Folia Hort. 34(1) (2022): 65–83

DOI: 10.2478/fhort-2022-0007

FOLIA HORTICULTURAE

Published by the Polish Society for Horticultural Science since 1989

ORIGINAL ARTICLE | Open access | http://www.foliahort.ogr.ur.krakow.pl

Comparative RNA-Seq analysis to understand anthocyanin biosynthesis and regulations in *Curcuma alismatifolia*

Yuan-Yuan Li[1], Xiao-Huang Chen[2], Hui-Wen Yu[1], Qi-Lin Tian[1], Luan-Mei Lu[1,]*

[1] Provincial Key Laboratory of Landscape Plants with Fujian and Taiwan Characteristics, College of the Bioscience and Technology, Minnan Normal University, Zhangzhou 363000, China

[2] College of the Chemistry, Chemical Engineering and Environment, Minnan Normal University, Zhangzhou 363000, China

ABSTRACT

Anthocyanins play a critical role in flower colour pattern formation, and their biosynthesis is typically regulated by transcription factors (TFs). *Curcuma alismatifolia* is a well-known ornamental plant with colourful flowers. However, little is known about the genes that regulate anthocyanin accumulation in *C. alismatifolia*. In the present study, high-quality RNA was extracted from three flowering stages of 'Dutch Red' and the blossoming stage of 'Chocolate'. In all, 576.45 Mb clean data and 159,687 de-redundant sequences were captured. The Kyoto Encyclopedia of Genes and Genomes analysis showed that the pathways of phenylpropanoid biosynthesis, flavonoid biosynthesis, flavone and flavonol biosynthesis, and terpenoid backbone biosynthesis were the most enriched. Thirty unique isoforms were annotated as encoding enzymes or TFs involved in anthocyanin biosynthesis. Further analysis showed that the up-regulation of anthocyanin biosynthesis genes was associated with the red colour formation of 'Dutch Red', and their expression was induced at the initial flowering stage. The gene *flavonoid 3',5'-hydroxylase*, a key enzyme in the formation of delphinidin-based anthocyanins, reduced expression in 'Chocolate'. In addition, we identified totally 14 TFs including 11 MYB proteins and 3 WD proteins, which might play important roles in the regulation of anthocyanin biosynthesis. The quantitative Real-Time Polymerase Chain Reaction (qRT-PCR) results were generally consistent with the high-throughput sequencing results. Together, the results of our study provide a valuable resource for the regulatory mechanism of anthocyanin biosynthesis in *C. alismatifolia* and for the breeding of *Curcuma* cultivars with novel and charming flower colours.

Keywords: anthocyanin biosynthesis, *C. alismatifolia*, transcriptome analysis

INTRODUCTION

Curcuma alismatifolia is a beautiful exotic plant originally from Thailand and Cambodia (Khumkratok et al., 2012). It belongs to the ginger family, and has recently become popular as an ornamental around the word. Nowadays, a large number of cultivars, including 'Qiang Mai Pink', 'Qiang Mai Red', 'Dutch Red' and 'UB Snow 701', have been introduced in China. They are produced as commercial cut flowers or as potted plants in urban gardens (Li et al., 2021a).

In general, the inflorescence is the main ornamental part of *C. alismatifolia.* It consists of colourful and green bracts on the showy flowering stem. The pink to purple red/white distal bracts are more numerous than the basal green bracts. Several small flower buds with purple flag petals are borne on the green bracts (Figure 1A). Anthocyanins are widely known as nutraceuticals and are a group of soluble vacuolar pigments among leaves, flowers, fruits and roots (Mizuta et al., 2009). These kinds of

*Corresponding author
e-mail: 358070295@qq.com (Luan-Mei Lu).

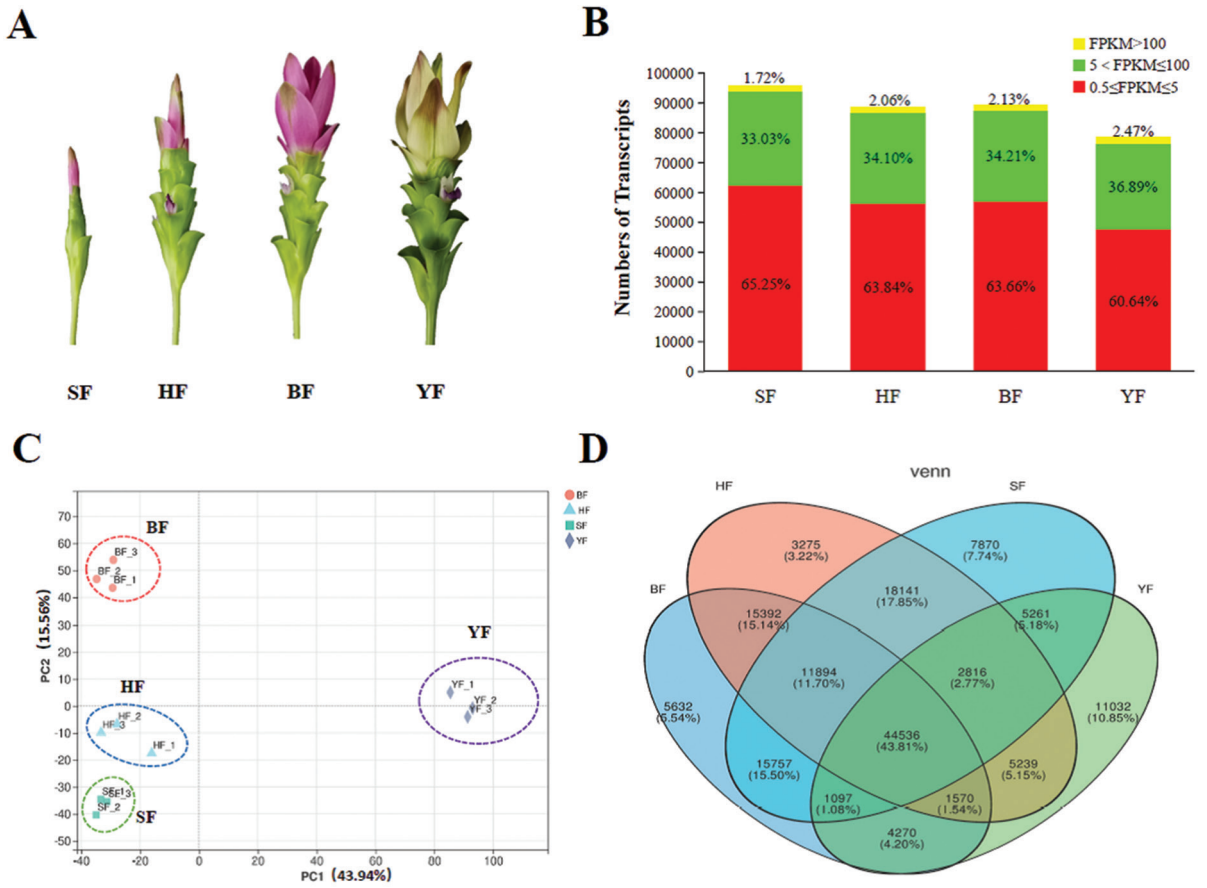

Figure 1. RNA-Seq data expression profiles in *C. alismatifolia* colour bracts. (A) Phenotypes of 'Dutch Red' at three typical development stages and 'Chocolate' at the blossomed stage. (B) Numbers of transcripts in the four bract samples. (C) Principal component analysis of the RNA-Seq data. (D) Venn diagram of the RNA-Seq data from the four bract samples. BF, blossomed flowering; HF, half-flowering; FPKM, fragments per kilobase of transcript per million mapped reads.

pigments exhibit a wide range of colours, from pink to blue-purple, and primarily consist of six common types of anthocyanidins, namely pelargonidin, cyanidin, delphinidin, peonidin, petunidin and malvidin. Both internal factors and exterior environments, such as different varieties and growth periods, can influence the distribution of anthocyanins (Zhao and Tao, 2015). The biosynthesis of anthocyanin is catalysed by a series of enzymes, comprising chalcone synthase (CHS), chalcone isomerase (CHI), dihydroflavonol 4-reductase (DFR) and anthocyanidin synthase (ANS) in the flavonoid biosynthetic pathways, and has been reported in *Arabidopsis* (Zhang et al., 2017), *Begonia semperflorens* (Wang et al., 2018) and *Narcissus pseudonarcissus* (Li et al., 2018). By promoting the activity of key enzymes, anthocyanin synthesis can be induced that leads to changed flower colours. Studies have also found that three main types of transcription factors (TFs), MYB, basic helix-loop-helix (bHLH) and WD40, can form an MBW complex to regulate the expressions of structural genes (Wei et al., 2019). In *C. alismatifolia,* a total of five anthocyanins (delphinidin 3-O-rutinoside, cyanidin 3-O-rutinoside, petunidin 3-O-rutinoside, malvidin 3-O-glucoside and malvidin 3-O-rutinoside) were identified as being responsible for the pink to red bracts (Koshioka et al., 2015). The genes, *CHS* and *DFR,* were successfully cloned from *C. alismatifolia* ('Qing Mai Pink') by Chanapan et al. (2017) and Petchang et al. (2017). However, the underlying molecular mechanisms that control anthocyanin synthesis in flower colour formation and in cultivars of *C. alismatifolia* with different colours are far from conclusive.

With the development of sequencing technology in recent years, high-throughput sequencing of mRNA is being used to provide specific expression information in particular species (He et al., 2020). This technology is additionally one of the effective approaches to detect functional genes in organisms without reference genome. Recently, assembly of *de-novo* transcriptomes was first performed in *C. alismatifolia,* identifying novel EST-SSR markers for *C.* species (Taheri et al., 2019). However, studies of the colour formation in the *C. alismatifolia* transcriptome are rare. In this study, we performed RNA-Seq analysis on the regulatory

networks of anthocyanin biosynthesis in 'flowers' (colour bracts) of the *C. alismatifolia* cultivars Dutch Red and Chocolate. The expression profiles in 'Dutch Red' at the different flower developmental stages were also investigated. We used this information to identify hub genes related to anthocyanin pathways and to analyse the differently expressed genes (DEGs) involved in anthocyanin biosynthesis and regulation. The goal was to acquire comprehensive knowledge of the anthocyanin regulatory network during *C. alismatifolia* flower formation and to provide a theoretical basis for the future breeding of *C. alismatifolia* cultivars with high ornamental and commercial values.

MATERIALS AND METHODS

Plant materials and growth

Rhizomes from *C. alismatifolia* cv. 'Dutch Red' and 'Chocolate' were obtained from a horticultural company (Jinluan) in China. Rhizomes were grown in a greenhouse within the campus of Minnan Normal University. During the experiment, the plants were grown at 25 ºC under 500 $\mu mol \cdot m^{-2} \cdot s^{-1}$. Colourful bracts from 'Dutch Red' were harvested at three continuous developmental stages [early flowering small flowers (SF), half-flowering (HF) and blossomed flowering (BF)]. The yellow bracts from 'Chocolate' were harvested at the peak flowering stage yellow flowers (YF). For each sample, three biological replicates (3 g each) were collected. These tissue samples were immediately subjected to RNA extraction.

RNA extraction, library construction and Illumina sequencing

High-quality total RNA from the tissue samples was isolated using the RNeasy mini Kit (Qiagen, Germany), following the manufacturer's instructions. gDNAase (provided by the Kit) was added to remove the genomic DNA. The purity and concentration of isolated RNA were checked by a 2100 Bioanalyser (Agilent Technologies, Inc., Santa Clara, CA, USA) and quantified using an ND-2000 (NanoDrop Thermo Scientific, Wilmington, DE, USA). Only high-quality RNA samples (OD260/280 = 1.8–2.2, OD260/230 ≥ 2.0, RIN ≥ 8.0, 28S:18S ≥ 1.0, >2 μg) were used to construct the sequencing library.

RNA purification, reverse transcription, library construction and sequencing were performed at Shanghai Majorbio Bio-pharm Biotechnology Co., Ltd. (Shanghai, China) according to the manufacturer's instructions (Illumina, San Diego, CA, USA). The RNA-Seq transcriptome libraries were prepared using an Illumina TruSeqTM RNA sample preparation Kit (San Diego, CA, USA). Poly(A) mRNA was purified from total RNA using oligo-dT-attached magnetic beads and then fragmented by fragmentation buffer. Taking these short fragments as templates, double-stranded cDNA was synthesised using a SuperScript double-stranded cDNA synthesis kit (Invitrogen, CA, USA) with random hexamer primers (Illumina). Then the synthesised cDNA was subjected to end-repair, phosphorylation and 'A' base addition according to Illumina's library construction protocol. Libraries were size selected for cDNA target fragments of 200–300 bp on 2% Low Range Ultra Agarose followed by PCR amplified using Phusion DNA polymerase (New England Biolabs, Boston, MA, USA) for 15 PCR cycles. After quantifying by TBS380, two RNA-Seq libraries were sequenced in single lane on an Illumina Hiseq NovaSeq 6000 sequencer (Illumina, San Diego, CA, USA) for 2 × 150 bp paired-end reads.

De novo assembly and functional annotation

The raw reads obtained from HiSeq2000 sequencing were filtered by fastp software (v0.19.6): reads containing adaptors, reads with more than 5% unknown nucleotides and low-quality reads with more than 20% of bases with a quality value ≤20 were deleted (Wang, 2007). Then all clean reads were assembled with Trinity (release 2013-02-25) with the following settings: inchworm: -K –L 25; Chrysalis: -min_glue 2 -glue_factor 0.05 -min_iso_ratio 0.05 -kk 48 –strand -report_welds -max_mem_reads 1,000,000 (Grabherr et al., 2011). We then filtered out any transcripts with less than 1% of the per-component expression level using a script bundled with Trinity. ORFs were extracted using the Perl script bundled with Trinity (transcripts_to_best_scoring_ORFs.pl). To estimate the number of full length transcripts that had been assembled in our data sets we used the tool benchmarking universal single-copy orthologs (BUSCO) (v5.2.2), which is based on evolutionarily informed expectations of gene content, with default settings. HMMER is then run to assign a score to the candidate amino acid sequence (Seppey et al., 2019). Finally, Trinity connects the transcripts and obtains sequences that cannot be extended on either end. Such sequences are defined as unigenes. When multiple samples from the same species are sequenced, the unigenes from each sample's assembly can be further analysed through sequence splicing and the removal of redundancy using sequence clustering software to acquire non-redundant unigenes with the longest length possible. In the final step, transcripts and unigenes were separately searched against the National Center for Biotechnology Information (NCBI) protein non-redundant (NR), the Clusters of Orthologous Groups of Proteins (COG) and Swiss-Prot databases using DIAMOND (v0.8.37.99) to perform function annotations with a significant threshold E-value of 10^{-5}. BLAST2GO (v2.5.0) (http://www.blast2go.com/b2ghome) program was used to get GO annotations of unigenes for describing biological processes, molecular functions and cellular components (Conesa et al., 2005). Metabolic pathway analysis was performed using KOBAS (v2.1.1) based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Ogata et al., 1999).

Differential expression analysis and functional enrichment

To identify DEGs between two different samples, the gene expression level was calculated using fragments per kilobase of exon per million mapped reads (FPKM). RNA-Seq by expectation-maximization (RSEM) (http://deweylab.biostat.wisc.edu/rsem/) (Dewey and Li, 2011) was used to quantify gene and isoform abundances. R statistical package software EdgeR (Empirical analysis of Digital Gene Expression in R, http://www.bioconductor.org/packages/2.12/bioc/html/edgeR.html) (Robinson et al., 2010) was utilised for DEGs between two samples (Anders et al., 2013). The threshold for the *p*-value was determined by the false-discovery rate (FDR). Unigenes with FDR≤0.001 and ratio of FPKMs of the two samples larger than 2 (genes for which FPKM <1 were filtered) were considered to be DEGs in this study. In addition, functional-enrichment analysis including GO and KEGG were performed to identify those DEGs that were significantly enriched in GO terms and metabolic pathways at Bonferroni-corrected *p*-value ≤0.05 compared with the whole-transcriptome background. GO functional enrichment and KEGG pathway analysis were carried out by Goatools (https://github.com/tanghaibao/Goatools) and KOBAS (http://kobas.cbi.pku.edu.cn/home.do) (Xie et al., 2011).

qRT-PCR analysis

Primers used for qRT-PCR assays were designed with the online Primer3 (v0.4.0) software and they amplified PCR products that varied from 80 bp to 200 bp. The housekeeping gene *ACTIN* (Li et al., 2021b) was used as internal reference control for normalisation. qRT-PCR assays were performed as described previously (Li et al., 2019). The relative expression of the genes was calculated using the $2^{-\Delta Ct}$ method.

RESULTS

Transcriptome analysis in colourful bracts of 'Dutch Red' and 'Chocolate'

To obtain a more detailed understanding of the anthocyanin regulatory network within different varieties and different flowering developments, four samples, derived from the red bracts from the SF, HF and BF stages of 'Dutch Red', and the yellow bracts from BF stage of 'Chocolate' (YF), were used as materials for RNA sequencing (Figure 1A). Each sample was performed with three biological replicates; thus a total of 12 sequenced libraries were constructed.

In total, approximately 576.45 Mb of raw read data were generated from Illumina HiSeq sequencing ranging from 45.24 Mb to 52.98 Mb in each sample (Table S1 in Supplementary Materials). After raw read filtering, a total of 566,269,472 clean reads (the number of total clean nucleotides was 83,999,856,857 nt; the Q20 and GC percentages were 98.31% and 49.25%, respectively) were obtained from the 12th libraries (SF-1, SF-2, SF-3; HF-1, HF-2, HF-3; BF-1, BF-2, BF-3; YF-1, YF-2, YF-3). Then, clean reads were assembled into 159,687 transcripts and 69,453 unigenes, reaching an averagely mapped ratio of 73% in samples. The assembled transcript length ranged from 500 bp to over 4,500 bp with an average of 1,236 bp and N50 of 1,830 bp (Table 1). Other statistical results obtained from the *C. alismatifolia* transcriptome sequencing and assemblies are also listed in Table 1. The accession numbers of all raw data in the Short Read Archive (SRA) Sequence Database in the NCBI are listed in Table S1 in Supplementary Materials.

These transcripts offer a potential source for the identification of functional genes. Subsequently, reads with an FPKM value <0.5 were removed, and 95,625 (mean from three replicates, the same below), 88,479, 89,429 and 78,537 transcripts were found in SF, HF, BF and YF, respectively. On average, 63% of the transcripts were in the 0.5–5 FPKM range, and 35% of the transcripts were in the 5–100 FPKM range (Figure 1B; Table S2 in Supplementary Materials). A principal component analysis showed that there were highly correlated transcriptome characteristics between the biological replicates of each sample (Figure 1C). A total of 44,536 transcripts were shared between samples (Figure 1D).

The transcript and unigene sequences were first aligned against the NR and Swiss-Prot databases using BlastN and BlastX searches with an E-value less than10^{-5}. This analysis indicated that 104,303 transcripts (65.32% of the total transcripts) and 34,189 unigenes (49.23% of the total unigenes) were matched to the NR database (Table S2 in Supplementary Materials), while 82,193 transcripts (51.47%) and 26,589 unigenes (38.28%) were matched to the Swiss-Prot database (Table S3 in Supplementary Materials). The annotated transcripts and unigenes in the COG, GO and KEGG databases are also listed in Table S3 in Supplementary Materials. However, 53,694 transcripts (33.62%) and 34,270 unigenes (49.34%) could not be matched to any functions included in these five databases.

Comparisons of differently expressed genes between different development stages and varieties

Comparison of gene expression levels in varieties and different development stages revealed a total of 15,449 DEGs (BF vs. YF, 11,000 DEGs; BF vs. HF, 4,761 DEGs; BF vs. SF, 7,742 DEGs; HF vs. SF, 1,535 DEGs), and only 1% (215 DEGs) of the DEGs between varieties and different development stages are the same (Figure 2A). Interestingly, the number of down-regulated DEGs was greater than the number of up-regulated DEGs in red bracts of blooming flowers of 'Dutch Red', as indicated from the comparisons of BF vs. HF (1,643 DEGs up-regulated; 3,118 DEGs down-regulated) and BF vs. SF (3,591 DEGs up-regulated; 4,151 DEGs down-regulated); however, more genes were significantly up-regulated in the bracts of 'half flowering' of 'Dutch

Table 1. Summary of assembly results for 'Dutch Red' at three development stages and 'Chocolate' at the blossomed stage.

Features	SF	HF	BF	YF
Total raw reads (Mb)	45.24	48.46	52.98	45.47
Total clean reads (Mb)	44.47	47.55	52.10	44.63
Total clean bases (Gb)	6.59	7.05	7.75	6.61
Clean reads Q20 (%)	98.27	98.23	98.34	98.40
Clean reads Q30 (%)	94.63	94.56	94.83	95.01
Clean reads (pair reads)	22.24	23.78	26.05	22.31
Mapped reads	16,378,830	17,511,187	19,200,458	16,316,551
Mapped ratio (%)	73.66	73.66	73.69	73.13
Total number of transcripts	159,687			
Total number of unigenes	69,453			
Total sequence base	197,307,177			
Average length of transcripts	1,236			
N50 value of transcripts	1,830			
E90N50 value of transcripts	1,921			
GC (%)	42.49			
TransRate score	0.30			
BUSCO score	60.1% (3.5%)			

Q20: The rate of bases whose quality is greater than 20; Q30: The rate of bases whose quality is greater than 30; N50: A weighted median statistic in which 50% of the total length is contained in unigenes greater than or equal to this value; E90N50: A weighted median statistic in which 50% of the total length is contained in the top 90% unigenes greater than or equal to this value; GC (%): The percentage of G and C bases in all unigenes; TransRate: Score the assembly results with transrate; BUSCO: Score the assembly integrity with BUSCO.
BF, blossomed flowering; HF, half-flowering.

Red', as indicated from the comparisons of HF vs. SF (1,198 DEGs up-regulated; 337 DEGs down-regulated) (Figure 2A).

KEGG analysis of the identified DEGs showed most enrichment in the following pathways: 'Phenylpropanoid biosynthesis' (map00940), 'Plant-pathogen interaction' (map04626), 'Plant hormone signal transduction' (map04075), 'Flavonoid biosynthesis' (map00941), 'Starch and sucrose metabolism' (map00500), 'Flavone and flavonol biosynthesis' (map00944) and 'Terpenoid backbone biosynthesis' (map00900) (Figure 2B). We noted that among the DEGs associated with flavonoid, those involved in anthocyanin biosynthesis, and transcripts of *shikimate O-hydroxycinnamoyltransferase* (HCT, EC2.3.1.133), the enzyme which is of great importance in plant secondary metabolite production, are more enriched in the initial blossoming stage when compared with the blossomed stage of 'Dutch Red'. Those proteins, involved in stress resistance and antioxidant defence, were more enriched in the half and fully blossomed stages when compared with the initial blossoming stage. In addition, most of the transcripts that participated in photosynthesis were more enriched in the yellow bracts of 'Chocolate' when compared with those in 'Dutch Red'.

Heat maps of the clustering analysis results also suggested that genes related to the 'Flavonoid biosynthesis', 'Flavone and flavonol biosynthesis' and 'Phenylpropanoid biosynthesis' pathways were significantly up-regulated in the early stage bracts of 'Dutch Red' when compared with those in the half and fully blossomed stages (Figure 3A). Analysis of DEGs related to 'Signal transduction' revealed that auxin-responsive proteins were screened as the most variable gene family, and auxin-responsive small auxin up RNA (SAUR) genes were important in the early development stage of red bracts. Meanwhile, IAA transcripts were mainly produced in both the half and fully blossomed stages when compared with the initial blossoming stage. Besides, the transcripts of *3-ketoacyl-CoA synthase* (XP_009405939.1), a key enzyme in fatty acid biosynthesis, were observed exclusively enriched in an early development stage compared with the half and fully blossomed stages (Figure 3A). From the clustering analysis of blossoming stages between 'Chocolate' and 'Dutch Red', DEGs related to 'flavonoid biosynthesis' and 'Photosynthesis' were more up-regulated in the yellow bracts of 'Chocolate' (Figure 3B). DEGs in the 'Porphyrin and chlorophyll metabolism' pathway were significantly enriched in 'Chocolate'. DEGs in 'Phenylproanoid biosynthesis' showed that peroxidase transcripts were significantly enriched in the red bracts of 'Dutch Red', which was suggestive of a stress response in the blossomed bracts of 'Dutch Red' (Figure 3B).

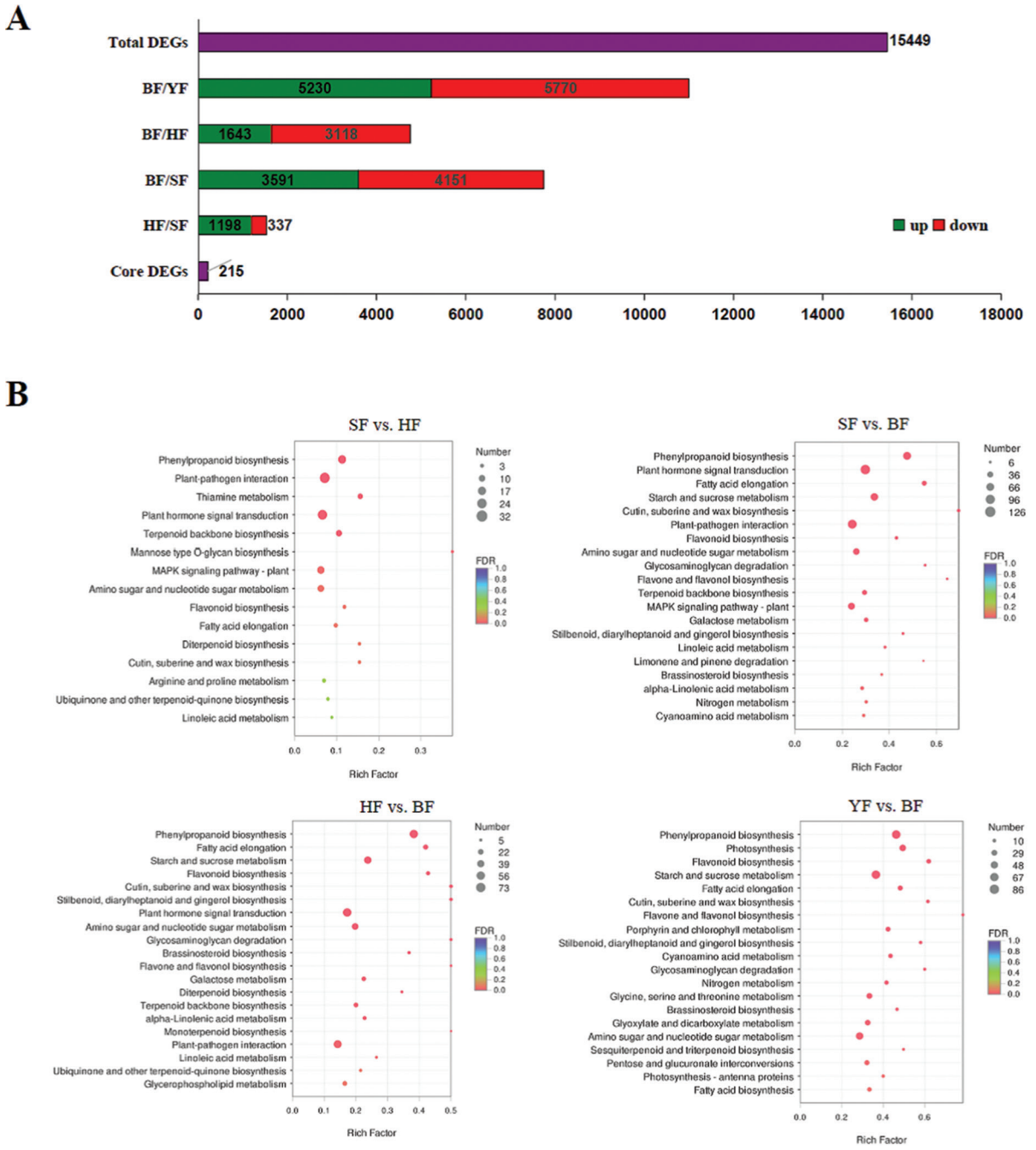

Figure 2. (A) Distributions and (B) KEGG analysis of DEGs identified from pairwise comparisons between different developmental stages and between varieties. The FDR value is the multiple hypothesis test-corrected *p*-value. The FDR value is in the range of [0–1]. The closer that number is to 0, the more significant the enrichment. The rich factor refers to the ratio of the number of genes among the DEGs located in a number of pathways to the total number of genes in the pathway entries in all of the annotated genes. The greater the rich factor, the greater the degree of enrichment. BF, blossomed flowering; DEGs, differentially expressed genes; FDR, false-discovery rate; HF, half-flowering; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Identification of candidate genes related to anthocyanin biosynthesis

By annotation in the public database, more than 30 genes affiliated to nine structural gene families were predicted to participate in the anthocyanin biosynthesis pathway (Figure 4). Anthocyanin biosynthesis started with the conversion of phenylalanine to coumarate-CoA by phenylalanine ammonia lyase (PAL), cinnamate-4-hydroxylase (C4H) and 4-coumarate: CoA ligase (4CL). There were seven unigenes annotated as *PAL* and 12 unigenes annotated as *4CL*, among which *DN13924_c2* and *DN18026_c0* had a significantly lower expression in the blossomed stage of the 'Dutch Red'. Subsequently, dihydroflavonol was formatted by coumarate-CoA

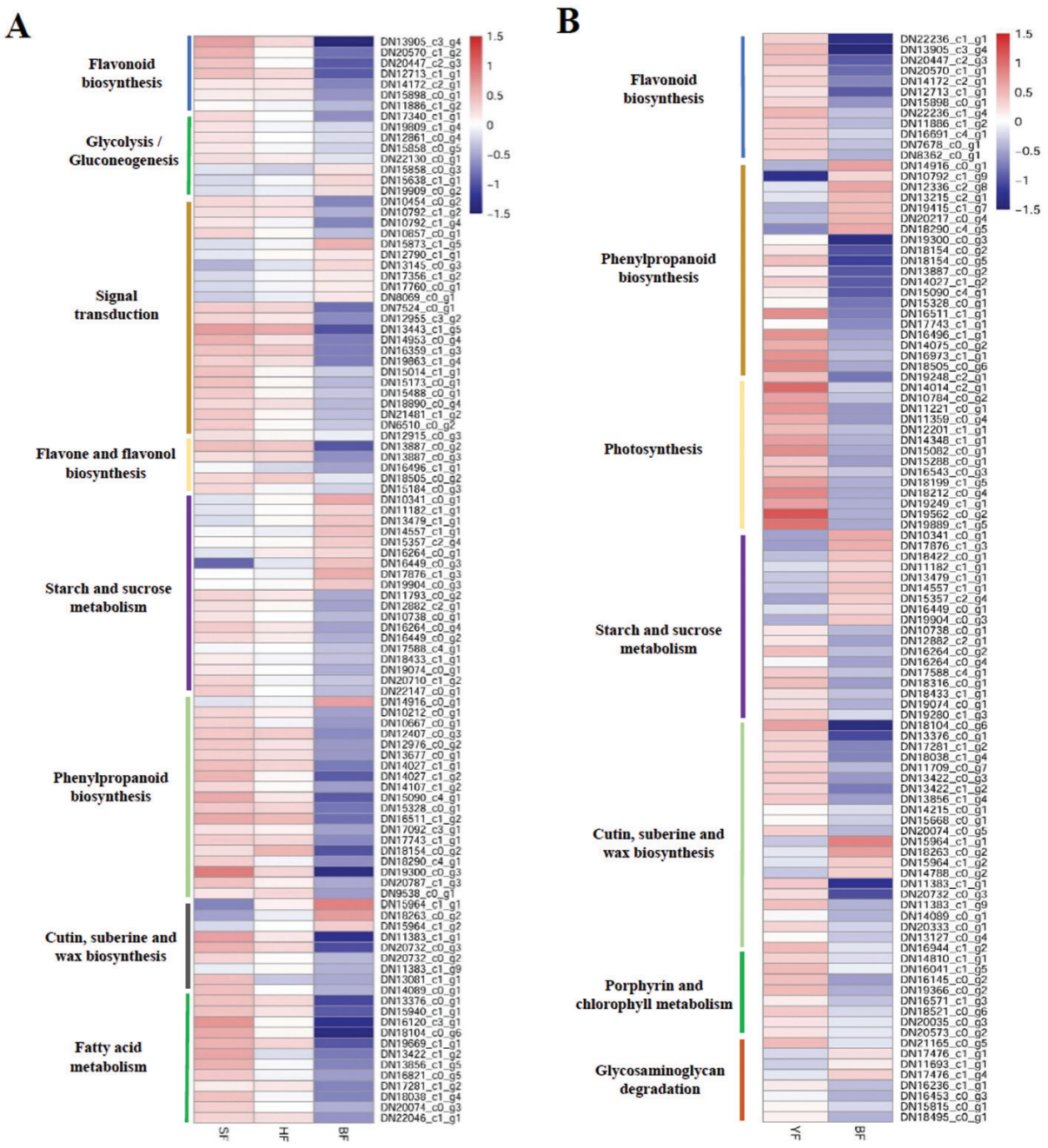

Figure 3. Comparisons of DEGs. (A) Heatmap comparison of DEGs in the most enriched KEGG pathways during the three development stages. (B) Heatmap comparison of DEGs in the most enriched KEGG pathways between 'Chocolate' and 'Dutch Red'. DEGs, differentially expressed genes; KEGG, Kyoto Encyclopedia of Genes and Genomes.

and malonyl-CoA catalysed by CHS, CHI, flavanone-3-hydroxylase (F3H) and flavonoid 3′,5′-hydroxylase (F3′5′H), which were the key enzymes in the metabolism of anthocyanin. The formation of various anthocyanidins by dihydroflavonols was then catalysed by DFR and ANS. Among them, *DN14172_c2*, *DN16691_c4* and *DN17923_c1* encode *F3H, DFR* and *CHS*, respectively (Table 2). The first two were the same genes as that submitted previously in the *C. alismatifolia* by Kriangphan et al. (2009), while *DN17923_c1* was closely related to *CHS* (AEU17693.1) gene in *C. longa* (Resmi and Soniya, 2012). They all showed significantly higher expression in the bracts of 'Chocolate' (FPKM as 1440.4, 570.8 and 2667.3, respectively) (Figure 4; Table 2). Another three anthocyanin structural genes, *DN10857_c0, DN16064_c1* and *DN12990_c1*, have 70.9%, 84.2% and 84.0%, respectively, nucleic acid similarity with *CHI* (XP_009384766.1), *F3′5′H* (XP_009411862.1) and *ANR* (XP_009412668.1) in *Musa acuminate*. *DN10857_c0* was expressed significantly lower in the blossomed stage of 'Dutch Red' (FPKM as 38.0) than those in the bracts of SF, HF and YF (FPKM as 501.8, 344.7 and 561.0, respectively). *DN16064_c1* was expressed significantly lower in the bracts of YF than those of 'Dutch Red'. A total of 14 genes

were identified in the clade with glucosyltransferase-like family with different patterns of expression (Figure 4). Among them, *DN12119_c3* and *DN13592_c1* encode the anthocyanidin/anthocyanin 3-O-glucosyltransferases, which catalyse unstable anthocyanidin into anthocyanin. They are the last key enzymes in the anthocyanin biosynthetic pathway. *DN12119_c3, DN16301_c1* and *DN13065_c1* had significantly lower expressions in BF than that in SF and HF, while other unigenes showed no significant divergence in different stages or cultivars (Table 2; Figure 4).

Identification of candidate regulators

In addition to the structural genes in the anthocyanin biosynthetic pathway, TFs also play important roles in flower colour development through regulating the temporal and spatial expression of structural genes. Many DEGs identified in the KEGG database were annotated as TFs, including *MYB, AR2/ERF, C2C2, bHLH* and *WD*. Their abundance levels and detailed information are shown in Figure 5A and Table S4 in Supplementary Materials. *MYB superfamily* and *WD-repeat* transcripts were the most abundant found to regulate the flower development and flower formation (Figure 5B). *DN17014_c0* was annotated as expansin-A15, which belongs to the MYB superfamily, and *DN13707_c0* was annotated as WD-repeat–containing protein. They were the two important TFs that expressed highly during the whole flower developments (Table 3). Further comparison showed that *DN17014_c0* shares sequence similarity with the TF *MYB105-like* gene, and the average value was about 95%. *DN11279_c3* and *DN19744_c1* annotated as MYB-related proteins are highly expressed during initial flower development in 'Dutch Red' (FPKM as 105.1 and 160.6, respectively), but during the blossomed stages they were significantly decreased. This suggests that *DN11279_c3* and *DN19744_c1* were more likely the floral activators of 'Dutch Red'. Another MYB protein,

Table 2. Putative structural genes in anthocyanin biosynthesis identified from DEGs.

Unigene	Annotation	FPKM-SF	FPKM-HF	FPKM-BF	Log_2Ratio (BF/SF)	FPKM-YF	Log_2Ratio (BF/YF)
DN17923_c1_g3	CHS-like protein	2,593.0	1,906.1	185.4	−3.80	2,667.3	−3.8
DN16695_c0_g5	CHS	80.4	760.5	298.4	1.9	2,187.2	−2.9
DN10857_c0_g4	Probable chalcone – flavonone isomerase 3 isoform X1 (CHI)	501.8	344.7	38.0	−3.7	561.0	−3.9
DN15898_c0_g1	Chalcone – flavonone isomerase (CHI)	152.8	139.5	20.0	−2.9	237.0	−3.6
DN14172_c2_g1	F3H	976.4	886.8	232.7	−2.1	1,440.4	−2.6
DN22055_c1_g3	Flavonol synthase/F3H	5.1	7.3	11.9	1.2	1.4	3.1
DN22236_c1_g1	Flavonol synthase/F3H	296.2	125.1	3.0	−6.6	257.7	−6.4
DN16064_c1_g1	Flavonoid 3′,5′-hydroxylase 1-like (F3′5′H)	1,422.9	1,389.6	1,138.9	−0.3	64.0	4.2
DN8362_c0_g1	Flavonoid 3′,5′-hydroxylase 1-like (F3′5′H)	6.6	14.1	2.3	−1.5	20.0	−3.1
DN14193_c0_g5	DFR	7.5	8.8	2.4	−1.6	18.5	−2.9
DN16691_c4_g1	DFR	67.8	259.7	117.8	−0.8	570.8	−2.3
DN12990_c1_g6	Anthocyanidin reductase (ANR)	64.7	331.3	90.8	0.5	323.2	−1.8
DN13257_c0_g1	Anthocyanidin reductase (ANR)	27.8	249.5	85.3	1.6	277.8	−1.7
DN12119_c3_g2	Anthocyanidin 3-O-glucosyltransferase 7-like flavonoid 3-O-glucosyltransferase (UFGT)	191.51	222.9	58.1	−1.7	250.3	0.38
DN13065_c1_g1	Anthocyanin 3′-O-beta--glucosyltransferase-like (UFGT)	3.74	2.6	0.24	−4.0	11.2	−5.54
DN13592_c1_g3	Anthocyanidin 3-O-glucosyltransferase-like (UFGT)	155.8	42.5	3.1	−5.67	196.8	−6.0

BF, blossomed flowering; CHI, chalcone isomerase; CHS, chalcone synthase; DEGs, differentially expressed genes; DFR, dihydroflavonol 4-reductase; F3H, flavanone-3-hydroxylase; F3′5′H, flavonoid 3′,5′-hydroxylase; FPKM, fragments per kilobase of transcript per million mapped reads; HF, half flowering.

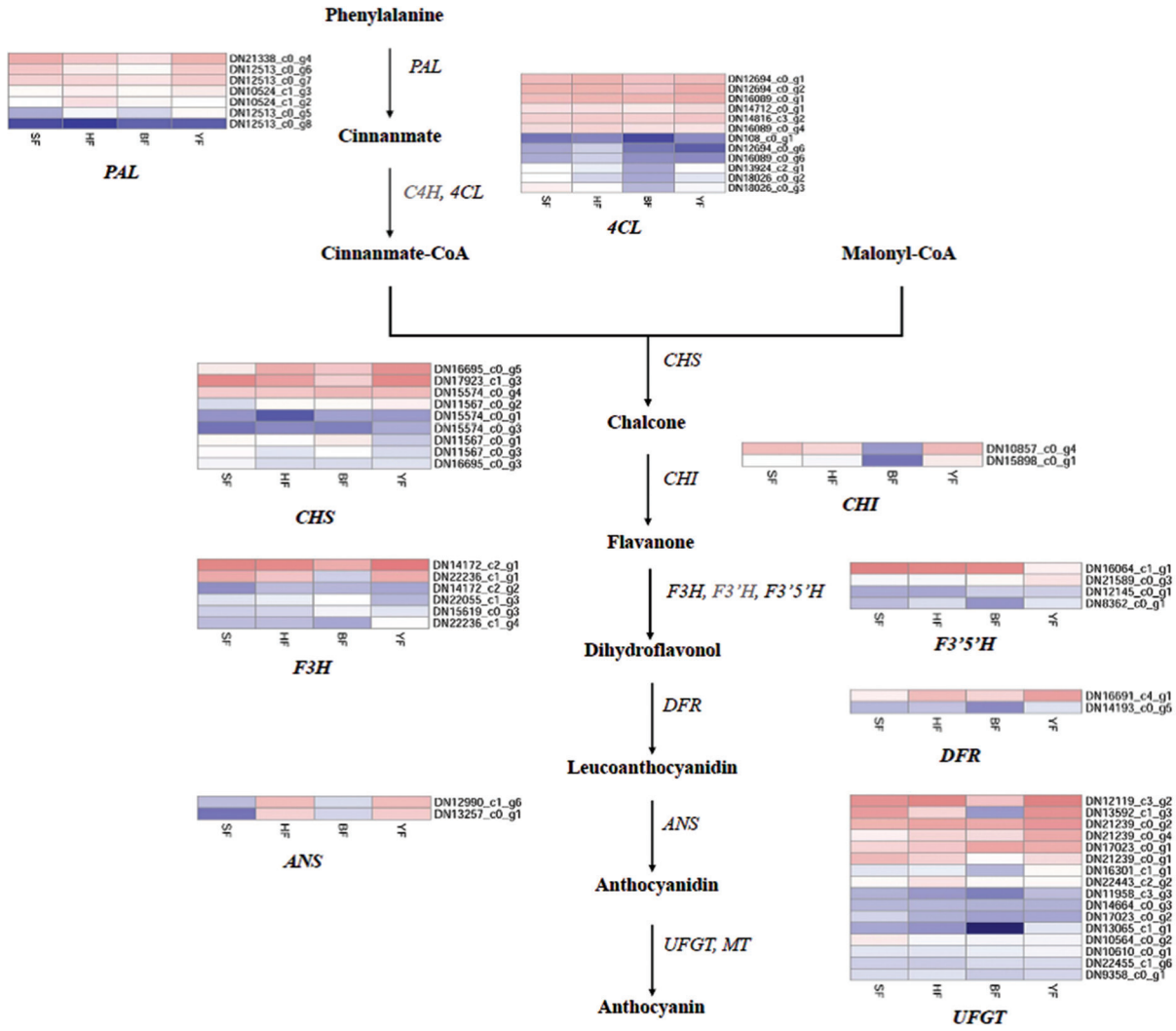

Figure 4. The anthocyanin biosynthesis process with its core metabolites and enzymes and the expression levels of core enzyme genes. Enzyme names and expression patterns are indicated at the side of each step. Colour boxes from left to right represent unigenes showing lower or higher expression level in the colour bracts of 'Dutch Red' of SF, HF, BF and YF, respectively. ANS, anthocyanidin synthase; BF, blossomed flowering; CHI, chalcone isomerase; CHS, chalcone synthase; C4H, cinnamate-4-hydroxylase; DFR, dihydroflavonol 4-reductase; F3H, flavanone-3-hydroxylase; F3′5′H, flavonoid 3′,5′-hydroxylase; HF, half-flowering; PAL, phenylalanine ammonia lyase.

DN18449_c0, annotated as trihelix TF GT-3a, was exclusively expressed highly in the blossomed bracts of 'Dutch Red' (FPKM as 116.7), but was significantly lower in the blossom bracts of 'Chocolate' (Table 3). A comparison of the MYB TFs between 'Dutch Red' and 'Chocolate' showed that *DN21970_c4*, annotated as divaricata-like MYB gene, had significantly high expression in the blossomed bracts of both 'Dutch Red' and 'Chocolate' (Table 3), indicating that *DN21970_c4* was possibly the common TF in regulating later floral development in *C. alismatifolia*.

qRT-PCR analysis validated the differently expressed gene DEGs of the anthocyanin pathway

To evaluate the reliability of RNA-Seq analysis using Illumina sequencing, 14 candidate genes in the anthocyanin biosynthesis and regulations were selected for qRT-PCR test in the four bracts (SF, HF, BF and YF). The oligonucleotide primer sequences used to amplify these transcripts for each gene and related unigene are listed in Table S5 in Supplementary Materials. In general, our qRT-PCR results show a high degree of consistency with the RNA-Seq results (Figure 6). Expression patterns of six (42.85%) genes (*F3H1, F3H2, MYB4, WD2, MYB5* and *DFR*) had almost similar expression patterns but with very small partial inconsistencies compared with the RNA-Seq results. Seven (50%) genes (*CHS2, CHI, F35H, MYB1, MYB2, MYB3* and *WD1*) fit well with the RNA-Seq results across all four samples. The gene *CHS1* expression was up-regulated significantly in the yellow 'Chocolate' bracts, which was inconsistent with our FPKM database. It is rational and acceptable that there are certain differences in direct comparisons between the RNA-Seq and qRT-PCR results due to

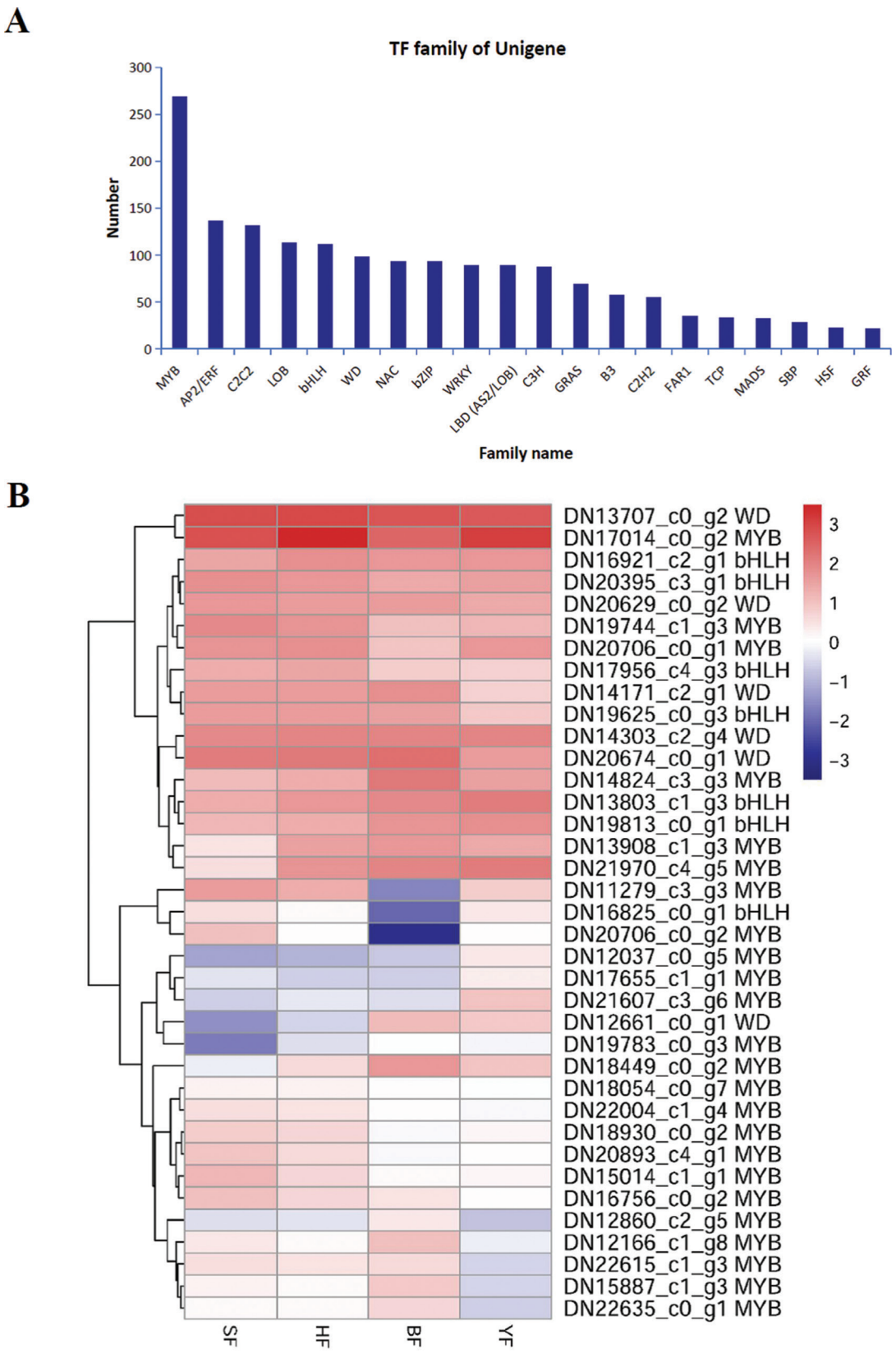

Figure 5. Distribution and selection of key differently expressed TFs associated with anthocyanin biosynthesis. (A) Distribution of TF family of unigenes. (B) Clustering heat maps of significantly enriched differently expressed TFs including *MYB, bHLH* and *WD* during three development stages and between species 'Chocolate' and 'Dutch Red'. TF, transcription factors.

Table 3. Differently expressed genes in transcription factor families of MYB and WD40.

Unigene	Annotation	FPKM-SF	FPKM-HF	FPKM-BF	FPKM-YF	Regulation
DN13707_c0_g2	WD-repeat protein	306.8	283.4	244.7	226.5	Whole flowering periods
DN14303_c2_g4	WD-repeat protein	90.3	97.0	97.2	101.5	
DN17014_c0_g2	MYB_superfamily	613.5	1,695.1	531.6	718.1	
DN20674_c0_g1	WD-repeat protein	121.4	131.1	136.2	67.3	
DN15014_c1_g1	MYB_superfamily	66.3	27.2	13.4	11.8	Early flowering periods
DN11279_c3_g3	MYB_superfamily	105.1	64.9	0.8	38.9	
DN19744_c1_g3	MYB_superfamily	160.6	132.0	45.9	53.1	
DN12166_c1_g8	MYB_superfamily	16.3	10.9	41.8	6.4	Blossom periods in 'Dutch Red'
DN12860_c2_g5	MYB_superfamily	4.5	5.2	22.7	2.8	
DN15887_c1_g3	MYB_superfamily	14.1	10.7	36.5	2.8	
DN22635_c0_g1	MYB_superfamily	7.1	11.2	26.8	3.3	
DN12661_c0_g1	WD-repeat protein	1.9	5.5	36.5	36.2	Blossom periods
DN21970_c4_g5	MYB_superfamily	15.3	121.8	220.9	298.7	

FPKM, fragments per kilobase of transcript per million mapped reads; HF, half-flowering.

different normalisation methods and other technical biases.

DISCUSSION

Anthocyanin, a family of water-soluble bioactive flavonoids, is considered to be rich in deep-coloured flowers and the content was found to have a correlation with colour (Pavel et al., 2018). The flavonoid biosynthesis, as a branch of the phenylpropanoid pathway, is responsible for the production of anthocyanins, isoflaonoid and flavonols (Tohge et al., 2017). In this study, the comparative transcriptome was used to explore the mechanisms of colour formation of 'Dutch Red' and the colour difference between 'Dutch' and 'Chocolate'. Our results suggested that the deepening of bract colour in *C. alismatifolia* was mostly associated with the up-regulated activity of structural enzymes in anthocyanin biosynthesis pathway at the initial flowering stage. The activity of these structural enzymes in the blossomed 'Chocolate' bracts was contrary to those in 'Dutch Red', and we examined different expressing patterns of MYB TF, *DN12166_c1, DN12860_c2, DN15887_c1 and DN22635_c0*, between 'Chocolate' and 'Dutch Red', which indicated that MYB factors might have participated in regulating anthocyanin biosynthesis and colour formation between different *C. alismatifolia* varieties. However, their real roles are difficult to define due to the lack of a transformation system in *C. alismatifolia*.

For anthocyanin biosynthesis pathways, as is shown in Figure 4, when phenylalanine is transformed into coumarate-CoA, it will be introduced into the anthocyanin biosynthesis pathway by CHS and CHI. CHS is one of the key and rate-limiting enzymes of phenylpropanoid pathway which plays superior roles in the production of anthocyanin. Singh and Kumaria (2020) showed that *CHS* expression was positively correlated with flavonoid and anthocyanin contents from the orchid plant. Accumulation of anthocyanin constituents was also quantified and found to corroborate well with the expression levels of *CHS* in plants of eggplant (Wu et al., 2020) and *Grewia asiatica* (Wani et al., 2017). CHI is the second rate-limiting and the first reported enzyme involved in the biosynthetic pathway of anthocyanin. Over-expression of *CHI* enhanced the flavonoid production in *D. cambodian*a and tobacco (Zhu et al., 2021). *CHI* suppression in the tissues of herbaceous peony resulted in the impairing of anthocyanin accumulation (Wu et al., 2018). In this study, up-regulated genes account for the majority of genes related to anthocyanin biosynthesis pathways, such as *PAL, 4CL, CHS* and *UFGT*, while their expressions were relatively higher at the beginning of flowering. *CHI* expression was significantly down-regulated at the full flowering stage of 'Dutch Red', but not at the full flowering stage of 'Chocolate'. This result implies that significant up-regulation of anthocyanin-related genes from the beginning of flowering may be one of the reasons for the deepening

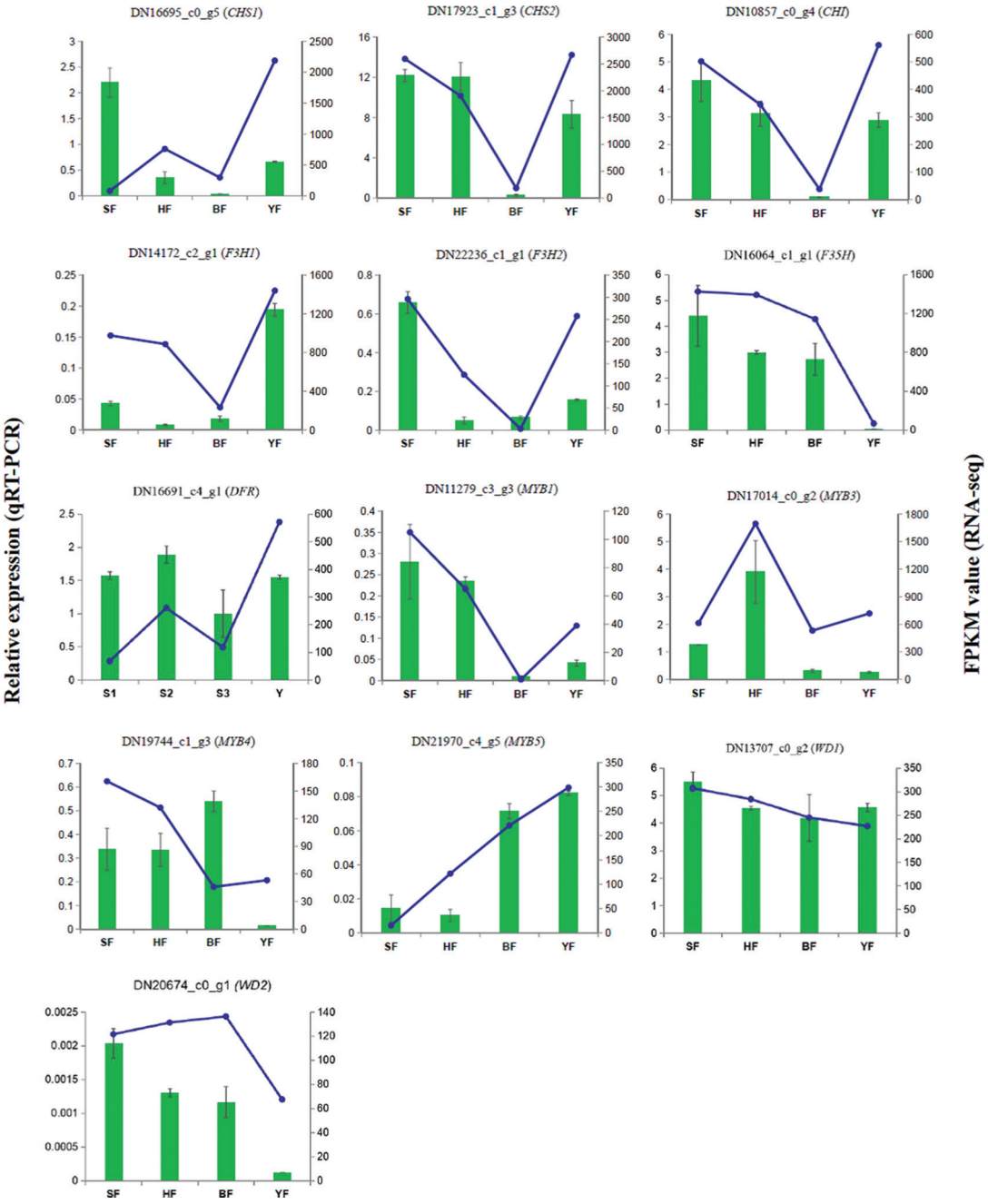

Figure 6. qRT-PCR validations of 14 putative genes involved in anthocyanin biosynthesis and regulations. The histograms represent expression determined by qRT-PCR (left *y*-axis), while lines represent expression by RNA-Seq in FPKM values (right *y*-axis). The *x*-axis in each chart represents the SF, HF, BF and YF, respectively. For qRT-PCR assays, the mean was calculated from three biological replicates. For RNA-Seq, each point is the mean of three biological replicates. BF, blossomed flowering; HF, half-flowering.

of red *C. alismatifolia* bracts, but the yellow formation mechanism in bracts of 'Chocolate' is another way of regulation, which is hard to define in this study. *F3′5′H* plays a dominant role in the formation of delphinidin-based anthocyanins. Previous studies (Li et al., 2021b; Nguyen et al., 2021) have shown that *F3′5′H* plays key roles in flower colour regulation. Over-expression of *F3′5′H* significantly increased the total flavonoid content in petals of *Aconitum carmichaelii*, and flowers of the transgenic lines of cyclamen showed modified colour and this correlated positively with the loss of endogenous *F3′5′H* transcript (Boase et al., 2010). In this study, most of the genes related to anthocyanin biosynthesis were up-regulated in YF (Table 2; Figure 4); however, the expression of *F3′5′H* was significantly down-regulated, which was consistent with the results of Boase et al. (2010). Reduced contents of F3′5′H proteins would make less dihydroflavonol introduced into the anthocyanin production, which is unfavourable for anthocyanin accumulation.

In plants, the structural genes of the flavonoid biosynthetic pathway are significantly regulated at the level of transcription. TFs of *R2R3-MYB, bHLH* and *MBW* complex, which activate the transcription of anthocyanins pathway, were widely studied. In previous studies, 13 TFs (*MYB, bHLH, WD40*, etc,) which showed strongly relating to flavonoid biosynthesis had been screened from the transcriptome analysis of *Paeonia* by Guo et al. (2019). Three *MYB* genes and two *bHLH* genes were strikingly down-regulated in the white flowers of a tobacco mutant (Jiao et al., 2020). WD-repeat protein 5 (WDR5), as a member of the WD40 protein family, is found widely involved in epigenetic regulation, and associated with controlling long noncoding RNAs and TFs (Chen et al., 2021; Dölle et al., 2021). In this study, nine *MYB* genes and one *WD* gene *(DN12661_c0)* were also found in different expression patterns that regulate anthocyanin biosynthesis. *MYB105* belongs to the S21 subfamily of the MYB family, and is involved in the biosynthesis of volatiles and flavonoid metabolites (Chen et al., 2015). In this study, expression of one MYB superfamily *(DN17014_c0_g2)* highly homologous to *MYB105* was significantly higher in the half-blossoming flowers than that in the buds, which is consistent with the comparative transcriptomic analysis of rose studied by Feng et al. (2021). No significant DEGs encoding bHLH proteins were found in comparisons with YF, SF, HF and BF. In conclusion, the regulation of MYB TF is more likely the main factor in the regulation of anthocyanin biosynthesis in *C. alismatifolia* colourful bracts.

CONCLUSION

To summarise, we applied the RNA-Seq technology to perform transcript analysis of the three flowering stages of 'Dutch Red' (SF, HF and BF), and the yellow 'Chocolate', YF. Our results revealed that the deepening of *C. alismatifolia* colour was associated with the up-regulation of anthocyanin biosynthesis at the initial flowering stage. In addition, we examined the down-regulation of the gene *F3′5′H,* an important enzyme for the synthesis of different basic anthocyanins, in the yellow 'Chocolate'. This gives clues to how the yellow-green appearance is formed; however, more research is still needed to verify this. In this paper, we also exhibited the differently expressed patterns of MYB TFs associated with anthocyanin-related gene expressions. Above all, our transcriptome analysis provided valuable molecular information for the colour formation of *C. alismatifolia*.

ACKNOWLEDGEMENTS

None.

FUNDING

This work was supported by the Zhangzhou Natural Science Fund (ZZ2020J10), the Regional Development Project of Fujian, China (2020N3010) and the Provincial Education and Science Foundation of Fujian, China (JAT200294).

AUTHOR CONTRIBUTIONS

L.Y.Y. and C.X.H. designed and performed experiments, analysed data and wrote the paper. Y.H.W. and T.Q.L. performed experiments. L.L.M. designed experiments and polished the paper.

CONFLICT OF INTEREST

All authors declare that no conflict of interest exists.

REFERENCES

Anders, S., McCarthy, D. J., Chen, Y., Okoniewski, M., Smyth, G. K., Huber, W., and Robinson, M. D. (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature Protocols, 8*(9), 1765–1786, doi: 10.1038/nprot.2013.099.

Boase, M. R., Lewis, D. H., Davies, K. M., Marshall, G. B., and Deroles, S. C. (2010). Isolation and antisense suppression of flavonoid 3', 5'-hydroxylasemodifies flower pigments and colour in cyclamen. *BMC Plant Biology*, *10*, 107, doi: 10.1186/1471-2229-10-107.

Chanapan, S., Tontiworachai, B., Deewatthanawong, R., and Suwanagul, A. (2017). Cloning and sequence analysis of chalcone synthase gene in *Curcuma alismatifolia*. *Acta Horticulture*, *1167*, 299–304, doi: 10.17660/ActaHortic.2017.1167.43.

Chen, L. Y., Bernhardt, A., Lee, J., and Hellmann, H. (2015). Identification of Arabidopsis MYB56 as a Novel Substrate for CRL3BPM E3 Ligases. *Molecular Plant*, *8*(2), 242–250, doi: 10.1016/j.molp.2014.10.004.

Chen, X., Xu, J. J., Wang, X. H., Long, G. L., You, Q. D., and Guo, X. K. (2021). Targeting WD Repeat-Containing Protein 5 (WDR5): a medicinal chemistry perspective. *Journal of Medicinal Chemistry*, *64*(15), 10537–10556, doi: 10.1021/acs.jmedchem.1c00037.

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*(18), 3674–3676, doi: 10.1093/bioinformatics/bti610.

Dewey, C. N., and Li, B. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*(1), 323, doi: 10.1186/1471-2105-12-323.

Dölle, A., Adhikari, B., Krmer, A., Weckesser, J., and Knapp, S. (2021). Design, synthesis, and evaluation of wd-repeat-containing protein 5 (wdr5) degraders. *Journal of Medicinal Chemistry*, *64*(15), 10682–10710, doi: 10.1021/acs.jmedchem.1c00146.

Feng, D., Zhang, H., Qiu, X., Jian, H., Wang, Q., Zhou, N., Ye, Y., Lu, J., Yan, H., and Tang, K. (2021). Comparative transcriptomic and metabonomic analysis revealed

the relationships between biosynthesis of volatiles and flavonoid metabolites in Rosa rugosa. *Ornamental Plant Research*, *1*, 5, doi: 10.48130/OPR-2021-0005.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., and Zeng, Q. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652, doi: 10.1038/nbt.1883.

Guo, L. P., Wang, Y. J., Teixeira Da Silva, J. A., Fan, Y. M., and Yu, X. N. (2019). Transcriptome and chemical analysis reveal putative genes involved in flower color change in *Paeonia* 'Coral Sunset'. *Plant Physiology and Biochemistry*, *138*, 130–139, doi: 10.1016/j.plaphy.2019.02.025.

He, C., Liu, X., Teixeira Da Silva, J. A., Liu, N., Zhang, M. Z., and Duan, J. (2020). Transcriptome sequencing and metabolite profiling analyses provide comprehensive insight into molecular mechanisms of flower development in dendrobium officinale (*Orchidaceae*). *Plant Molecular Biology*, *104*, 529–548, doi: 10.1007/s11103-020-01058-z.

Jiao, F., Zhao, L., Wu, X., Song, Z., and Li, Y. (2020). Metabolome and transcriptome analyses of the molecular mechanisms of flower color mutation in tobacco. *BMC Genomics*, *21*(1), 611, doi: 10.1186/s12864-020-07028-5.

Khumkratok, S., Boongtiang, K., Chutichudet P., and Pramaul, P. (2012). Geographic distributions and ecology of ornamental *Curcuma* (Zimgiberaceae) in Northeastern Thailand. *Pakistan Journal of Biological Sciences*, *15*, 929–935, doi: 10.3923/pjbs.2012.929.935.

Koshioka, M., Umegaki, N., Boontiang, K., Pornchuti, W., Thammasiri, K., Yamaguchi, S., Tatsuzawa, F., Nakayama, M., Tateishi, A., and Kubota, S. (2015). Anthocyanins in the bracts of *Curcuma* species and relationship of the species based on anthocyanin composition. *Natural Product Communications*, *10*(3), 453–456, doi: 10.1177/1934578×1501000320.

Kriangphan, N., Sailam, P., Sucharitakul, K., Rakmit, R., Leelapon, O., Roytrakul, S., Teerakathiti, T., Apisitwanich, S., and Chanvivattana, Y. (2009). Molecular characterization of the colour pigment biosynthesis pathways in *Curcuma* flowers. BIOTEC Central Research Unit, National Center for Genetic Engineering and Biotechnology, Paholyothin, Pathumthani 12120, Thailand.

Li, M., Cao, Y., Debnath, B., Yang, H., Kui, X., and Qiu, D. (2021a). Cloning and expression analysis of *Flavonoid 3', 5'-hydroxylase* gene from *Brunfelsia acuminata*. *Genes*, *12*, 1086, doi: 10.3390/genes12071086.

Li, Y. Y., Chen, X. H., Yu, H. W., Tian, Q. L., and Lu, L. (2021b). Identification and characterization of *CONSTANS*-like genes from *Curcuma a lismatifolia*. *Horticulture Environment and Biotechnology*, *62*, 279–286, doi: 10.1007/s13580-020-00314-x.

Li, X., Tang, D., Du, H., and Shi, Y. (2018). Transcriptome sequencing and biochemical analysis of perianths and coronas reveal flower color formation in *Narcissus pseudonarcissus*. *International Journal of Molecular Sciences*, *19*(12), 4006, doi: 10.3390/ijms19124006.

Li, Y. Y., Chen, X. H., Xue, C., Zhang, H., and Wang, D. Z. (2019). Proteomic response to rising temperature in the marine cyanobacterium *Synechococcus* grown in different nitrogen sources. *Frontiers in Microbiology*, *10*, 1976, doi: 10.3389/fmicb.2019.01976.

Li, Y. Y., Tian, Q. L., Yu, H. W., and Lu, L. M. (2021). Progress towards a molecular-level understanding of *Curcuma alismatifolia*. *European Journal of Horticultural Science*, *86*(3), 328–334, doi: 10.17660/eJHS.2021/86.3.12.

Mizuta, D., Ban, T., Miyajima, I., Nakatsuka, A., and Kobayashi, N. (2009). Comparison of flower color with anthocyanin composition patterns in evergreen azalea. *Scientia Horticulturae*, *122*(4), 594–602, doi: 10.1016/j.scienta.2009.06.027.

Nguyen, T. N. L., Hoang, T. T. H., Nguyen, H. Q., Tu, Q. T., Tran, T. H., Lo, T. M. T., Vu, T. T. T, and Chu, H. M. (2021). *Agrobacterium tumefaciens* - mediated genetic transformation and overexpression of the flavonoid 3′5′-hydroxylase gene increases the flavonoid content of the transgenic *Aconitum carmichaelii* Debx. *In Vitro Cellular & Developmental Biology - Plant*, *58*, 93–102, doi: 10.1007/s11627-021-10190-4.

Ogata, H., Goto, S, Sato, K., Fujibuchi, W., and Kanehisa, M. (1999). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *27*(1), 29–34, doi: 10.1093/nar/27.1.29.

Pavel, Z., Elena, A., Anna, L., Natalia, T., and Anatoly, V. (2018). Anthocyanin composition and content in rye plants with different grain color. *Molecules*, *23*(4), 948, doi: 10.3390/molecules23040948.

Petchang, R., Buddharak, P., Chundet, R., and U-Kong, W. (2017). Cloning of DFR gene in *Curcuma alismatifolia* 'Chiang Mai Pink' and Agrobacterium-mediated transformation. *Research Journal of Biotechnology*, *12*(3), 1.

Resmi, M. S., and Soniya, E. V. (2012). Molecular cloning and differential expressions of two cDNA encoding Type III polyketide synthase in different tissues of *Curcuma longa* L. *Gene*, *491*(2), 278–283, doi: 10.1016/j.gene.2011.09.025.

Robinson, M. D., Mccarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140, doi: 10.1093/bioinformatics/btp616.

Seppey, M., Manni, M., and Zdobnov, E. M. (2019). BUSCO: assessing genome assembly and annotation completeness. In: M. Kollmar (Ed.) *Gene prediction. Methods in molecular biology*. New York, NY, USA: Humana, doi: 10.1007/978-1-4939-9173-0_14.

Singh, N., and Kumaria, S. (2020). Molecular cloning and characterization of chalcone synthase gene from *Coelogyne ovalis* Lindl. and its stress-dependent expression. *Gene*, *76*, 145104, doi: 10.1016/j.gene.2020.145104.

Taheri, S., Abdullah, T. L., Rafii, M. Y., Harikrishna, J. A., Werbrouck, S. P. O., Teo, C. H., Sahebi, M., and Azizi, P. (2019). De novo assembly of transcriptomes, mining, and development of novel EST-SSR markers in *Curcuma alismatifolia* (Zingiberaceae family) through Illumina sequencing. *Scientific Reports*, *9*(1), 3047, doi: 10.1038/s41598-019-39944-2.

Tohge, T., De Souza, L. P., and Fernie, A. R. (2017). Current understanding of the pathways of flavonoid biosynthesis in model and crop plants. *Journal of Experimental Botany*, *68*(15), 4013–4028, doi: 10.1093/jxb/erx177.

Wang, G. (2007). P50-M high-throughput and automatic plasmid DNA preparation with SeqPrep technology. *Journal of Biomolecular Techniques*, *18*(1), 17–18.

Wang, J., Guo, M., Li, Y., Wu, R., and Zhang, K. (2018). High-throughput transcriptome sequencing reveals the role of anthocyanin metabolism in Begonia semperflorens under high light stress. *Photochemistry and Photobiology*, *94*(1), 105–114, doi: 10.1111/php.12813.

Wani, T. A., Pandith, S. A., Gupta, A. P., Chandra, S., Sharma, N., and Lattoo, S. K. (2017). Molecular and functional characterization of two isoforms of chalcone synthase and their expression analysis in relation to flavonoid constituents in *Grewia asiatica* L. *PLoS ONE*, *12*(6), e0179155, doi: 10.1371/journal.pone.0179155.

Wei, Z., Cheng, Y., Zhou, C., Li, D., Gao, X., Zhang S., and Chen, M. (2019). Genome-wide identification of direct targets of the TTG1–bHLH–MYB complex in regulating trichome formation and flavonoid accumulation in *Arabidopsis Thaliana*. *International Journal of Molecular Sciences*, *20*, 5014, doi: 10.3390/ijms20205014.

Wu, X., Zhang, S., Liu, X., Shang, J., and Zhang, D. (2020). Chalcone synthase (CHS) family members analysis from eggplant (*Solanum melongena* L.) in the flavonoid biosynthetic pathway and expression patterns in response to heat stress. *PLoS One*, *15*(4), e0226537, doi: 10.1371/journal.pone.0226537.

Wu, Y. Q., Zhu, M. Y., Jiang, Y., Zhao, D. Q., and Tao, J. (2018). Molecular characterization of chalcone isomerase (CHI) regulating flower color in herbaceous peony (*Paeonia lactiflora Pall.*). *Journal of Integrative Agriculture*, *17*(1), 122–129, doi: 10.1016/S2095-3119(16)61628-3.

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., Lei, K., Ge, G., Li, C. Y., and Wei, Y. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research*, *39*, 316–322, doi: 10.1093/nar/gkr483.

Zhang, X. H., Zheng, X. T., Sun, B. Y., Peng, C. L., and Chow, W. S. (2017). Over-expression of the *CHS* gene enhances resistance of *Arabidopsis* leaves to high light. *Environmental and Experimental Botany, 154,* 33–43, doi: 10.1016/j.envexpbot.2017.12.011.

Zhao, D. Q., and Tao, J. (2015). Recent advances on the development and regulation of flower color in ornamental plants. *Frontiers in Plant Science*, *6*, 261, doi: 10.3389/fpls.2015.00261.

Zhu, J., Zhao, W., Li, R., Guo, D., Li, H., Wang, Y., Mei, W., and Peng, S. (2021). Identification and characterization of Chalcone isomerase genes involved in flavonoid production in *Dracaena cambodiana*. *Frontiers in Plant Science*, *12*, 226, doi: 10.3389/fpls.2021.616396.

Received: November 9, 2021; accepted: March 13, 2022

SUPPLEMENTARY MATERIALS

Table S1. RNA sequencing data and corresponding quality control.

Sample	Raw reads	Raw bases	Clean reads	Clean bases	Error rate (%)	Q20 (%)	Q30 (%)	GC content (%)	Accession number
BF_1	54,170,030	8,179,674,530	53,271,940	7,920,660,194	0.0243	98.33	94.81	48.94	SRR17783056
BF_2	55,251,576	8,342,987,976	54,405,312	8,096,016,364	0.0241	98.41	94.99	48.87	SRR17783055
BF_3	49,533,034	7,479,488,134	48,632,006	7,241,216,477	0.0244	98.29	94.7	49.33	SRR17783054
HF_1	47,378,590	7,154,167,090	46,432,354	6,880,585,840	0.0248	98.14	94.35	50.69	SRR17783059
HF_2	47,325,446	7,146,142,346	46,485,250	6,893,639,691	0.0244	98.27	94.67	49.13	SRR17783058
HF_3	50,665,694	7,650,519,794	49,743,012	7,381,231,384	0.0244	98.28	94.67	48.55	SRR17783057
SF_1	43,599,698	6,583,554,398	42,859,044	6,355,814,159	0.0247	98.16	94.36	48.96	SRR17783064
SF_2	44,423,394	6,707,932,494	43,751,414	6,485,935,396	0.024	98.48	95.15	48.33	SRR17783063
SF_3	47,696,422	7,202,159,722	46,810,850	6,927,255,232	0.0247	98.16	94.38	48.97	SRR17783060
YF_1	45,005,912	6,795,892,712	44,227,366	6,543,730,785	0.0241	98.4	95	49.34	SRR17783053
YF_2	47,249,416	7,134,661,816	46,353,918	6,861,606,053	0.0241	98.41	95.05	50.28	SRR17783062
YF_3	44,148,530	6,666,428,030	43,297,006	6,412,165,282	0.0241	98.39	94.99	49.7	SRR17783061

Read Number: Pair-end Reads number in Clean Data; Base Number: Total base number in Clean Data; GC Content: The percentages of G and C in Clean Data constitute the total base; Q20/30: The percentage of bases in which Clean Data quality is greater than or equal to 20 or 30; Accession number: Accession number of all raw data in the SRA Sequence Data base in the NCBI.

BF, blossomed flowering; HF, half-flowering; NCBI, National Center for Biotechnology Information; SRA, short read archive.

Table S2. Numbers of transcripts in the 12 sequenced libraries.

seq_id	FPKM (<0.5)	FPKM (0.5–5)	FPKM (5–100)	FPKM (>100)	SUM (³0.5)
SF_1	65,107	61,397	31,536	1,647	94,580
SF_2	62,631	63,656	31,774	1,626	97,056
SF_3	64,446	62,140	31,444	1,657	95,241
HF_1	80,443	49,508	27,811	1,925	79,244
HF_2	67,574	59,459	30,864	1,790	92,113
HF_3	65,607	60,672	31,696	1,712	94,080
BF_1	67,310	59,403	31,170	1,804	92,377
BF_2	66,614	60,394	30,896	1,783	93,073
BF_3	76,848	51,186	29,570	2,083	82,839
YF_1	76,226	51,191	30,451	1,819	83,461
YF_2	84,464	45,418	27,806	1,999	75,223
YF_3	82,758	46,325	28,622	1,982	76,929

BF, blossomed flowering; FPKM, fragments per kilobase of transcript per million mapped reads; HF, half-flowering.

Table S3. Functional annotation of transcripts and unigenes in databases.

	Transcript number (%)	Unigene number (%)
NR	104,303 (0.6532)	34,189 (0.4923)
Swiss-Prot	82,193 (0.5147)	26,589 (0.3828)
Pfam	73,262 (0.4588)	24,101 (0.347)
COG	25,088 (0.1571)	7,537 (0.1085)
GO	71,758 (0.4494)	23,229 (0.3345)
KEGG	47,453 (0.2972)	14,838 (0.2136)
Total_anno	1,05,993 (0.6638)	35,183 (0.5066)
Total	1,59,687 (1)	69,453 (1)

COG, Clusters of Orthologous Groups of Proteins; KEGG, Kyoto Encyclopedia of Genes and Genomes; NR, non-redundant.

Table S4. Detailed annotation of differentially expressed MYBs and WD proteins.

Unigene	Hit_name	Description	Identity (%)	Similarity (%)	KO	Paths	Swiss-Prot
DN13707_c0_g2	XP_009415403.1	WD-repeat protein	88	95.3	–	–	gi|74698589|sp|Q9Y7K5.2|YGI3_SCHPO
DN14303_c2_g4	ONM38806.1	WD-repeat protein	76.8	85.4	K10752	–	gi|22096353|sp|O22607.3|MSI4_ARATH
DN17014_c0_g2	XP_009404081.1	MYB_superfamily	88	91.6	–	–	gi|115502385|sp|O80622.2|EXP15_ARATH
DN20674_c0_g1	XP_010931565.1	WD-repeat protein	79.7	91.5	–	–	gi|75318693|sp|O80775.2|WDR55_ARATH
DN15014_c1_g1	XP_009385043.1	MYB_superfamily	54.9	63.7	K14491	map04075	gi|75323583|sp|Q6H805.1|ORR24_ORYSJ
DN11279_c3_g3	XP_017696697.1	MYB_superfamily	95.6	97.8	K09422	–	gi|75317981|sp|O22264.1|MYB12_ARATH
DN19744_c1_g3	XP_009418439.1	MYB_superfamily	76.3	82.4	–	–	gi|75330977|sp|Q8S9H7.1|DIV_ANTMA
DN12166_c1_g8	XP_016205675.1	MYB_superfamily	86.2	92.6	K09422	–	gi|56749347|sp|Q8LPH6.1|MYB86_ARATH
DN12860_c2_g5	XP_023531843.1	MYB_superfamily	49.8	59.8	K09422	–	gi|75333993|sp|Q9FKL2.1|MYB36_ARATH
DN15887_c1_g3	XP_009408601.1	MYB_superfamily	44.7	55	–	–	gi|75338846|sp|Q9ZQ85.2|EFM_ARATH
DN22635_c0_g1	PSS31516.1	MYB_superfamily	66.9	74.4	K09422	–	gi|75335856|sp|Q9M2Y9.1|RAX3_ARATH
DN12661_c0_g1	XP_009389455.1	WD-repeat protein	72.9	78.9	K14963	–	gi|82232080|sp|Q5M786.1|WDR5_XENTR
DN21970_c4_g5	XP_009405310.1	MYB_superfamily	59.5	66.4	–	–	gi|75330977|sp|Q8S9H7.1|DIV_ANTMA

Table S5. Primers used for qRT-PCR analysis.

Genes	Sequences (5′-3′)	Product size
ANR 12990	CAGCTGAAGCAGATGCAGAG	147
	CTCCGGATTCTCCGACATTA	
ANR 13257	TCACATCATCCTCACCGTGT	103
	AAGGGGCTGGAGGAGATTTA	
CHI 10857	ACGGAGTTCTTCCAGAGCAA	165
	TTTCTGCATCATCTCCACCA	
CHI 15898	GCTCGAACACCTCCTTGAAC	155
	CGACAAGTTCACGAGGATCA	
CHS 16695	TGAGGGAGAACCCGAATATG	161
	GCAGAAGACGAGGTGTGTGA	
CHS 17923	ATGGCCTGAAGATGGATCAG	186
	GCGTCCTCTTCATTCTCGAC	
DFR 16691	TGCCCGTTTAACTCATTTCC	101
	ACTGTGGAGAAGGAGCAGGA	
DFR 14193	GTGGTGTTCACCTCCTCGAT	189
	CGCTAGCTCTACCCCCTTCT	
F3′5′H 8362	CTCCTCCTCCGCTACCTTCT	160
	GGTACATGATGGGGCCATAG	
F3′5′H 16064	ATGTGGAGGTTGCAGAGCTT	80
	GCGTACGAAGGACAGGACAT	
F3H 14172	ACGTTGATTGGACCGAACAT	186
	CTGCGAGATCTACACGGACA	
F3H 22055	GCCTCTTGCATAGCTTCACC	83
	GGTGATCCTGCCAACTCATT	
F3H 22236	AAGGAGAAGTACGCGTCCAA	184
	TGTATTCCTCGTTCGCCTTC	
MYB 11279	TTTCTCCCATGCTGCTTTCT	160
	TCACAATGCCACGGTTAAGA	
MYB 12860	ACCCTATCGTCGAACCACTG	174
	TCGACCAAGAGAGCCAACTT	
MYB 15887	CATCTTCCTCCTTGGAGCTG	138
	TCTGGATCTGCGAGGAAAGT	
MYB 18449	TTCCACTTGGACTTGCACTG	110
	AAGAGCTTCACGGAGACGAA	
MYB 12166	GCCGTCCAAACACATCTTCT	85
	GATTTGGTCGAGCCAGAGAG	
MYB 15014	ATTTCTGCAGATGGCTTGCT	99
	ATCAATTGGGCATCGAGAAG	
MYB 17014	CTACGGAGGAGGATGGATGA	93
	AAGTTGCCATACCCACAAGC	
MYB 19744	CGAAGAACACTGCACTGGAA	190
	TATCCACTGCCTCCTCCATC	
MYB 21970	GCACCGGTAAGTGCGTTAAT	159
	TTCGTTTCCATCCATCCATT	
MYB 22635	GTCGCAATAACCGACCATCT	125
	ATCATCTGCAGCCTCTTCGT	
UFGT 12119	TCGGCGATCAAGGTAGATTC	137
	GGATGAAGCTGAGCAACTC	
UFGT 13065	CGGAGGTAGCCATCAACAAT	143
	ATCCCACATGGTGGTCTTGT	
UFGT 13592	CTGGAGCTAAGGAGCAAGGA	118
	TACCAGGGGTAATGGGATGA	
WD 12661	CAGTAGCCATGGCAGCTACA	109
	ATGGCGGATGAGTACCAAAG	
WD 13707	CCAAAGAAGTCCAGCTCCTG	91
	AATTTTTGGCAATGGCAGAG	
WD 14303	CAGGCGATGTGAAGAATTGA	
	GCAACTGATGGGGGTAAAGA	
WD 20674	TTCGTCCCTTTGAGTTTGCT	
	ATGGGCTGTAACACGTAGCC	

CHI, chalcone isomerase; CHS, chalcone synthase; DFR, dihydroflavonol 4-reductase; F3H, flavanone-3-hydroxylase.