

Long-read sequencing of *Chrysanthemum morifolium* cv. 'Hangju' transcriptome reveals flavonoid biosynthesis and regulation

Tao Wang

Nanjing Agricultural University

Feng Yang

Nanjing Agricultural University

Qiaosheng Guo (✉ gqs@njau.edu.cn)

Institute of Chinese Medicinal Materials, Nanjing Agricultural University

Qingjun Zou

Nanjing Agricultural University

Wenyan Zhang

Nanjing Agricultural University

Lin Zuo

Nanjing Agricultural University

Research article

Keywords: *Chrysanthemum morifolium*; transcriptome; SMRT-Seq; flavonoid

Posted Date: December 3rd, 2019

DOI: <https://doi.org/10.21203/rs.2.17990/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The inflorescence of *Chrysanthemum morifolium* cv. 'Hangju' has been widely used in China due to its antioxidant and anti-inflammatory properties. The biosynthesis and regulation of flavonoids, a group of bioactive components, in *C. morifolium* are poorly understood. Transcriptome sequencing is an effective method for obtaining transcript information. Therefore, single-molecule real-time (SMRT) sequencing was performed to obtain the full-length genes involved in flavonoid biosynthesis and regulation in *C. morifolium*. **Results:** High-quality RNA was extracted from the inflorescence of *C. morifolium* at different developmental stages and used to construct two libraries (0-5 kb and 4.5-10 kb) for sequencing. Finally, 125,532 non-redundant isoforms with a mean length of 2,009 bp were obtained. Of these, 2,083 transcripts were annotated to pathways related to flavonoid biosynthesis, and 56 isoforms were annotated as *CHS*, *CHI*, *F3H*, *F3'H*, *FNS*, *FLS*, *DFR* and *ANS* genes. Based on gene expression levels at different stages, we predicted the major genes involved in flavonoid biosynthesis. By phylogenetic analysis, we found two candidate MYB transcription factors (*CmMYBF1* and *CmMYBF2*) activating flavonol biosynthesis. **Conclusions:** Based on the full-length transcriptomic data and further quantitative analysis, the major genes involved in flavonoid biosynthesis and regulation in *C. morifolium* were predicted in our study. The results provide a valuable theoretical basis for the introduction and cultivation of *C. morifolium* cv. 'Hangju'.

Background

Dried inflorescence of *Chrysanthemum morifolium* Ramat. (Asteraceae) have been widely used in China for thousands of years and have significant medicinal and economic value. According to the Chinese Pharmacopoeia, medicinal *C. morifolium* can be mainly classified into five varieties, namely, 'Hangju', 'Boju', 'Chuju', 'Gongju' and 'Huaiju'. Many pharmacological studies have reported that medicinal *C. morifolium* has antioxidative [1-3], cardiovascular protective [4, 5], antimicrobial [6] and anti-inflammatory [7, 8] activities, which are mainly due to its active components, such as flavonoids. The *C. morifolium* variety 'Hangju' has been cultivated for centuries and includes several genotypes, with the genotype 'Zaoyang' having the largest cultivated area and yield [9]. Furthermore, the genotype 'Zaoyang' from Tongxiang County (the traditional cultivation area) is considered superior to genotypes from other cultivation areas due to its high luteolin content [10], which suggests that flavonoids are the vital components in genuine medicinal *C. morifolium*.

Flavonoids are a large group of secondary metabolites in plants whose biosynthetic pathways are well established as useful models for studying metabolic regulation. The genes involved in the flavonoid biosynthesis pathways are mainly divided into early and late biosynthetic genes according to their location in the pathways. Early biosynthetic genes (EBGs) including *CHS*, *CHI*, *F3H*, *F3'H* and *FLS* catalyse flavonol biosynthesis, whereas late biosynthetic genes (LBGs: *DFR*, *ANS* and *ANR*) lead to both anthocyanin and proanthocyanidin biosynthesis [11-13]. Other genes such as *FNS* and *IFS* can also catalyse flavone and isoflavone biosynthesis [14, 15]. Furthermore, the transcription factors regulating such biosynthesis have been identified in many species. *AtMYB11*, *AtMYB12* and *AtMYB111* in

Arabidopsis thaliana activate the expression of LEGs to produce flavonol [16, 17], and high expression of *MdMYB1* and *MdMYB10* increases the anthocyanin content in *Malus × domestica* [18]. However, flavonoid biosynthesis and regulation in *C. morifolium* are poorly understood because of the lack of a reference genome database. Thus, transcriptome sequencing offers an alternative approach for obtaining gene information. Second-generation sequencing (SGS) was used to identify the differentially expressed genes in *C. morifolium* under stress or in different tissues [19-22]. However, the short-length reads generated by SGS decrease the accuracy of transcript assembly, especially for species without genome information. Single-molecule real-time (SMRT) sequencing, as a third-generation sequencing (TGS) technology, has the advantage of generating high-throughput long reads and thus avoids the inaccuracy resulting from assembling short sequences [23, 24].

In this study, SMRT sequencing was performed to obtain full-length transcripts for better understanding flavonoid biosynthesis and regulation in *C. morifolium*. The key genes were screened based on annotation analysis, and their expression profiles were analysed at different developmental stages. Furthermore, the candidate MYB transcription factors activating flavonol biosynthesis were identified, which not only provide an understanding of flavonoid biosynthesis and regulation in *C. morifolium* but also will advance the introduction and cultivation of *C. morifolium*.

Results

The full-length transcriptome of *C. morifolium*

To identify the transcripts in *C. morifolium* during the developmental stages, SMRT sequencing technology was used to obtain the complete and full-length transcriptomes. In total, we obtained 15.56 Gb and 13.93 Gb of raw data from the 0-4.5K and 4.5-10K libraries, respectively. Approximately 591,513 and 560,454 ROIs (reads of insert) were recognized, with mean lengths of 3,328 and 3,587 bp, respectively (Table 1). The clean data were uploaded to the Sequence Read Archive (SRA) with accession numbers SRR10054190 and SRR10054977. More than 50% of the reads were classified as full-length (FL) non-chimeric reads, and fewer than 5% of the reads were classified as chimeric reads, which indicated that the SMRTbell libraries were of high quality (Supplementary figure 1). After clustering and polishing, 129,610 and 112,142 HQ isoforms were finally merged into 125,532 non-redundant isoforms with a mean length of 2,009 bp and an N50 of 2,274 bp (Table 1).

Functional annotation

All non-redundant isoforms were searched against 7 functional databases, and approximately 115,807 (92.25%) transcripts were annotated in at least one functional database (Figure 2A). There were 107,700 (85.79%) isoforms assigned to NR databases, and most of them had high similarity to genes in *Helianthus annuus* (60.13%), *Cynara cardunculus* var. *scolymus* (24.18%), *Chrysanthemum × morifolium* (1.66%) and other species (Figure 2B). Approximately 33,901 (27.01%) transcripts were assigned to 53 subcategories in the GO database, and the largest subcategories of the three primary GO categories were cellular process, cell and catalytic activity (Supplementary figure 2). We also obtained annotations for

91,179 (72.63%) isoforms from the KEGG database, 2,411 of which were associated with the Biosynthesis of other secondary metabolites subgroup. Furthermore, we found that 1,447, 480, 54 and 102 transcripts were annotated as belonging to the Phenylpropanoid biosynthesis, Flavonoid biosynthesis, Anthocyanin biosynthesis, Isoflavonoid biosynthesis, and Flavone and flavonol biosynthesis pathways, respectively (Figure 2C).

The candidate genes involved in the flavonoid biosynthesis pathway

Combining the annotation results from seven databases, 143 isoforms were annotated as *PAL*, *C4H*, *4CL*, *CHS*, *CHI*, *F3H*, *F3'H*, *FLS*, *FNS*, *DFR* and *ANS* genes in the flavonoid biosynthesis pathways (Table 2). Every gene had many isoforms, especially those in the *PAL* and *4CL* gene families, with more than 30 isoforms. Interestingly, no sequences were similar to *F3'5'H*, which catalyses the substrate of blue anthocyanins. This finding is consistent with the lack of blue varieties of medicinal *C. morifolium*. *ANR* and *LAR* genes also were not found in the database, which suggested a lack of proanthocyanins in medicinal *C. morifolium*. Further analysis of the genes involved in luteolin and quercetin biosynthesis revealed that the isoforms belonged to 5 *CHS* genes, 3 *CHI* genes, 1 *F3H* gene, 1 *F3'H* gene, 1 *FNS* gene and 2 *FLS* genes based on sequence similarity (Supplementary table 1).

Differential expression of genes involved in luteolin and quercetin biosynthesis

Luteolin and quercetin are the marker compounds of 'Hangju' in the Chinese Pharmacopoeia; thus, the *CHS*, *CHI*, *F3H*, *F3'H*, *FLS* and *FNS* genes were selected and analysed. The expression profile of every gene at the different flowering stages is shown in Figure 3. Most of the genes had the highest expression level at the bud stage (stage A), which is an important stage for flavonoid biosynthesis. Expression of *CHS1* was significantly higher than that of the other *CHS* genes at every stage, while *CHS4* and *CHS5* were barely expressed. Therefore, *CHS1* might be a major gene controlling flavonoid production in the inflorescence of *C. morifolium*. Similarly, *CHI1*, *CHI3* and *FLS1* may also be key genes. Interestingly, the *CHS2* gene had a high expression level at stage A but was not expressed at the other stages. The change in expression level suggested that *CHS2* has significant stage-specific expression.

The MYBs regulating flavonoid biosynthesis in *C. morifolium*

Based on their protein domains, a total of 4,068 isoforms were identified as transcription factors, and MYB was the largest family, with approximately 444 isoforms (Figure 4A). After alignment with MYB transcription factors isolated from *A. thaliana*, we found that 17 isoforms were clustered into R2R3-MYB subgroups 4-7 (Supplementary figure 3). The members of all four subgroups also aligned with other MYBs isolated from *Malus × domestica*, *Vitis vinifera* and other species (Figure 4B); the information on the genes is provided in Supplementary table 2. Finally, three isoforms were clustered into R2R3 MYB subgroup 7, members of which are regarded as activators of flavanol biosynthesis. Five isoforms were grouped into subgroup 6, members of which have an active function in anthocyanin biosynthesis. Nine isoforms were grouped into subgroup 4, members of which repress the biosynthesis of phenylpropanoids, such as lignin, phenolic acid and flavonoids. However, no isoforms were clustered with subgroup 5,

members of which activate the biosynthesis of proanthocyanidins. This is consistent with the result that no isoforms were annotated as *LAR* and *ANR* genes. Furthermore, we found that isoforms 90494 and 90874 had high similarity to a single gene, which was not similar to isoform 65995; therefore, we inferred that two genes activated flavanol biosynthesis. In addition, the isoforms in subgroups 4 and 6 were assigned to 2 genes and 1 gene, respectively, based on sequence similarity (Supplementary table 3). When comparing the expression levels of these two genes in the inflorescence, we found that *CmMYBF1* was more highly expressed than *CmMYBF2* at every stage (Figure 4C) and *CmMYBF1* had an expression profile similar to those of *FLS* genes. Therefore, *CmMYBF1* probably activates flavonoid biosynthesis in the inflorescence of *C. morifolium*.

Discussion

C. morifolium, as a traditional medicinal plant, has a long culture history in China, and its inflorescence has high medicinal value due to its flavonoids. However, there is a poor understanding of flavonoid biosynthesis and regulation in *C. morifolium*. To identify genes involved in flavonoid biosynthesis and regulation with long sequences, full-length transcriptome sequencing of the inflorescence during flowering was performed. By comparison with 63,854 unigenes with a mean length of 741 bp isolated from 'Chuju' by SGS technology, we found 125,532 isoforms in 'Hangju' with a significantly longer mean length of 2,009 bp. We found more genes in the flavonoid biosynthesis pathways in this cultivar than reported in 'Chuju' [25]. SMRT sequencing was clearly more accurate than SGS technology in the process of obtaining transcript information. Furthermore, TGS is also an effective method for detecting alternative splicing (AS) events [26, 27]. In this study, we found that many isoforms shared high similarity with some fragment inserts or losses. Therefore, these isoforms were perhaps the results of AS of one gene.

According to the gene annotations, we found 5 *CHS* and 3 *CHI* genes in the database for *C. morifolium*, fewer than the 17 *CHS* and 8 *CHI* genes identified in the genome of *Chrysanthemum nankingense* [28]. Moreover, no *LAR* or *ANR* genes also were detected in *Chrysanthemum nankingense*, which suggests that proanthocyanidins are not the functional component in *C. morifolium* and *C. nankingense*. In addition, *C. morifolium* has a gene number similar to that in other species of Asteraceae. For instance, sunflower also has 3 *CHI* genes, 2 *FLS* genes and 1 *F3H* gene, similar to *C. morifolium* [28]. In our previous study, one *FLS* gene was reported in *C. morifolium* to catalyse dihydroquercetin to quercetin [29]. In addition, we found that another *FLS* gene was highly similar to the predicted gene in *Lactuca sativa* and *Nicotiana tabacum*, which needs to be verified in future studies. At the different stages, the genes had a high expression level at the bud stage, *CHS*, *CHI*, *F3H*, *F3'H* and *FLS* genes also highly expressed in budding of 'Chuju' [25].

In the MYB transcription factor analysis, the 5 transcripts in subgroup 6 were annotated as the *CmMYB6* gene (No.: AKP06190), which reportedly functions as the activator of anthocyanin biosynthesis in *C. morifolium* [30]. The ternary MYB-bHLH-WD40 (MBW) transcriptional regulator was recognized to regulate flavonoid biosynthesis, especially anthocyanin and proanthocyanidin biosynthesis [31, 32]. However, no factors in *C. morifolium* were clustered with bHLH (AtEGL3, AtGL3 and TT8), which interacts

with MYB to regulate anthocyanin biosynthesis (Supplementary figure 4) [33, 34]. In addition, no genes were annotated as the *CmbHLH2* gene (No.: KT724056), which reportedly functions in the activation of anthocyanin biosynthesis in *C. morifolium* for ornamentation [35]. *TT8* also promotes proanthocyanidin biosynthesis [31], and no genes were similar to *TT8* result in the small amounts of anthocyanins and proanthocyanidins produced in 'Hangju'. Therefore, it reliable to find candidate transcription factors via phylogenetic analysis. Accordingly, we found two genes belonging to SG7 MYBs, which are important factors activating flavonol biosynthesis. Both the CmMYBF1 and CmMYBF2 proteins had a GR[V/T]SR[C/S][V/K][A/M]K motif, which is the typical motif of SG7 MYB transcription factors [36]. Furthermore, *CmMYBF2* gene had a sequence similar to that of the *CmMYB3* gene, which has been proposed to regulate flavonoid biosynthesis but not promote *DRF* gene expression in *C. morifolium* for ornamentation [30]; thus, CmMYBF2 very likely regulates flavonol biosynthesis. However, CmMYBF1 had a higher expression level than CmMYBF2, and CmMYBF1 might also be a transcription factor regulating flavonol biosynthesis. Thus, which transcription factor is the most important needs to be further verified.

Conclusion

In our study, the TGS platform (PacBio Sequel) was used to obtain the full-length transcriptome of *C. morifolium* cv. 'Hangju'. A total of 29.49 Gb of raw data and 125,532 non-redundant isoforms were obtained. After gene annotation, 144 isoforms involved in the pathway of flavonoid biosynthesis were detected. We also predicted the major genes participating in luteolin and quercetin biosynthesis by expression analysis. The candidate genes had similar expression patterns, with the highest expression level at the bud stage. Furthermore, the MYB transcription factors activating flavonol biosynthesis were screened and analysed. All of these results provide a theoretical basis for further studying the quality of 'Hangju' and will advance its introduction and cultivation.

Methods

Plant material and RNA extraction

The 'Zaoyang' genotype of *Chrysanthemum morifolium* cv. 'Hangju' was formally identified by Professor Qiaosheng Guo and collected from the germplasm resource preservation centre at Nanjing Agricultural University. The plant material used in our study is recognized by the Department of Agriculture of Zhejiang Province, China, but no voucher specimens have been deposited in public herbaria. The inflorescence at different developmental stages (stage A – stage E) were collected and frozen at -80°C for subsequent experiments (Figure 1) [10, 37]. Total RNA was isolated from the samples by using the TaKaRa MiniBEST Plant RNA Extraction Kit (TaKaRa, China). The quality of RNA was assessed on a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA) and 1% agarose gel. Then, equal amounts of high-quality RNA from every sample were pooled together for PacBio sequencing analysis.

PacBio SMRTbell library preparation and sequencing

A total of 800-1,000 ng of RNA was used to synthesize cDNA by a Clontech SMARTer PCR cDNA Synthesis Kit with oligo-dT primers. After large-scale PCR amplification, the double-stranded cDNA was size selected into 0-5 kb and 4.5-10 kb fractions by the BluePippin system (Sage Science, USA). Additional PCR was performed post size selection to produce adequate templates, and then, size-selected cDNA was ligated with a blunt adapter for SMRTbell library construction. Finally, a total of two SMRT cell sequencing reactions were performed by a PacBio Sequel sequencer (BGI-Shenzhen, China).

Bioinformatics analysis of isoform sequences

Reads of insert were generated by removing adapters and artefacts from raw SMRT sequencing data, and they were further classified into FL non-chimeric, FL chimeric, non-FL, or short reads based on the presence of 5'/3' primer and polyA-tail sequences. Then, FL non-chimeric transcripts were clustered into consensus sequences by the ICE (Interactive Clustering and Error Correction) algorithm and polished by Quiver. Finally, high-quality consensus isoforms from each library were merged, and CD-HIT-EST was used to remove redundant isoforms based on sequence similarity. The isoform sequences were aligned to the non-redundant protein (NR), non-redundant nucleotide (NT), Gene Ontology (GO), Clusters of Orthologous Groups (COG), Kyoto Encyclopaedia of Genes and Genomes (KEGG), SwissProt, and InterPro databases with Blast, Blast2GO and InterProScan 5. TransDecoder was used to identify the candidate coding areas and extract the longest ORF. Then, the homologous protein sequences were searched against SwissProt and Hmmscan to predict the coding regions.

Expression analysis of flavonoid biosynthesis genes in *C. morifolium*

According to functional annotation analysis, genes involved in the luteolin and quercetin biosynthesis pathways were selected (including *CHS*, *CHI*, *F3H*, *F3'H*, *FLS*, and *FNS*). Then, the predicted CDSs were used to design primers (Supplementary table 4), and their expression profiles at different developmental stages were analysed by Q-PCR with relative quantification.

Phylogenetic analysis of the MYB transcription factor family

The ORFs of isoforms were searched against the PlantTFDB databases by hmmsearch to identify MYB transcription factors. Then, the protein sequences of putative MYBs were aligned with 168 MYB transcription factors isolated from *A. thaliana* by MAFFT. The output file from MAFFT was used to construct a phylogenetic tree by RAXML with the maximum likelihood method and 1000 bootstraps. Then, candidate transcripts were further aligned with the MYB transcription factors regulating flavonoid biosynthesis in *A. thaliana* and other plants. The expression levels of candidate MYB genes were also analysed by Q-PCR.

List of abbreviations

SMRT-Seq: Single-molecule real-time sequencing; EBGs: Early biosynthetic genes; LBGs: Late biosynthetic genes; SGS: Second-generation sequencing; TGS: Third-generation sequencing; ROI: Reads of insert; NR: non-redundant protein; NT: non-redundant nucleotide; GO: Gene Ontology; COG: Clusters of Orthologous

Groups; KEGG: Kyoto Encyclopaedia of Genes and Genomes; ICE: Interactive Clustering and Error Correction; PAL: Phenylalanine ammonia-lyase; C4H: Cinnamate 4-hydroxylase; 4CL: 4-coumarate–CoA ligase; CHS: Chalcone synthase; CHI: Chalcone isomerase; F3H: Flavanone 3-hydroxylase; F3'H: Flavonoid 3'-hydroxylase; F3'5'H: Flavonoid 3',5'-hydroxylase; FNS \boxtimes : Flavone synthase \boxtimes ; IFS: Isoflavone synthase; FLS: Flavonol synthase; DFR: Dihydroflavonol 4-reductase; ANS: Anthocyanidin synthase; ANR: Anthocyanidin reductase; LAR: Leucoanthocyanidin reductase; MBW: MYB-bHLH-WD40; SG: Subgroup. ORF: Open read frame; AS: Alternative splicing. FL: Full length.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available due to the data was analyzing for other research, but are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by grants from the National Nature Science Foundation of China (No. 81503180) in the design of the study and collection of data. The Fundamental Research Funds for the Central Universities (No. KYZ201608 and KJQN201643) were used in supporting the analysis and interpretation of data and manuscript writing.

Authors' contributions

TW, FY, and QG conceived and designed the experiments. TW and FY performed the experiments and written the original draft. TW, FY, and QZ analyzed the data and make figures and tables. WZ and LZ contributed to cultivation and collection of experimental materials. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

References

1. Cui G, Niu YR, Wang H, Dong J, Yuki H, Chen SZ: **Rapid Isolation and Identification of Active Antioxidant Ingredients from Gongju Using HPLC-DAD-ESI-MSn and Postcolumn Derivatization.** *J Agric Food Chem* 2012, **60**(21):5407-5413.
2. He DX, Ru XC, Wen L, Wen YC, Jiang HD, Bruce IC, Jin J, Ma X, Xia Q: **Total flavonoids of Flos Chrysanthemi protect arterial endothelial cells against oxidative stress.** *J Ethnopharmacol* 2012, **139**(1):68-73.
3. Zheng CP, Dong Q, Chen HJ, Cong QF, Ding K: **Structural characterization of a polysaccharide from Chrysanthemum morifolium flowers and its antioxidant activity.** *Carbohydr Polym* 2015, **130**:113-121.
4. Gao T, Zhu ZY, Zhou X, Xie ML: **Chrysanthemum morifolium extract improves hypertension-induced cardiac hypertrophy in rats by reduction of blood pressure and inhibition of myocardial hypoxia inducible factor-1alpha expression.** *Pharm Biol* 2016, **54**(12):2895-2900.
5. Jin HF, Liu XW, Tang YM, Tang LJ, Wang YL, Du CQ: **Effects of total flavones from Dendranthema morifolium on vasoconstriction and proliferation of vascular smooth muscle cells.** *Mol Med Rep* 2016, **13**(1):989-993.
6. Kuang CL, Lv D, Shen GH, Li SS, Luo QY, Zhang ZQ: **Chemical composition and antimicrobial activities of volatile oil extracted from Chrysanthemum morifolium Ramat.** *J Food Sci Technol-Mysore* 2018, **55**(7):2786-2794.
7. Ukiya M, Akihisa T, Yasukawa K, Kasahara Y, Kimura Y, Koike K, Nikaido T, Takido M: **Constituents of compositae plants. 2. Triterpene diols, triols, and their 3-O-fatty acid esters from edible chrysanthemum flower extract and their anti-inflammatory effects.** *J Agric Food Chem* 2001, **49**(7):3187-3197.
8. Tao JH, Duan JA, Jiang S, Feng NN, Qiu WQ, Ling Y: **Polysaccharides from Chrysanthemum morifolium Ramat ameliorate colitis rats by modulating the intestinal microbiota community.** *Oncotarget* 2017, **8**(46):80790-80803.
9. Shao QS, Guo QS, Deng YM, Guo HP: **A comparative analysis of genetic diversity in medicinal Chrysanthemum morifolium based on morphology, ISSR and SRAP markers.** *Biochem Syst Ecol* 2010, **38**(6):1160-1169.
10. Wang T, Guo QS, Mao PF: **Flavonoid accumulation during florescence in three Chrysanthemum morifolium Ramat cv. 'Hangju' genotypes.** *Biochem Syst Ecol* 2014, **55**:79-83.
11. Vogt T: **Phenylpropanoid biosynthesis.** *Mol Plant* 2010, **3**(1):2-20.
12. Weisshaar B, Jenkins GI: **Phenylpropanoid biosynthesis and its regulation.** *Current Opinion in Plant Biology* 1998, **1**:251-257.
13. Deng Y, Lu S: **Biosynthesis and Regulation of Phenylpropanoids in Plants.** *Critical Reviews in Plant Sciences* 2017, **36**(4):257-290.

14. Righini S, Rodriguez EJ, Berosich C, Grotewold E, Casati P, Ferreyra MLF: **Apigenin produced by maize flavone synthase I and II protects plants against UV-B-induced damage.** *Plant Cell Environ* 2019, **42**(2):495-508.
15. Jung W, Yu O, Lau SMC, O'Keefe DP, Odell J, Fader G, McGonigle B: **Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes.** *Nat Biotechnol* 2000, **18**(2):208-212.
16. Mehrtens F, Kranz H, Bednarek P, Weisshaar B: **The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis.** *Plant Physiol* 2005, **138**(2):1083-1096.
17. Stracke R, Ishihara H, Huet 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):660-677.
18. Espley RV, Hellens RP, Putterill J, Stevenson DE, Kutty-Amma S, Allan AC: **Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10.** *Plant J* 2007, **49**(3):414-427.
19. Wang K, Bai ZY, Liang QY, Liu QL, Zhang L, Pan YZ, Liu GL, Jiang BB, Zhang F, Jia Y: **Transcriptome analysis of chrysanthemum (*Dendranthema grandiflorum*) in response to low temperature stress.** *BMC Genomics* 2018, **19**:19.
20. Xu YJ, Gao S, Yang YJ, Huang MY, Cheng LN, Wei Q, Fei ZJ, Gao JP, Hong B: **Transcriptome sequencing and whole genome expression profiling of chrysanthemum under dehydration stress.** *BMC Genomics* 2013, **14**:15.
21. Zhao N, Li CW, Yan YJ, Cao W, Song AP, Wang HB, Chen SM, Jiang JF, Chen FD: **Comparative Transcriptome Analysis of Waterlogging-Sensitive and Waterlogging-Tolerant Chrysanthemum morifolium Cultivars under Waterlogging Stress and Reoxygenation Conditions.** *Int J Mol Sci* 2018, **19**(5):21.
22. Liu H, Sun M, Du DL, Pan HT, Cheng TR, Wang J, Zhang QX: **Whole-Transcriptome Analysis of Differentially Expressed Genes in the Vegetative Buds, Floral Buds and Buds of Chrysanthemum morifolium.** *PLoS One* 2015, **10**(5):24.
23. Rhoads A, Au KF: **PacBio Sequencing and Its Applications.** *Genomics Proteomics Bioinformatics* 2015, **13**(5):278-289.
24. Sharon D, Tilgner H, Grubert F, Snyder M: **A single-molecule long-read survey of the human transcriptome.** *Nat Biotechnol* 2013, **31**(11):1009-1014.
25. Yue J, Zhu C, Zhou Y, Niu X, Miao M, Tang X, Chen F, Zhao W, Liu Y: **Transcriptome analysis of differentially expressed unigenes involved in flavonoid biosynthesis during flower development of Chrysanthemum morifolium 'Chuju'.** *Sci Rep* 2018, **8**(1):13414.
26. Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, Ben-Hur A, Reddy AS: **A survey of the sorghum transcriptome using single-molecule long reads.** *Nat Commun* 2016, **7**:11706.
27. Cheng B, Furtado A, Henry RJ: **Long-read sequencing of the coffee bean transcriptome reveals the diversity of full-length transcripts.** *Gigascience* 2017, **6**(11):1-13.

28. Song C, Liu YF, Song AP, Dong GQ, Zhao HB, Sun W, Ramakrishnan S, Wang Y, Wang SB, Li TZ *et al*: **The Chrysanthemum nankingense Genome Provides Insights into the Evolution and Diversification of Chrysanthemum Flowers and Medicinal Traits.** *Mol Plant* 2018, **11**(12):1482-1491.
29. Zou Q-J, Wang T, Guo Q-S, Zhang W-Y, Wang R, Wang Y-H, Xiao Y-M: **Prokaryotic expression and in vitro enzyme activity analysis of flavonol synthase in Chrysanthemum morifolium cv. 'Hangju'.** *Zhongguo Zhong yao za zhi* 2018, **43**(17):3471-3476.
30. Liu X-f, Xiang L-l, Yin X-r, Grierson D, Li F, Chen K-s: **The identification of a MYB transcription factor controlling anthocyanin biosynthesis regulation in Chrysanthemum flowers.** *Scientia Horticulturae* 2015, **194**:278-285.
31. Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L: **TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in Arabidopsis thaliana.** *Plant J* 2004, **39**(3):366-380.
32. Xu W, Dubos C, Lepiniec L: **Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes.** *Trends Plant Sci* 2015, **20**(3):176-185.
33. Gonzalez A, Zhao M, Leavitt JM, Lloyd AM: **Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings.** *Plant J* 2008, **53**(5):814-827.
34. Zhou LL, Shi MZ, Xie DY: **Regulation of anthocyanin biosynthesis by nitrogen in TTG1-GL3/TT8-PAP1-programmed red cells of Arabidopsis thaliana.** *Planta* 2012, **236**(3):825-837.
35. Xiang LL, Liu XF, Li X, Yin XR, Grierson D, Li F, Chen KS: **A Novel bHLH Transcription Factor Involved in Regulating Anthocyanin Biosynthesis in Chrysanthemums (Chrysanthemum morifolium Ramat.).** *PLoS One* 2015, **10**(11):e0143892.
36. Stracke R, Werber M, Weisshaar B: **The R2R3-MYB gene family in Arabidopsis thaliana.** *Current Opinion in Plant Biology* 2001, **4**(5):447-456.
37. Wang T, Zhu ZB, Guo QS, Mao PF: **Variation in major flavonoids glycosides and caffeoylquinic acids during florescence of three Chrysanthemum morifolium Ramat cv. 'Hangju' genotypes.** *Biochem Syst Ecol* 2013, **47**:74-79.

Tables

Table 1 Statistics of transcriptome data

Library	B01 (0-5K)	G01 (4.5-10K)
Cell Number	1	1
Total Raw Reads	606,174	575,319
Total Base (GB)	15.56	13.93
Reads of Insert	591,513	560,454
Read Bases of Insert (bp)	1,823,985,219	2,010,084,287
Mean Read Length of Insert (bp)	3,084	3,587
HQ Isoforms Number	129,610	112,142
Total Isoforms Number	125,532	
Mean Read Length of Isoforms (bp)	2,009	
N50 (bp)	2,274	

Table 2 The Isoforms annotated into the flavonoid biosynthesis pathways

Gene name	Number of Isoforms	Isoforms name
PAL	45	Isoform_52833, Isoform_47763, Isoform_43454, Isoform_62542, Isoform_67724, Isoform_102200, Isoform_108441, Isoform_81892, Isoform_68061, Isoform_25536, Isoform_27366, Isoform_28938, Isoform_35394, Isoform_27213, Isoform_26269, Isoform_110994, Isoform_26254, Isoform_122495, Isoform_121987, Isoform_121852, Isoform_57291, Isoform_48658, Isoform_45373, Isoform_55744, Isoform_76695, Isoform_79590, Isoform_94611, Isoform_88194, Isoform_68496, Isoform_90394, Isoform_95170, Isoform_77862, Isoform_102606, Isoform_22996, Isoform_33822, Isoform_29134, Isoform_38700, Isoform_39651, Isoform_112668, Isoform_102808, Isoform_28794, Isoform_123285, Isoform_43973, Isoform_50840, Isoform_50978
C4H	9	Isoform_61275, Isoform_63752, Isoform_89464, Isoform_51120, Isoform_51919, Isoform_73975, Isoform_62853, Isoform_53493, Isoform_55308
4CL	33	Isoform_43749, Isoform_59715, Isoform_42120, Isoform_44780, Isoform_79349, Isoform_89705, Isoform_110940, Isoform_70800, Isoform_80544, Isoform_74817, Isoform_33855, Isoform_104805, Isoform_40088, Isoform_46710, Isoform_40758, Isoform_45456, Isoform_48895, Isoform_64143, Isoform_42313, Isoform_49751, Isoform_52298, Isoform_91185, Isoform_73851, Isoform_75666, Isoform_113605, Isoform_19191, Isoform_48574, Isoform_39980, Isoform_40022, Isoform_123654, Isoform_57545, Isoform_46926, Isoform_50728
CHS	16	Isoform_69087, Isoform_88884, Isoform_70059, Isoform_71639, Isoform_67309, Isoform_90473, Isoform_71339, Isoform_70758, Isoform_72050, Isoform_79285, Isoform_69249, Isoform_57251, Isoform_58137, Isoform_71025, Isoform_63828, Isoform_84387
CHI	8	Isoform_101012, Isoform_107006, Isoform_107139, Isoform_108103, Isoform_108228, Isoform_111718, Isoform_91714, Isoform_92657
F3H	3	Isoform_82489, Isoform_122564, Isoform_84881
F3'H	6	Isoform_50027, Isoform_74567, Isoform_86056, Isoform_54616, Isoform_108195, Isoform_124366
FLS	2	Isoform_82400, Isoform_83438
FNS II	13	Isoform_52845, Isoform_102292, Isoform_74378, Isoform_84302, Isoform_54094, Isoform_49130, Isoform_75446, Isoform_83978, Isoform_112139, Isoform_52828, Isoform_94894, Isoform_101125, Isoform_110531
DFR	5	Isoform_34792, Isoform_82000, Isoform_76056, Isoform_77859, Isoform_119850
ANS	3	Isoform_86095, Isoform_83405, Isoform_100339

Figures

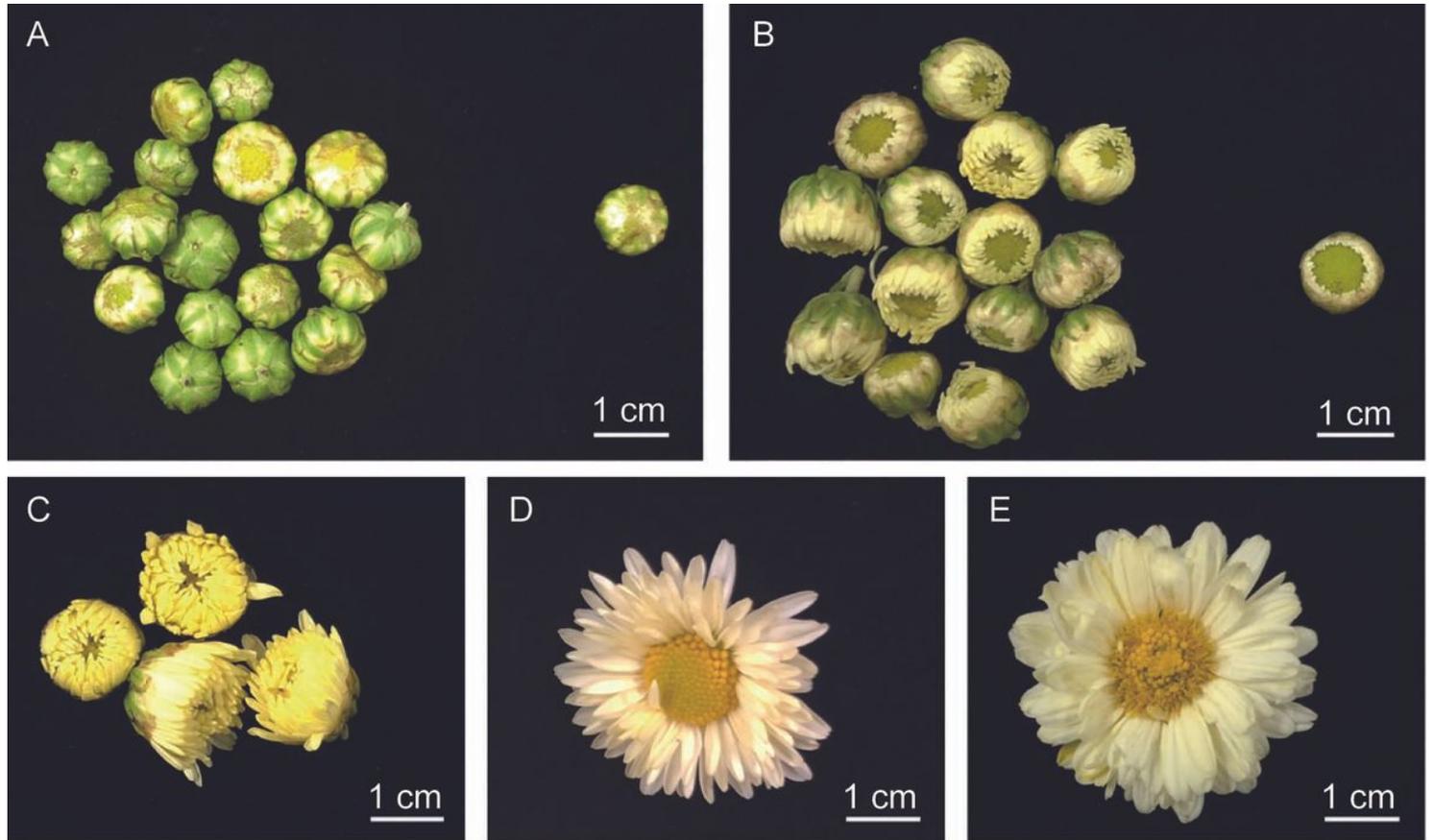


Figure 1

The inflorescence of *C. morifolium* at five different developmental stages. a The inflorescence at stage A (bud). b The inflorescence at stage B. c The inflorescence at stage C. d The inflorescence at stage D. e The inflorescence at stage E.

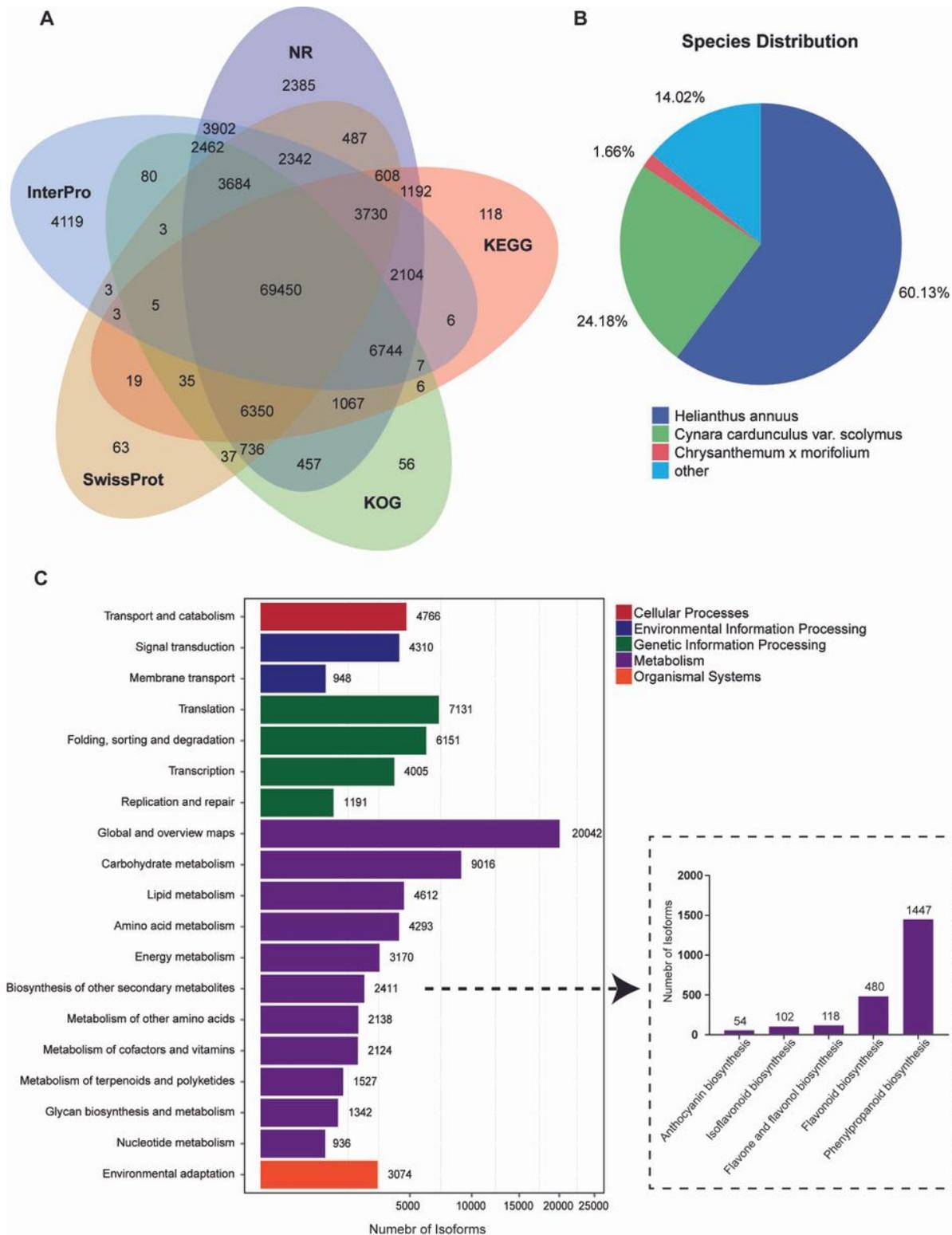


Figure 2

The gene annotation of isoforms from the *C. morifolium*. a The venn diagram of isoforms annotation between NR, KOG, KEGG, Swissprot and Interpro database. b The distribution of NR annotated species, each transcript was annotated to only one species with the highest similarity score. c The functional distribution of KEGG annotation, the isoforms annotated to the pathway related to flavonoid biosynthesis was showed in the dashed box.

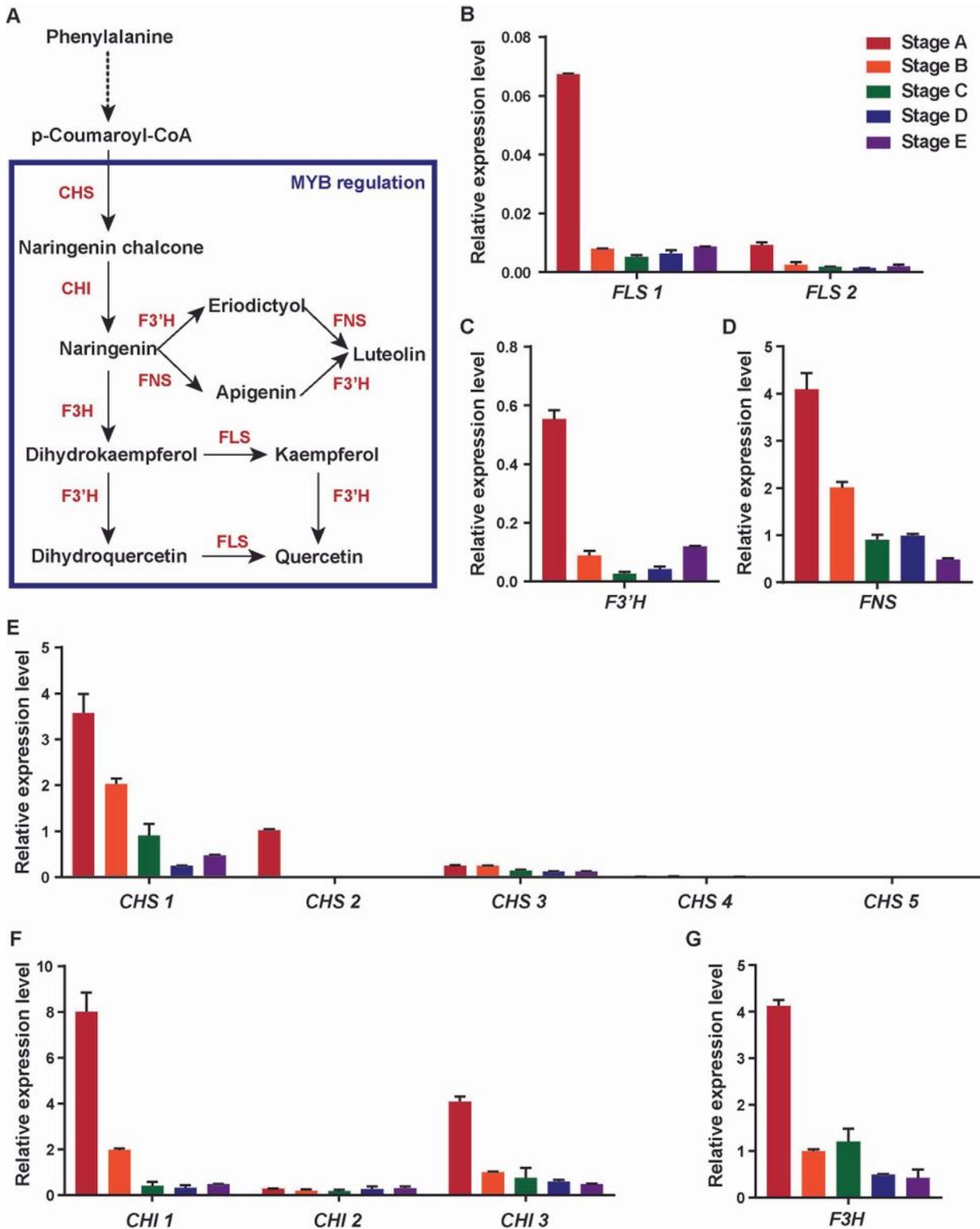


Figure 3

The expression levels of genes involved in flavonoid biosynthesis. a The pathway of luteolin and quercetin biosynthesis. b-g The relative expression levels of FLS, F3'H, FNS, CHS, CHI and F3H gene at five developmental stages, respectively.

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

- [SupplementaryFigure4.pdf](#)
- [SupplementaryFigure2.pdf](#)
- [SupplementaryTable.xlsx](#)
- [SupplementaryFigure1.pdf](#)
- [SupplementaryFigure3.pdf](#)