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; MYB

Posted Date: October 1st, 2019

DOI: <https://doi.org/10.21203/rs.2.15455/v1>

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

[Read Full License](#)

Abstract

Abstract Background : The capitulum of *Chrysanthemum morifolium* cv. 'Hangju' has been widely used in China for antioxidant and anti-inflammatory. Flavonoids as one of the bioactive components in *C. morifolium* have a poor understanding in their biosynthesis and regulation. Nowadays, transcriptome sequencing as an effective method was used in capturing the transcripts information. So, single-molecule real-time (SMRT) sequencing was performed to obtain the full length of genes involved in flavonoid biosynthesis and regulation in *C. morifolium*. **Results :** The high-quality RNA was extracted from the capitulum of *C. morifolium* at different development stages, and it was constructed into two libraries (0-5 kb and 4.5-10 kb) for sequencing. Finally, 125,532 non-redundant isoforms with mean length of 2,009 bp were captured. Of which, 2,083 transcripts were annotated in the pathway related to the flavonoid biosynthesis and 56 isoforms were annotated as *CHS*, *CHI*, *F3H*, *F3'H*, *FNS*, *FLS*, *DFR* and *ANS* genes. Based on the gene expression level at different stages, we predicted the major genes involved in the flavonoid biosynthesis. And we found two candidate MYB factors (*CmMYBF1* and *CmMYBF2*) activating the flavonol biosynthesis by phylogenetic analysis. **Conclusions :** Based on the full-length transcriptome 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 introduction and cultivation of *C. morifolium* cv. 'Hangju'.

Background

Flos Chrysanthemi is the dried capitulum of *Chrysanthemum morifolium* Ramat. (Asteraceae) has been widely used in China for thousands of years, which has a significant medicinal and economic value. According to the Chinese Pharmacopoeia, Flos Chrysanthemi could be mainly classified into five varieties 'Hangju', 'Boju', 'Chuju', 'Gongju' and 'Huaiju'. Many pharmacological studies reported that it possesses various functions of antioxidation [1-3], cardiovascular protection [4, 5], antimicrobial [6] and anti-inflammatory [7, 8], which are benefit from the mainly active components, like flavonoids. The variety the 'Hangju' of Flos Chrysanthemi, which has been cultivated for centuries and developed into several genotypes, and the genotype 'Zaoyang' has the largest cultivated area and yield among all the genotypes [9]. Furthermore, the genotype 'Zaoyang' from Tongxiang county (the traditional area) is considered to be advantage in higher contents of luteolin than the other cultivated areas [10], which suggest that flavonoids are the vital components in genuine medicinal materials.

Flavonoids is a large group of secondary metabolites in plants, whose biosynthetic pathway is well established as a useful model for studying metabolic regulation. The genes involved in the pathway were mainly divided into early and late biosynthetic genes. Early biosynthetic genes (EBGs) including *CHS*, *CHI*, *F3H*, *F3'H* and *FLS* could catalyze the flavonol biosynthesis, whereas late biosynthetic genes (LBGs: *DFR*, *ANS* and *ANR*) lead to both anthocyanin and proanthocyanidin biosynthesis [11-13]. There are also other genes like *FNS* and *IFS* could catalyze the flavone and isoflavone biosynthesis [14, 15]. Furthermore, the transcriptional factors regulating the biosynthesis were identified in many species. The *AtMYB11*, *AtMYB12* and *AtMYB111* in *Arabidopsis thaliana* could activate the expression of LEGs to produce

flavonol [16, 17], the high expression of *MdMYB1* and *MdMYB10* increase the content of anthocyanin in *Malus × domestica* [18]. However, the flavonoid biosynthesis and regulation are poorly understood in *C. morifolium* because of lacking reference genome database. So, the transcriptome sequencing offered an alternative approach to obtain genes information. Of which, second-generation sequencing (SGS) was usually used to identified the differentially expressed genes in *C. morifolium* under stress or at different tissues [19-22]. However, short read lengths generated by SGS decrease the accuracy of transcripts assembly, especially for the species without genome information. Single-molecule real-time (SMRT) sequencing as the third-generation sequencing (TGS) technology has an advantage of generating high-throughput long reads, so it avoids the inaccuracy resulting from assembling short sequences [23, 24].

In this study, SMRT sequencing was preformed to get the full-length transcripts for better understanding the flavonoid biosynthesis and regulation in *C. morifolium*. The key genes were screened based on the annotation analysis, and their expressional profile were also analyzed in different development stages. Furthermore, the candidate MYB factors activating the flavonol biosynthesis were identified, which not only provide an understanding of flavonoid biosynthesis and regulation in *C. morifolium*, also better for the introduction and cultivation of *C. morifolium*.

Results

The full-length transcriptome of *C. morifolium*

To identify the transcripts in *C. morifolium* during the development stages, SMRT sequencing technology was used to capture the complete and full-length transcriptome. Totally, we obtained 15.56 Gb and 13.93 Gb raw data from 0-4.5K and 4.5-10K libraries, respectively. About 591,513 and 560, 454 ROIs (Reads of insert) were recognized with mean length of 3,328 and 3,587 bp (Table 1). The clean data was upload to Sequence Read Archive (SRA) with accession SRR10054190 and SRR10054977. Of which, more than 50% reads were classified as full-length (FL) non-chimeric reads, less than 5% reads were classified into chimeric reads, which indicated the SMRTbell libraries have a high quality (Supply figure 1). After clustered and polished, 129,610 and 112,142 HQ isoforms were finally merged into 125,532 non-redundant isoforms with mean length of 2,009 bp and N50 of 2,274 bp (Table 1).

Functional annotation

All non-redundant isoforms were searched against to 7 functional databases, and about 115,807 (92.25%) transcripts could be annotated with at least one functional database (Figure 2A). There were 107,700 (85.79%) isoforms were assigned to NR databases, and most of them have high similarity to the genes in *Helianthus anuus* (60.13%) *Cynara cardunculus* var. *scolymus* (24.18%), *Chrysanthemum × morifolium* (14.02%) and other species (Figure 2B). About 33,901 (27.01%) transcripts could be assigned in GO database within 53 subcategories, and the largest subcategories of three primary GO categories were cellular process, cell and catalytic activity (Supply figure 2). We also assigned 91,179 (72.63%) isoforms into KEGG database, of which 2,411 isoforms were classified in to the subgroup of Biosynthesis of other secondary metabolites. Furthermore, we found 1,447, 480, 54 and 102 transcripts were annotated

in the pathway of Phenylpropanoid biosynthesis, Flavonoid biosynthesis, Anthocyanin biosynthesis, Isoflavonoid biosynthesis, Flavone and flavonol biosynthesis, 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 pathway (Table 2). And every gene has many isoforms, especially for the *PAL* and *4CL* gene families having more than 30 isoforms. Interesting, no sequence was similar to F3'5'H, which could catalyze to be as the substrate of blue anthocyanin. So, it corresponds the fact that the *C. morifolium* have no blue varieties. And *ANR* and *LAR* genes were also not found in the database, which suggest there were probably no proanthocyanins in the *C. morifolium*. Further analyzing the gene involved in the luteolin and quercetin biosynthesis, the isoforms were classified into 5 *CHS*, 3 *CHI*, 1 *F3H*, 1 *F3'H*, 1 *FNS* and 2 *FLS* genes based on the sequence similarity, (Supply table 1).

Different expression of genes involved in luteolin and quercetin biosynthesis

luteolin and quercetin as the marker compounds of 'Hangju' in the Chinese Pharmacopoeia, so *CHS*, *CHI*, *F3H*, *F3'H*, *FLS* and *FNS* genes were selected and analyzed. The expression profile of every gene in the different flowering stages were shown in the Figure 3. The most of genes had the highest expression level at the stage of buds (Stage A), which would be the important stage for flavonoid biosynthesis. For *CHS* genes, *CHS1* expressed significantly higher than others at every stage, while *CHS4* and *CHS5* barely expressed. So, *CHS1* might be major gene control the flavonoid at the flower of *C. morifolium*. Similarly, *CHI1*, *CHI3* and *FLS1* also maybe the key genes. Interesting, *CHS2* gene had a high expression level at Stage A, while not expressed at other stages. The change of expression level suggested that *CHS2* have significant expression specificity.

The MYB regulating the flavonoid biosynthesis in *C. morifolium*

Based on the protein domain, a total of 4,068 isoforms were identified as transcription factors, and the MYB was the largest family with about 444 isoforms (Figure 4A). After aligning with MYB factors isolated from *Arabidopsis thaliana*, we found 17 isoforms were clustered into R2R3-MYB subgroup 4-7 (Supply figure 3). All of the four subgroups members were also aligned with other MYB isolated from *Malus × domestica*, *Vitis vinifera* and other species (Figure 4B), the information of the genes were listed in the Supply table 2. Finally, there were 3 isoforms were clustered into R2R3 MYB subgroups 7, which were regarded as the activator of flavanol biosynthesis pathways. And 5 isoforms were grouped into subgroups 6 having an active function of anthocyanin biosynthesis. There were 9 isoforms were grouped into subgroup 4, which could repress the biosynthesis of phenylpropanoid metabolism, like lignin, phenolic acid and flavonoid. However, no isoforms were clustered into subgroup 5, which could activate the biosynthesis of proanthocyanidin. It is consistent with the results that no genes were annotated to *LAR* and *ANR*. Furthermore, we found that Isoform 90494 and 90874 have high similarity to same one gene different from Isoform 65995, so we infer that there were two genes activate the flavanol

biosynthesis. Meanwhile, the isoforms in subgroup 4 and 6 were classified into 2 and 1 genes based on the sequence similarity (Supply table 3). When comparing the expression level of these two genes in the capitulum, we found *CmMYBF1* expressed highly than *CmMYBF2* in every stage (Figure 4C), and *CmMYBF1* have a similar expression profile with *FLS* genes. So, *CmMYBF1* was a probable gene activating the flavonoid biosynthesis in the capitulum of *C. morifolium*.

Discussion

C. morifolium as the traditional medicinal plant has a long culture history in China, whose capitulum have a high medicinal activity benefit from the flavonoids. However, there were poor understanding of flavonoid biosynthesis and regulation in *C. morifolium*. To obtain genes involved in the flavonoid biosynthesis and regulation with long sequences, full-length transcriptome sequence was performed in the capitulum during the flowering. Comparing the 63,854 unigenes with mean length of 741 isolated from 'Chuju' by SGS technology, we found 125,532 isoforms in 'Hangju' with mean length of 2,009 bp, which was significantly long. And we found more genes in flavonoid biosynthesis pathway than 'Chuju'[25]. It obviously that SMRT sequencing was more accurate than SGS technology in the process of getting transcripts information. Furthermore, TGS was also an effective method to detect [alternative splicing \(AS\)](#) events [26, 27]. In this study, we found there were many isoforms have high similarity with some fragments insert or loss. So, these isoforms perhaps were the AS of one gene.

According the gene annotation, we found 5 *CHS* and 3 *CHI* genes in the database of *C. morifolium*, which were less than the 17 *CHS* and 8 *CHI* genes identified in the genome of *Chrysanthemum nankingense* [28]. However, there were also no *LAR* and *ANR* genes like our results, which suggest that proanthocyanidins was not the functional component in the *C. morifolium* and *C. nankingense*. And *C. morifolium* have similar genes number with other species in Asteraceae. For instance, sunflower also have 3 *CHI*, 2 *FLS* and 1 *F3H* genes same as *C. morifolium* [28]. In our previous study, there was one *FLS* gene was reported in *C. morifolium* have a function of catalyzing the dihydroquercetin to quercetin [29]. Besides, we found another *FLS* gene have high similarity with predicted gene in *Lactuca sativa* and *Nicotiana tabacum*, it needs to verified at our next work. In the different stages, the genes have similar expression tendency with the genes in 'Chuju' [25], and have a high expression level at bud stage.

In the MYB factor analysis, the 5 transcripts in subgroup 6 were annotated as *CmMYB6* gene (No: [AKP06190](#)), which was reported function as the activator of anthocyanin biosynthesis in *C. morifolium* [30]. To date, the ternary MYB-bHLH-WD40 (MBW) transcriptional regulator were recognized regulating the flavonoids biosynthesis, especially for anthocyanin and proanthocyanidin biosynthesis [31, 32]. And no factors were clustered with bHLH (AtEGL3, AtGL3 and TT8) interacting with MYB to regulating anthocyanin biosynthesis (Supply figure 4) [33, 34]. Besides, no gene were annotated to the *CmbHLH2* gene (No: KT724056), which was reported had a function of activating anthocyanin biosynthesis in *C. morifolium* for ornament [35]. The TT8 also could promoting proanthocyanidin biosynthesis [31], the reason of no gene was similar to *TT8* were a small amount of anthocyanin and proanthocyanidin in 'Hangju'. So, it reliable to find a candidate factor by phylogenic analysis. According the way, we found two

genes belong to SG7 MYB, which were the important factors activating the flavonol biosynthesis. Both of *CmMYBF1* and *CmMYBF2* protein had a GR[V/T]SR[C/S][V/K][A/M]K motif, which was the typical motif of SG7 MYB factors [36]. Furthermore, *CmMYBF2* had a similar sequence with *CmMYB3* gene, which was proposed have function of regulating flavonoid biosynthesis but not promoting the *DRF* gene expression in *C. morifolium* for ornament [30], so it very likely regulate flavonol biosynthesis. But *CmMYBF1* have a higher expression level than *CmMYBF2*, *CmMYBF1* also might be the transcription factor regulating flavonol biosynthesis. So, which one was the major factor needs to be further verified.

Conclusions

In our study, the TGS platform (PacBio Sequel) was used to capture the full-length transcriptome of *C. morifolium* cv. 'Hangju'. A total of 29.49 Gb raw data and 125,532 de-redundant isoforms were obtained. After gene annotation, there were 144 isoforms involved in the pathway of flavonoids biosynthesis. We also predicted the major gene participated in the luteolin and quercetin biosynthesis by expression analysis. The candidate genes had a similar expression pattern with the highest expression level at the bud stage. Furthermore, the MYB factors activating flavonol biosynthesis were screened and analyzed. All of these results provided a theoretical basis for further studying the quality formation of 'Hangju' and also help for 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 germplasm resource preserving center at Nanjing Agricultural University. And the plant material in our study was recognized by Department of Agriculture of Zhejiang Province, China, but there was no voucher specimen deposited in a publicly available herbarium. The capitulum at different development stages (Stage A – Stage E) were collected and frozen in the -80°C for next 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 Nanodrop 2000 Spectrophotometer and 1% agarose gel. Then, equal amounts of high-quality RNA from every samples were pooled together for PacBio sequencing analysis.

PacBio SMRTbell library preparation and sequencing

The 800-1000 ng RNA was used to synthesized cDNA by 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, MA, USA). Additional PCR were preformed post size-selection to produce adequate templates, then size-selected cDNA was ligated with blunt adapter for SMRTbell library construction. Finally, a total of two SMRT cells Sequencing reactions were performed by the PacBio Sequel sequencer (BGI-Shenzhen, China).

Bioinformatics analysis of isoform sequencing

Reads of insert were generated by removing adapters and artefacts from raw SMRT sequencing data, and they were further classified into FL non-chimeric, 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 sequence by ICE (Iterative 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 the redundancy 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 InterProScan5. The Transdecoder was used to identify the candidate coding area and extract the longest ORF, then the Pfam protein homologous sequences was searched by blast to SwissProt and Hmmscan to predict the coding region.

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

According the functional annotation analysis, genes involved in luteolin and quercetin biosynthesis pathway were selected (including *CHS*, *CHI*, *F3H*, *F3'H*, *FLS*, *FNS*). Then the predicted CDS sequences were used to designed primers (Supply table 4), and their expression profile at different developmental stages were analyzed by Q-PCR with relative quantification.

Phylogenetic analysis of MYB transcription factor family

The ORF of isoforms were searched against the PlantfDB databases by hmmsearch to identify MYB transcript factors. Then the protein sequences of putative MYB were aligned with 168 MYB factors isolated from *Arabidopsis thaliana* by MAFFT. The out file of MAFFT were used to construct phylogenetic tree by RAxML in maximum likelihood method with 1000 bootstraps. Then candidate transcripts were further aligned with the MYB factors regulating the flavonoid biosynthesis in *Arabidopsis thaliana* and other plants. The expression level of candidate MYB genes were also analyzed by Q-PCR.

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: Iterative 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: Flavone synthase; 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 vasocontraction 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, Huep G, Barsch A, Mehrrens 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 = Zhongguo zhongyao zazhi = China journal of Chinese materia medica* 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 in the flavonoid biosynthesis pathway

Gene name	Numuber 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 capitulum of *C. morifolium* at 5 different development stages. a The capitulum at Stage A (bud). b The capitulum at Stage B. c The capitulum at Stage C. d The capitulum at Stage D. e The capitulum 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. c The functional distribution of KEGG annotation, the isoforms annotated in the pathway related to flavonoid biosynthesis was showed in the dashed box.

Figure 3

The expression level of genes involved in flavonoid biosynthesis. a The pathway of luteolin and quercetin biosynthesis. b The relative expression level of FLS at 5 development stages. c The relative expression level of F3'H at 5 development stages. d The relative expression level of FNS at 5 development stages. e The relative expression level of CHS at 5 development stages. f The relative expression level of CHI at 5 development stages. g The relative expression level of F3H at 5 development stages.

Figure 4

The analysis of MYB factors regulating the flavonoid biosynthesis. a The statistics of isoforms annotated into transcript factors. b The phylogenetic analysis of MYB factors clustered into R2R3-MYB subgroup 4-7. c The relative expression level of MYB in SG7 at 5 development stages.

Supplementary Files

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

- [SupplyFigure2.pdf](#)
- [SupplyFigure4.pdf](#)
- [SupplyFigure3.pdf](#)
- [SupplyTable.xlsx](#)
- [SupplyFigure1.pdf](#)