Transcriptome analysis of the molecular mechanism of *Chrysanthemum* flower color change under short-day photoperiods

Wei Dong, Mangmang Li, Zhongai Li, Shuailei Li, Yi Zhu, Hongxu, Zicheng Wang

PII: S0981-9428(19)30492-9

DOI: https://doi.org/10.1016/j.plaphy.2019.11.027

Reference: PLAPHY 5942

To appear in: *Plant Physiology and Biochemistry*

Received Date: 8 October 2019

Accepted Date: 16 November 2019

Please cite this article as: W. Dong, M. Li, Z. Li, S. Li, Y. Zhu, Hongxu, Z. Wang, Transcriptome analysis of the molecular mechanism of *Chrysanthemum* flower color change under short-day photoperiods, *Plant Physiology et Biochemistry* (2019), doi: https://doi.org/10.1016/j.plaphy.2019.11.027.

Transcriptome analysis of the molecular mechanism of *Chrysanthemum* flower color change under short-day photoperiods

Wei Dong, Mangmang Li, Zhongai Li, Shuailei Li, Yi Zhu, Hongxu Ding, Zicheng Wang*

School of Life Science, Henan University, Plant Genetics Laboratory, Kaifeng, Henan 475000, People's Republic of China

*Corresponding author e-mail address: wzc@henu.edu.cn

Contributors e-mail address: dongw@henu.edu.cn

Other authors' e-mail addresses: 854976036@qq.com; Zhongai Li: 42062796@qq.com; Shuailei Li: 1109569406@qq.com; Yi Zhu: 276150673@qq.com; Xuhong Ding: 2387446208@qq.com

Abstract

Chrysanthemum [*Dendranthema morifolium* Tzvel.] is an ornamental plant grown under long-term artificial cultivation conditions. In production, early *Chrysanthemum* blossoms are often promoted by artificial short-day treatment. However, we found that the flower colour of *Chrysanthemum* blossoms induced by artificial short-day treatment was lighter than those induced by the natural photoperiod. To explore the intrinsic mechanism of colour fading in flowers, we performed full-length transcriptome sequencing of *Chrysanthemum morifolium* cv. 'Jinbeidahong' using single-molecule real-time sequencing and RNA-sequencing under natural daylight (ND) and short daylight (SD) conditions. The clustered transcriptome sequences were assigned to various databases, such as NCBI, Swiss-Prot, Gene Ontology and so on. The comparative results of digital gene expression analysis revealed that there were differentially expressed transcripts (DETs) in the four stages under ND and SD conditions. In addition, the expression patterns of anthocyanin biosynthesis structural genes were verified by quantitative real-time PCR. The major regulators of the light signalling *ELONGATED HYPOCOTYL5* genes were markedly upregulated under ND conditions. The patterns of anthocyanin accumulation were consistent with the expression patterns of *CHI1* and *3GT1*. The results showed that the anthocyanin synthesis is tightly regulated by the photoperiod, which will be useful for molecular breeding of *Chrysanthemum*.

Keywords

Chrysanthemum; full-length transcriptome sequencing; functional annotation; artificial short-day treatment; anthocyanin

Background

Chrysanthemum [*Dendranthema morifolium* (Ramat.) Tzvel.] is a perennial root plant in the family Compositae. In China, *Chrysanthemum* is one of the top 10 most famous flowers as well as one of the four major cut flowers. The international demand and economic benefits of *Chrysanthemum* cut flowers are very high, and in China, the *Chrysanthemum* cultivation area, production, and sales have always been at the forefront of the global cut flower industry (Lv et al., 2016). *Chrysanthemum* is an ornamental plant that was bred through long-term artificial selection. This plant has more than 30,000 cultivars, approximately 3,000 of which are grown in China.

In production, early *Chrysanthemum* blossoms are often promoted by artificial short-day (SD) treatment. After vegetative growth reaches a certain stage, proper shading treatment is carried out every day. The duration of daylight, namely, the photoperiod, changes regularly with different seasons (Yang et al., 2018). Photoperiodic lighting can promote flowering of long-day plants and inhibit flowering of

SD plants (Meng and Runkle, 2018). However, we found that the flower colour of *Chrysanthemum* blossoms induced by artificial SD treatment was lighter than those induced by natural day (ND) treatment. Flower colour is a significant feature that determines the commercial value of *Chrysanthemum* varieties. Pink and purplish-red ray petal colours are all derived from anthocyanins (Ohmiya, 2018). Environmental factors affect the anthocyanin biosynthesis pathway related genes (Guo et al., 2008). Light conditions are one of these key factors. In the ray petals, the sharp decrease in anthocyanins is due to the shading of *Chrysanthemum* flowers (Hong et al. 2015). And the expression of anthocyanin biosynthesis pathway genes, including *flavanone 3-hydroxylase* (*F3H*), *anthocyanidin synthase* (*ANS*), *dihydroflavonol 4-reductase* (*DFR*), *UDP-glucose-flavonoid 3-O-glucosyltransferase* (*3GT*) genes were significantly inhibited, as well as transcription factor of anthocyanin biosynthesis, *MYB domain 5-1* (*MYB5-1*), *basic helix-loop-helix 24* (*bHLH24*), *cryptochrome1a* (*CRY1a*), *CONSTITUTIVE PHOTOMORPHOGENIC* (*COP1*) and *ELONGATED HYPOCOTYL5* (*HY5*) (Hong et al., 2016). However, little is known about how the relevant processes are regulated at the molecular level in *Chrysanthemum*. It is very necessary to study the molecular mechanism in depth.

Chrysanthemums are an allohexaploid (2n = 6× = 54). Because the genome of *Chrysanthemum* is large in size and complex in structure, it is very difficult to perform full genome sequencing. Currently, there are no available allohexaploid *Chrysanthemum* genome sequences, which limits in-depth studies into the molecular biology of this species. With the continuous progress of biotechnology, PacBio single-molecule real-time (SMRT) sequencing technology has been widely used. SMRT sequencing can generate approximately 20-Kb reads, spanning a complete transcript from the 5′-end to the 3′-poly (A) tail, allowing for accurate identification of isoforms and precise analysis of alternative splicing, fusion genes, transposable elements (TEs) and allele expression (Fang et al., 2012). The application of SMRT transcriptome sequencing has been reported in *Chrysanthemum morifolium* 'Fenditan' and *Chrysanthemum morifolium* 'Yuuka' (Liu et al., 2015; Ren et al., 2016). Analyze the quantity of gene expression levels using RNA sequencing to gain a comprehensive understanding of biological pathways (Wang et al., 2015; Lowe et al., 2017).

To further understand the mechanism of anthocyanin accumulation under SD and ND conditions in *Chrysanthemum*, the ray florets of *C. morifolium* 'Jinbeidahong', one traditional cultivar, were analysed using transcriptomic methods and PacBio SMRT sequencing technology. These results will be very useful for improving the understanding of the light-induced anthocyanin biosynthesis molecular mechanisms. Furthermore, this study provides the basis for molecular breeding theory for modifying the colour of *Chrysanthemum*. Finally, the high-quality sequencing data provide valuable reference genes of *Chrysanthemum* for follow-up studies.

Results

Comparative analysis of anthocyanin accumulation in 'Jinbeidahong' four stage

Four sequential developmental stages are divided into budding (BD), bud breaking (BB), early blooming (EB) and full blooming (FB) (Fig. 1 a). And the flower cuttings were collected from these four stages. The anthocyanin levels that accumulated in each stage of *Chrysanthemum* development under ND conditions were much higher than those under SD conditions. The highest levels of anthocyanin accumulation appeared in the FB stage under ND conditions (Fig. 1 b). Inflorescence

diameter and fresh weight in each stage of *Chrysanthemum* development under ND conditions were much higher than those under SD conditions (Fig. 1 c,d).

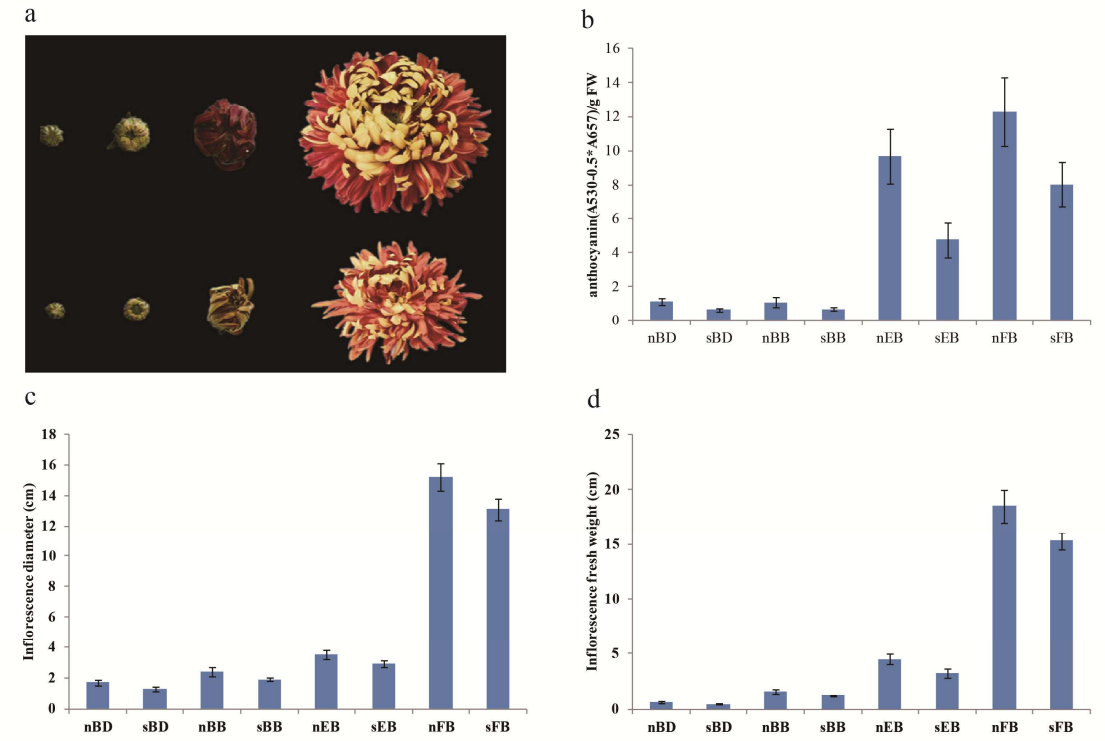


Fig. 1 a C. *morifolium* 'Jinbeidahong' under ND conditions (up) and SD conditions (down): BD stage, BB stage, EB stage and FB stage. b Anthocyanin content analysis of C. *morifolium* 'Jinbeidahong' at the BD, BB, EB and FB stages. c Inflorescence diameter of C. *morifolium* 'Jinbeidahong' at the BD, BB, EB and FB stages. d Inflorescence fresh weight of C. *morifolium* 'Jinbeidahong' at the BD, BB, EB and FB stages. Each column represents the mean of three independent measurements. Error bars represent the SD of the mean values.

Structure analysis of the full-length transcriptome

The three different cDNA libraries of leaf, petiole, stem, bud, flower, and root samples were sequenced using the PacBio RS II system, and cDNA insert sizes of 1-2 kb, 2-3 kb and 3-6 kb were prepared (Fig. 2). The sequencing results showed that the average single-molecule ratio in the ZMW wells was 62.4%. After stringent quality checks and data cleaning, 60 GB of data were obtained and the average proportion of clean reads was approximately 93%. Amount of 842,913 reads of insert (ROI) and 415,387 full-length non-chimeric reads (FLNC) were obtained. The FLNC reads were clustered and corrected to generate 245,033 non-redundant cluster consensus sequences following the iterative clustering error (ICE) correction algorithm of the IsoSeqCluster module in SMRT Analysis v2.3. A secondary clustering was performed to reduce redundancy using Cd-hit-est (v4.6). Finally, 89,477 FLNC reads were obtained. The complete evaluation of the de-duplicated transcriptome was performed by BUSCO (Fig. 3). The proportion of complete and single copies was approximately 21%. The proportion of complete and duplicated copies was approximately 38%. The proportion of fragments was approximately 4%. Finally, 37% of the copies were missing.

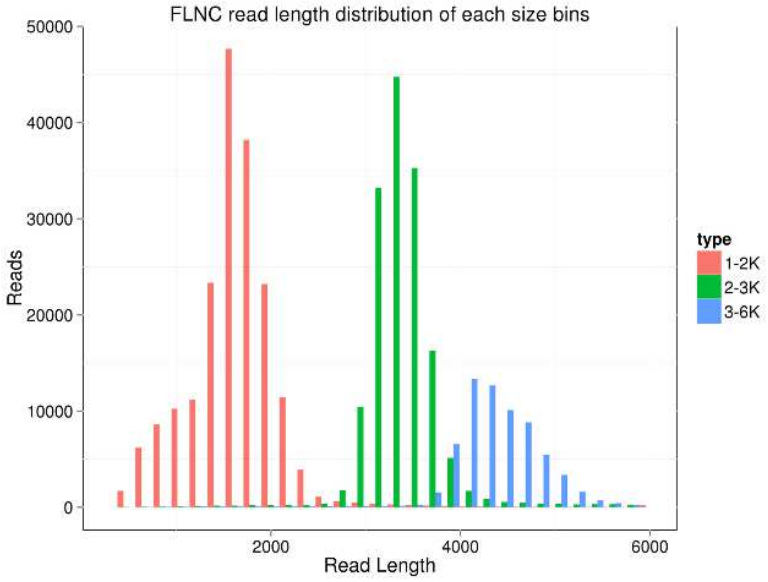


Fig. 2 FLNC read length distribution of each size bin.

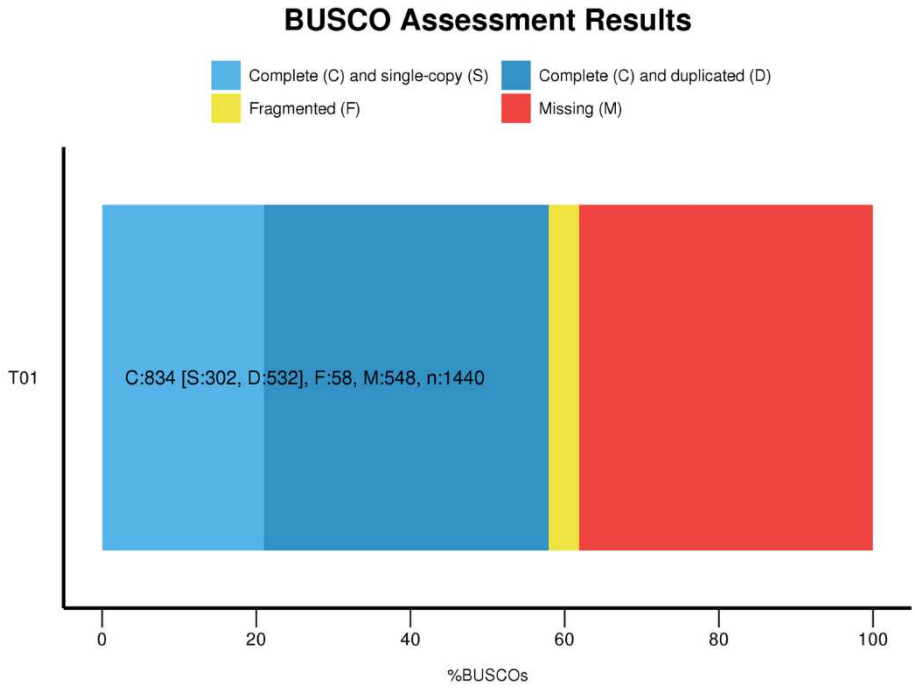


Fig. 3 BUSCO assessment results.

Gene transcription produces precursor mRNA, which can be spliced in many ways. Different exons are selected to produce different mature RNAs, which can be translated into different proteins representing the diversity of biological traits. The results of all-vs-all BLAST analysis showed a total of 9249 candidate variable splicing events. In particular, two unigenes *ELONGATED HYPOCOTYL5* (*HY5*), T01_cb9752_c45742/f3p1/2628 and T01_cb19426_c42556/f12p3/3429, were identified as variable splicing events involved in anthocyanin biosynthesis. Non-coding RNA was predicted in the high-quality non-redundant FLNC reads as described in Methods. A total of 7,314 long non-coding RNAs were predicted, accounting for 5.2% of the total reads. And no anthocyanin biosynthesis structural gene was predicted. The prediction results are available in Supplemental 1.

Functional annotation

Of the 89,477 unigenes, 83,654, 66,608, 52,187 37,685, 55,469, 69,072, 38,455 and 82,217 were aligned to the NCBI non-redundant (NR), Swiss-Prot, Gene Ontology (GO), Clusters of Orthologous Groups (COG), euKaryotic Orthologous Groups (KOG), Protein family (Pfam) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases, respectively. A total of 83,654 genes (93.5% of the total) were aligned to the Swiss-Prot, Protein Information Resource (PIR), Protein Research Foundation (PRF), and Protein Data Bank (PDB) databases, and the protein data was translated from CDS data from GenBank and RefSeq (Fig. 4). The top scores were associated with *Vitis vinifera* (13.94%), followed by *Sesamum indicum* (9.54%) and *Coffea canephora* (7.33%). The GO was used for gene annotation and analysis, which included molecular functions, cellular components and biological processes three ontologies. In view of the NR annotation, 52,187 unigenes were classified into 52 functional GO categories. A total of 53769 unigenes were classified into 16 cellular component categories, 86952 unigenes were classified into 17 molecular function categories, and 136031 unigenes were classified into 19 biochemical process categories (Fig. 5). A eukaryote-specific version of the COG tool was used to identify orthologous and paralogous proteins. Based on the COG databases, 84,850 unigenes were classified into 25 functional classes. The top scores were associated with *Replication, recombination and repair* (12.29%), followed by *Transcription* (11.68%) and *Signal transduction mechanisms* (10.71%) (Fig. 6).

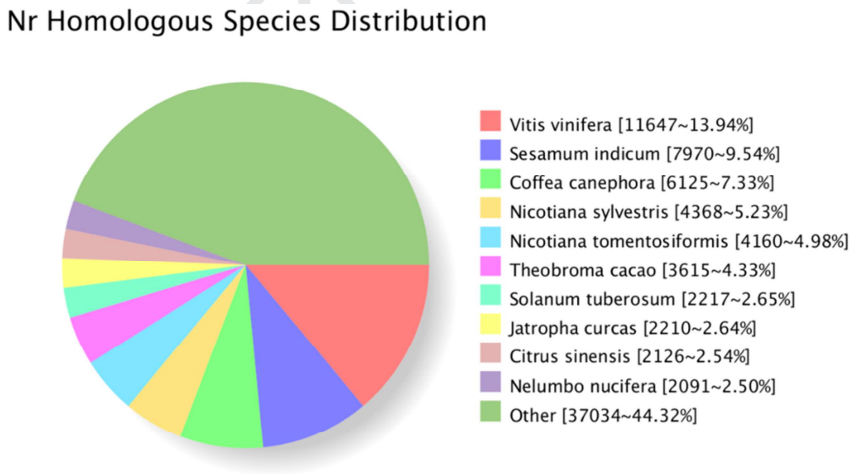


Fig. 4 NR homologous species distribution of the 83,654 unigenes.

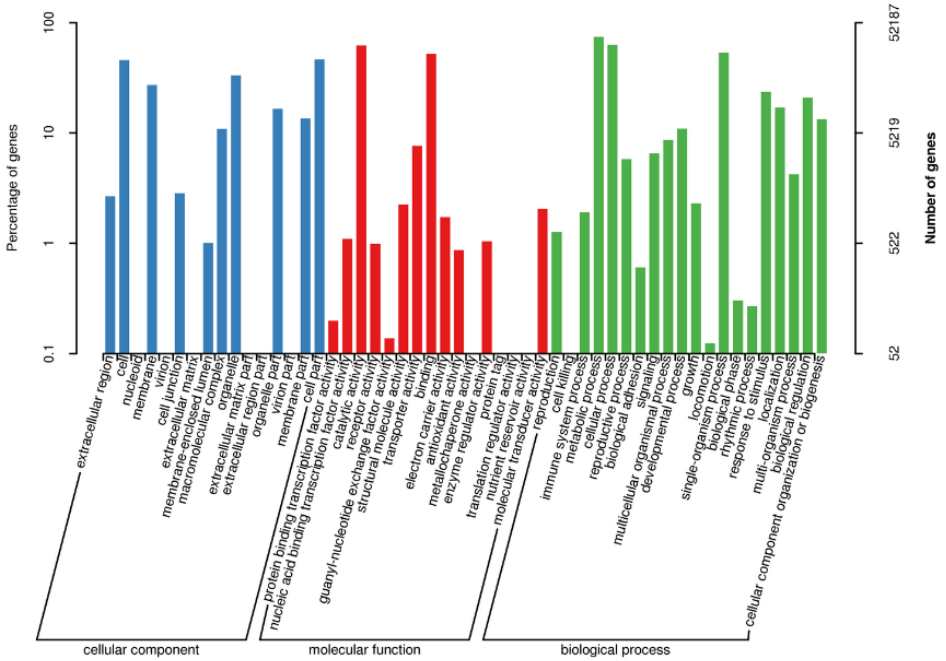


Fig. 5 GO classification. The three main GO categories are (left to right): cellular component, molecular function and biological process.

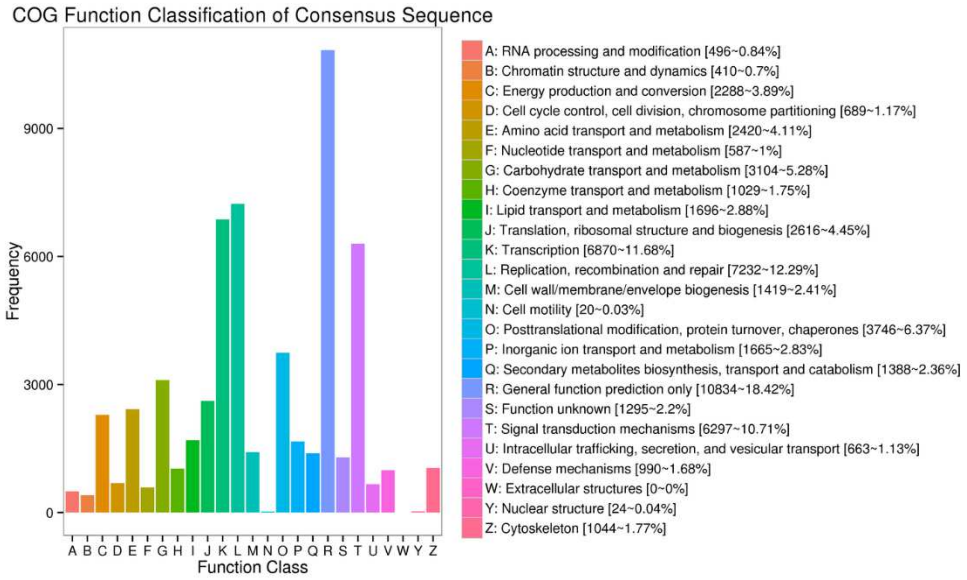


Fig. 6 COG classification in *C. morifolium* ‘Jinbeidahong’. The 25 categories of COG classifications are shown.

Comparative analysis of digital gene expression profiling in ‘Jinbeidahong’ four stages

In order to study the effect of photoperiod on anthocyanins in C. *morifolium* ‘Jinbeidahong’ at the BD, BB, EB and FB stage, digital gene expression (DGE) profiling was also performed. The results show that the raw reads was 22,504,773, 22,100,052, 38,897,346, 23,335,517, 25,910,326, 28,471,699 and 26,105,006 generated from the eight samples, respectively.

Subsequently, DETs were identified between the samples grown under ND and SD conditions and

the BD, BB, EB and FB stages. There were 974 DETs (431 upregulated and 543 downregulated) in the BD stage, 2,434 DETs (639 upregulated and 1,795 downregulated) in the BB stage, 6,808 DETs (2,987 upregulated and 3,821 downregulated) in the EB stage, and 1,202 DETs (610 upregulated and 592 downregulated) in the FB stage (Fig. 7). As shown in the figures, the highest number of DETs during flower development of C. *morifolium* 'Jinbeidahong' was observed for EB-stage unigenes.

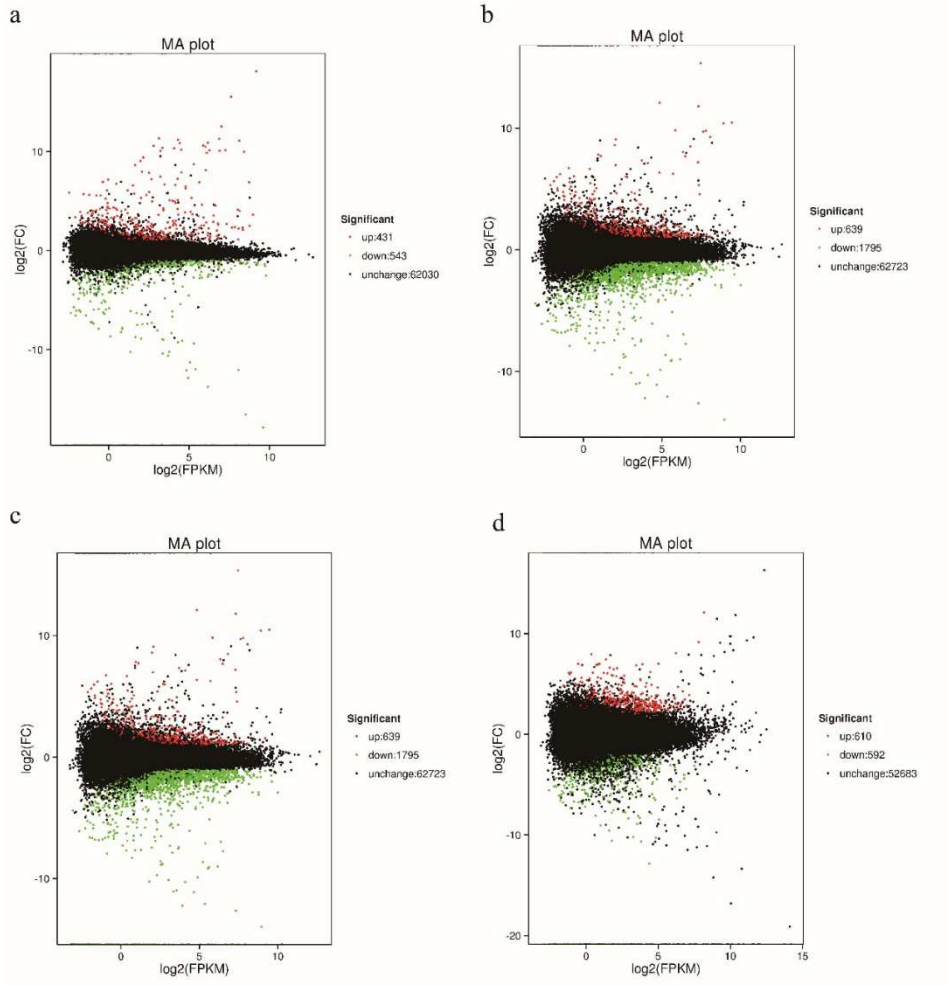


Fig. 7 MA plot of upregulated and downregulated genes in the BD stage (a), BB stage (b), EB stage (c) and FB stage (d).

Different publicly databases are used to classify the functions of DETs. For example, 310, 466, 343, 395, 816, 596, 696, 722, were aligned against the COG, GO, KEGG, KOG, Pfam, Swiss-Prot, eggNOG and NR databases, respectively (Table 1). All the transcriptome unigenes were used as background, obviously enriched GO terms were obtained for the DETs in four stage. In the BD stage, the DETs between the samples under ND conditions and those under SD conditions were significantly enriched in rhythmic process included in biological process, extracellular region included in cellular component and electron carrier activity included in molecular function. In the BB stage, the DETs were significantly enriched in rhythmic process included in biological process, supramolecular complex included in cellular component and antioxidant activity included in molecular function. In the EB stage, the DETs were significantly enriched in rhythmic process included in biological process, extracellular region included in cellular component and electron carrier activity included in molecular function. In the FB stage, the DETs were significantly enriched in extracellular region included in cellular

component and transporter activity included in molecular function (Supplementary Fig. 1). In the analysis of KEGG pathway, the prevailing pathways were as follows: protein processing in endoplasmic reticulum (8.16%) in the BD stage between the samples under ND conditions and SD conditions, carbon metabolism (7.69%) in the BB stage, carbon metabolism (7.69%) in the EB stage, and starch and sucrose metabolism in the FB stage (Supplementary Fig. 2).

Table 1 Number of differentially expressed transcripts in annotations

DEG_Set	Annotated	COG	GO	KEGG	KOG	Pfam	Swiss-Prot	eggNOG	nr
T1	918	310	466	343	395	816	596	696	722
T2	2327	1044	1544	1136	1254	2090	1824	2045	2086
T3	6519	2910	4386	2981	3700	5728	5291	6136	6254
T4	1172	491	693	478	619	999	868	1006	1022

The statistical enrichment of differentially expressed genes (DEGs) in KEGG pathways was tested using KOBAS software. The figures show the first 20 paths with the lowest significant Q values (Fig. 8). The over-presentation in the BD stage was associated with taurine and hypotaurine metabolism between the samples under ND conditions and SD conditions, while in the BB stage this was associated with thiamine metabolism, in the EB stage this was associated with the photosynthesis antenna, and in the FB stage this was associated with anthocyanin biosynthesis.

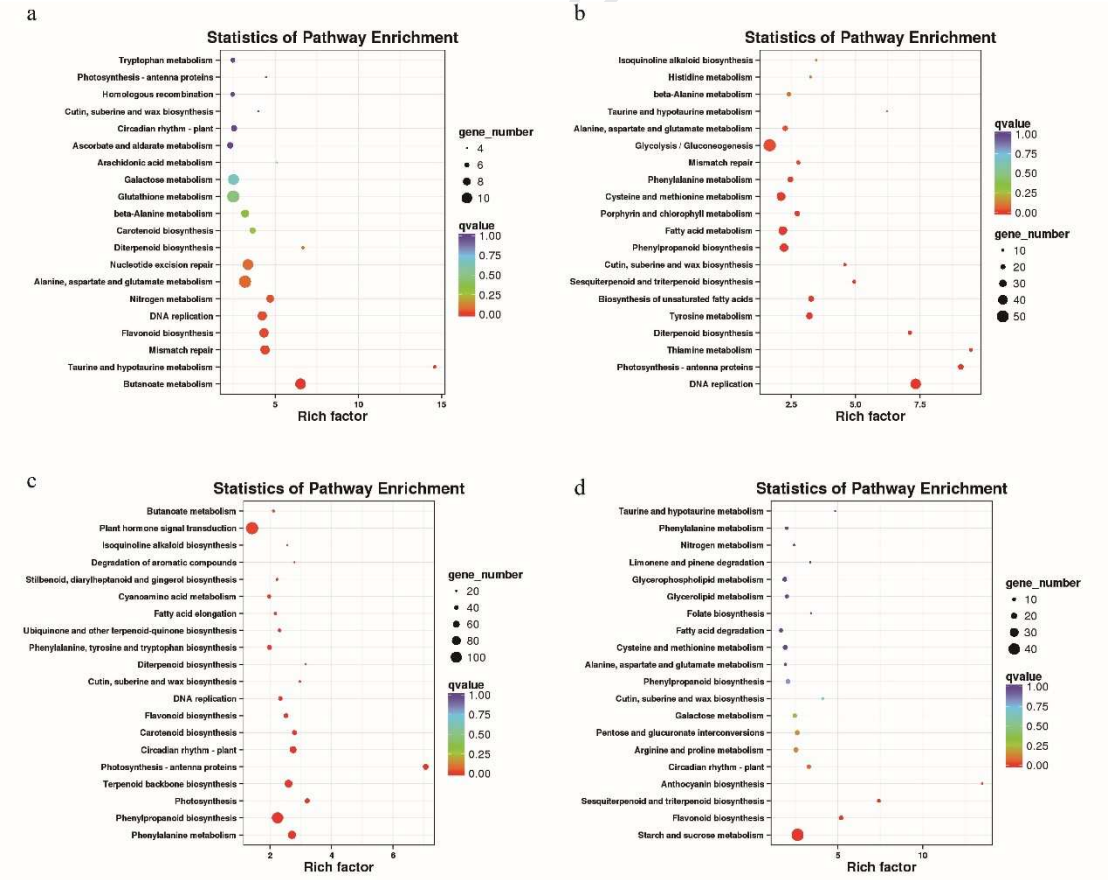


Fig. 8 KEGG pathway enrichment scatter map of differentially expressed transcripts in the BD stage (a), BB stage (b), EB stage (c) and FB stage (d).

Characterization of functional genes involved in anthocyanin biosynthesis

To explore the molecular basis of the difference in flower colours between the samples under ND and

SD conditions, we identified important functional genes involved in the *C. morifolium* 'Jinbeidahong' anthocyanin biosynthetic pathway. The genes for the anthocyanin biosynthetic pathway consist of structural genes and regulatory genes (Tanaka et al., 2008). The results of bioinformatic analysis show that many unigenes were determined as homologous sequences of the structural genes in anthocyanin biosynthetic pathway. The anthocyanin synthesis of upstream genes includes twenty-three *chalcone synthase* (*CHS*), seventeen *chalcone isomerase* (*CHI*), twenty-five *flavanone 3-hydroxylase* (*F3H*), twenty-eight *flavonoid 3'-hydroxylase* (*F3'H*), twenty-four *flavonoid 3',5'-hydroxylase* (*F3'5'H*). The anthocyanin synthesis of downstream genes includes thirteen *dihydroflavonol 4-reductase* (*DFR*), fifteen *anthocyanidin synthase* (*ANS*) and thirty-one *UDP-glucose-flavonoid 3-O-glucosyltransferase* (*3GT*) genes (Table 2). Regarding transcription factors involved in the regulation of anthocyanin biosynthesis, the basic helix-loop-helix (bHLH), MYB domain (MYB) and WD40-repeat domain-containing proteins (WD40) proteins have been widely demonstrated to activate or repress transcription of the anthocyanin structural genes. A total of 942 MYB-related homologous family members, including 350 bHLH proteins and 305 WD40 proteins, were identified. Photoreceptor and regulator of light signalling genes, including 44 *cryptochrome* (*CRYs*), 44 *phytochrome* (*PHYs*), *CONSTITUTIVE PHOTOMORPHOGENIC* (*COP1*), 41 *CONSTANS* (*CO*) and *ELONGATED HYPOCOTYL5* (*HY5*), were identified. The results of DEG analysis showed that the structural genes expression was very different during flower development among the samples. Interestingly, two unigenes (unigene ID: T01_cb15047_c1/f1p0/927 and T01_cb15047_c6/f1p0/815, a member of *CH1* homologies) were always significantly upregulated in the BD, BB and EB stages under ND conditions compared to their expression levels under SD conditions. The highest CHI transcript levels were observed at the EB stage under ND conditions, while the lowest were observed at the BB stage under SD conditions. All homologous members of the 3GT family were significantly upregulated in the four stages. Two unigenes (unigene ID: T01_cb8439_c28/f13p5/1745 and T01_cb8439_c27/f10p4/1748, a member of *3GT* homologies) were significantly upregulated during the four stages under ND conditions compared to their expression levels under SD conditions. The highest *3GT* transcript levels were detected at the BD stages under both conditions. One unigene (unigene ID: T01_cb5961_c8/f1p1/3157, a member of *F3H* homologies) that was significantly downregulated was expressed in the BD, BB and EB stages. The highest *F3H* transcript levels were detected at the EB stages under both conditions. The highest *HY5* transcript levels were detected at the BD stages under SD conditions and FB stages under ND conditions. Four unigenes (unigene ID: T01_cb19426_c30649/f1p1/3155, T01_cb5508_c11326/f1p0/3944, T01_cb19426_c42556/f12p3/3429, T01_cb9752_c45742/f3p1/2628) that were significantly downregulated in the BB, EB and FB stages. The highest *F3H* transcript levels were detected at the EB stages under SD conditions. One unigene (unigene ID: T01_cb5961_c8/f1p1/3157) that was significantly downregulated was expressed in C. *morifolium* 'Jinbeidahong' under ND conditions.

Table 2 Homologous sequences of the major structural genes involved in anthocyanin biosynthesis

Gene name	Number	Unigene ID
CHS	23	T01_cb10302_c44/f1p1/1479,T01_cb10302_c207/f1p0/1765,T01_cb10302_c207/f1p0/1765,T01_cb10302_c207/f1p0/1765,T01_cb10302_c207/f1p0/1765,T01_cb10302_c236/f1p5/1464,T01_cb10302_c236/f1p5/1464,T01_cb10302_c62/f1p0/1536,T01_cb9002_c2/f4p2/1441,T01_cb9002_c7/f1

		p1/1418,T01_cb9002_c9/f1p0/1345,T01_cb9002_c11/f1p0/1440,T01_cb10302_c43/f2p1/1465,T01_cb10302_c47/f1p3/1370,T01_cb10302_c64/f1p4/1453,T01_cb10302_c61/f1p0/1571,T01_cb10302_c63/f1p1/1420,T01_cb11086_c2/f1p0/1391,T01_cb10302_c232/f12p1/1382,T01_cb10302_c59/f1p2/1431,T01_cb14962_c0/f1p0/2020,T01_cb18732_c2/f1p0/531,T01_cb16039_c1/f1p0/1377
CHI	17	T01_cb5189_c1/f2p0/3307,T01_cb6001_c4/f1p0/1481,T01_cb5189_c4/f1p0/2355,T01_cb18734_c3/f1p0/768,T01_cb15047_c1/f1p0/927,T01_cb15047_c6/f1p0/815,T01_cb15047_c3/f1p0/855,T01_cb15047_c5/f1p0/907,T01_cb15047_c4/f1p0/943,T01_cb5189_c0/f9p0/979,T01_cb15047_c2/f1p0/874,T01_cb5189_c2/f2p0/933,T01_cb6001_c0/f1p0/1790,T01_cb5189_c5/f1p0/1724,T01_cb18734_c2/f5p0/759,T01_cb5508_c22988/f1p0/3343,T01_cb5189_c7/f1p0/980,T01_cb18354_c1/f1p0/1022
F3H	25	T01_cb9752_c4098/f1p0/1275,T01_cb5961_c21/f1p0/1458,T01_cb18379_c1/f1p0/1201,T01_cb11471_c17/f10p1/1300,T01_cb5961_c25/f3p2/1376,T01_cb11303_c6/f1p0/1689,T01_cb5961_c11/f1p4/1357,T01_cb9752_c44692/f2p0/1512,T01_cb11303_c0/f2p0/1337,T01_cb5961_c19/f1p0/2146,T01_cb5961_c26/f5p2/1270,T01_cb19426_c37265/f1p0/4299,T01_cb5705_c16/f1p0/1741,T01_cb5961_c10/f1p2/1470,T01_cb5961_c16/f1p1/1020,T01_cb11471_c16/f3p1/1330,T01_cb5508_c5156/f1p0/3331,T01_cb13983_c2/f2p0/1301,T01_cb5961_c20/f1p0/1436,T01_cb13983_c5/f1p0/1225,T01_cb11471_c10/f1p1/910,T01_cb5961_c8/f1p1/3157,T01_cb13983_c3/f2p0/1257,T01_cb16192_c5/f1p2/1309, T01_cb5961_c12/f1p0/2014,T01_cb11471_c14/f1p0/367
F3'H	28	T01_cb6480_c1/f1p0/3486,T01_cb5657_c21/f1p1/1646,T01_cb7984_c4/f1p1/1840,T01_cb16591_c3/f1p0/1859,T01_cb5657_c5/f3p3/1722,T01_cb16591_c9/f1p1/1786,T01_cb19426_c37199/f2p0/3558, T01_cb8246_c1/f2p0/1739,T01_cb16591_c2/f1p3/1818,T01_cb5657_c13/f1p1/1672,T01_cb5657_c9/f1p3/1778,T01_cb7984_c2/f1p0/1811,T01_cb5325_c0/f2p0/3269,T01_cb7984_c1/f1p0/1851,T01_cb7984_c5/f1p0/1619,T01_cb5657_c23/f13p6/1733,T01_cb5657_c7/f3p4/1708,T01_cb16591_c7/f1p1/1897,T01_cb12342_c4/f1p1/1824,T01_cb12342_c6/f1p0/1807,T01_cb5657_c0/f38p6/1704,T01_cb5657_c24/f1p3/1716,T01_cb12342_c1/f2p0/1724,T01_cb5508_c14914/f1p0/4979,T01_cb6480_c0/f1p0/2264,T01_cb5657_c2/f1p3/1667,T01_cb5657_c10/f1p1/3227,T01_cb16591_c6/f1p3/1790
F3'5'H	24	T01_cb8246_c16/f1p0/1936,T01_cb7984_c4/f1p1/1840,T01_cb8246_c4/f1p0/1724,T01_cb14652_c8/f1p0/1751,T01_cb8246_c12/f1p0/1551,T01_cb8246_c1/f2p0/1739,T01_cb9430_c1/f2p0/1694, T01_cb16591_c2/f1p3/1818,T01_cb14652_c1/f4p1/1668,T01_cb9517_c4/f2p1/1748,T01_cb8246_c9/f1p0/1685,T01_cb8673_c2/f2p0/1849,T01_cb7984_c2/f1p0/1811,T01_cb7984_c1/f1p0/1851,T01_cb9517_c17/f1p2/1233,T01_cb14652_c13/f6p2/1801,T01_cb8246_c2/f2p0/1751,T01_cb15969_c1/f1p0/1758,T01_cb14652_c6/f1p3/1817,T01_cb8246_c6/f1p0/1040,T01_cb8246_c5/f1p0/1812,T01_cb8246_c3/f2p0/1722,T01_cb8246_c0/f2p1/1952,T01_cb14665_c0/f2p0/1751
DFR	13	T01_cb11761_c1/f12p1/1292,T01_cb11761_c12/f1p1/1642,T01_cb11761_c8/f1p0/1466,T01_cb11761_c17/f1p1/1747,T01_cb11761_c16/f1p0/1350,T01_cb11761_c19/f1p1/1568,T01_cb11761_c7/f1p1/1647,T01_cb11761_c24/f1p1/1578,T01_cb11761_c5/f3p0/1415,T01_cb15530_c9/f1p0/1582,T01_cb15530_c8/f1p0/1700,T01_cb11761_c21/f3p1/1591,T01_cb11761_c20/f1p1/1643
ANS	15	T01_cb10429_c7/f1p0/1275,T01_cb11303_c1/f2p0/1207,T01_cb11303_c5/f1p0/1326,T01_cb11471_c7/f1p0/1005,T01_cb11471_c5/f1p1/993,T01_cb10429_c7/f1p0/1275,T01_cb10429_c9/f1p0/1346,T01_cb11303_c8/f1p0/1171,T01_cb11303_c3/f1p0/920,T01_cb11303_c7/f1p0/1483,T01_cb13983_c4/f1p0/1172,T01_cb17823_c0/f2p0/1175,T01_cb10429_c10/f39p4/1251,T01_cb10429_c8/f1p0/1326,T01_cb10429_c10/f39p4/1251
3GT	31	T01_cb5533_c47/f2p2/1365,T01_cb15075_c0/f3p0/1655,T01_cb15075_c4/f1p0/1652,T01_cb12322_c0/f5p1/1583,T01_cb8439_c27/f10p4/1748,T01_cb13952_c80/f1p1/1059,T01_cb8439_c28/f13p5/1745,T01_cb5533_c41/f4p2/1497,T01_cb5533_c137/f1p2/1618,T01_cb5533_c131/f1p1/1624,T01_cb15075_c3/f1p0/1959,T01_cb9752_c49692/f2p1/1578,T01_cb13952_c86/f13p1/1708,T01_cb5533_c47/f2p2/1365,T01_cb13259_c0/f7p0/1634,T01_cb9261_c4/f1p0/2049,T01_cb5533_c30/f12p3/1826,T01_cb15904_c1/f1p1/1724,T01_cb10756_c11/f1p0/1076,T01_cb5533_c170/f1p1/1734,T01_cb8439_c24/f1p0/2273,T01_cb16924_c0/f3p0/1712,T01_cb12322_c1/f2p0/1695,T01_cb13259_c2/f1p0/1644,T01_cb13952_c85/f43p4/1731,T01_cb5533_c125/f1p2/1772,T01_cb5533_c165/f6p4/1820,T01_cb8439_c16/f1p0/1832,T01_cb5533_c121/f1p2/1772,T01_cb5533_c140/f8p0/1648,T01_cb13952_c14/f4p1/1788,T01_cb13259_c4/f1p0/1689

An expression heatmap was constructed based on the expression of the identified DEGs of the anthocyanin pathway (Fig. 9). Sixty-two unigenes encoding eight enzymes showed large changes during flower development, including thirteen *CHS*, nine *CHI*, six *F3H*, seven *F3'H*, six *F3'5'H*, six *DFR*, one *ANS* and fourteen *3GT* genes.

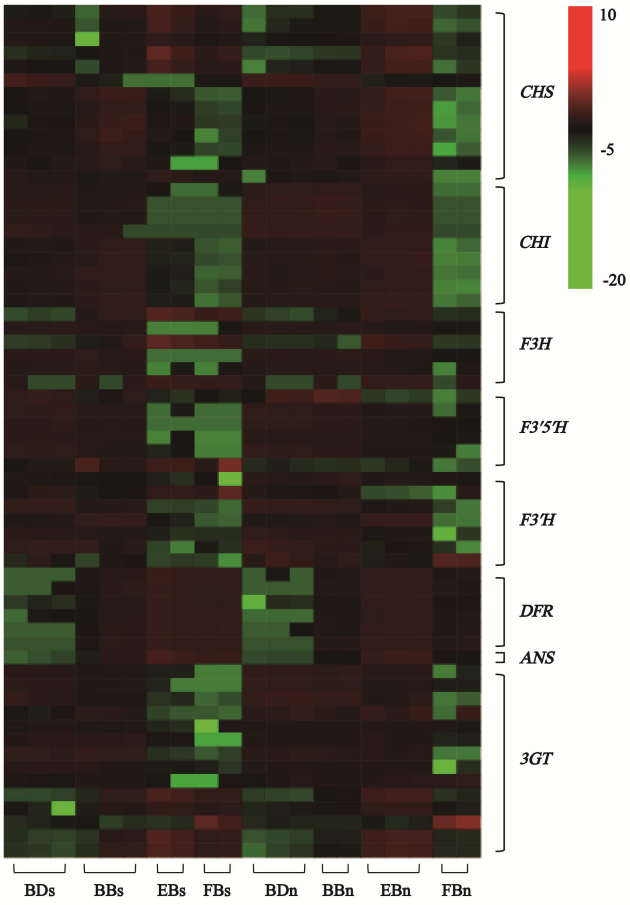
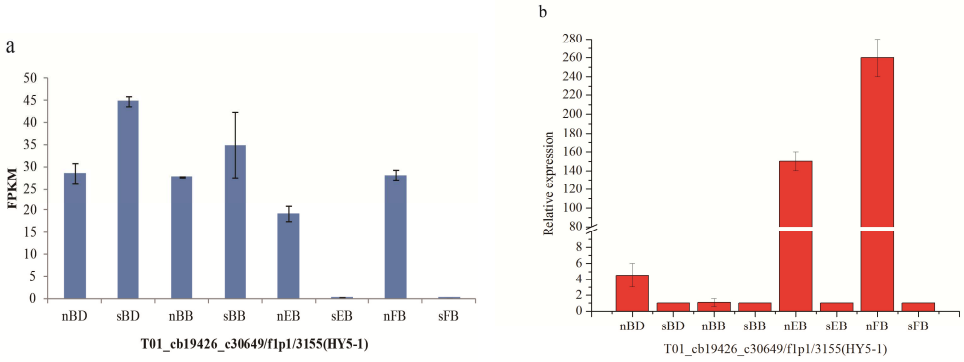
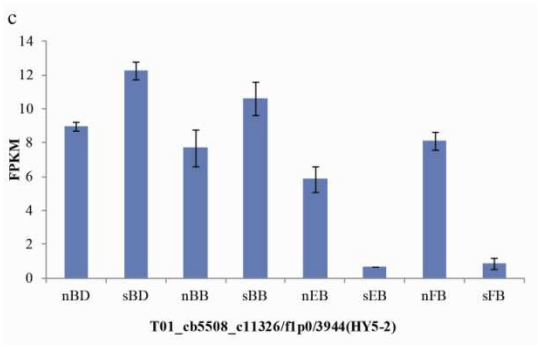


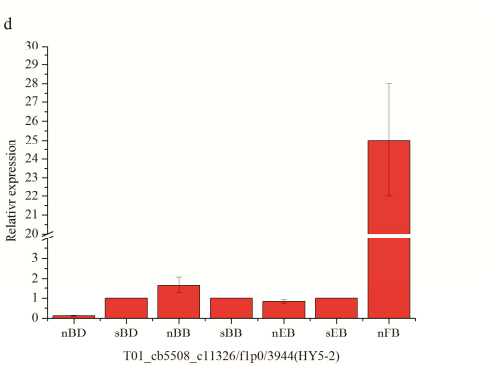
Fig. 9 Heatmap of the anthocyanin structural gene expression levels in the BD, BB, EB and FB stages. The colour scale indicates a relative fold-change, with red indicates high expression and green low expression.

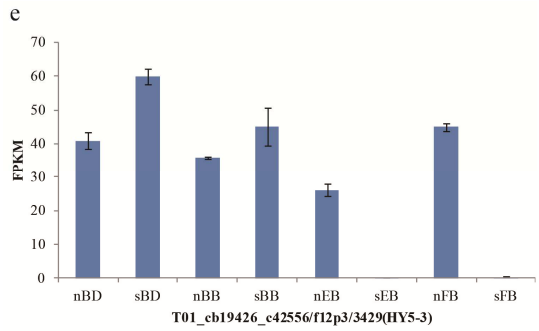
Validation of gene expression related to anthocyanin biosynthesis

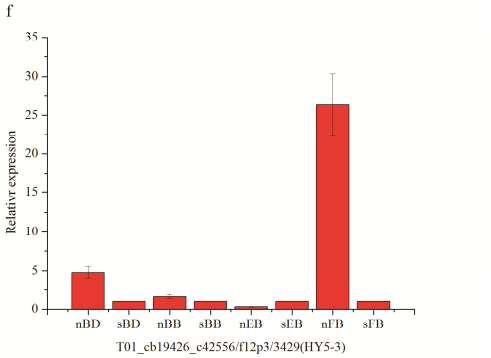
Among the sixty-two unigenes, six unigenes participating in the anthocyanin biosynthesis pathway were always significantly upregulated in the four stages under ND conditions compared to the expression levels under SD conditions. To confirm the accuracy and reproducibility of the Illumina RNA-Seq results, quantitative real-time PCR (qRT-PCR) was employed (Wang et al., 2017; Li et al., 2019). As shown in Fig. 10, the patterns of anthocyanin accumulation were consistent with the expression patterns of *HY5-1*, *HY5-2*, *HY5-3*, *HY5-4*, *CHI1* and *3GT1*. In contrast, *F3H* showed opposite patterns with respect to anthocyanin accumulation.

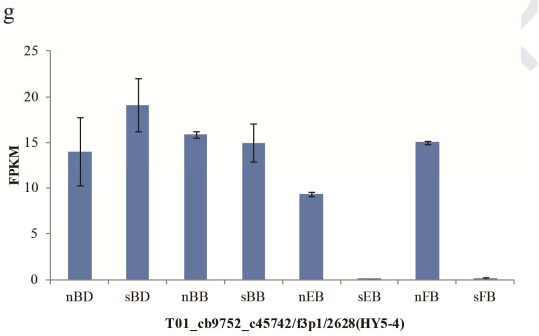


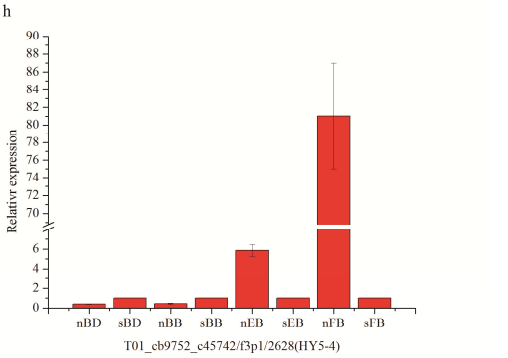

c
FPKM
nBD sBD nBB sBB nEB sEB nFB sFB
T01_cb5508_c11326/f1p0/3944(HY5-2)

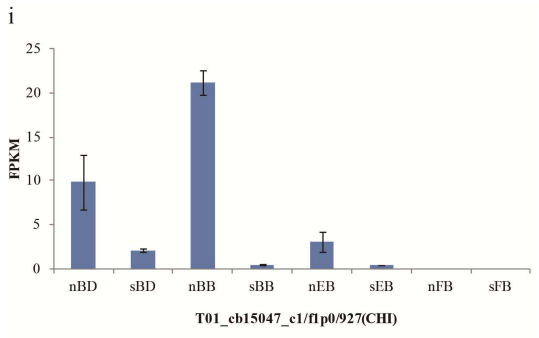

d
Relativr expression
nBD sBD nBB sBB nEB sEB nFB
T01_cb5508_c11326/f1p0/3944(HY5-2)

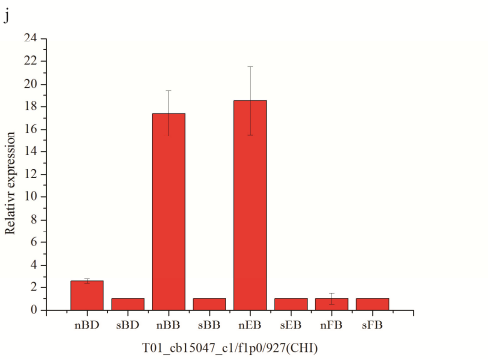

e
FPKM
nBD sBD nBB sBB nEB sEB nFB sFB
T01_cb19426_c42556/f12p3/3429(HY5-3)


f
Relativr expression
nBD sBD nBB sBB nEB sEB nFB sFB
T01_cb19426_c42556/f12p3/3429(HY5-3)


g
FPKM
nBD sBD nBB sBB nEB sEB nFB sFB
T01_cb9752_c45742/f3p1/2628(HY5-4)


h
Relativr expression
nBD sBD nBB sBB nEB sEB nFB sFB
T01_cb9752_c45742/f3p1/2628(HY5-4)


i
FPKM
nBD sBD nBB sBB nEB sEB nFB sFB
T01_cb15047_c1/f1p0/927(CHI)


j
Relativr expression
nBD sBD nBB sBB nEB sEB nFB sFB
T01_cb15047_c1/f1p0/927(CHI)

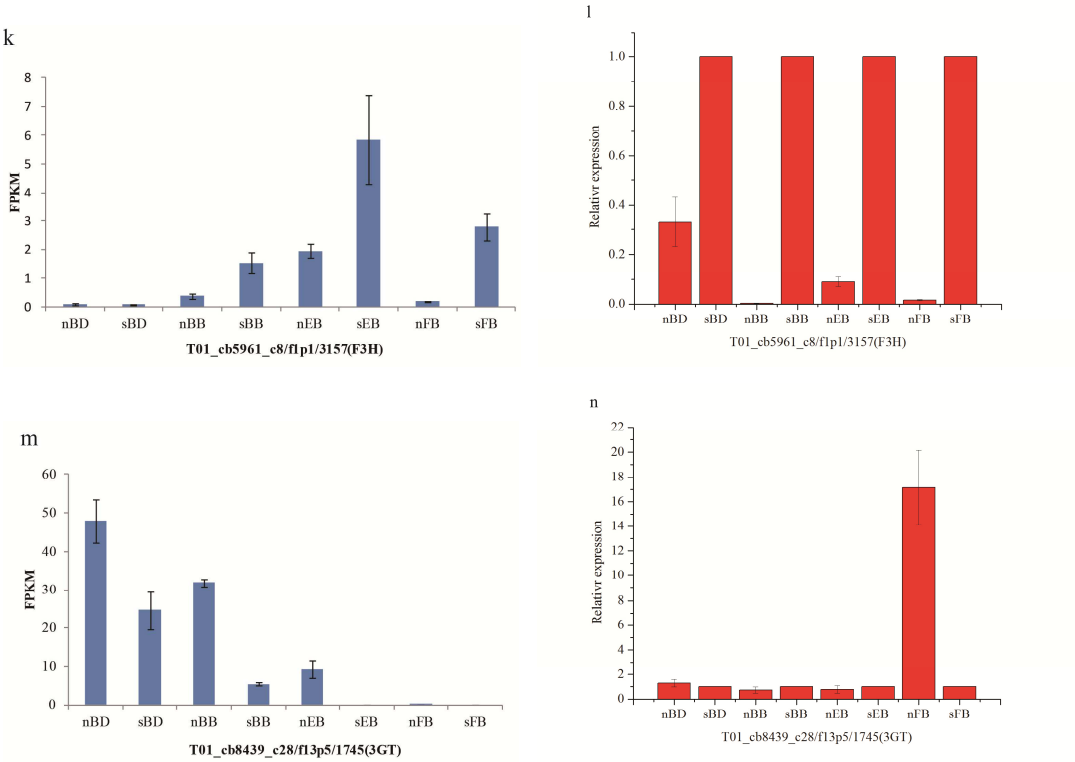


Fig. 10 The expression levels and patterns of *HY5-1* (a, b), *HY5-2* (c, d), *HY5-3* (e, f), *HY5-4* (g, h), *CHI1* (I, j), *F3H* (k, l) and *3GT1* (m, n) were confirmed by RNA-Seq technology and qRT-PCR during the BD stage, BB stage, EB stage and FB stage of *C. morifolium* 'Jinbeidahong'. Each column represents the mean of three independent measurements. Error bars represent the SD of the mean values.

Discussion

In production, artificial SD treatment is often used to promote early flowering of *chrysanthemum*s. However, we found that the flower colour of C. *morifolium* 'Jinbeidahong' blossoms induced by artificial short-day treatment was lighter than those induced by the ND photoperiod.

For the full-length transcriptome sequencing of *Chrysanthemum morifolium* cv. 'Jinbeidahong', approximately 60 GB of high-quality data were obtained, and 415,387 FLNC reads with an average length of 2786 bp were obtained. Finally, a total of 89,477 high-quality non-redundant FLNC reads were obtained. Based on the COG, GO, KEGG, KOG, Pfam, Swiss-Prot, eggNOG and NR databases, 84,850 unigenes were predicted for a specific or general function, corresponding approximately to all of the unigenes. This number of unigenes was higher than those found for C. *morifolium* Ramat. cv 'Chuju', C. *morifolium* 'Fenditan', C. *morifolium* 'Yuuka' and C. *lavandulifolium*, which were analyzed by the identical transcriptome strategy (Wang et al., 2014; Liu et al., 2015; Ren et al., 2016; Yue et al., 2018). For the DGE profiling of each flower sample, approximately 191.04 GB of high-quality data were obtained, and the average Q30 was greater than 90%. In this study, we characterized the transcriptomes of the traditional cultivar C. *morifolium* 'Jinbeidahong' using SMRT transcript sequencing. Compared to experimental assays for transcript tags assembled from short RNA-Seq reads, full-length transcripts can greatly improve the accuracy of genome annotation and transcriptome features (Dong et al., 2015). GSFLX 454 sequencing was used to read expressed

sequence tags (ESTs) of C. *morifolium* 'Jinbeidahong'. The average sequence lengths of these four cDNA libraries were 358 bp, 585 bp, 457 bp and 470 bp, respectively. The 3,243,586 reads were assembled after trimming and filtering processes, and only 39,221 contigs were finally obtained (Sasaki et al., 2017). Illumina paired-end sequencing technology was used to sequence transcriptomes of C. *morifolium* 'Jinbeidahong'. The 91,367 unigenes with an average length of 739 bp were assembled from 15.4 GB of reads (Liu et al., 2015). The sequences obtained do not fully cover the entire structure of each retroelement and transposon, including terminal repeat sequences and the protein-coding genes needed for transposition. Furthermore, the transcript tags derived from RNA-Seq may exhibit misassembly of the reads transcribed from highly repetitive regions or very similar members of multigene families in polyploid plants of *Chrysanthemum* that often harbour a large number of nearly identical homoeologous gene sets. The 76,428 raw reads were retained as singletons and further preprocessed into 41,368 clean reads for C. boreale, and the 97,947 raw reads were preprocessed into 57,035 clean reads for C. *morifolium*. The average length was 480 bp by GSFLX 454 sequencing (Won et al., 2017).

Our result is 2.5 times that for C. *morifolium*, which is also a hexaploid species. SMRT sequencing errors were random, so the subreads from the same ZMW hole could correct each other to obtain circular consensus sequences (CCSs). The minimum accuracy of the CCSs was 90%. On the other hand, the FLNC reads were clustered and corrected to generate non-redundant cluster consensus sequences following the ICE correction algorithm. The minimum accuracy of polished isoforms was 99%. This can reduce gene integration.

Subsequently, DEG profiling was performed in the flowers from the BD, BB, EB and FB stages grown under ND and SD conditions. Due to lack of *Chrysanthemum* genomic information, the RNA-Seq sequencing data were aligned against the unigenes assembled from our previous transcriptome analyses. As mentioned above, DETs were identified between the samples under ND conditions and those under SD conditions. The comparative results revealed that most DEGs (2,987 upregulated and 3,821 downregulated) were present in the EB stage. The over-presentation genes in the KEGG pathways were associated with taurine and hypotaurine metabolism in the BD stage, thiamine metabolism in the BB stage, the photosynthesis antenna in the EB stage and anthocyanin biosynthesis in the FB stage between the samples under ND conditions and those under SD conditions. Undoubtedly, the over-presentation genes might indicate their significant regulatory function during *Chrysanthemum* development.

The anthocyanin biosynthetic pathway in *Chrysanthemum* ray petals has been determined (Ohmiya et al., 2018). The steps catalysed by *CHS*, *CHI*, *F3H*, *F3'H*, *F3'5'H*, *DFR*, *ANS* and *3GT* lead to the production of different anthocyanin subgroups by modifying the molecular skeleton and/or backbone (Yue et al., 2018). In this study, twenty-three *CHS*, seventeen *CHI*, twenty-five *F3H*, twenty-eight *F3'H*, twenty-four *F3'5'H*, thirteen *DFR*, fifteen *ANS* and thirty-one *3GT* homologous unigenes were identified (Table 3). The results of DEG profiling show that 62 DEGs were predicted to be key structural genes involved in anthocyanin biosynthesis. Six of them were significantly upregulated during the four stages under ND conditions (Fig. 10). We suggest that a dramatic increase of transcripts encoding *HY5-1*, *HY5-2*, *HY5-3*, *HY5-4*, *CHI1* and *3GT1* probable provide more anthocyanin

biosynthesis substrate, which is mainly responsible for higher anthocyanin accumulation in flowers under ND conditions than SD conditions. In orange-flowered gentian (Gentiana lutea L. var. aurantiaca), the petals accumulated higher relative levels of *CHI* at stage S3 than the lutea petals, while *3GT* had higher relative levels at both stages (S3 and S5) (Berman et al., 2016). Silencing of the endogenous *CHI* was highly effective in reducing the chi gene and anthocyanin accumulation in floral tissues of Petunia (Keykha et al., 2016).

Anthocyanin accumulation is light-dependent in most plants. *HY5* is the major regulator of light signalling in plants. *HY5* positively regulates anthocyanin biosynthesis by activating the transcription of the anthocyanin biosynthetic genes (Shin et al., 2007). In this study, the anthocyanin accumulation of C. *morifolium* 'Jinbeidahong' blossoms in different stages induced by SD treatment were all lower than those induced by ND treatment. In addition to the transcripts encoding *HY5-1*, *HY5-2*, *HY5-3*, *HY5-4*, *CHI1* and *3GT1*, in terms of transcript levels, *SmHY5*, which binds the promoters of *SmCHS* and *SmDFR*, was upregulated by light in eggplant (Jiang et al., 2016). In tomato, SlHY5 directly recognized and bound to the G-box and ACGT-containing element in the promoters of anthocyanin biosynthesis genes (Liu et al., 2018). In pear, PpHY5 directly bound to the promoters of the anthocyanin biosynthesis genes *PpCHS*, *PpDFR*, *PpANS* and *PpMYB10* (Tao et al., 2018). In particular, *HY5-3* and *HY5-4* were two different variable splicing transcripts in C. *morifolium* 'Jinbeidahong'. Alternate splicing is an important process to increase genetic and functional diversity in an organism (Treutlein et al., 2014). We propose that a sharp increase in the variable splicing transcripts of *HY5-4* might be foremost responsible for the higher anthocyanin accumulation of C. *morifolium* 'Jinbeidahong' blossoms in different stages induced by SD treatment than those induced by ND treatment.

Temperature is crucial for the accumulation of anthocyanins. In this study, the anthocyanin contents of the flowers from the EB and FB stages under ND conditions were much higher than those under SD conditions. One of the reasons for this finding is that the temperature in the late stage of the ND conditions was lower than that in the late stage of the SD conditions. Low temperature promoted anthocyanin synthesis, and high temperature inhibited accumulation of the same (Dar et al., 2019). *3GT1* was significantly upregulated in the EB and FB stages under ND conditions compared to the expression level under SD conditions. The expression of the *AcCHIL* gene in pink-red petals of A. eriantha was markedly increased by low temperature but decreased by heat stress (Yang et al., 2017). MdbHLH3 activates *MdDFR* and *MdUFGT* expression in response to low temperature, which resulting in anthocyanin accumulation in apples (Xie et al., 2012). Therefore, it is likely that overexpression of *CHI1* and *3GT1* of C. *morifolium* 'Jinbeidahong' could increase the anthocyanin content. Moreover, to advance florescence and keep petal colour unchanged, the genetic transformation method that employs an Agrobacterium-mediated overexpression of *CHI1* and *3GT1* should be used in *Chrysanthemum* production.

Furthermore, we observed that anthocyanin biosynthesis is also regulated at the epigenetic level in plants. In the present study, DNA methyltransferases were significantly upregulated in the four stages under ND conditions compared to expression levels under SD conditions. In our previous study, it was found that DNA methylation in the SD treatment group was significantly lower than that in the control group (Li et al., 2016). The DNA methyltransferase inhibitor 5-aza-2-deoxycytidine (5-aza-dC)

increased the anthocyanin content in bagged fruits (Ma et al., 2018). H3K9 methylation, H3K27 methylation, and DNA methylation are generally thought to be indicators of condensed heterochromatin (He et al., 2014). The histone H3K9 demethylase gene *JMJ25JMJ25* directly affects MYB182 expression by altering the histone methylation status of its chromatin and causing DNA methylation, resulting in repression of anthocyanin accumulation (Fan et al., 2018). *ANS* and *F3H* exhibited upregulated expression in apple mutants, and differences were observed in the methylation patterns of their promoters (Jiang et al., 2019).

To sum up, the single-molecule real-time sequencing and RNA-Seq analysis of C. *morifolium* 'Jinbeidahong' under ND and SD conditions revealed that the anthocyanin biosynthesis is tightly regulated by the photoperiod. The *HY5-1*, *HY5-2*, *HY5-3*, *HY5-4*, *CHI1* and *3GT1* genes are significant in anthocyanin biosynthesis accumulation, which could provide a basis for molecular breeding. With this perspective, we are currently generating *HY5-1*, *HY5-2*, *HY5-3*, *HY5-4*, *CHI1* and *3GT1* overexpression and knock-out mutant lines of *Chrysanthemum* to examine gene expression and function. It is likely that the use of transgenic breeding to overexpress *HY5-1*, *HY5-2*, *HY5-3*, *HY5-4*, *CHI1* and *3GT1* in *Chrysanthemum* could improve the flower colour induced by artificial SD treatment in production.

Conclusions

The single-molecule real-time sequencing and RNA-Seq analysis of C. *morifolium* 'Jinbeidahong' under ND and SD conditions revealed that the anthocyanin biosynthesis is tightly regulated by the photoperiod. The *HY5-1*, *HY5-2*, *HY5-3*, *HY5-4*, *CHI1* and *3GT1* genes are significant in anthocyanin biosynthesis accumulation, which could provide a basis for molecular breeding. The high-quality sequencing data provide valuable reference genes for follow-up studies of *Chrysanthemum*.

Methods

Plant material

C. morifolium 'Jinbeidahong' is a Chinese traditional flower, and plants were preserved in our laboratory, Henan University, Plant Germplasm Resources and Genetic Engineering, for eleven years. Plants were grown under ND conditions in an experimental field of Henan University at Kaifeng, Henan Province, People's Republic of China (N34°82′, E114°30′). The cultivar has floral competence at the 10-leaf stage and a limited inductive photoperiod of 43 days under SD conditions. Leaf, petiole, stem, bud, flower, and root samples were collected from *C. morifolium* 'Jinbeidahong' plants once a week until the complete flower appeared. The samples were subsequently rapidly frozen in liquid nitrogen and kept at -80 °C for RNA extraction and pigment measurement.

RNA extraction and anthocyanin quantification

Total RNA samples were isolated from leaf, petiole, stem, bud, flower, and root samples using a commercial kit (Takara Biotechnology, Dalian, China). The integrity of the RNA was determined with an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, California). Only the total RNA samples with RIN values ≥ 8 were used for constructing the cDNA libraries with PacBio sequencing.

Flower petals were collected from budding (BD), bud breaking (BB), early blooming (EB) and full blooming (FB), four sequential developmental stages. Due to the requirements of biological replicates, the flowers petals from different three plants were sampled. Total RNA samples were isolated from

flower samples using a commercial kit (Takara Biotechnology, Dalian, China). The integrity of the RNA was determined with an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, California). The total RNA samples were used for RNA-Seq quantification analysis. The anthocyanin content was quantified using the method of Fuleki (1988).

PacBio single-molecule real-time sequencing, alternative splicing, lncRNA analysis, and functional annotation

Total RNA about 1 μg was reverse transcribed into cDNA using the SMARTer PCR cDNA Synthesis Kit, which has been optimized for preparing high-quality, full-length cDNAs (Takara Biotechnology, Dalian, China), followed by size fractionation (1-2 k, 2-3 k and 3-6 k) using the BluePippin™ Size Selection System (Sage Science, Beverly, MA). Each SMRT bell library was constructed using 1.5 μg of size-selected cDNA with the Pacific Biosciences DNA Template Prep Kit 2.0. The binding of SMRT bell templates to polymerases was conducted using the DNA/Polymerase Binding Kit P6 and v2 primers. Sequencing was carried out on the Pacific Bioscience RS II platform using C4 reagents with 240 min movies.

Raw reads were processed into error-corrected reads of insert (ROIs) using the Iso-seq pipeline with minFullPass=0 and minPredictedAccuracy=0.75. Next, full-length, non-chimeric (FLNC) transcripts were determined by searching for the polyA tail signal and the 5’ and 3’ cDNA primers in ROIs. ICE (Iterative Clustering for Error Correction) was used to obtain consensus isoforms, and FL consensus sequences from ICE were polished using Quiver. High-quality FL transcripts were classified with the criterion of post-correction accuracy above 99%. Iso-Seq high-quality FL transcripts were used for removing redundancy with cd-hit (identity > 0.99).

We used Iso-SeqTM data directly to run all-vs-all BLAST searches with high identity settings. BLAST alignments that met all criteria were considered products of candidate AS events: there should be two HSPs (High-scoring Segment Pairs) in the alignment. Two HSPs have the same forward/reverse direction, within the same alignment, and one sequence should be continuous or have a small "Overlap" size (smaller than 5 bp); the other sequence should be distinct and show an "AS Gap", and the continuous sequence should nearly completely align to the distinct sequence. The AS gap should be larger than 100 bp and at least 100 bp away from the 3'/5' end.

Four computational approaches include CPC/CNCI/CPAT/Pfam/ were combined to sort non-protein coding RNA candidates from putative protein-coding RNAs in the transcripts. Putative protein-coding RNAs were filtered out using a minimum length and exon number threshold. Transcripts with lengths more than 200 nt and have more than two exons were selected as lncRNA candidates and further screened using CPC/CNCI/CPAT/Pfam that have the power to distinguish the protein-coding genes from the non-coding genes.

Functional annotation was performed by aligning the FLNC reads to selected databases and methods as follows. The unigenes were aligned to the NR, KOG, and KO databases using NCBI-BLAST 2.2.29+ (Koonin et al., 2004; Deng et al., 2006). Based on the NR protein annotation results, Gene Ontology (GO) enrichment analysis of the differentially expressed genes (DEGs) was implemented by the GOseq R package-based Wallenius non-central hyper-geometric distribution, which can adjust for gene length bias in DEGs (Ashburner et al., 2000). KEGG (Kanehisa et al., 2004) is a database

resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (http://www.genome.jp/kegg/).

RNA-Seq and quantification of gene expression levels

The expression levels of the unigenes in the eight samples were calculated by RNA-Seq quantification analysis. Raw data (raw reads) of fastq format were first processed through in-house Perl scripts. In this step, clean data (clean reads) were obtained by removing reads containing adaptors, reads containing ploy-N and low-quality reads from raw data. At the same time, the Q20, Q30, GC content and sequence duplication level of the clean data were calculated. All downstream analyses were based on clean data with high quality. The transcriptomes of the samples were used as a reference to screen the clean reads. Only reads with a perfect match or one mismatch were further analysed and annotated based on the reference transcriptome. Tophat2 tools were used to map the reference genome.

Quantification of gene expression levels was estimated by the fragments per kilobase of transcript per million fragments mapped. For the samples with biological replicates, prior to differential gene expression analysis, for each sequenced library, the read counts were adjusted by the edgeR program package through one scaling normalized factor. Differential expression analysis of two samples was performed using the EBSeq R package. The resulting FDR (false discovery rate) was adjusted using the PPDE (posterior probability of being DE). The FDR < 0.05 and |log2 (fold-change) | ≥1 were set as the threshold for significant differential expression.

Verification by qRT-PCR analysis

RNA was purified, and first-strand cDNA was synthesized as described above (Dong et al., 2018). The resulting cDNA samples were analysed by qRT-PCR in a 25 μl reaction volume containing 12.5 μl SYBR Premix Ex Taq II (TaKaRa), which was performed on a LightCycler 480II real-time PCR detection system (Roche). For RT-PCR, the specific primers were designed according to the gene sequences of C. *morifolium* 'Jinbeidahong'. The real-time PCR primers were listed in Table 3. For normalizing gene expression data, the *Actin* gene was included as a control (Fu et al., 2013).

Table 3 Primers used for Real time PCR

Gene name	Forward primers（5'to 3'）	Reverse primers（5'to 3'）
HY5-1	AAAGGTTTTCGGCGTGTTTCGGCTC	TGATTTGCACTGCGGTCTTCGTTCC
HY5-2	GTTGGTGCGAGGGAGTTGGTGATG	TTGGTTTCCTTTTAGGCCGCGGAG
HY5-3	GGTGTTCAATAGAGGCTAAAGAAAGCGAGA	CTTGGCTACATGATTTTGGATGGAAGATAC
HY5-4	GGACGTATAGCTTTTCAAGCACTCTTCTCAAG	CGACAAAAACAGGTAAAACATATAGCACCCAC
CHI	GAATCCGTCTTCGTTCCACCGTCCTT	CCACATCCCAACTATCTTTTCTGCCG
F3H	TAGCCAGAGGTGGAGATAAAGGGGGT	TAGCATCAAGAGCGGAGCCAGAACAT
3GT	CCACCTTTACTCCCCATCGGACCAT	CCCTATAGCAGCCGCTTCCATCACC
Actin	ACAACTGCTGAACGGGAAAT	TCATAGACGGCTGGAAAAGG

List of abbreviations

SMRT: single-molecule real-time; CCSs: circular consensus sequences; FLNC: full-length non-chimeric; kb: kilobase; GO: Gene Ontology; NR: NCBI non-redundant; KOG: euKaryotic

Orthologous Groups; KEGG: Kyoto Encyclopedia of Genes and Genomes; ICE: iterative clustering error; LTR: long terminal repeat.

Declarations

Ethics approval and consent to participate

The experiments in this study comply with the current laws of China.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

Full-length transcriptome sequencing was supported by the National Natural Science Foundation of China (2018YFD1000403 and 31372090). The RNA-Seq was supported by the National Natural Science Foundation of China (2018YFD1000403 and 31801882). Real-time PCR was supported by the National Natural Science Foundation of China (31801882) and a Postdoctoral Research Grant in Henan Province (001701031).

Availability of data and materials

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (Wang et al., 2017) in the BIG Data Center (BIG Data Center Members, 2018), Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under accession number CRA 001531, which is publicly accessible at http://bigd.big.ac.cn/gsa.

Authors' contributions

WD and ML contributed equally to this work. WD, ML and ZW designed the experiments and wrote the manuscript. ML, ZL, SL, YZ and XD grew the traditional *Chrysanthemum* cultivar 'Jinbeidahong'. WD, ML and ZW performed the full-length transcriptome analysis. All authors have read and approved the manuscript.

Acknowledgements

Not applicable.

Reference

Ashburner M, Ball C A, Blake J A, et al. Gene ontology: tool for the unification of biology. *Nat Genet*. 2000, 25(1): 25-29.

Berman J, Sheng Y, Gómez Gómez L, et al. Red Anthocyanins and Yellow Carotenoids Form the Color of Orange-Flower Gentian (Gentiana lutea L. var. aurantiaca). *PLoS One*. 2016, 11(9):e0162410.

BIG Data Center Members. Database Resources of the BIG Data Center in 2019. Nucleic Acids Res 2019, 47(D1):D8-D14.

Dara J A, Wania A A, Ahmedc M, et al. Peel colour in apple (Malus×domestica Borkh.): An economic quality parameter in fruit market. *Sci Hortic*. 2019, 244:50-60.

Deng YY, Li JQ, Wu S F, Zhu YP, et al. Integrated NR Database in Protein Annotation System and Its Localization. *Computer Engineering*. 2006, 32(5):71-74.

Dong L, Liu H, Zhang J, et al. Single-molecule real-time transcript sequencing facilitates common wheat genome annotation and grain transcriptome research. *BMC Genomics*. 2015, 16:1039.

Dong W, Wu D, Li G, et al. Next-generation sequencing from bulked segregant analysis identifies a dwarfism gene in watermelon. *Sci Rep*. 2018, 8(1):2908.

Fan D, Wang X, Tang X, et al. Histone H3K9 demethylase JMJ25 epigenetically modulates anthocyanin biosynthesis in poplar. *Plant J*. 2018, 96(6):1121-1136.

Fang G, Munera D, Friedman DI, et al. Genome-wide mapping of methylated adenine residues in pathogenic Escherichia coli using single-molecule real-time sequencing. *Nat Biotechnol*. 2012, 30(12): 10.

Fuleki F. Quantitative methods for anthocyanins. *J Food Sci*. 1988, 33:72-77.

Guo J, Han W, Wang MH. Ultraviolet and environmental stresses involved in the induction and regulation of anthocyanin biosynthesis: a review. *Afr J Biotechnol*. 2008, 7(25):4966-4972.

He S, Yan S, Wang, et al. Comparative analysis of genome-wide chromosomal histone modification patterns in maize cultivars and their wild relatives. *PLoS One*. 2014, 9(5):e97364.

Hong Y, Yang LW, Li ML, Dai SL. Comparative analyses of light-induced anthocyanin accumulation and gene expression between the ray florets and leaves in *chrysanthemum*. *Plant Physiol Biochem*. 2016, 103:120-132.

Hong Y, Tang X, Huang H, Zhang Y, Dai S. Transcriptomic analyses reveal species-specific light-induced anthocyanin biosynthesis in *chrysanthemum*. *BMC Genomics*. 2015, 16: 202.

Jiang M, Ren L, Lian H. Novel insight into the mechanism underlying light-controlled anthocyanin accumulation in eggplant (Solanum melongena L.) *Plant science*. 2016, 249:46-58.

Kanehisa M, Goto S, Kawashima S, et al. The KEGG resource for deciphering the genome. *Nucleic Acids Res*. 2004, 32:277-280.

Keykha F, Bagheri A, Moshtaghi N, Bahrami AR, Sharifi A. RNAi-induced silencing in floral tissues of Petunia hybrida by agroinfiltration: a rapid assay for chalcone isomerase gene function analysis, *Cell Mol Biol*. 2016, 62(10):26-31.

Koonin EV, Fedorova ND, Jackson JD, et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome biol*. 2004, 5(2): R7.

Li P, Zhang F, Chen S, et al. Genetic diversity, population structure and association analysis in cut *chrysanthemum* (*Chrysanthemum morifolium* Ramat.). *Mol Genet Genomics*. 2016, 291(3):1117-1125.

Li S, Li M, Li Z, et al. Effects of the silencing of CmMET1 by RNA interference in chrysanthemum (Chrysanthemum morifolium). *Plant Biotechnol Rep*. 2019, 13(1):63-72.

Liu C, Chi C, Jin L. The bZip transcription factor HY5 mediates CRY1a-induced anthocyanin biosynthesis in tomato. *Plant cell environ*. 2018, 41(8):1762-1775.

Liu H, Sun M, Du D, et al. Whole-Transcriptome Analysis of Differentially Expressed Genes in the Vegetative Buds, Floral Buds and Buds of *Chrysanthemum morifolium*. *PLoS One*. 2015, 10, e0128009 .

Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. *PLoS Comput Biol*. 2017, 13, e1005457.

Lv SH, Xi L, Nie J, Wen C, Yuan CQ, Ma N, Zhao L. Isolation and expression analysis of auxin efflux transporter CmPIN1 in *Chrysanthemum* (*Chrysanthemum morifolium* 'Jinba'), *Journal of China Agricultural University*. 2016, 21(3):58-66.

Ma C, Liang B, Chang B, et al. Transcriptome Profiling Reveals Transcriptional Regulation by DNA Methyltransferase Inhibitor 5-Aza-20-Deoxycytidine Enhancing Red Pigmentation in Bagged "Granny Smith" Apples (Malus domestica). *Int J Mol Sci*. 2018, 19:3133.

Meng Q, Runkle ES. Regulation of flowering by green light depends on its photon flux density and involves cryptochromes. *Physiol Plant*. 2018. [Epub ahead of print]

Ohmiya. Molecular mechanisms underlying the diverse array of petal colors in *chrysanthemum* flowers. *Breeding Sci*. 2018, 68: 119-127.

Ren L, Liu T, Cheng Y, et al. Transcriptomic analysis of differentially expressed genes in the floral transition of the summer flowering *chrysanthemum*. *BMC Genomics.* 2016,17, 673.

Sasaki K, Mitsuda N, Nashima K, et al. Generation of expressed sequence tags for discovery of genes responsible for floral traits of *Chrysanthemum* morifolium by next-generation sequencing technology. *BMC Genomics*. 2017, 18(1):683.

Shin J, Park E, Choi G. PIF3 regulates anthocyanin biosynthesis in an HY5-dependent manner with both factors directly binding anthocyanin biosynthetic gene promoters in Arabidopsis. *Plant J*. 2007, 49, 981-994.

Tanaka Y, Sasaki N, Ohmiya A.Plant pigments for coloration. *Plant J*. 2008, 54: 733-749.

Tao, Ruiyan; Bai, Songling; Ni, Junbei, The blue light signal transduction pathway is involved in anthocyanin accumulation in 'Red Zaosu' pear . *Planta*. 2018, 248(1):37-48.

Treutlein B, Gokce O, Quake SR, Sudhof TC. Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. *Proc Natl Acad Sci U S A.* 2014,111(13):E1291-E12919.

Wang D, Yang C, Dong L, et al. Comparative transcriptome analyses of drought-resistant and -susceptible Brassica napus L. and development of EST-SSR markers by RNA-Seq. *J Plant Biol*. 2015, 58(4):259-269.

Wang P, Yang C, Chen H, et al. Transcriptomic basis for drought-resistance in Brassica napus L. *Sci Rep*. 2017, 7:40532.

Wang Y, Huang H, Ma YP, Fu J. Construction and *de novo* characterization of a transcriptome of *Chrysanthemum lavandulifolium*: analysis of gene expression patterns in floral bud emergence. *Plant Cell Tiss Org.* 2014, 116, 297-309.

Wang Y, Song F, Zhu J, et al. GSA: Genome Sequence Archive. Genomics Proteomics Bioinformatics. 2017 Feb; 15(1): 14-18.

Won SY, Kwon SJ, Lee TH, et al. Comparative transcriptome analysis reveals whole-genome duplications and gene selection patterns in cultivated and wild *Chrysanthemum* species. *Plant Mol Biol*. 2017, 95(4-5):451-461.

Xie XB, Li S, Zhang RF, et al. The bHLH transcription factor MdbHLH3 promotes anthocyanin accumulation and fruit colouration in response to low temperature in apples. *Plant Cell Environ*. 2012, 35: 1884-1897.

Yang LW, Wen XH, Fu JX, Dai SL. ClCRY2 facilitates floral transition in *Chrysanthemum* lavandulifolium by affecting the transcription of circadian clock-related genes under short-day photoperiods. *Hortic Res.* 2018, 5:58.

Yang M， Li J， Ye C, et al. Characterization and expression analysis of a chalcone isomerase-like gene in relation to petal color of Actinidia chrysantha. *Biologia*. 2017, 72(7): 753-763.

Yue J, Zhu C, Zhou Y, et al. Transcriptome analysis of differentially expressed unigenes involved in anthocyanin biosynthesis during flower development of Chrysanthemum morifolium 'Chuju'. *Sci Rep*. 2018, 8: 13414.

Highlights:

- The flower colour of *Chrysanthemum* blossoms induced by artificial short-day treatment was lighter than those induced by the natural photoperiod.
- The clustered sequences were used for functional annotation (GO, NR, KEGG, and KOG), identification and distribution analysis of simple sequence repeat loci, and prediction of coding sequences, transposons and retrotransposons by full-length transcriptome sequencing of *Chrysanthemum*.
- Transcriptome analysis of *C. morifolium* 'Jinbeidahong' revealed that the biosynthesis of anthocyanin is tightly regulated by the photoperiod.

Authors' contributions:

WD and ML contributed equally to this work. WD, ML and ZW designed the experiments and wrote the manuscript. ML, ZL, SL, YZ and XD grew the traditional *Chrysanthemum* cultivar 'Jinbeidahong'. WD, ML and ZW performed the full-length transcriptome analysis. All authors have read and approved the manuscript.

Declaration of interests

☑ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: