

RESEARCH

Open Access



Integrated metabolomic and transcriptomic analysis reveals biosynthesis mechanism of flavone and caffeoylquinic acid in chrysanthemum

Chenfei Lu^{1,2†}, Xiaoyun Yan^{1,2†}, Haohao Zhang^{1,2}, Taowei Zhong^{1,2}, Aijun Gui^{1,2}, Yuchen Liu^{1,2}, Lanying Pan^{3*} and Qingsong Shao^{1,2*}

Abstract

Background *Chrysanthemum morifolium* ‘HangBaiJu’, a popular medicinal and edible plant, exerts its biological activities primarily through the presence of flavones and caffeoylquinic acids (CQAs). However, the regulatory mechanism of flavone and CQA biosynthesis in the chrysanthemum capitulum remains unclear.

Results In this study, the content of flavones and CQAs during the development of chrysanthemum capitulum was determined by HPLC, revealing an accumulation pattern with higher levels at S1 and S2 and a gradual decrease at S3 to S5. Transcriptomic analysis revealed that *CmPAL1/2*, *CmCHS1/2*, *CmFNS*, *CmHQT*, and *CmHCT* were key structural genes in flavones and CQAs biosynthesis. Furthermore, weighted gene co-expression correlation network analysis (WGCNA), k-means clustering, correlation analysis and protein interaction prediction were carried out in this study to identify transcription factors (TFs) associated with flavone and CQA biosynthesis, including MYB, bHLH, AP2/ERF, and MADS-box families. The TFs CmERF/PTI6 and CmCMD77 were proposed to act as upstream regulators of *CmMYB3* and *CmbHLH143*, while CmMYB3 and CmbHLH143 might form a complex to directly regulate the structural genes *CmPAL1/2*, *CmCHS1/2*, *CmFNS*, *CmHQT*, and *CmHCT*, thereby controlling flavone and CQA biosynthesis.

Conclusions Overall, these findings provide initial insights into the TF regulatory network underlying flavones and CQAs accumulation in the chrysanthemum capitulum, which laid a theoretical foundation for the quality improvement of *C. morifolium* ‘HangBaiJu’ and the high-quality development of the industry.

Keywords Flavone, Caffeoylquinic acid, Biosynthesis, Transcriptome, Transcriptional regulation, Chrysanthemum

[†]Chenfei Lu and Xiaoyun Yan Co-first author.

*Correspondence:
Lanying Pan
lanyingpan@zjsru.edu.cn
Qingsong Shao
sqszjfc@126.com

¹Zhejiang Provincial Key Laboratory of Resources Protection and Innovation of Traditional Chinese Medicine, Zhejiang A&F University, Hangzhou 311300, China

²State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Hangzhou 311300, China

³Shulan International Medical College, Zhejiang Shuren University, Hangzhou 310015, China



Introduction

Chrysanthemum morifolium ‘HangBaiJu’ is a traditional medicinal and edible plant in China, which is favored by consumers because of its disease-preventing and health-improving functions [1–3]. The capitulum, as the medicinal and edible part of chrysanthemum, is rich in flavones and caffeoylquinic acids (CQAs). Flavones and CQAs are the main active ingredients of chrysanthemum, exhibiting diverse biological activities such as anti-inflammatory [4], antibacterial [5], antioxidant [6], and anti-diabetes [7]. They are also important indicators for evaluating the quality of *C. morifolium* ‘HangBaiJu’ [8, 9]. Clarifying the metabolic regulation mechanism of flavone and CQA biosynthesis can lay the foundation for quality enhancement and future metabolic engineering of chrysanthemum.

Flavones and CQAs are one of the most essential components of flavonoids and phenolic acids, respectively [10, 11]. Currently, the metabolic genes of flavonoids and phenolic acids (Fig. 1) have been extensively studied in plants [12, 13]. It was recorded that flavonoid and phenolic acid metabolism both belong to the phenylalanine pathway [14], and the key enzymes are phenylalanine ammonia-lyase (PAL), 4-coumarate-CoA ligase (4CL), and cinnamate 4-hydroxylase (C4H), which catalyze the synthesis of p-coumaroyl-CoA from phenylalanine [13]. Subsequently, p4-Coumaroyl-CoA is catalyzed by a variety of different enzymes to form flavonoids and phenolic acids, respectively. In the pathway of flavonoid

biosynthesis, p4-Coumaroyl-CoA is catalyzed to flavanone (naringenin) by chalcone synthase (CHS) and Chalcone isomerase (CHI). Naringenin acts as an important precursor for the synthesis of many flavonoid subclasses such as flavones, flavanols, anthocyanins, proanthocyanidins, and isoflavone, then ultimately flows into multiple branching pathways of different flavonoid biosynthesis [13]. Subsequent initiation of phenolic acid biosynthesis comes from 4-Coumaroyl-CoA catalyzed by hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase (HQT), Hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase (HCT) and p-Coumarate 3-hydroxylase (C3H) to finally obtain CQAs such chlorogenic acid and Di-caffeoylquinic acid [15]. Several studies have declared that the expression level of metabolic genes is closely connected to flavones and CQAs accumulation in plants. However, the specific transcriptional regulation controlling the expression of these structural genes has remained extremely restricted in chrysanthemum.

There are a large number of TF families participating in plant biological development, adversity stress and secondary metabolism, such as the MYB family, the bHLH family, the MADS-box family, the NAC family, and the ERF/AP2 family. Numerous studies have shown the significance of MYB and bHLH TFs in regulating flavonoid and phenolic acid biosynthesis, and they often form complexes to exercise their functions [16]. For instance, EsMYBA1, a MYB TF in *Epimedium sagittatum*, interacted with bHLH TF involved in the flavonoid synthesis

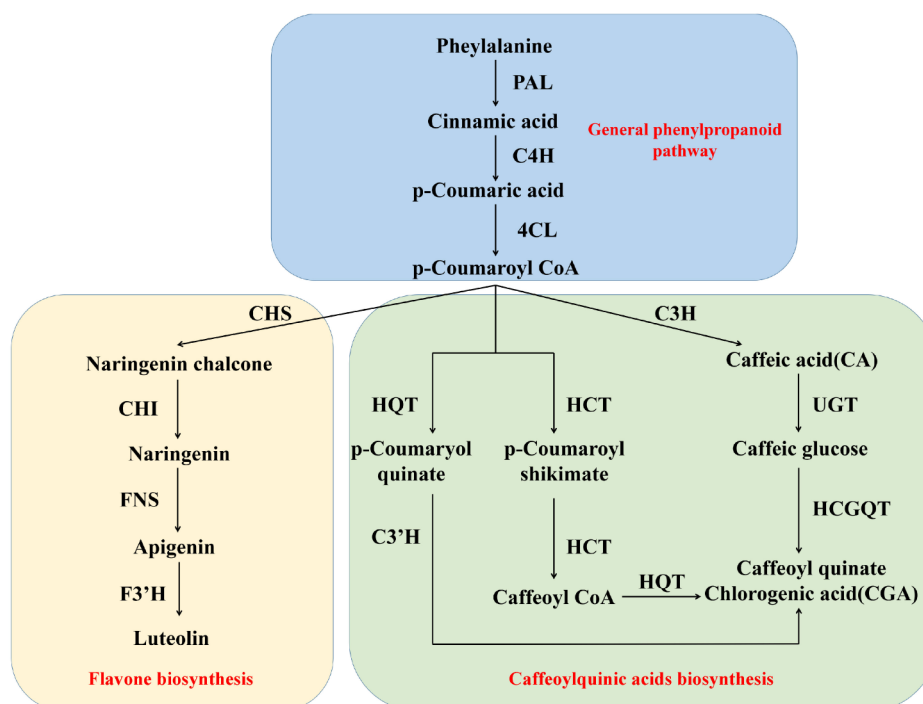


Fig. 1 Metabolic pathways of flavone and caffeoylquinic acid (CQA)

pathway and activated the promoters of genes activating the flavonoid synthesis pathway, thereby promoting anthocyanin biosynthesis [17]. In *Salvia miltiorrhiza*, the MYB TF SmMYB52 and the bHLH TFs SmbHLH148 and SmMYC2 both up-regulated Salviaolic acid B biosynthesis pathway genes and thereby promoted the accumulation of Salviaolic acid B [18]. Currently, CmMYB3 in chrysanthemum has been reported to be involved in flavonol biosynthesis [19], while CmMYB6 [20], CmMYB9a [21] and CmMYB3-like [22] have been reported to participate in anthocyanin biosynthesis. In contrast, the MYB and bHLH TFs that regulate the biosynthesis of flavone and CQA in chrysanthemum are presently unclear.

MADS-box TFs play a key role in the regulation of capitulum development [23], and have also been reported to be involved in the regulation of flavonoid biosynthesis. In tomato (*Solanum lycopersicum*), the MADS-box TF SITAGL1 was found to regulate flavonoid biosynthesis by modulating the key rate-limiting enzyme 3-Dehydroquinate dehydratase/shikimate dehydrogenase [24]. Ethylene is an important hormone during flower development, and AP2/ERF TFs play an important role in respond to ethylene signals [25, 26]. The development of flowers is accompanied by the change of metabolites, so AP2/ERF TFs also have a certain relationship with plant secondary metabolism. At present, some studies have demonstrated the regulatory roles of AP2/ERF TFs in flavonoid and phenolic acid biosynthesis. Zhao et al. [27] identified three AP2/ERF TFs in citrus that activated the expression level of *CHI* gene to regulate the content of flavonoids. SmERF115, an AP2/ERF TF in *Salvia miltiorrhiza*, was able to promote phenolic acid production by regulating the structural genes of the phenolic acid biosynthesis pathway, including *PAL3* and *4CL5* [28]. However, the MADS-box and AP2/ERF TFs affecting metabolism in chrysanthemum are currently unknown.

In this study, high performance liquid chromatography (HPLC), an analytical technique for separating and quantifying chemicals, was used to determine the contents of flavonoids and CQAs at different developmental stages,

which revealed the dynamic changes of flavones and CQAs during the capitulum development. Meanwhile, the molecular mechanisms regulating flavone and CQA biosynthesis during the development of chrysanthemum capitulum were explored by transcriptome sequencing, utilizing weighted gene co-expression correlation network analysis (WGCNA), K-means and correlation analysis. Our study improved the understanding of the flavone and CQA biosynthesis pathways and provided a theoretical foundation for future studies on metabolic engineering of flavones and CQAs in chrysanthemum.

Materials and methods

Plant materials

C. morifolium 'HangBaiJu' planted in Zhejiang A & F University was used as the material for the experiment. The development of chrysanthemum capitulum was divided into five stages (S1-S5) (Fig. 2). S1 was described as when the ray florets had barely outgrown the bract; S2 was defined as when the ray florets had outgrown the bract partially; S3 was identified as when the ray florets completely outgrew the bract; S4 was defined as when the ray florets were fully opened; and S5 was described as when the outer whorl of ray florets began to decay [29]. Chrysanthemum capitulum from these five growth stages were selected for active components contents determination, comparative transcriptome analysis, WGCNA analysis, K-means and qRT-PCR analysis.

Extraction and contents analysis of total flavonoids and polyphenol

0.25 mg of extraction solution, 1.2 mL of water and 0.75 μL NaNO_3 (5%) were mixed for 6 min. Subsequently, 0.15 μL of AlCl_3 (10%) was added and the mixture was allowed to react for 5 min. Then, 0.5 mL of NaOH (5%) and 0.275 mL of water were added to the mixture. Absorbance was measured at 510 nm. Each stage of sample extraction was repeated three times. The total flavonoid content (TFC) was expressed as quercetin equivalent per gram of sample dry weight. [30].



Fig. 2 Differentiation stages of *C. morifolium* 'HangBaiJu'. S1 was described as when the ray florets had barely outgrown the bract; S2 was defined as when the ray florets had outgrown the bract partially; S3 was identified as when the ray florets completely outgrew the bract; S4 was defined as when the ray florets were fully opened; and S5 was described as when the outer whorl of ray florets began to decay

1.25 g dried flower powder was added to 25 mL of 80% methanol and placed in a 25 °C shaker (150 RPM) for 24 h to obtain the methanol extract. Samples from each stage were extracted three times. After filtration, total polyphenolic content (TPC) was quantified with the Folin-Ciocalteu reagent [31]. TPC was reported as mg of equivalent tannic acid per gram of sample on a dry basis.

Extraction and HPLC analysis of the main flavones and CQAs

0.10 g chrysanthemum powder was extracted via ultrasonic extraction with 5.00 mL 60% CH₃OH at room temperature, and repeated three times per sample. The extraction solution and standard solution including Chlorogenic acid, Apigenin-7-O-glucoside, Luteolin-7-O-glucoside and 3,5-Di-caffeoylquinic were filtered through 0.45 µm filters for HPLC analysis. The HPLC separation was performed on a Waters 2695 series HPLC system using a WondaSil-C18 column (4.6 mm × 250 mm, 5 µm), sentry guard column was used at flow rate of 1.0 mL/min. The column oven temperature was set at 25°C. The mobile phase consisted of a combination of A (0.1% formic acid in water) and B (0.1% formic acid in acetonitrile). The gradient was varied linearly from 10 to 26% B (v/v) in 40 min, 65% B at 70 min, and finally to 100% B at 71 min and held at 100% B to 75 min. The DAD was set at 350 nm for real-time monitoring of the peak intensity [32].

RNA-seq and exploration of differential expression genes (DEGs) of five stages

In this study, RNA-seq of *C. morifolium* 'HangBaiJu' at the five stages was performed, and samples from each stage were repeated three times.

Total RNA was extracted using the RNAsimple Total RNA kit (Tiangen Biotech, China), and treated with DNaseI to eliminate genomic DNA contamination. Total RNA yield (ng/µL) and purity (260:280 wavelength ratios) was measured by using Nanodrop2000 (Eppendorf, USA) instrument. 2 µg RNA from each of the samples were used for the synthesis of single stranded cDNA by Bio-Rad iScript™ (Hercules, CA, USA). Then cDNA libraries were constructed using Hieff NGS Ultima Dual-mode mRNA Library Prep Kit for Illumina (Yeasen, China) and sequenced on the Illumina NovaSeq 6000 sequencing platform (Illumina, San Diego, CA, USA) by Biomarker Technologies Corporation (Beijing, China). The cDNA libraries sequenced by Illumina ranged in size from 200 bp to 400 bp. Raw reads were evaluated qualitatively by the Fastp tool (v0.23.1). Low quality bases (Q<20) and adapter sequences were trimmed and clean reads were screened for errors and corrected accordingly. Clean reads were obtained and then assembled using Trinity (version: v2.5.1, major parameter: -min_contig_length

200 -group_pairs_distance 500 -min_kmer_cov 1) [33]. Short reads were first assembled into contigs with no gaps, and the reads were mapped back to the contigs. Obtained contigs clustered into sets of connected scaffolds (The contigs fulfill the following conditions will be clustered into one scaffold: (a) Perfect overlap of k-1 bases. (b) A minimal number of reads spanned between two contig junctions. (c) (k-1)/2 bases mapped backed to both ends of (k-1)mer junction). Finally, De Bruijn was applied here to recognize unigenes in the scaffolds. The Bowtie (<https://bowtie-bio.sourceforge.net/index.shtml>) [34] was used to map reads back to the unigenes. According to the comparison results, the unigene expression was calculated and normalized to Transcript per Million Mapped Reads (FPKM) through RSEM [35]. The unigene sequences were aligned to the NR, Swiss-Prot, COG, KOG, eggNOG4.5, and KEGG databases using DIAMOND software (version: v2.0.4) [36] to obtain annotations.

The transcript abundance of unigenes was estimated via the fragments per kilobase of transcript per million mapped reads (FPKM) [35]. DEGs analysis with the false discovery rate (FDR)<0.01 and the fold change (FC)≥2 was performed using DESeq2 [37].

WGCNA and K-means clustering analysis

In order to screen genes involved in the biosynthesis of flavone and CQA in *C. morifolium* 'HangBaiJu', WGCNA analysis was performed on BMKCloud (www.biocloud.net) through "WGCNA" package under R environment (<https://horvath.genetics.ucla.edu/html/Coexpression-Network/Rpackages/WGCNA/>) First of all, the gene expression file is converted into a suitable format, and the soft thresholding power (β value) was filtered based on the calculation of scale-free topological fit index and mean connectivity. The best β value was confirmed with a scale-free fit index bigger than 0.85 as well as the highest mean connectivity by performing a gradient test from 1 to 30. After that, the topological overlap matrix (TOM) was constructed by calculating the topological overlap between pairwise genes, and hierarchical clustering analysis was performed. The co-expression relationships among different modules were analysed and modules with high similarity were merged at the threshold of 0.25 [38]. Subsequently, the modules showing the highest correlation with the content of flavones and CQAs were identified, and the genes in the modules were subjected to KEGG analysis.

K-means cluster analysis was carried out on BMKCloud (www.biocloud.net) through R package (cluster package, Biobase package, Q-value package). Eleven clusters were chosen for grouping the 46,100 unigenes in this study. Those 11 clusters were used to identify centroids, defined as an average point specific to a cluster of points.

With centroids moved into a position such that an optimum separation of objects into groups occurred, clustering was run for 1,000 repetitions using squared-Euclidian distance as the algorithm to measure pairwise distance between data points. Having 11 clusters fit the dataset so that all the values were affiliated with a specific cluster [39].

Sequence analysis and phylogenetic tree construction of TFs

For investigating the potential biological functions, phylogenetic tree of MYB, bHLH and MADS-box was constructed. Sequence alignments were performed using the ClustalW mode in Molecular Evolutionary Genetics Analysis version7 (MEGA7), and phylogenetic trees were subsequently constructed according to the neighbor-joining statistical method. Tree nodes were evaluated using the bootstrap method for 1000 replicates, and branches corresponding to partitions reproduced in less than 50% of the bootstrap replicates were condensed into single branches. Evolutionary distances were computed using the p-distance method and expressed in units of amino acid differences per site.

Real-time quantitative PCR analysis (Verification of DEGs) and co-expression network construction

Seven key structural genes and nine TFs involved in flavone and CQA biosynthesis were selected to validate the accuracy of the transcriptome results through qRT-PCR analysis. Total RNA was extracted by SteadyPure Plant RNA Extraction Kit (AG21019), and was repeated three times for each sample. First-strand cDNA was synthesized using a Evo M-MLVMix Kit with gDNA Clean for qPCR (AG11728). qRT-PCR was carried out with SYBR Green Premix Pro Taq HS qPCR Kit (AG11701) on a CFX96 Touch Real-Time PCR Detection System (Bio-rad). 35 S was used as a reference gene to normalize the expression levels between samples. The relative expression levels of genes were calculated using the $2^{-\Delta\Delta Cq}$ method [40]. Primers were listed in Table S2.

The Protein-Protein Interaction (PPI) network was constructed using the STRING database (<https://string-db.org/>) and the PPI network was visualised and landscaped using Cytoscape v3.10.0 software [41].

Statistical analysis

All experimental data was measured with three biological replicates. One-way ANOVA analysis based on Duncan's multiple comparison test was conducted using SPSS 26.0 statistical software ($P < 0.05$). All data were presented as "mean \pm standard error". Different letters were added to indicate the significant differences. The figures were generated by the graphPad Prism9.

Results

Flavones and CQAs contents and compositional analysis of chrysanthemum capitulum at five developmental stages

C. morifolium 'HangBaiJu' is rich in flavones and CQAs compounds, which has a variety of biological activities such as anti-inflammation, antibacterial and antioxidant [42, 43]. In order to clarify the variations in active substances in the chrysanthemum capitulum at different developmental stages, total content of flavonoids (TFC, including flavone, flavonol and anthocyanin, etc.) and polyphenols (TPC, including CQAs, coumarins, flavonoids, etc. [44]) were determined by ultraviolet spectrophotometry. As depicted in Fig. 3A, TFC was higher in S1 (17.43 ± 0.975 mg/g DW) and S2 (16.95 ± 3.956 mg/g DW), and lower in other stages. Different from the trend of flavonoids content, TPC increased initially and then decreased with the development of chrysanthemum capitulum (Fig. 3B), reaching its peak in S2 (60.16 ± 10.440 mg/g DW).

Luteolin-7-O-glucoside, Apigenin-7-O-glucoside, Chlorogenic acid, and 3, 5-Di-caffeoylquinic acid are the main components of the biological activities of *C. morifolium* 'HangBaiJu' [45]. They are also served as standards for judging the quality of *C. morifolium* 'HangBaiJu' in *Pharmacopoeia of the people's Republic of China*. HPLC was used to analyse the contents of Luteolin-7-O-glucoside, Apigenin-7-O-glucoside, Chlorogenic acid and 3, 5-Di-caffeoylquinic in different stages of chrysanthemum capitulum (Fig. 3C). It was observed that chlorogenic acid concentration initially increased with the development of capitulum, peaking at S2 (3.09 ± 0.353 mg/g DW), while 3, 5-Di-caffeoylquinic and Apigenin-7-O-glucoside similarly reached a maximum at S2, with values of 14.42 ± 0.616 mg/g DW and 2.38 ± 0.212 mg/g DW, respectively, and gradually decreased with the development of chrysanthemum capitulum. The content of Luteolin-7-O-glucoside was the highest in S1 (1.48 ± 0.089 mg/g DW) and gradually decreased with the development of capitulum. These results indicated that the contents of active ingredients were relatively higher in S1 and S2, and progressively declined in the S3 to S5 as the chrysanthemum capitulum developed.

3.2 Sequencing, assembly, functional annotation of the transcriptome and expression analysis of flavone and CQA metabolic genes in five stages.

The active substances are catalyzed by metabolic enzymes. In this study, RNA-seq of the five stages of capitulum was performed to further explore the accumulation mechanism of flavone and CQA in *C. morifolium* 'HangBaiJu' (Fig. 4A). A total of 110.52 Gb clean data was obtained by sequencing 15 samples, Q30 (base correct recognition rate of 99.9%) was 91.17-95.43% (Table S3). Following assembly, 49,231 Unigenes were obtained, with an N50 of 2,029 bp (Table S4). A total of 33,858 Unigenes

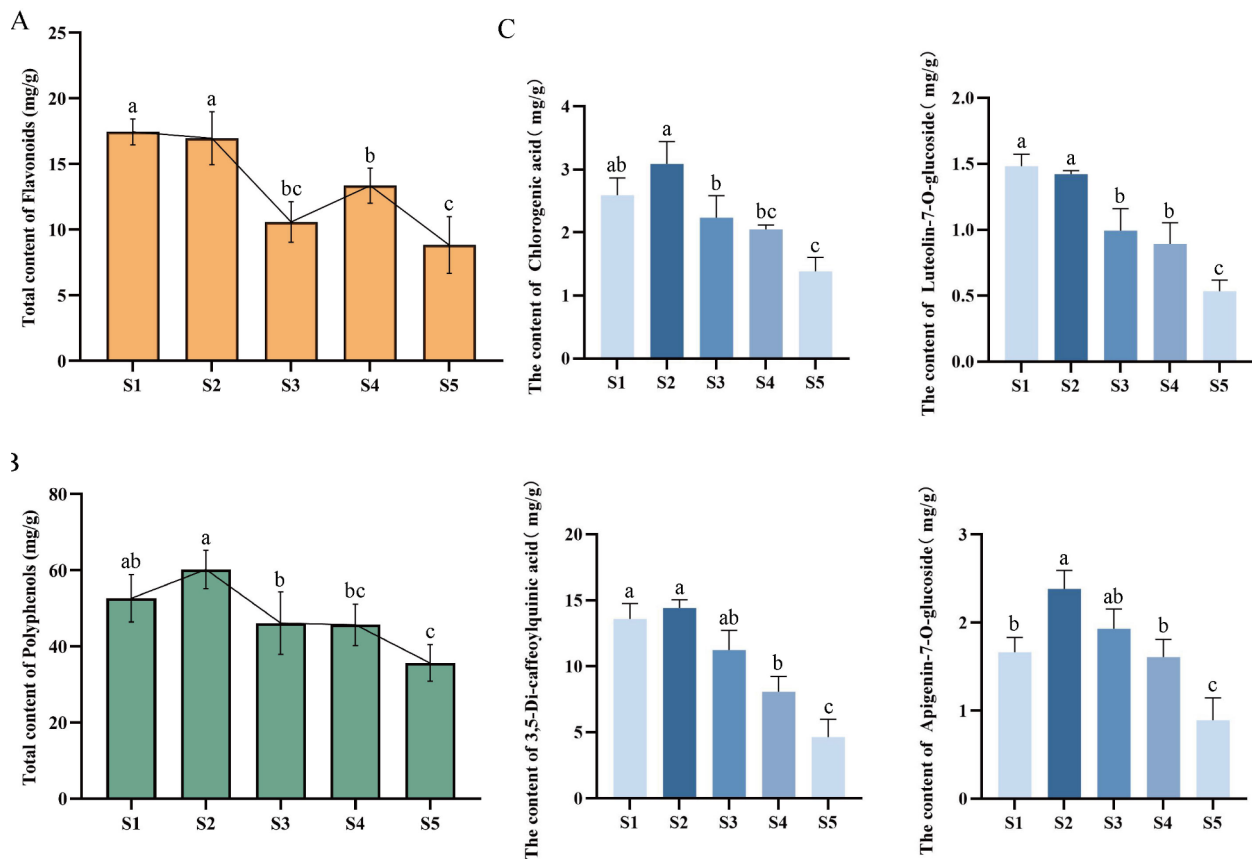


Fig. 3 Contents of active components in five stages of chrysanthemum capitulum. **(A)** Total content of flavonoids. **(b)** Total content of polyphenols. **(c)** Contents of Chlorogenic acid, Luteolin-7-O-glucoside, 3, 5-Di-caffeoylquinic and Apigenin-7-O-glucoside. The significance of difference is calculated by one-way ANOVA analysis based on Duncan's multiple comparison test. Different letters (a-c) between two averages indicate a significant difference ($P < 0.05$) between the two averages, and the presence of the same letters between two averages indicates no significant difference ($P \geq 0.05$) between two averages

with annotation information were finally obtained, of which 16,664 (49.22%) Unigenes were annotated to the KOG database. 20,424 (60.32%) Unigenes were annotated in Pfam database, 20,628 (60.93%) Unigenes were annotated in Swissprot database, and 33,129 (97.85%) Unigenes were annotated in Nr database (Fig. 4B).

A total of 33,858 Unigenes with annotation information were finally obtained, of which 16,664 (49.22%) Unigenes were annotated to the KOG database. 20,424 (60.32%) Unigenes were annotated in Pfam database, 20,628 (60.93%) Unigenes were annotated in Swissprot database, and 33,129 (97.85%) Unigenes were annotated in Nr database (Fig. 4B). Among them, cellular component functional group has the most proteins related to intracellular and cellular anatomical entities. In the molecular function group, it is noted that the number of Unigenes related to binding and catalytic activity is the largest. The proteins related to cellular process, metabolic process and biological regulation were the most in the functional group of biological processes. According to COG annotations, 7,626 (22.52%) Unigenes were classified into

26 functional groups (Fig. S1B). Additionally, a total of 20,542 (60.67%) Unigenes were mapped to 136 KEGG pathways (Fig. S1C).

To investigate whether these significant differences in flavones and CQAs accumulation among the five stages were caused by the different expression of their structural genes, the FPKM of relevant genes were thoroughly exploited based on transcriptome database. We have acquired 5 members of *PAL* genes, 1 member of *C4H* gene, 8 members of *4CL* genes, 3 members of *CHS* genes, 7 members of *CHI* genes, 1 member of *FNS* gene, 3 members of *F3'H* genes, 4 members of *HQT* genes, 1 member of *C3'H* gene and 5 members of *HCT* genes (Fig. 4C). In order to screen out the most important structural genes, the correlation analysis was conducted between the FPKM of structural genes and metabolite content (Fig. S2), and genes with high correlation with metabolites were screened: *CmPAL* (TRINITY_DN45389_c0_g1, TRINITY_DN144576_c0_g1), *CmCHS* (TRINITY_DN52214_c0_g1, TRINITY_DN124567_c0_g1), *CmFNS* (TRINITY_DN8097_c0_g2), *CmHQT*

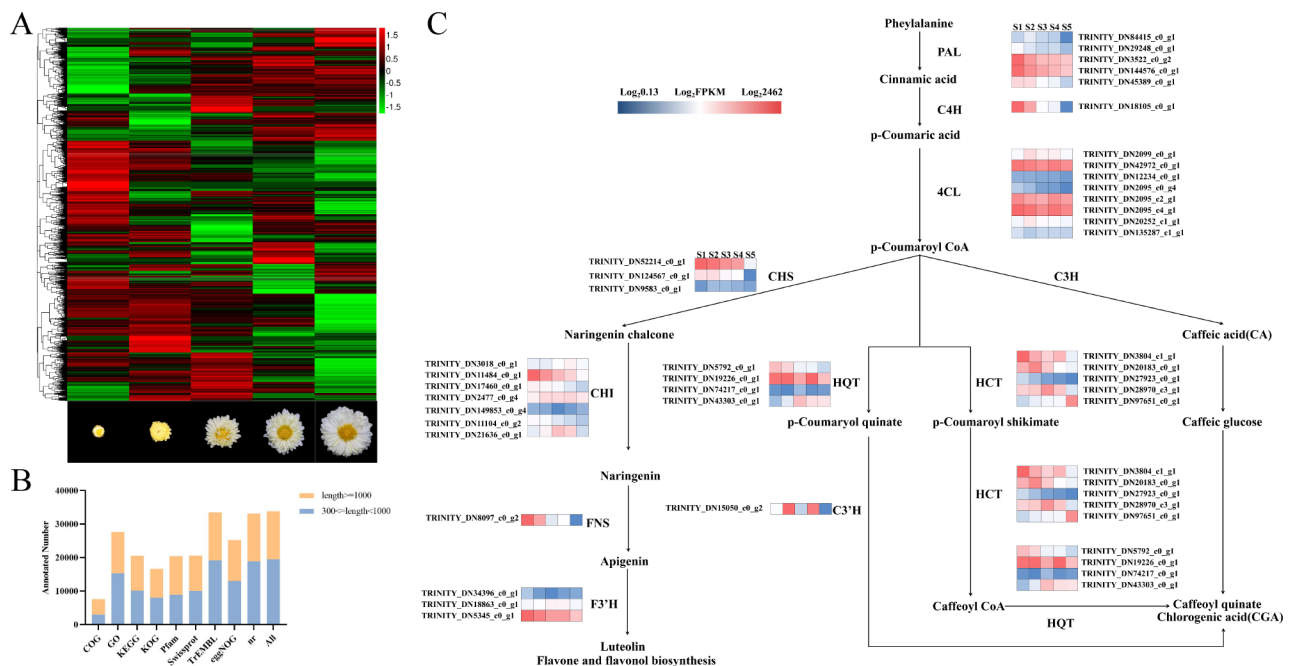


Fig. 4 Expression of the structural genes of the flavone and CQA metabolic pathway at five stages according to transcriptome database. **(A)** FPKM heatmaps of five stages by RNA-seq. The heatmap shows high expression in red and low expression in green. The color of the heatmap from red to green indicates the amount of gene expression from high to low. **(B)** Statistics on the number of Unigenes in different databases. **(C)** Expression of the structural genes of the flavone and caffeoylquinic acid metabolic pathway. The color of the heatmap from red to blue indicates the amount of gene expression from high to low

(TRINITY_DN5792_c0_g1), and *CmHCT* (TRINITY_DN3804_c1_g1, TRINITY_DN20183_c0_g1). The expression of structural genes is intimately related to TFs, so which TFs regulating these key structural genes need to be investigated.

Screening for TFs regulating flavones and CQAs accumulation through WGCNA

TFs act as switches that control the expression of downstream genes. Here, WGCNA analysis was carried out to screen for TFs that might regulate flavones and CQAs metabolism (Fig. 5A). According to the expression patterns, the obtained genes were divided into 13 modules (Fig. S3). Eigengenes (the first principal component in the module) represented the gene expression profile of the whole module. Based on previous studies, S1 and S2 were the key stages for accumulation of flavones and CQAs. By comparing the correlation coefficients between different modules and samples, it was found that the module sienna3, which was highly correlated with the accumulation of flavones and CQAs (Fig. 5B), contained a large number of key metabolic genes. We have speculated that the genes in this module might be closely related to the biosynthesis of flavone and CQA.

Further analysis revealed that sienna3 module contained 5,011 genes, with abundant expression in S1 and S2 (Fig. 5C). KEGG functional analysis showed that 57

genes were enriched in the phenylpropanoid biosynthesis pathway (ko00940), and 32 genes were enriched in the flavonoid biosynthesis pathway (ko00941) (Fig. 5D). Furthermore, several TFs belonging to basic bHLH, AP2/ERF, MYB, WRKY and MADS-box TF families were identified (Fig. 5E), suggesting their potential roles in regulating flavone and CQA metabolism.

Screening of key upstream TFs regulating flavone and CQA metabolism by K-means

Besides WGCNA analysis, K-means clustering analysis was utilized to further explore TFs that may affected flavones and CQAs metabolism. K-means cluster analysis divides all DEGs into 11 clusters, illustrating the overall expression pattern of DEGs (Fig. 6A). The genes in cluster 1, cluster 7, cluster 8 and cluster 10 were expressed at high levels at S1. Among them, the expression of genes in cluster10 gradually decreased from S2 to S5, aligning with the accumulation pattern of active ingredients. Therefore, we hypothesized that the genes within cluster10 might be involved in flavone and CQA biosynthesis. Several metabolic genes such as *CmPAL*, *CmCHS* and *CmFNS* were identified to be concentrated in cluster10. There were also a significant number of TFs associated with flavone and CQA biosynthesis, including 16 members of the MYB TFs, 12 members of the bHLH TFs, and some other TFs (Fig. 6B) found in cluster10.

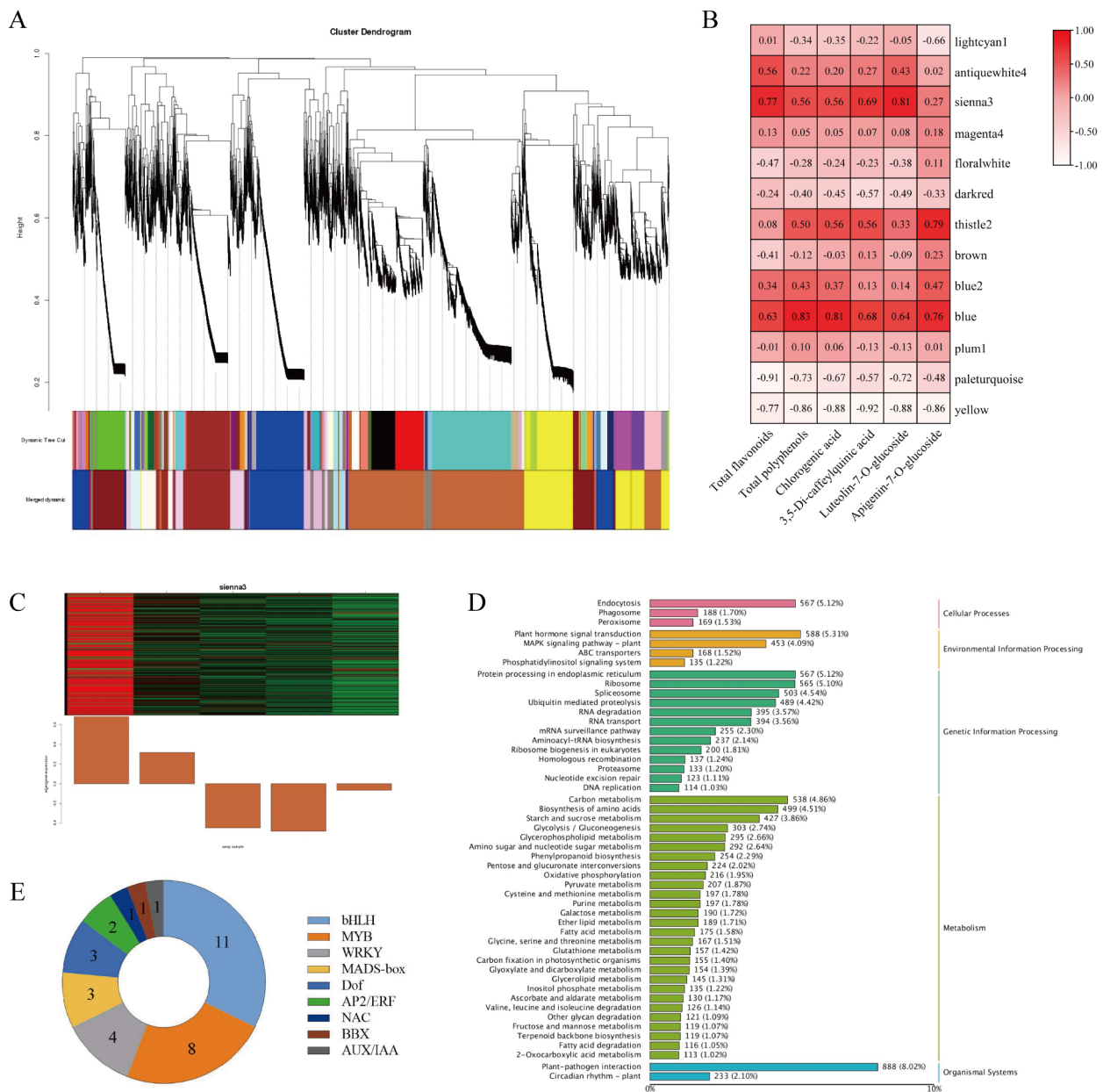


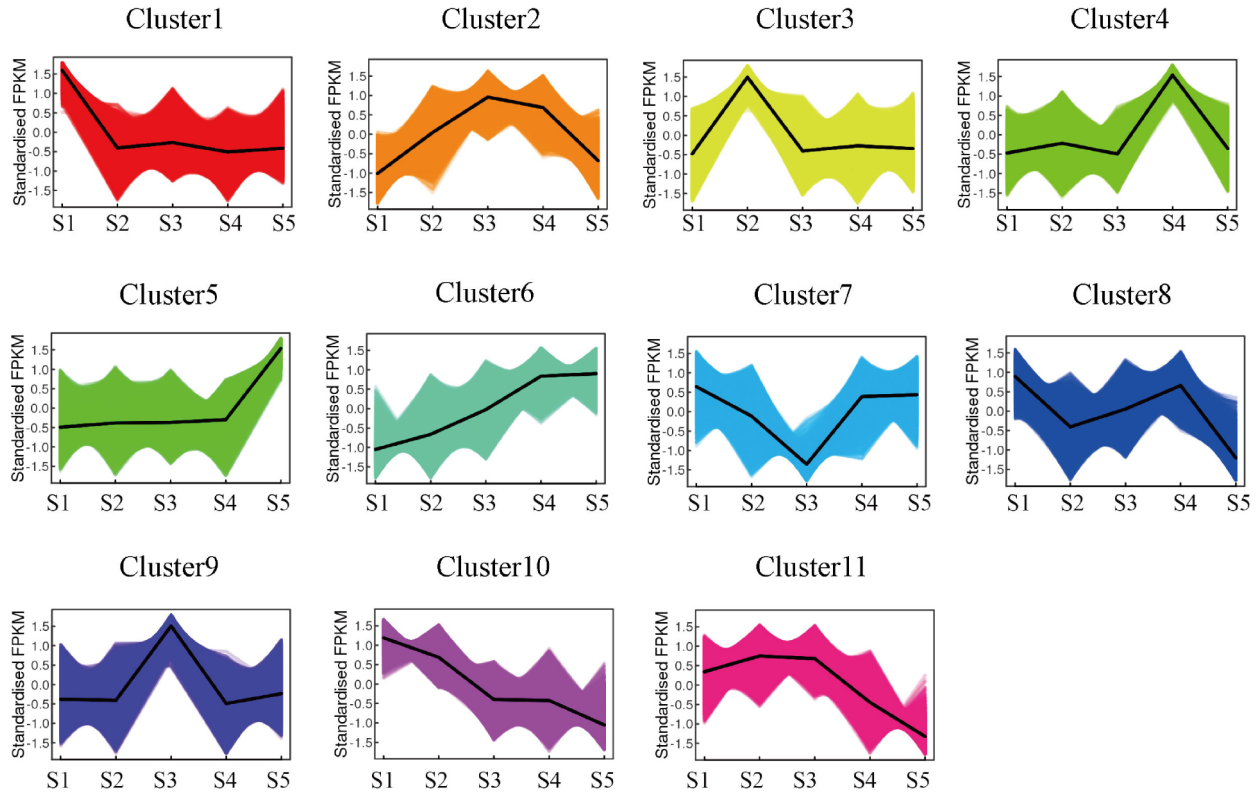
Fig. 5 WGCNA of all Unigenes in the transcriptome and the expression pattern and annotation of genes in the sienna3 module. **(A)** WGCNA of all Unigenes in the transcriptome. **(B)** Module-trait weight correlations and corresponding p-values. The color scale shows module-trait correlation from low (white) to high (red). **(C)** Expression pattern of eigengenes in the sienna3 module. **(D)** KEGG pathways significantly enrich genes in the sienna3 module. **(E)** Statistics on the number of important TFs in the sienna3 module

To further explore potential TFs which regulated flavone and CQA biosynthesis in chrysanthemum capitulum, a comprehensive analysis of the Sienna3 module in WGCNA and the cluster10 TFs in K-means was conducted. Generally, 21 MYB TFs, 19 bHLH TFs, 3 MADS-box TFs, 2 AP2/ERF TFs, 3 NAC TFs, and 2 BBX TFs were present in the sienna3 module or cluster10 (Table S5). It was hypothesized that these TFs may be involved in flavone and CQA biosynthesis by regulating the expression of these structural genes.

Sequence analysis and phylogenetic tree construction of MYB, bHLH and other TFs

As MYB and bHLH play crucial roles in flavones and CQAs metabolism. To explore the evolutionary relationship between the screened MYBs and the MYB TFs in Arabidopsis and other species, phylogenetic trees were constructed (Fig. S4, Fig. 7A), which showed that CmMYB1 (TRINITY_DN4766_c2_g1) exhibited high homology with AtMYB3 belonging to subgroup 4. Subgroup 4 of MYB TFs have been reported to be

A



B

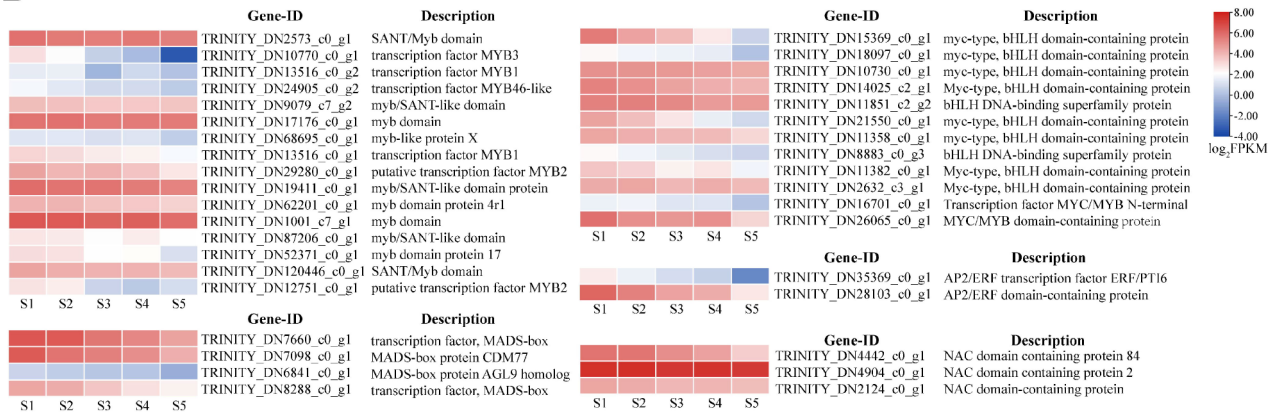


Fig. 6 The eleven clusters of DEGs with different expression patterns and TFs in cluster10

transcriptional repressors of flavonoid biosynthesis [46]. CmMYB3 (TRINITY_DN10770_c0_g1) were clustered with AtMYB8, which belonged to subgroup 7 and was responsible for regulating flavonoid and phenolic acid biosynthesis [47]. Additionally, through constructing the bHLH phylogenetic trees as above (Fig. S5, Fig. 7B), we found that CmMYC2 (TRINITY_DN30658_c0_g1) clustered with AtbHLH4 belonging to subgroup IIIe, and CmbHLH143 (TRINITY_DN14025_c2_g1) was grouped with AtbHLH8 in subgroup VIIa, both subgroups have been reported to be involved in the regulation

of flavonoid biosynthesis [48]. Sequence comparison (Fig. S6) and motif analysis (Fig. 7C, D) revealed that CmMYB3 and CmMYB1 presented a highly conserved R2R3 domain. In addition, a typical bHLH conservative region was observed in the N-terminal region of the CmMYC2 and CmbHLH143 protein. Furthermore, evolutionary trees were constructed for three protein sequences annotated as MADS-box by NR (Fig. S7), and we have found that CDM77 (TRINITY_DN7098_c0_g1) was classified as a class E MADS-domain TF belonging to SEP subfamily. AGL15 (TRINITY_DN7660_c0_g1), a

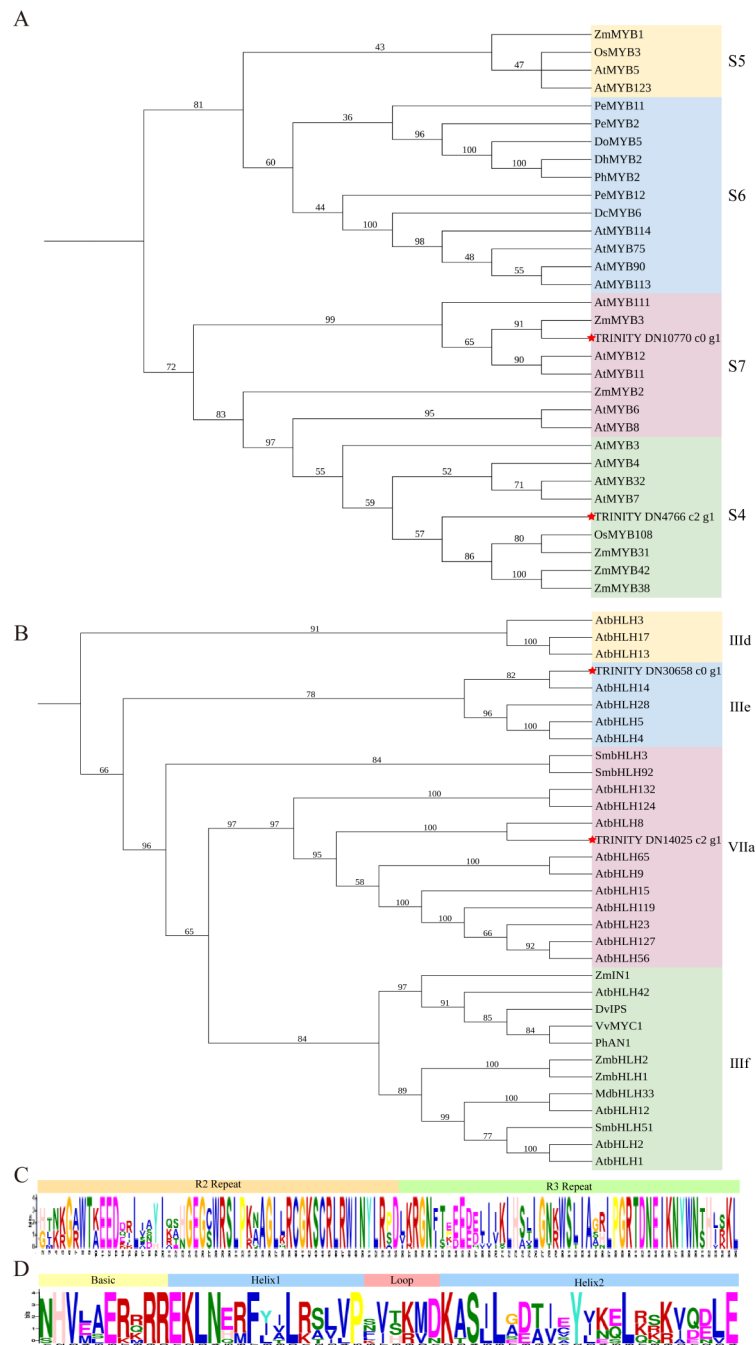


Fig. 7 Phylogenetic analysis (**A, B**) and Motif prediction (**C, D**) of MYB and bHLH TFs

class C MADS-domain TE, was a member of AGL17 subfamily. MADS23 (TRINITY_DN8288_c0_g1) was a class B TF which belongs to the PI subfamily. These results suggested that these TFs associated with capitulum development of chrysanthemum may also related to the regulate accumulation of flavone and CQA [49]. Additionally, based on sequence alignment, it was also found that TRINITY_DN35369_c0_g1, which is annotated as

an AP2/ERF TF, may function as a transcriptional activator for flavonoid and phenolic acid metabolism.

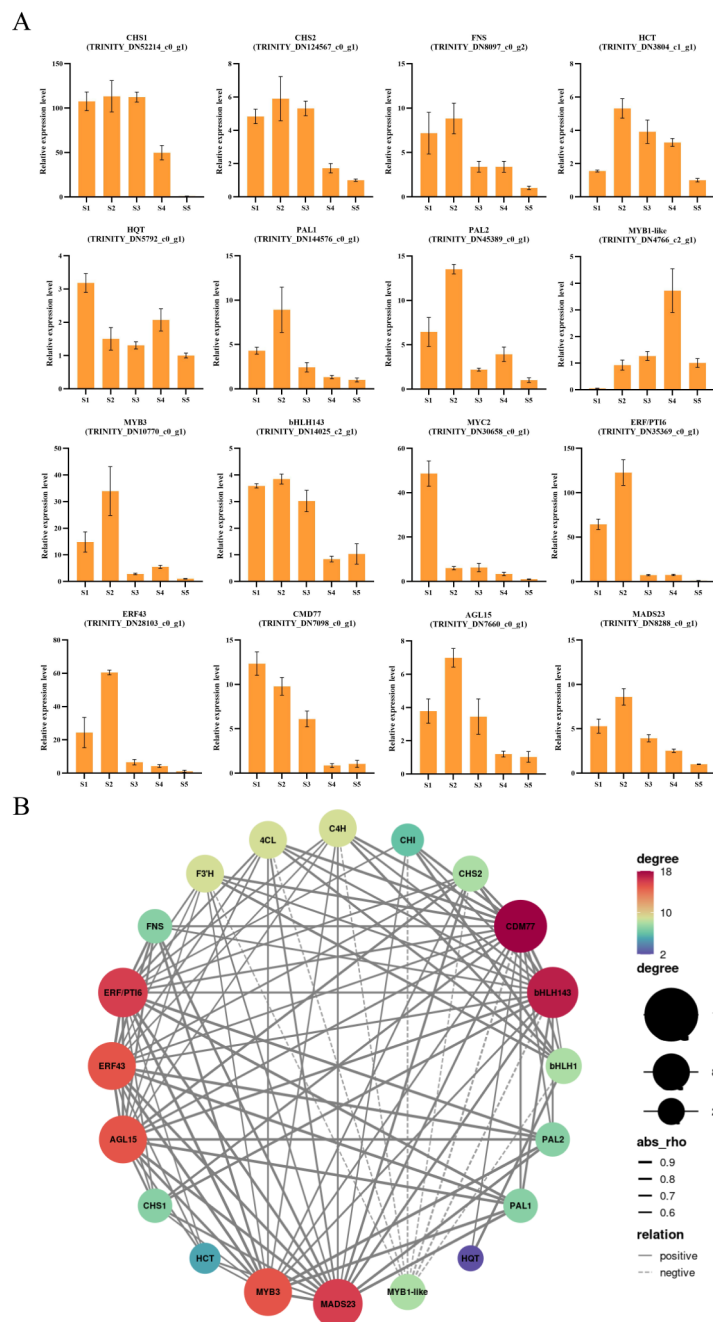
qRT-PCR expression validation and co-expression network construction of important structural genes and TFs

Based on the results of transcriptome and phylogenetic analysis, it was hypothesized that the above upstream TFs (CmMYB1-like, CmMYB3, CmMYC3, CmbHLH143, CmCDM77, CmAGL17, CmMADS23, CmERF43, and

CmERF/PT16) regulated the expression level of metabolism genes, thereby altering the accumulation of flavones and CQAs during chrysanthemum capitulum development. In order to clarify the expression patterns of structural genes and TFs, qRT-PCR analyses were carried out in capitulum at different developmental stages in this study. The results (Fig. 8A) demonstrated that the expression pattern of the most structural genes and TFs were consistent with the changes in the trends of the active components, showing an initial increase, reaching

a maximum at S2, followed by a gradual decrease with capitulum development. However, the overall trend of gene expression of *CmMYB1-like* was opposite to that of flavones and CQAs, which might be due to the fact that *CmMYB1-like* belongs to the Subgroup 4 of MYBs that have the role of negative regulation of flavonoid synthesis.

To explore the most important TFs regulating flavone and CQA biosynthesis from the TFs obtained by the above screening, pearson correlation analysis and



co-expression network construction were performed (Fig. 8B). The results revealed that *CmMYB3*, *CmbHLH143*, *CmERF/PTI6* and *CmCDM77* were significantly positively correlated with the structural genes, respectively ($P > 0.9$). Previous studies have shown that MYB and bHLH synergized to form a complex to regulate flavonoid and phenolic acids biosynthesis pathways [16]. *CmMYB3* and *CmbHLH143*, exhibiting similar expression patterns, were strongly correlated with structural genes such as *CmPAL1/2*, *CmCHS1/2*, *CmFNS*, *CmHQT* and *CmHCT*. The interaction between MYB3 and bHLH143 was predicted by STRING (Fig. S8), suggesting that CmMYB3 and CmbHLH143 may form a complex that regulates flavone and CQA biosynthesis at different growth stages of chrysanthemum capitulum. *CmERF/PTI6* and *CmCDM77* have showed highly significant positive correlations ($P > 0.9$) with *CmMYB3* and *CmbHLH143*, respectively, indicating that CmCDM77 and CmERF/PTI6 may regulate the biosynthesis of flavone and CQA during capitulum development as upstream TFs of MYB3 and bHLH143.

Discussion

Accumulation patterns of flavones and CQAs in chrysanthemum capitulum at different developmental stages

Flavone and CQA compounds are the main active ingredients in *C. morifolium* 'HangBaiJu', exhibiting various biological activities. In this study, we observed that Flavone and CQA compounds, such as Chlorogenic acid, Luteolin-7-O-glucoside, Apigenin-7-O-glucoside and 3, 5-Di-caffeoylquinic, were existed in all five stages of chrysanthemum capitulum. Besides, the content of these active ingredients were higher in S1 and S2, and gradually declined in S3-S5 (Fig. 3C). A similar pattern was observed for flavonoids and phenolic compounds during the development of rose flowers [50, 51].

Based on transcriptome sequencing results, a significant number of differentially expressed structural genes involved in the biosynthesis pathways of flavone and CQA were discovered in chrysanthemum. Different expression level of structural genes contribute to changes in the levels of active components. Correlation analysis revealed the key rate-limiting enzyme genes, such as *CmPAL1/2*, *CmCHS1/2*, *CmFNS*, *CmHQT* and *CmHCT*, involved in the metabolism of flavone and CQA in chrysanthemum (Fig. 4C, S2). Among them, *CmPAL1/2* is the rate-limiting enzyme in the phenylalanine metabolic pathway for the synthesis of precursor substances such as flavonoids and phenolic acids [52]. Subsequently, the synthesis of flavone and CQA (chlorogenic acid and its derivatives) is carried out in the presence of *CmCHS1/2* and *CmHCT*, respectively [53]. Besides, flavanones are converted to flavones by the flavonoid-directed branching

metabolizing enzyme *CmFNS* [54], and *CmHQT* acts as a key rate-limiting enzyme in CQA metabolism [55]. Thus, *CmPAL1/2*, *CmCHS1/2*, *CmFNS*, *CmHQT* and *CmHCT* were considered as the key enzyme genes for improving the quality of chrysanthemum. Currently, the study of HQT in the phenolic acid synthesis pathway is relatively rare, and further research on its enzymatic function is needed.

CmMYB3-CmbHLH143 TF complex may regulate flavone and CQA compositions in chrysanthemum capitulum

MYB and bHLH TFs are widely present in plants and have been proven to play key regulatory roles in the biosynthesis of secondary metabolites [56]. Numerous studies have demonstrated the involvement of MYB and bHLH TFs in the biosynthesis of flavonoids and phenolic acids. For instance, ThMYB4 and ThMYB7 in *Tetragium hemsleyanum* promoted flavonoid biosynthesis by regulating the structural genes *ThCHS* and *ThCHI* [57]. Overexpressed *GhMYB18* in cotton and found that it activates the phenylpropanoid signaling pathway, thereby promoting the synthesis of flavonoids and phenolic compounds [58]. EbbHLH80 in *Erigeron breviscapus* has been identified as a regulator of the expression of certain structural genes in the flavonoid biosynthetic pathway, influencing flavonoid accumulation [59]. In *Taraxacum antungense* Kitag, TabHLH1 bound directly to the bHLH-binding motifs of *proTaHQT2* and *proTa4CL*, thereby regulating chlorogenic acid and luteolin synthesis [9]. In this study, WGCNA and k-means analysis were conducted to identify multiple TFs associated with flavone and CQA biosynthesis during the development of chrysanthemum capitulum. Through phylogenetic analysis, qRT-PCR, and co-expression network construction, two TFs, CmMYB3 and CmbHLH143, belonging to the subgroup 7 and subgroup VIIa, respectively, exhibited high correlation with the expression patterns of the metabolic genes (Figs. 7 and 8B). Both subgroups have been reported to be involved in the biosynthesis of flavonoids and phenolic acids [48, 60].

Additionally, MYB often forms complexes with bHLH and WD40 proteins to activate specific gene transcription [54]. For example, NtMYB330 in tobacco (*Nicotiana tabacum* L.) interacted with flavonoid-related bHLH TF NtAn1b and WDR protein NtAn11-1, forming a complex that strongly activates the transcription of *NtDFR1*, a gene involved in the phenylalanine synthesis pathway [61]. Protein interaction prediction suggested a potential interaction between CmMYB3 and CmbHLH143 in the study (Fig. S8). It is hypothesized that CmMYB3 and CmbHLH143 regulated the expression of structural genes through forming MYB-bHLH complexes, thereby participating in the biosynthesis and accumulation of flavone and CQA (Fig. 9).

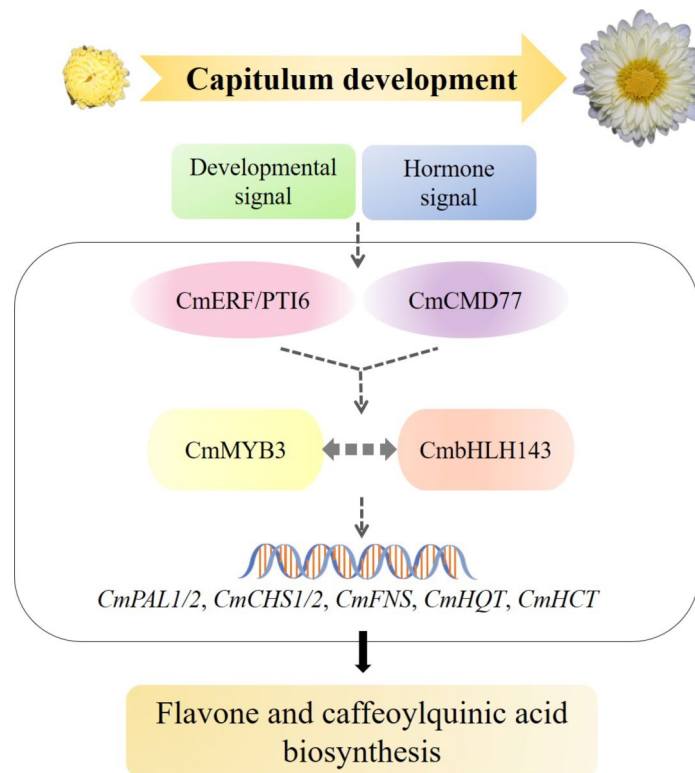


Fig. 9 Mechanisms of flavone and CQA accumulation during the development of chrysanthemum capitulum. The dotted lines represent the predicted interaction relationship or hypothetical regulatory network

AP2/ERF and MADS-box TFs regulate MYB and bHLH TFs expression

AP2/ERF and MADS-box TFs are crucial TFs found in plants, involving various biological processes. AP2/ERF TFs have been reported to play a critical role in plant growth and development, including in response to environmental stress [62]. MADS-box TFs have been extensively studied for their key roles in floral development [63]. In recent years, several studies have indicated that AP2/ERF and MADS-box TFs are also involved in the regulation of plant secondary metabolite biosynthesis by directly regulating the expression of MYB and bHLH TFs. For instance, The AP2/ERF TF FaRAV1 in strawberry was found to directly bind to the *FaMYB10* promoter and promote anthocyanin synthesis [64]. In *Eriobotrya japonica*, EjAP2-1 indirectly regulated lignin biosynthesis by acting as an upstream regulator of *EjMYB1* and *EjMYB2*, which are associated with lignin biosynthesis [65]. The MADS-box TF OAGL6-2 was found to form the protein complex with OAP3-1 in orchids for regulating the expression of *PeMYB11* to form floral patches [66]. Besides, a MADS-box TF SVP3 in kiwifruit (*Actinidia* spp.) was identified as an upstream regulator of *MYB110a*, involved in anthocyanin synthesis [67]. In our study, two TFs CmERF/PTI6 and CmCDM77, from the AP2/ERF family and the MADS-box family were identified through WGCNA, k-means, and correlation

analysis, respectively. The expression patterns of *CmERF/PTI6* and *CmCDM77* showed a consistent trend with *CmMYB3* and *CmbHLH143* (Fig. 8A). Besides, CmERF/PTI6 and CmCDM77 played a central role in gene regulation network (Fig. 8B). We speculate that CmERF/PTI6 and CmCDM77 may act as upstream regulators of *CmMYB3* and *CmbHLH143*, indirectly modulating the expression level of metabolic genes.

Our study provides a preliminary prediction of the regulatory roles of AP2/ERF and MADS-box TFs on MYB and bHLH TFs, constructing a regulatory network (Fig. 9) for flavone and CQA biosynthesis in chrysanthemum capitulum. Elucidating this regulatory mechanism will contribute to a better understanding of the regulatory network of flavones and CQAs metabolism in plants, providing novel strategies for the synthesis and utilization of plant secondary metabolites. However, there are still many questions such as interaction mechanisms and regulatory networks that require further research and validation in the future.

Conclusion

In conclusion, our study revealed the accumulation pattern of flavones and CQAs in the development of chrysanthemum capitulum, with higher levels in S1 and S2, followed by a gradual decrease in S3 to S5 through HPLC. Transcriptome analysis identified key structural genes,

including *CmPAL1/2*, *CmCHS1/2*, *CmFNS*, *CmHQT* and *CmHCT*, participated in the biosynthesis of flavone and CQA. The TFs involved in the regulation of flavones and CQAs were identified by WGCNA and K-means analysis. Correlation analysis and co-expression network construction were used to further screen TFs and construct the regulatory network of flavonoid biosynthesis in *Chrysanthemum*. Specifically, The TFs *CmERF/PTI6* and *CmCMD77* were identified as potential upstream regulators of *CmMYB3* and *CmbHLH143*. Moreover, *CmMYB3* and *CmbHLH143* were hypothesized to form a complex to directly regulate the structural genes *CmPAL1/2*, *CmCHS1/2*, *CmFNS*, *CmHQT*, and *CmHCT*, thereby controlling flavone and CQA biosynthesis. These results have provided valuable insights for further research on the metabolic engineering of chrysanthemum flavones and CQAs, and supplied a in-depth understanding of the molecular mechanisms underlying flavones and CQA biosynthesis in higher plants.

Abbreviations

CQA	Caffeoylquinic acid
HPLC	High Performance Liquid Chromatography
PAL	Phenylalanine ammonia-lyase
4CL	4-coumarate-CoA ligase
C4H	Cinnamate 4-hydroxylase
CHS	Chalcone synthase
CHI	Chalcone isomerase
HQT	Hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase
HCT	Hydroxycinnamoyl-CoA shikimate/quinic hydroxycinnamoyl transferase
C3H	p-Coumarate 3-hydroxylase
WGCNA	Weighted gene co-expression correlation network analysis
TFC	Total flavonoid content
TPC	Total polyphenolic content
FPKM	Fragments per kilobase of transcript per million mapped reads
PPI	Protein-Protein Interaction

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10676-6>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5

Acknowledgements

Not applicable.

Author contributions

C.L. performed a part of experiments, designed the experiments and technical route, and mainly revised the manuscript. X.Y. drafted the manuscript and was responsible for the revision of manuscripts. H.Z. and T.Z. performed the majority of experiments and data analysis. A.G. and Y.L. participated in a part of experiments and helped to collect samples. L.P. and Q.S. revised the manuscript, conceived the project, and final approved of the manuscript. All authors read and approved the final manuscript.

Funding

This work was performed under the National Natural Science Foundation of China (32301645), General Project of Education Department of Zhejiang Province (Y202249814), Scientific Research Foundation of Zhejiang A&F University (2022LFR112), and National Training Program of Innovation and Entrepreneurship for Undergraduates (202310341069).

Data availability

Data is provided within the manuscript or supplementary information files.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 1 April 2024 / Accepted: 30 July 2024

Published online: 03 August 2024

References

1. He J, Zhang C, Zhou L, He Y. Simultaneous determination of five micro-components in *Chrysanthemum morifolium* (Hangbaiju) using near-infrared hyperspectral imaging coupled with deep learning with wavelength selection. *Infrared Phys Technol*. 2021;116:103802.
2. Wang Y, Wang M, Li Y, Wu A, Huang J. Effects of arbuscular mycorrhizal fungi on growth and nitrogen uptake of *Chrysanthemum morifolium* under salt stress. *PLoS ONE*. 2018;13(4):e0196408.
3. Chen J. Essential role of medicine and food homology in health and wellness. *Chin Herb Med*. 2023;15(3):347–8.
4. Li Y, Liu XJ, Su SL, Yan H, Guo S, Qian DW, Duan JA. Evaluation of anti-inflammatory and antioxidant Effects of *Chrysanthemum* Stem and Leaf extract on zebrafish inflammatory bowel Disease Model. *Molecules*. 2022;27(7).
5. Hodaie M, Rahimmalek M, Arzani A. Variation in bioactive compounds, antioxidant and antibacterial activity of Iranian *Chrysanthemum morifolium* cultivars and determination of major polyphenolic compounds based on HPLC analysis. *J Food Sci Technol*. 2021;58(4):1538–48.
6. Khan IA, Xu W, Wang D, Yun A, Khan A, Zongshuai Z, Ijaz MU, Yiqun C, Hussain M, Huang M. Antioxidant potential of *chrysanthemum morifolium* flower extract on lipid and protein oxidation in goat meat patties during refrigerated storage. *J Food Sci*. 2020;85(3):618–27.
7. Wang Z, Bai Z, Yan J, Liu T, Li Y, Xu J, Meng X, Bi Y. Anti-diabetic effects of linarin from *Chrysanthemi Indici Flos* via AMPK activation. *Chin Herb Med*. 2022;14(1):97–103.
8. Zhang Y, Guo C, Hu J, Liu F, Fu S, Guo X, Chen Q, Zhang L, Zhu L, Hou X. Effects of 6-Benzylaminopurine combined with Prohexadione-Ca on yield and quality of *Chrysanthemum morifolium* Ramat Cv. *Hangbaiju Agric*. 2023;13(2):444.
9. Liu Q, Li L, Cheng H, Yao L, Wu J, Huang H, Ning W, Kai G. The basic helix-loop-helix transcription factor TabHLH1 increases chlorogenic acid and luteolin biosynthesis in *Taraxacum antungense* Kitag. *Hortic Res*. 2021;8(1):195.
10. Pagliuso D, Palacios Jara CE, Grandis A, Lam E, Pena Ferreira MJ, Buckeridge MS. Flavonoids from duckweeds: potential applications in the human diet. *RSC Adv*. 2020;10(73):44981–8.
11. Liu W, Li J, Zhang X, Zu Y, Yang Y, Liu W, Xu Z, Gao H, Sun X, Jiang X, et al. Current advances in naturally occurring caffeoylquinic acids: structure, Bioactivity, and synthesis. *J Agric Food Chem*. 2020;68(39):10489–516.
12. Marchiosi R, dos Santos WD, Constantini RP, de Lima RB, Soares AR, Finger-Teixeira A, Mota TR, de Oliveira DM, Foletto-Felipe MP, Abrahão J, et al. Biosynthesis and metabolic actions of simple phenolic acids in plants. *Phytochem Rev*. 2020;19(4):865–906.
13. Nabavi SM, Šamec D, Tomczyk M, Milella L, Russo D, Habtemariam S, Sutar I, Rastrelli L, Daglia M, Xiao J, et al. Flavonoid biosynthetic pathways in plants: versatile targets for metabolic engineering. *Biotechnol Adv*. 2020;38:107316.

14. Dong NQ, Lin HX. Contribution of phenylpropanoid metabolism to plant development and plant-environment interactions. *J Integr Plant Biol*. 2021;63(1):180–209.
15. Shi M, Huang F, Deng C, Wang Y, Kai G. Bioactivities, biosynthesis and biotechnological production of phenolic acids in *Salvia miltiorrhiza*. *Crit Rev Food Sci Nutr*. 2019;59(6):953–64.
16. Naik J, Misra P, Trivedi PK, Pandey A. Molecular components associated with the regulation of flavonoid biosynthesis. *Plant Science: Int J Experimental Plant Biology*. 2022;317:111196.
17. Huang W, Sun W, Lv H, Luo M, Zeng S, Pattanaik S, Yuan L, Wang Y. A R2R3-MYB transcription factor from *Epimedium Sagittatum* regulates the flavonoid biosynthetic pathway. *PLoS ONE*. 2013;8(8):e70778.
18. Yang R, Wang S, Zou H, Li L, Li Y, Wang D, Xu H, Cao X. R2R3-MYB transcription factor SmMYB52 positively regulates biosynthesis of Salivianolic Acid B and inhibits Root Growth in *Salvia miltiorrhiza*. *Int J Mol Sci*. 2021;22(17).
19. Yang F, Wang T, Guo Q, Zou Q, Yu S. The CmMYB3 transcription factors isolated from the *Chrysanthemum morifolium* regulate flavonol biosynthesis in *Arabidopsis thaliana*. *Plant Cell Rep*. 2023;42(4):791–803.
20. Tang M, Xue W, Li X, Wang L, Wang M, Wang W, Yin X, Chen B, Qu X, Li J, et al. Mitotically heritable epigenetic modifications of CmMYB6 control anthocyanin biosynthesis in *chrysanthemum*. *New Phytol*. 2022;236(3):1075–88.
21. Wang Y, Zhou LJ, Wang Y, Geng Z, Liu S, Chen C, Chen S, Jiang J, Chen F. CmMYB9a activates floral coloration by positively regulating anthocyanin biosynthesis in *chrysanthemum*. *Plant Mol Biol*. 2022;108(1–2):51–63.
22. Zhou L, Liu S, Wang Y, Wang Y, Song A, Jiang J, Chen S, Guan Z, Chen F. CmMYB3-like negatively regulates anthocyanin biosynthesis and flower color formation during the post-flowering stage in *Chrysanthemum morifolium*. *Hortic Plant J*. 2024;10(1):194–204.
23. Won SY, Jung JA, Kim JS. Genome-wide analysis of the MADS-Box gene family in *Chrysanthemum*. *Comput Biol Chem*. 2021;90:107424.
24. Wang R, Liu K, Tang B, Su D, He X, Deng H, Wu M, Bouzayen M, Grierson D, Liu M. The MADS-box protein SITAGL1 regulates a ripening-associated SIDQD/SDH2 involved in flavonoid biosynthesis and resistance against *Botrytis Cinerea* in post-harvest tomato fruit. *Plant J*. 2023;115(6):1746–57.
25. Iqbal N, Khan NA, Ferrante A, Trivellini A, Francini A, Khan M. Ethylene role in plant growth, development and senescence: interaction with other phytohormones. *Front Plant Sci*. 2017;8:235913.
26. Feng K, Hou X-L, Xing G-M, Liu J-X, Duan A-Q, Xu Z-S, Li M-Y, Zhuang J, Xiong A-S. Advances in AP2/ERF super-family transcription factors in plant. *Crit Rev Biotechnol*. 2020;40(6):750–76.
27. Zhao C, Liu X, Gong Q, Cao J, Shen W, Yin X, Grierson D, Zhang B, Xu C, Li X, et al. Three AP2/ERF family members modulate flavonoid synthesis by regulating type IV chalcone isomerase in citrus. *Plant Biotechnol J*. 2021;19(4):671–88.
28. Sun M, Shi M, Wang Y, Huang Q, Yuan T, Wang Q, Wang C, Zhou W, Kai G. The biosynthesis of phenolic acids is positively regulated by the JA-responsive transcription factor ERF115 in *Salvia miltiorrhiza*. *J Exp Bot*. 2019;70(1):243–54.
29. Lu C, Pu Y, Liu Y, Li Y, Qu J, Huang H, Dai S. Comparative transcriptomics and weighted gene co-expression correlation network analysis (WGCNA) reveal potential regulation mechanism of carotenoid accumulation in *Chrysanthemum × morifolium*. *Plant physiology and biochemistry: PPB*. 2019;142:415–428.
30. Zhang DY, Yao XH, Duan MH, Wei FY, Wu GH, Li L. Variation of essential oil content and antioxidant activity of *Lonicera* species in different sites of China. *Ind Crops Prod*. 2015;77:772–9.
31. Gharibi S, Tabatabaei BES, Saeidi G, Goli SAH, Talebi M. Total phenolic content and antioxidant activity of three Iranian endemic *Achillea* species. *Ind Crops Prod*. 2013;50:154–8.
32. Lin L-Z, Harnly JM. Identification of the phenolic components of *chrysanthemum* flower (*Chrysanthemum morifolium* Ramat). *Food Chem*. 2010;120(1):319–26.
33. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29(7):644–52.
34. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25.
35. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28(5):511–5.
36. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12(1):59–60.
37. Love MI, Huber W, Anders S. Moderated estimation of Fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
38. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
39. MacQueen J. Some methods for classification and analysis of multivariate observations. Berkeley, University of California Press. 1967;1:281–297.
40. Schmitgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc*. 2008;3(6):1101–8.
41. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
42. Chen L, Liu Y, Huang X, Zhu Y, Li J, Miao Y, Du H, Liu D. Comparison of Chemical constituents and Pharmacological effects of different varieties of *Chrysanthemum Flos* in China. *Chem Biodivers*. 2021;18(8):e2100206.
43. Huang Y, Xu L, Zhang F, Liu Y, Wang Y, Meng F, Li S, Cheng X, Bi Y. Preparation and pharmacokinetics in vivo of linarin solid dispersion and liposome. *Chin Herb Med*. 2022;14(2):310–6.
44. Chon SU. Total polyphenols and bioactivity of seeds and sprouts in several legumes. *Curr Pharm Design*. 2013;19(34):6112–24.
45. Gong J, Chu B, Gong L, Fang Z, Zhang X, Qiu S, Wang J, Xiang Y, Xiao G, Yuan H et al. Comparison of Phenolic Compounds and the Antioxidant Activities of Fifteen *Chrysanthemum morifolium* Ramat cv. 'Hangbaiju' in China. *Antioxidants* (Basel, Switzerland). 2019;8(8).
46. Rodrigues JA, Espley RV, Allan AC. Genomic analysis uncovers functional variation in the C-terminus of anthocyanin-activating MYB transcription factors. *Hortic Res*. 2021;8(1):77.
47. Liu J, Osbourn A, Ma P. MYB Transcription Factors as regulators of Phenylpropanoid metabolism in plants. *Mol Plant*. 2015;8(5):689–708.
48. Zhao L, Gao L, Wang H, Chen X, Wang Y, Yang H, Wei C, Wan X, Xia T. The R2R3-MYB, bHLH, WD40, and related transcription factors in flavonoid biosynthesis. *Funct Integr Genom*. 2013;13(1):75–98.
49. Schilling S, Pan S, Kennedy A, Melzer R. MADS-box genes and crop domestication: the jack of all traits. *J Exp Bot*. 2018;69(7):1447–69.
50. Schmitzer V, Osterc G, Veberic R, Stampar F. Correlation between chromaticity values and major anthocyanins in seven *Acer palmatum* Thunb. Cultivars. *Sci Hort*. 2009;119(4):442–6.
51. Chen X, Baldermann S, Cao S, Lu Y, Liu C, Hirata H, Watanabe N. Developmental patterns of emission of scent compounds and related gene expression in roses of the cultivar *Rosa × Hybrida* Cv. 'Yves Piaget'. *Plant Physiol Biochem*. 2015;87:109–14.
52. Zhao T, Li R, Yao W, Wang Y, Zhang C, Li Y. Genome-wide identification and characterisation of phenylalanine ammonia-lyase gene family in grapevine. *J Hortic Sci Biotechnol*. 2021;96(4):456–68.
53. Chen Y, Yi N, Yao Sb, Zhuang J, Fu Z, Ma J, Yin S, Jiang X, Liu Y, Gao L, et al. CsHCT-Mediated lignin synthesis pathway involved in the response of tea plants to biotic and Abiotic Stresses. *J Agric Food Chem*. 2021;69(35):10069–81.
54. Li H, Li D, Yang Z, Zeng Q, Luo Y, He N. Flavones produced by Mulberry Flavone synthase type I constitute a Defense line against the Ultraviolet-B stress. *Plants* (Basel). 2020;9(2).
55. Wang H, Zheng X, Wu Y, Zhan W, Guo Y, Chen M, Bai T, Jiao J, Song C, Song S, et al. Transcriptome Analysis Identifies Genes Associated with Chlorogenic Acid Biosynthesis during Apple Fruit Development. *Horticultrae*. 2023;9(2):217.
56. Pratyusha DS, Sarada DVL. MYB transcription factors-master regulators of phenylpropanoid biosynthesis and diverse developmental and stress responses. *Plant Cell Rep*. 2022;41(12):2245–60.
57. Xu H, Dai X, Hu X, Yu H, Wang Y, Zheng B, Xu J, Wu X. Phylogenetic analysis of R2R3-MYB family genes in *Tetragium Hemsleyanum* Diels et Gilg and roles of ThMYB4 and ThMYB7 in Flavonoid Biosynthesis. *Biomolecules*. 2023;13(3).
58. Hu Z, Zhong X, Zhang H, Luo X, Wang Y, Wang Y, Liu T, Zhang Y, Wang X, An H, et al. GhMYB18 confers *Aphis gossypii* Glover resistance through regulating the synthesis of salicylic acid and flavonoids in cotton plants. *Plant Cell Rep*. 2023;42(2):355–69.
59. Gao Q, Song W, Li X, Xiang C, Chen G, Xiang G, Liu X, Zhang G, Li X, Yang S, et al. Genome-wide identification of bHLH transcription factors: Discovery of a candidate regulator related to flavonoid biosynthesis in *Erigeron Breviscapus*. *Front Plant Sci*. 2022;13:977649.

60. Stracke R, Ishihara H, Huelpe 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):660–77.
61. Zhao L, Song Z, Wang B, Gao Y, Shi J, Sui X, Chen X, Zhang Y, Li Y. R2R3-MYB transcription factor NtMYB330 regulates Proanthocyanidin biosynthesis and seed germination in Tobacco (*Nicotiana tabacum* L.). *Front Plant Sci.* 2021;12:819247.
62. Jiang Q, Wang Z, Hu G, Yao X. Genome-wide identification and characterization of AP2/ERF gene superfamily during flower development in *Actinidia Eriantha*. *BMC Genomics.* 2022;23(1):650.
63. Schilling S, Kennedy A, Pan S, Jermini LS, Melzer R. Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization. *New Phytol.* 2020;225(1):511–29.
64. Zhang Z, Shi Y, Ma Y, Yang X, Yin X, Zhang Y, Xiao Y, Liu W, Li Y, Li S, et al. The strawberry transcription factor FaRAV1 positively regulates anthocyanin accumulation by activation of FaMYB10 and anthocyanin pathway genes. *Plant Biotechnol J.* 2020;18(11):2267–79.
65. Zeng JK, Li X, Xu Q, Chen JY, Yin XR, Ferguson IB, Chen KS. EjAP2-1, an AP2/ERF gene, is a novel regulator of fruit lignification induced by chilling injury, via interaction with EjMYB transcription factors. *Plant Biotechnol J.* 2015;13(9):1325–34.
66. Hsu HF, Chen WH, Shen YH, Hsu WH, Mao WT, Yang CH. Multifunctional evolution of B and AGL6 MADS box genes in orchids. *Nat Commun.* 2021;12(1):902.
67. Wu R, Wang T, McGie T, Voogd C, Allan AC, Hellens RP, Varkonyi-Gasic E. Overexpression of the kiwifruit SVP3 gene affects reproductive development and suppresses anthocyanin biosynthesis in petals, but has no effect on vegetative growth, dormancy, or flowering time. *J Exp Bot.* 2014;65(17):4985–95.

Publisher's Note

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