis of needle and root of *Pinus Massoniana* in response to continuous drought stress. Plants. 2021;10(4):769.
25. Zhu JY, Xu QS, Zhao SQ, Xia XB, Yan XM, An YL, Mi XZ, Guo LX, Samarina L, Wei CL. Comprehensive co-expression analysis provides novel insights into temporal variation of flavonoids in fresh leaves of the tea plant (*Camellia sinensis*). Plant Science. 2020;290:110306.
26. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen, YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437(7057):376–80.
27. Wang F, Chen Z, Pei HM, Guo ZY, Wen D, Liu R, Song BX. Transcriptome profiling analysis of tea plant (*Camellia sinensis*) using oxford nanopore long-read RNA-Seq technology. Gene. 2021;769:145247.
28. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, deWinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murph D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. Real-time DNA sequencing from single polymerase Molecules. Science. 2009;323(5910):133–8.
29. Li XH, Yu KQ, Li FY, Lu WX, Wang Y, Zhang WZ, Bai YF. Novel method of full-length RNA-seq that expands the identification of non-polyadenylated RNAs using nanopore sequencing. Anal Chem. 2022;94(36):12342–51.
30. Wong LL, Razali SA, Deris ZM, Danish-Daniel M, Tan MP, Nor, SAM, Ma HY, Min W, Yantao L, Asaduzzaman M, Sung YY, Liu Z, Sorgeloos P, Van De Peer, Y., Afiqah-Aleng N. Application of second-generation sequencing (SGS) and third generation sequencing (TGS) in aquaculture breeding program. Aquaculture. 2022;548:737633.
31. Hou SY, Du W, Hao YR, Han YH, Li HY, Liu LL, Zhang KX, Zhou ML, Sun ZX. Elucidation of the regulatory network of flavonoid biosynthesis by profiling the metabolome and transcriptome in tartary buckwheat. J Agric Food Chem. 2021;69(25):7218–29.

32. Qian MJ, Wu HX, Yang C, Zhu W, Shi B, Zheng B, Wang SB, Zhou KB, Gao AP. RNA-Seq reveals the key pathways and genes involved in the light-regulated flavonoids biosynthesis in mango (*Mangifera indica* L.) peel. Front Plant Sci. 2023;13:1119384.
33. Ju YL, Wang WN, Yue XF, Xue W, Zhang YL, Fang YL. Integrated metabolomic and transcriptomic analysis reveals the mechanism underlying the accumulation of anthocyanins and other flavonoids in the flesh and skin of teinturier grapes. Plant Physiology and Biochemistry. 2023;197:107667.
34. Lu WL, Chen YP, Tang WJ, Li, Jun. Determination of main flavonoids in eagle tea by HPLC. Acta Universitatis Medicinalis Anhui. 2009;44(5):653–4.
35. Zuo WS, Xu J, Zhou CX, Gan LS. Simultaneous determination of five flavonoid compounds in leaves of *Adinandra nitida* by HPLC PAD. China Journal of Chinese Materia Medica. 2010;35(18):2406–9.
36. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644–52.
37. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9(1):559.
38. Chen CJ, Chen H, Zhang Y, Thomas HR, Frank MH, He YH, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Molecular Plant. 2020;13(8):1194–202.
39. 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.
40. Naik J, Rajput R, Pucker B, Stracke R, Pandey A. The R2R3-MYB transcription factor MtMYB134 orchestrates flavonol biosynthesis in *Medicago truncatula*. Plant MolBiol. 2021;106(1):157–72.
41. Kubina R, Iriti M, Kabała-Dzik A. Anticancer potential of selected flavonols: fisetin, kaempferol, and quercetin on head and neck cancers. Nutrients. 2021;13(3):845.
42. Yu JP, Ye H, Gu LQ. The flavonoids from *Litsea coreana*. Acta Scientiarum Naturalium Universitatis Sunyatseni. 2001;40(6):110-111+114.
43. Chen, YP, Cheng WM, Li J. Analyse on the chemical constituents from flavonoids of *Litsea coreana* L. Acta Universitatis Medicinalis Anhui. 2008;43(1):65–7.
44. Travisany D, Ayala-Raso A, Di Genova A, Monsalve L, Bernales M, Martínez JP, González-Agüero M, Defilippi B, Cherian S, Maass A, Fuentes L. RNA-seq analysis and transcriptome assembly of raspberry fruit (*Rubus idaeus* ¨Heritage¨) revealed several candidate genes involved in fruit development and ripening. Scientia Horticulturae. 2019;254:26–34.
45. Winkel-Shirley B. Flavonoid Biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiology. 2001;126(2):485–93.
46. Petit P, Granier T, d'Estaintot BL, Manigand C, Bathany K, Schmitter JM, Lauvergeat V, Hamdi S, Gallois B. Crystal structure of grape dihydroflavonol 4-reductase, a key enzyme in flavonoid biosynthesis. J Mol Biol. 2007;368(5):1345–57.

47. Wang FF, Su YL, Chen NZ, Shen SH. Genome-wide analysis of the UGT gene family and identification of flavonoids in *Broussonetia papyrifera*. Molecules. 2021;26(11):3449.
48. Li YT, Zhang J, Wang SJ, Zhang HE, Liu YC, Yang MS. Integrative transcriptomic and metabolomic analyses reveal the flavonoid biosynthesis of *Pyrus hopeiensis* flowers under cold stress. Horticultural Plant Journal. 2022;9(3):395–413.
49. Khan IA, Cao K, Guo J, Li Y, Wang Q, Yang XW, Wu JL, Fang WC, Wang LR. Identification of key gene networks controlling anthocyanin biosynthesis in peach flower. Plant Science. 2022;316:111151.
50. Peng X, Wu H, Chen HJ, Zhang YJ, Qiu D, Zhang ZY. Transcriptome profiling reveals candidate flavonol-related genes of Tetrastigma hemsleyanum under cold stress. BMC Genomics. 2019(1);20:687.
51. Feng HJ, Tian XH, Liu YC, Li YJ, Zhang XY, Jones BJ, Sun YQ, Sun J. Analysis of flavonoids and the flavonoid structural genes in brown fiber of upland cotton. PLoS ONE. 2013;8(3):e58820.
52. Medda S, Sanchez-Ballesta MT, Romero I, Dessena L, Mulas M. Expression of structural flavonoid biosynthesis genes in dark-blue and white myrtle berries (*Myrtus communis* L.). Plants. 2021(2);10:316.
53. Song SS, Tao Y, Gao LH, Liang HL, Tang D, Lin J, Wang YC, Gmitter FG, Li CF. An integrated metabolome and transcriptome analysis reveal the regulation mechanisms of flavonoid biosynthesis in a purple tea plant cultivar. Front Plant Sci. 2022;13:880227.
54. Zhang K, Sun Y, Li MN, Long RC. CrUGT87A1, a UDP-sugar glycosyltransferases (UGTs) gene from *Carex rigescens*, increases salt tolerance by accumulating flavonoids for antioxidation in *Arabidopsis thaliana*. Plant Physiology and Biochemistry. 2021;159:28–36.
55. Smita S, Katiyar A, Chinnusamy V, Pandey DM, Bansal KC. Transcriptional regulatory network analysis of MYB transcription factor family genes in rice. Front Plant Sci. 2015;6.
56. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. MYB transcription factors in *Arabidopsis*. Trends in Plant Science. 2010;15(10):573–81.
57. Wang XY, Tian L, Feng SJ, Wei AZ. Identifying potential flavonoid biosynthesis regulator in *Zanthoxylum bungeanum* Maxim. by genome-wide characterization of the MYB transcription factor gene family. J Integr Agric. 2022;21(7):1997–2018.
58. Sun ZX, Linghu B, Hou SY, Liu RH, Wang L, Hao YR, Han YH, Zhou ML, Liu LL, Li HY. Tartary buckwheat *FtMYB31* gene encoding an R2R3-MYB transcription factor enhances flavonoid accumulation in tobacco. J Plant Growth Regul. 2020;39(2):564–74.
59. Ravaglia D, Espley RV, Henry-Kirk RA, Andreotti C, Ziosi V, Hellens RP, Costa G, Allan AC. Transcriptional regulation of flavonoid biosynthesis in nectarine (*Prunus persica*) by a set of R2R3 MYB transcription factors. BMC Plant Biol. 2013;13(1):68.
60. Zhai R, Wang ZM, Zhang SW, Meng G, Song LY, Wang ZG, Li PM, Ma FW, Xu LF. Two MYB transcription factors regulate flavonoid biosynthesis in pear fruit (*Pyrus bretschneideri* Rehd.). EXBOTJ. 2016;67(5):1275–84.

61. Yao PF, Zhao HX, Luo XP, Gao F, Li CL, Yao HP, Chen H, Park SU, Wu Q. Fagopyrum tataricum *FtWD40* functions as a positive regulator of anthocyanin biosynthesis in transgenic tobacco. J Plant Growth Regul. 2017;36(3):755–65.
62. Sun QG, Jiang SH, Fang HC, Zhang TL, Wang N, Chen XS. Cloning of *MdNAC9* and functional of its regulation on flavonol synthesis. Acta Horticulturae Sinica. 2019;46(11):2073–81.
63. Lewis DR, Ramirez MV, Miller ND, Vallabhaneni P, Ray WK, Helm RF, Winkel BSJ, Muday GK. Auxin and ethylene induce flavonol accumulation through distinct transcriptional networks. Plant Physiology. 2011;156(1):144–64.

Figures

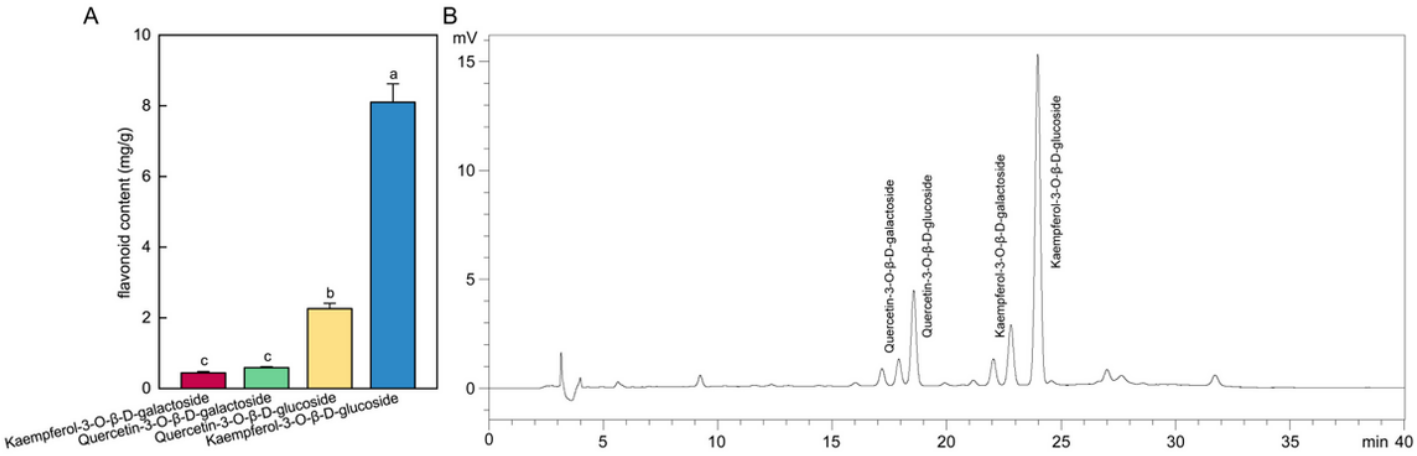

Figure 1

Flavonoid content analysis of *L. coreana* var. *sinensis.* (A) Flavonoid contents assessed by HPLC fingerprint analysis (mg/g dry weight). Different letters indicate significant differences ($P < 0.05$). (B) Typical HPLC chromatograms of *L. coreana* var. *sinensis.*

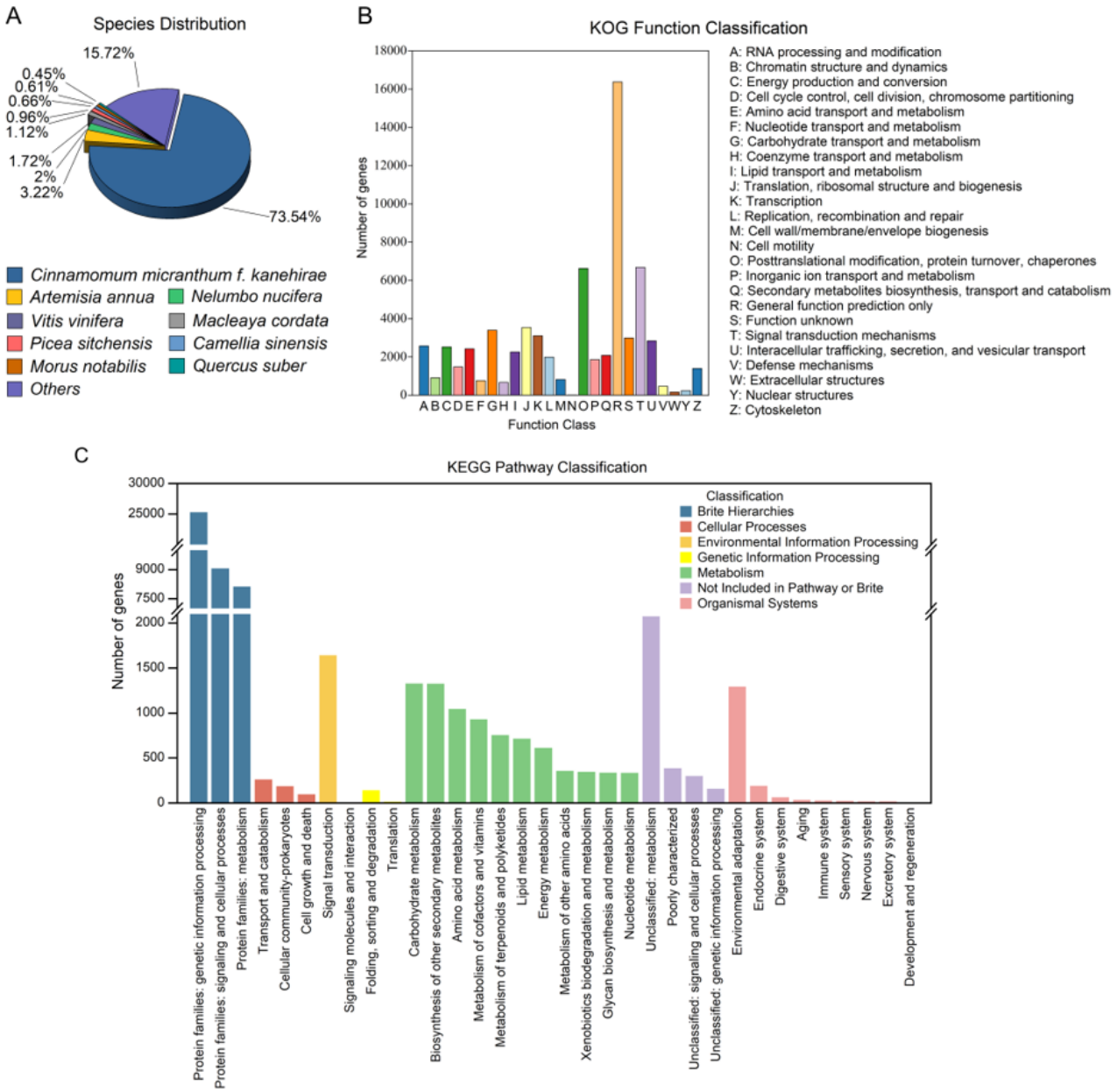

Figure 2

Gene annotation and functional classification. (A) Homologous species distribution annotated in the NR database. (B) The KOG categories of the unigenes. (C) Annotation in KEGG pathwayclassification.

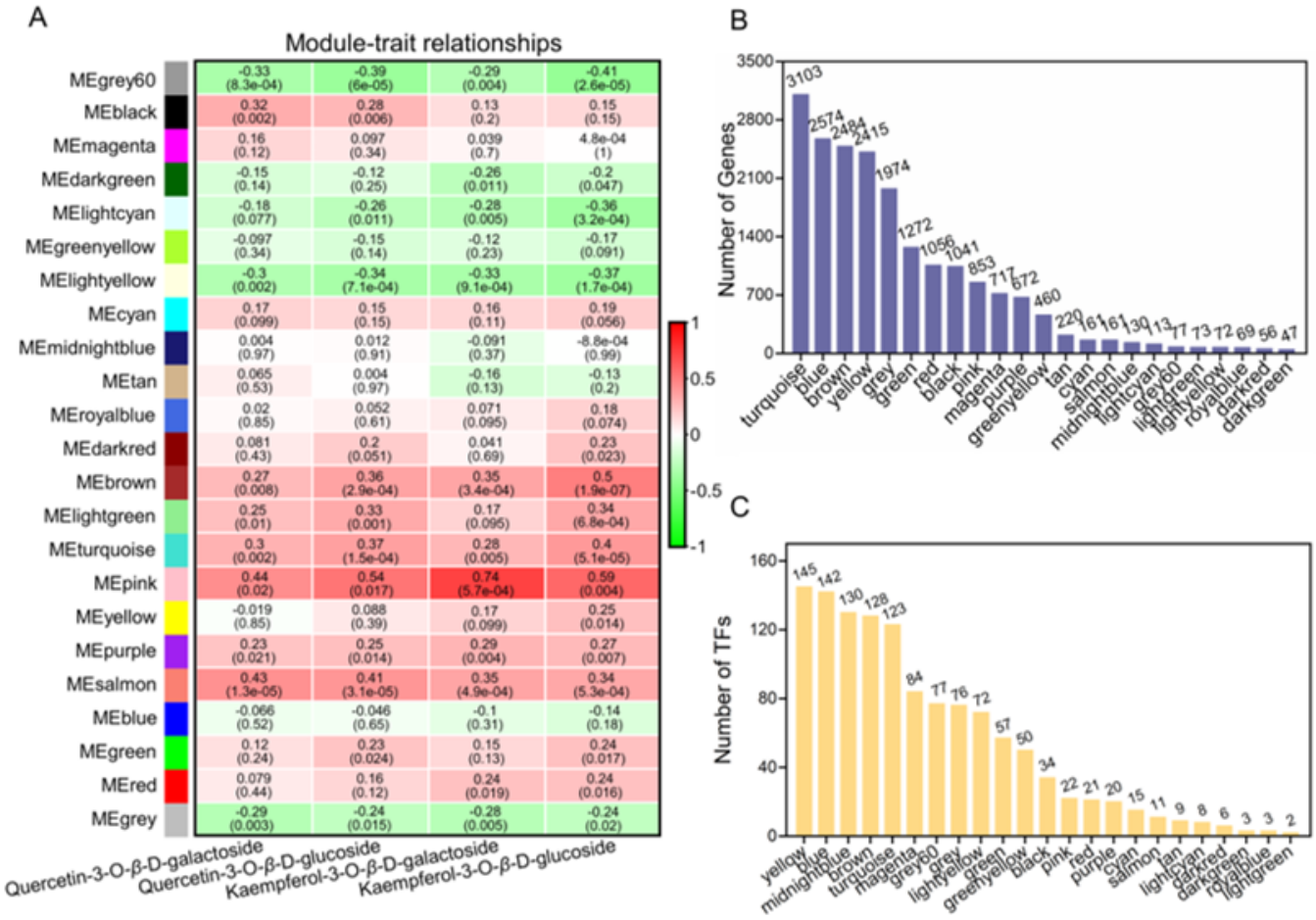

Figure 3

The results of WGCNA. (A) Correlation coefficient between flavonoids and module eigengenes with red and blue representing positive and negative correlations, respectively. Each column corresponds to K-3-gal, Q-3-glu, Q-3-gal and K-3-glu values, and each row represents a module. p-values are shown in parentheses. (B) The number of genes in each module. (C) The number of TFs in each module.

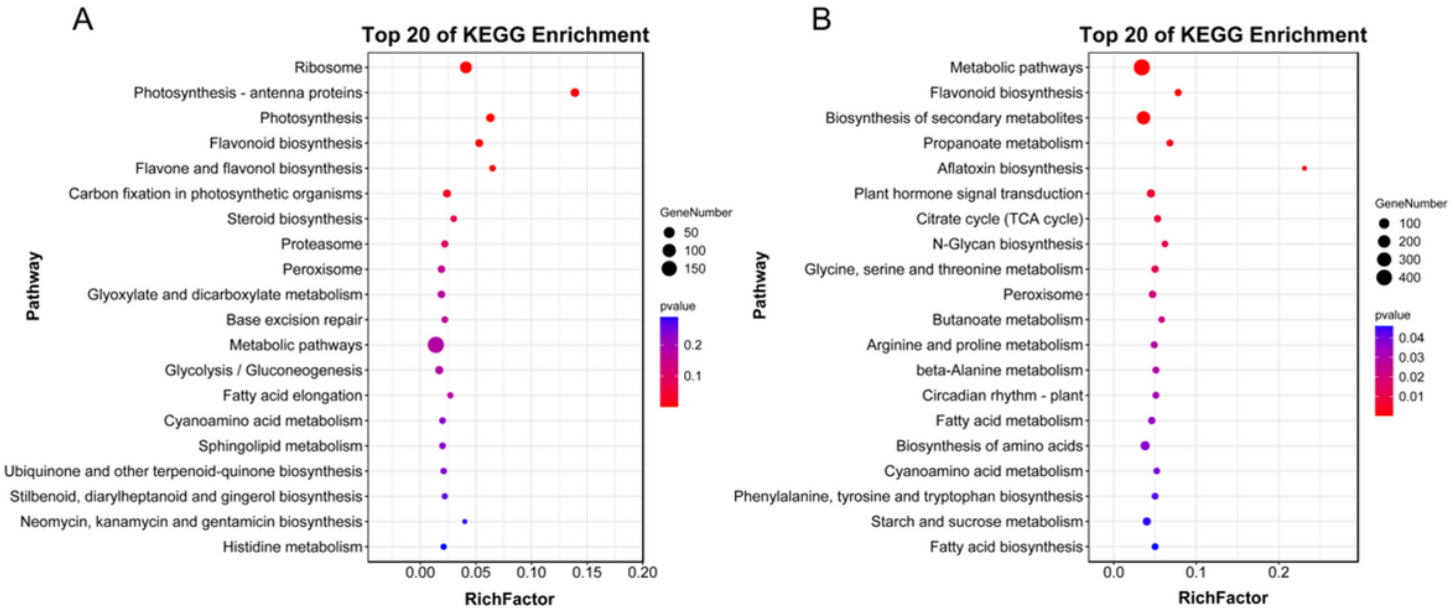

Figure 4

Bubble maps of top 20 KEGG-enriched pathways of genes in pink and turquoise modules. (A) pink module. (B) turquoise module. The x-axis and y-axis represent the enrichment factor and pathway name, respectively.

Figure 5

Hub gene visualization of the pink and turquoise modules. (A) pink module. (B) turquoise module. The darker the color, the greater the degree value .

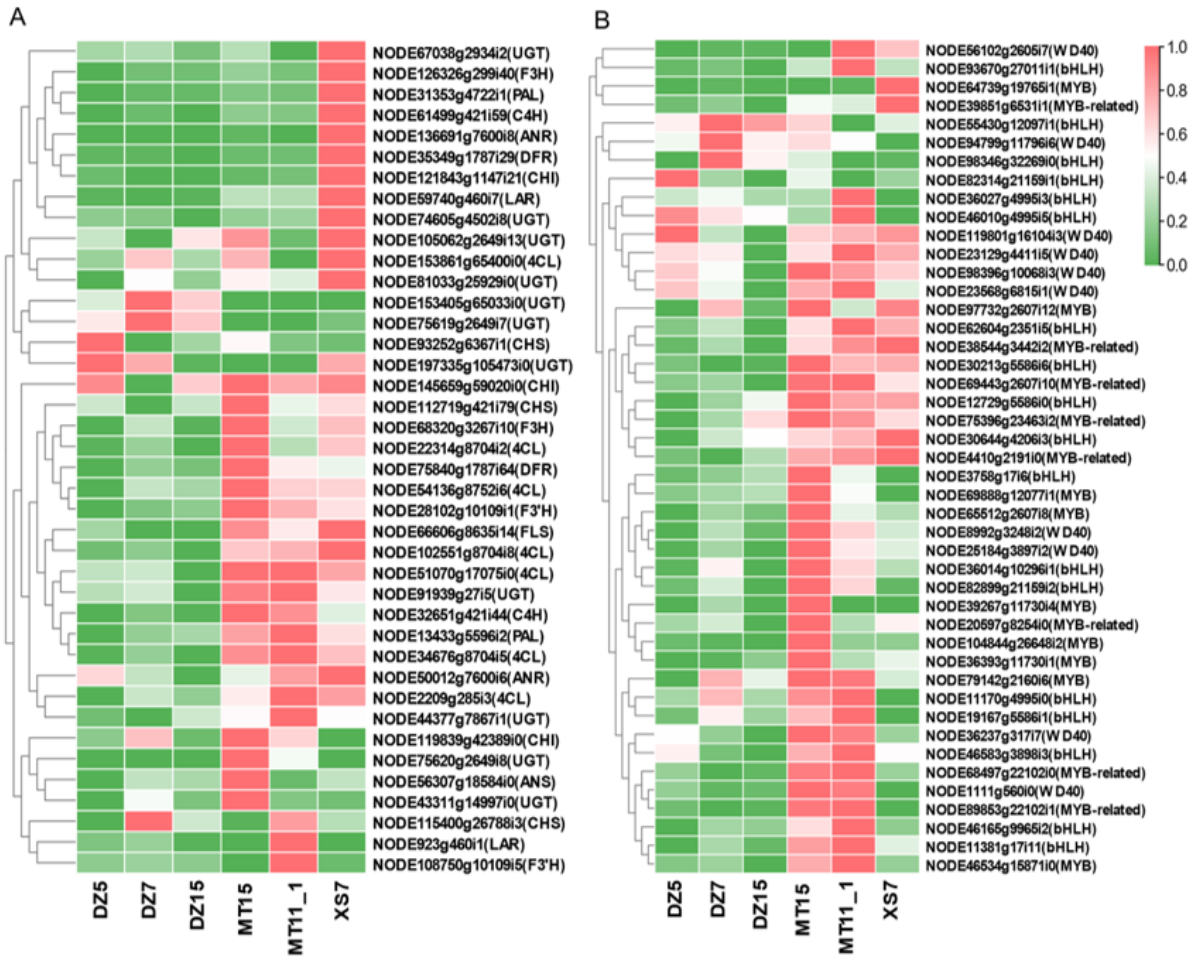

Figure 6

Expression patterns of genes involved in flavonoid biosynthesis pathways in pink and turquoise modules. (A) Structural genes expression levels. (B) TFs genes expression levels. The color scale from blue to red represents the TPM values from low to high. DZ5, DZ7, DZ15: the three samples with the lowest total amount of the K-3-gal, Q-3-glu, Q-3-gal and K-3-glu. MT15, MT11_1, XS7: the three samples with the highest total amount of the four K-3-gal, Q-3-glu, Q-3-gal and K-3-glu.

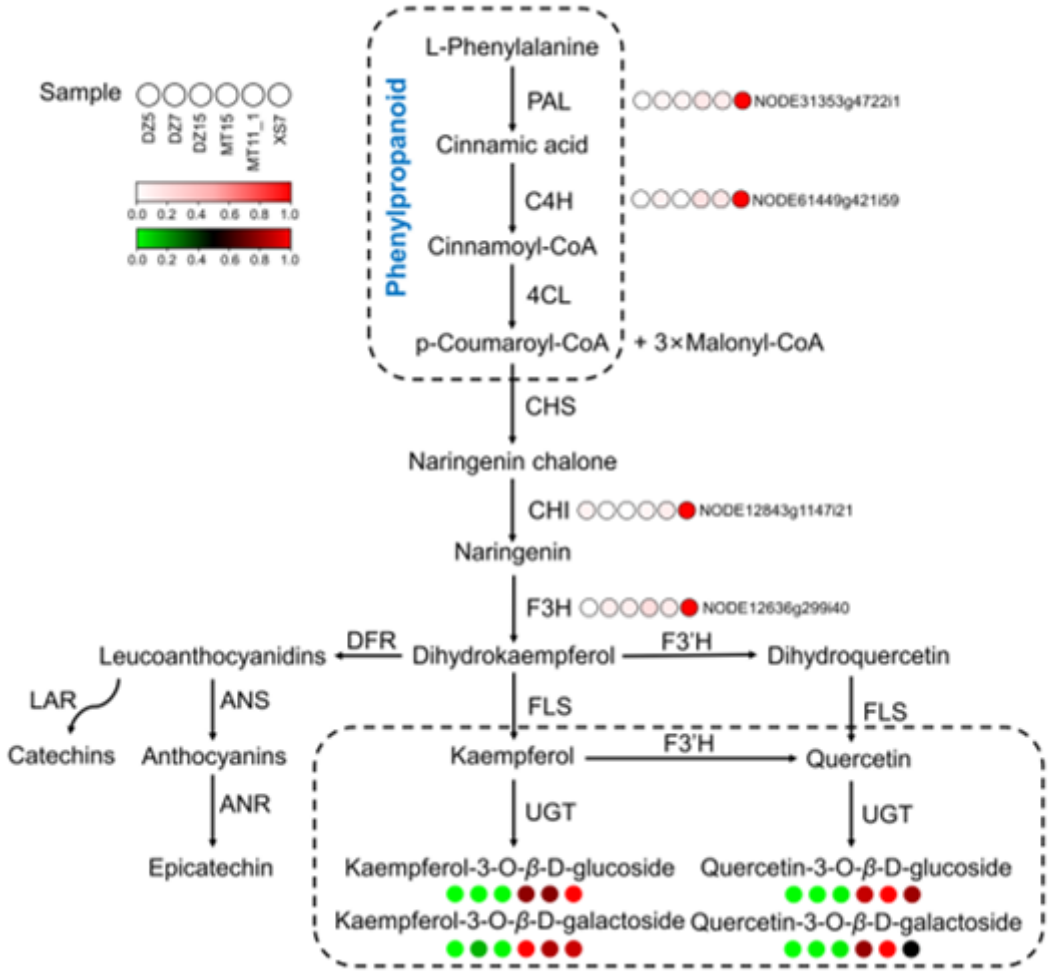

Figure 7

Visualization of transcriptional expression of putative key enzymes in the flavonoid biosynthesis pathway [40].

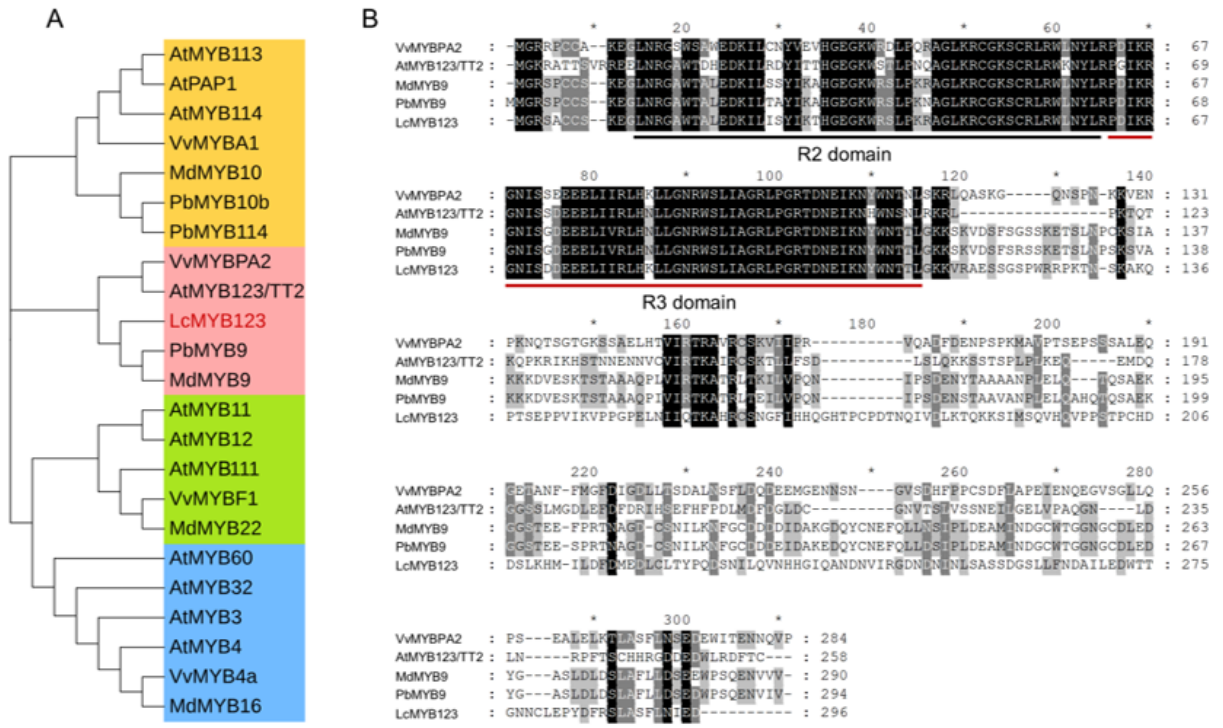

Figure 8

Phylogenetic analysis and protein sequence alignment of *LcMYB123*. (A) Phylogenetic analysis of flavonoid-regulating MYBs in different plants. (B) Multi-alignment of the protein sequence of MYBs from different species. The R2 and R3 domains were underlined. Accession numbers: AtMYB113 (AT1G66370.1), AtPAP1 (AAG42001), AtMYB114 (AT1G66380.1), VvMYBA1 (BAD18977), MdMYB10 (ACQ45201.1), PbMYB10b (KT601122), PbMYB114 (MF489219), MdMYB9 (ABB84757.1), AtMYB11 (NP_191820), AtMYB12 (NC_003071.7), AtMYB111 (AAK97396), VvMYBF1 (NP_001267930.1), MdMYB22 (AAZ20438.1), AtMYB60 (AT1G08810.1), AtMYB32 (AT4G34990.1), AtMYB3 (AEE30263.1), AtMYB4 (AT4G38620), VvMYB4a (ABL61515.1), MdMYB16 (HM122617).

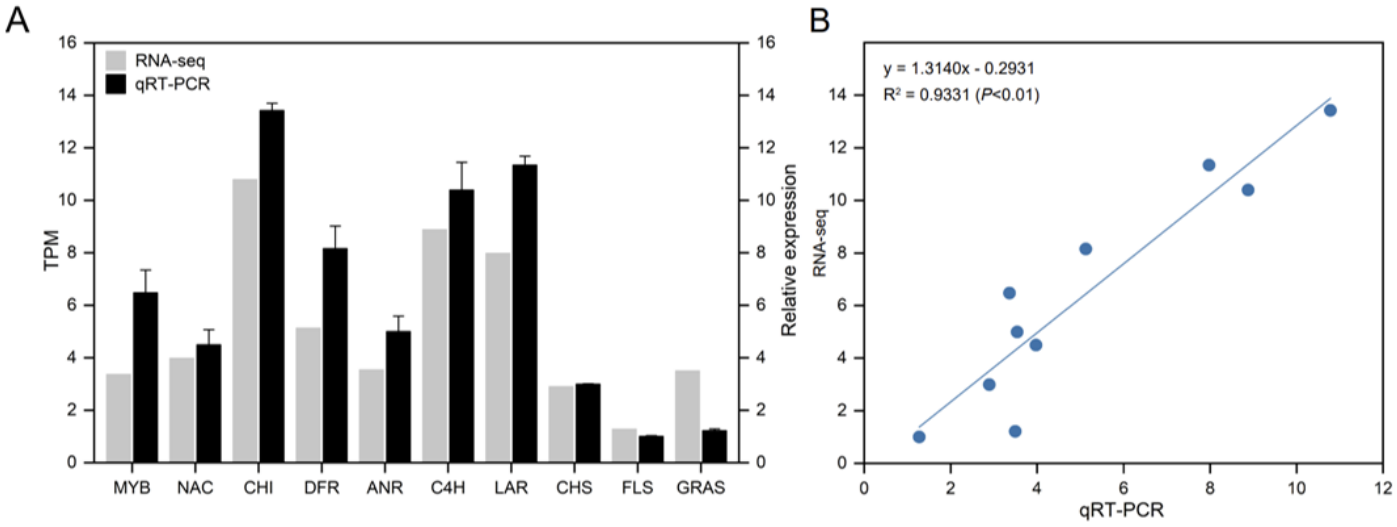

Figure 9

qRT-PCR verification. (A) Expression of 10 genes related to flavonoid biosynthesis. (B) Pearson's correlation of gene expression between RNA-seq and qRT-PCR.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.docx