Transcriptome sequencing and comparative analysis for mining genes related to flower color variation in Iris laevigata Fisch.

Dan Wang
Northeast Forestry University
Song Yu
Northeast Forestry University
Juan Yang
Northeast Forestry University
Ling Wang (wanglinghlj@126.com)
Northeast Agricultural University

Research article

Keywords: Iris laevigata, Transcriptome, Anthocyanin, Flavonoid, Flower color

Posted Date: January 16th, 2020

DOI: https://doi.org/10.21203/rs.2.21022/v1

Transcriptome sequencing and comparative analysis for mining genes related to flower color variation in *Iris laevigata* Fisch.

Dan Wang[1,2], Song Yu[3], Juan Yang[1] and Ling Wang[1,*]

Abstract

Background: *Iris laevigata* Fisch. (Iridaceae) is an important aquatic plant with dark blue flowers. *I. laevigata* var. *alba* Ling Wang, L. Su & F. J. Shang is especially rare, it has white perianths with white anthers with a blue tinge. *I. laevigata* var. *alba* provides a good reference material for revealing complex and variations in anthocyanin catabolism networks. Modern molecular biology technology will facilitate the study of the mechanisms underlying metabolite variation. However, molecular research on *I. laevigata* is limited due to the lack of sequence data.

Results: In this study, RNA-Seq was performed on *I. laevigata* and *I. laevigata* var. *alba* at flowering stage. Two libraries were sequenced using Illumina Hiseq 2000 platform. Clean data of each sample reached 7.01 Gb. A total of 64,537 unigenes were obtained after assembly, including 28,936 unigenes annotated to seven public databases. Then, 143 unigenes were putative homologs to color-related genes, including 1 up-regulated and 12 down-regulated unigenes. Combined with reverse transcript polymerase chain reaction (RT-PCR), a number of important color-related genes were tested. In perianths, 5

*Correspondence: wanglinghlj@126.com

1 College of Landscape Architecture, Northeast Forestry University,Harbin,150040,China

2 Forestry Research Institute of Heilongjiang Province,Harbin,150081,China

3 State Key Laboratory of Tree Genetics and Breeding, School of Forestry, Northeast Forestry University, Harbin 150040, China

flavonoids (4 anthocyanins and 1 flavone) were detected using HPLC at flowering stage. Then, the putative anthocyanin metabolic process of flowering stage and differential genes related to flower color variation were put forward. A new hypothesis about the absence of phenotypic color of *I. laevigata* var. *alba* was proposed. It was suggested that the loss of anthocyanin was due to the interaction of multiple genes. First, upstream metabolic fluxes were redistributed by up-regulated CHI. Second, synergism of down-regulated genes (F3H/DFR/ANS) and competition for substrates between DFR and FLS resulted in relatively low biological flux and multi-shunt of metabolic pathways. As a result, the anthocyanins content in *I. laevigata* var. *alba* is very limited, which leads to the variation of flower color phenotype.

Conclusions: This study provides mass sequence data by the deep sequencing of *I. laevigata* and *I. laevigata* var. *alba* for the first time. Combined with the detection and metabolite analysis of flavonoids, it will increase our understanding of the molecular mechanism of the phenotypic variation in *I. laevigata* var. *alba* and provide the basis for molecular breeding of unique flower colors.

Keywords: *Iris laevigata*, Transcriptome, Anthocyanin, Flavonoid, Flower color

Background

I. laevigata Fisch. (Iridaceae) belongs to the largest genus comprising more than 260 herbaceous species growing in the temperate regions of the northern hemisphere [1]. Previous studies have revealed that this species is widely distributed in northeastern China and has a white flower variety named *I. laevigata* var. *alba* [2] (Fig.1), which usually grows in shallow waters, marshes, and ponds [3].

It is an important genetic resource to understand the genetic composition and the anthocyanins of the species. In the late 1980's *I. laevigata* . was studied as a maternal parent for hybridization with related species because its early and perpetual flowering characteristics. Anthocyanins of colchicine-induced amphidiploids of *I. laevigata* x *I. ensata* were analyzed [4]. The *I. laevigata* anthocyanins were classified as a petunidin 3RGac5G-malvidin 3RGac5G type, which is related to the reddish purple and purple perianth colors [4-5]. In nature, the *I. laevigata* perianths color has evolved via interaction with evolving pollinating insects. The extreme color variation ranges from pure white to an intense purple color. It is reasonable to expect that such loss-of-color adaptations are relatively unconstrained because they can be achieved by many different processes [6-7]. The numerous tributaries of anthocyanin metabolism and products make it more difficult and unpredictable to distinguish the genetic cause of this color change, and, therefore, there are too many uncertainties to reach a definite conclusion as to the cause. RNA-seq provides an efficient and inexpensive alternative to genome-wide sequencing. Relative and uncertain factors will be identified as long as there is a genetic source of highly expressed tissues and appropriate quality control screening is performed. By choosing an appropriate method, not only the comparison of transcript levels, but also the detection of key metabolites and the metabolic flows can be determined in different varieties. *I. laevigata* var. *alba* provides useful genetic material to understand the synthetic pathway and variation mechanism of anthocyanins in this species.

Anthocyanin biosynthesis via the pathway of flavonoid metabolism has been well studied [8-12]. Anthocyanins are the main chromogenic substances that produce blue, purple and chestnut flowers in Iridaceae [13], as carotenoid and betalains generally produce yellow or red colors [14]. In the field of floral genetic engineering, the metabolic branches of anthocyanins, Cy, Del and Pel, are known to differ in the hydroxylation pattern of the B-ring [15-18]. Although anthocyanins have various structures,

there are six major anthocyanidins (chromophores of anthocyanins); pelargonidin, cyanidin, peonidin (3' O-methyl cyanidin), delphinidin, petunidin (3' O-methy delphinidin) and malvidin (3',5' O-dimethy delphinidin) [19-20]. Among these, malvidin and petunidin were regarded as major anthocyanins in *I. laevigata* and *I. ensata* [21]. However, the means by which the flux is controlled in the branched flavonoid pathway has remained largely unknown for *I. laevigata and I. laevigata* var. *alba*. Changes in flux in the three different branches of ABP may be caused by one of the following reasons: loss of function, changes in gene expression of encoding branching enzymes, or alteration of the substrate specificity making the enzyme unable to metabolize the specific precursor [22]. Therefore, it is very important to screen the key genes that control the branching pathway. In addition, flavones and flavonols also affect pigmentation [23-25], which changes the color of the pigment. For example, studies have shown that the co-staining effect of anthocyanin-flavonoids contributes to the expression of the gray-purple color of *Iris* species [26]. The intermolecular co-pigmentation of flavonoids and anthocyanins isolated from the *Iris* of the Netherlands causes the petals to be bluish [27].

In this study, the RNA-Seq project for *I. laevigata and I. laevigata* var. *alba* was carried out for the first time using Illumina (Illumina, Biomarker Technologies Inc.) sequencing technology. Combined with HPLC analysis, the pathways and products related to the anthocyanin metabolism of *I. laevigata* were revealed. Attempts were made to screen out the major metabolic branches affecting the perianth pigmentation and candidate genes for ABP blockade. In addition, this study will create new breakthroughs for the targeted breeding of *Iris*.

Results

The main color composition of *I. laevigatas*

To clarify the metabolic pathways of flower color variation in *I. laevigatas*, we compared the compounds and contents of flavonoids (Table 1). As speculated, *I. laevigatas* contain three anthocyanin components responsible for color pigmentation: Del, Cy and Pel. In contrast, no color anthocyanins or derivatives were detected in *I. laevigatas* var. *alba*. This suggests that the co-coloration of these three pigments (Del, Cy and Pel) forms the color phenotype of blue flowers. The total amount of the three anthocyanins in blue flowers is 84 times greater than white flowers. This illustrates that the loss of anthocyanins and the decrease in anthocyanin content are the important factors affecting pigment formation. It is worth noting, that in blue flowers Cy content is the primary anthocyanin, having approximately 1.2 and 1.5 times greater content than Pel and Del, respectively.

Core genes in the ABP were studied in detail. The results demonstrated that most of the uni-transcripts had significant changes in expression level. Most genes showed lower transcript abundance in white perianths than in blue perianths, except CHI (Fig.3). So naringenin, a key product in blue and white perianths, was tested to understand the pigment metabolic pathway. Surprisingly, naringenin appeared in blue flowers and was not detected in white flowers (Fig.2).

RNA-sequence and assembly

In order to analyze the metabolism of anthocyanins in different colors, two libraries were constructed with blue and white perianths during the full bloom period for high-throughput sequencing (Fig. 1B, D). The Clean Data of each sample reached 7.01 Gb, and the Q30 base percentage was 91.70% and above. The GC content of white flower was 51.98%, and the GC content of blue flower was 52.23%. Through the above sequencing quality control, high quality Clean Data can be obtained and analyzed in-depth.

Subsequently, 108,768 transcripts and 64,537 unigenes were recombined, and the N50 of transcript and

unigene (the maximum length of the nucleotide sequence covering 50% of all genes) was 1,399 nt and 1,015 nt, respectively. In total, 17,395 (26.95%) unigenes are between 200 nt and 300 nt; 20,340 (31.52%) are between 300 nt and 500 nt;14,741 (22.84%) are between 500 nt and 1000 nt; 8,283 (12.83%) are between 1000 nt and 2000 nt.3,778 (5.85%) above 2000 nt (Table 2).

Function annotation

By comparing the two species unigenes with seven databases (NR, Swiss-Prot, KOG, COG, GO, KEGG, and Pfam), approximately 44.84% of unigenes obtained were functional annotations. Among them, 28,282 unigenes (97.73% of all annotated unigenes) could be annotated to the NCBI NR database, while 17,879 (61.79%), 17,784 (61.46%), 16,123 (55.72%) unigenes could be annotated to Swiss-Prot, Pfam and KOG databases, respectively. We annotated 3,993 (13.80%), 7,499 (25.92%), 10,437 (36.07%) unigenes to GO, COG and KOG databases, respectively.

COG database is constructed based on the phylogenetic relationships of bacteria, algae and eukaryotes, and it can be used for direct homologous classification of gene products. There are 7,499 unigenes that aligned with 25 COG classifications. Of the 25 COG categories, the cluster for general function prediction was only (26.74%) representing the largest group, followed by replication, recombination and repair, transcription (13.2%), and signal transduction mechanisms (11.43%). Zero and a small percentage (lower than 0.1 %) of unigenes were assigned to the number of unigenes which involved extracellular structures and nuclear structure, respectively. It is worth noting that a large number of unigenes were assigned to posttranslational modification, protein turnover, chaperones (9.89%), translation, ribosomal structure and biogenesis (9.21%), carbohydrate transport and metabolism, (7.53%), amino acid transport and metabolism (7.01%), energy production and conversion (6.04%).

According to the results of expression, the white perianths have 1,908 differential genes relative to blue perianths, of which 818 are up-regulated and 1,090 are down-regulated (Additional file 2 Table S1). Then in the further functional annotation of the identified differentially expressed genes, a total of 1,458 functional genes were annotated. GO analysis showed that a total of 3,993 unigenes can be divided into three categories: biological processes, cellular components, and molecular functions (Additional file 1: Figure S1). A large number of differential genes are classified as biological processes and molecular functions. The biological process category mainly includes molecular processes (904, 22.6%) and cellular processes (97, 2.4%), while the molecular function category: consists of catalytic activity (132, 3.3%) and binding (2,744, 68.7%) (Additional file 3: Table S2). The 11 top-hit species based on NCBI NR annotation are shown in Fig.4. More than 60% of unigenes could be annotated with the sequences from the 3 species, such as *Elaeis guineensis*, *Phoenix dactylifera* and *Musa acuminata*. Among them, *Elaeis guineensis* has the highest homology (Fig.4).

GO functional enrichment analysis of DEGs

The GO database structure is divided into several levels. The lower the level, the more specific the functions represented by nodes. From the above figure, we can see the annotations of DEGs and all genes in the secondary functions of GO. The nodes with obvious differences between red and blue columns may be related to the differences. Figure 4 (Additional file 3) shows the gene enrichment of the secondary functions of GO in the background of DEGs and in the background of all genes. This reflects the status of the secondary functions of GO in the two backgrounds. The secondary functions with obvious proportion differences show that the enrichment trend of differentially expressed genes is different from that of all genes. Therefore, we can focus on whether this function is related to the differences. The results showed that metabolic, cellular, and single-organism processes were the most

significant enrichment GO terms under the biological process category. Catalytic activity, binding, and transporter activity were the most enriched in the molecular function category.

KEGG function enrichment analysis of DEGs

KEGG database is the main public database about pathways, which helps us to further understand the function of genes [28]. The annotation results of DEGs KEGG were classified according to the pathway types in KEGG. The classification chart is shown in Fig.5. Among these pathways, phenylpropanoid biosynthesis represented the greatest number of DEGs (22 DEGs, 7.48%, ko00940), followed by starch and sucrose metabolism (20 DEGs, 6.80%, ko00500), protein processing in endoplasmic reticulum (18 DEGs, 6.12%, ko04141), biosynthesis of amino acids (17 DEGs, 5.78%, ko01230), and fatty acid elongation (15 DEGs, 5.10%, ko00062). Numerous research studies have shown that changes in flower color are closely related to flavonoids [29-31]. As expected, five pathways related to color development were identified, namely phenylpropanoid biosynthesis (22 DEGs, 7.48%, ko00940), flavonoid biosynthesis (12 DEGs, 4.08%, ko00941), anthocyanin biosynthesis (1 DEG, 0.34%, ko00942), and flavone and flavonol biosynthesis (3 DEGs, 1.02%, ko00944).

Genes related to color development

In order to obtain the relevant gene information related to *I. laevigata* colore, all the node genes involved in the secondary metabolic pathways of anthocyanin flower material (anthocyanin synthesis pathway, flavone and flavonol synthesis pathway, flavonoid synthesis pathway) were extracted from the transcriptome database. A total of 143 unigenes in the *I. laevigata* transcriptome database were matched to these three metabolic pathways (Table 3). A comprehensive search was performed in the *I.*

laevigatas transcriptome annotation results according to the standard gene names and synonyms provided by the KEGG database (Table 3). The corresponding gene was then mapped to the reference pathway provided by KEGG according to the EC identification number corresponding to the annotation results. The data set includes more than 90% of the annotated sequence genes in the biosynthesis of flavonoids (Fig. 6-1), while the other two pathways cover only a small fraction (Fig. 6-2, 6-3). In many cases, changes in the accumulation of anthocyanins are related to the amount of expression of the genes encoded by the relevant enzymes [32-35]. The green box（Fig.6-1）is a significant down-regulated gene in white perianths, and the red box is a significantly up-regulated gene in white perianths. In the flavonoid biosynthesis pathway, there were significant changes in the expression levels of 12 key genes. Only one of the CHI genes was up-regulated in white perianths, and other genes were down-regulated. Subsequently, in order to verify the reliability of the transcriptome sequence data, we used the designed primers to quantitatively analyze the six core genes. The sequences of the six core genes are not only related to the color formation of *I. laevigatas*, but also their expression levels are different. The transcriptome sequencing data was reliable due to the high correlation coefficient between the transcriptome sequence and q-PCR (r^2=0.5198) (Fig.7). Our q-PCR results are basically consistent with the results obtained by the RNA-Seq method, and only the individual sequences in CHS, CHI and F3'H are slightly different from the sequencing results. The expression of individual sequences of these three genes was significantly increased in white flowers (Fig.8).

Discussion

According to Table 1, blue perianths contain three anthocyanin compounds responsible for color pigmentation: Cy, Pel and Del. In contrast, anthocyanins were absent except for a small amount of delphinidins in white *I. laevigatas* var. *alba*. Furthermore, the content of delphinidin in blue perianths is

22 times greater than in white perianths. These results proves that the delphinidin pigment pathway is not cut off in *I. laevigatas* var. *alba*, but decreased the metabolic flow. In order to study the biochemical basis of *I. laevigatas* var. *alba* lacking color phenotype, the metabolic spectra of perianths were compared, and the compounds related to color pigmentation are discussed. Previous studies have shown that the blue-to-white phenotypic changes in *I. laevigatas* are mainly due to the absence of anthocyanins. Meanwhile, naringenin which was key node compound in anthocyanin synthesis pathway was not detected in *I. laevigatas* var. *alba* (Fig.2). This implies that a certain portion of the pathway before naringenin formation may be blocked or shunted. The gene closely related to naringenin synthesis is chalcone isomerase (CHI). Naringenin is both a catalytic product of CHI and a catalytic substrate for F3'H, DFR, F3H. However, no naringenin is inconsistent with the significant up-regulation of CHI gene in white perianths in transcriptome sequencing data. In addition, naringenin is a necessary metabolite of the pelargonidin pigment synthesis pathway. Naringenin's absence in white flowers means the disruption site of pelargonidin was at naringenin or earlier in the pathway. The objective was to determine why CHI was significantly up-regulated, but naringenin was not detected. The pathway of naringenin must be blocked in some steps in white perianths. We analyzed and simulated the related metabolic pathways and structural genes.

Studies have shown that DFR has strict substrate specificity and the substrate specificity of DFR can be altered by minor changes in DFR [36]. CHI may have the same characteristics as DFR. It is understood that CHI is a chalcone isomerase. CHIs ultilize pinocembrin chalcone and pinocembrin chalcone. When CHI gene is overexpressed, CHI may be easier to catalyze other chalcone(such as pinocembrin chalcone) rather than naringenin chalcone. Naringenin chalcone is the precursor to naringenin. This is also an important factor in the synthesis of Pel, because there is no substrate naringenin. Therefore, it

can be inferred that the Pel pathway is blocked upsteam. At this point, metabolic flow has to be diverted to Caffeoyl-CoA pathway to produce Del and Cy anthocyanins. In view of the down-regulated expression of downstream genes (F3H, DFR, ANS), the metabolic flux decreases step by step. Finally, only a small amount of delphinidin was detected. In addition the CHI gene does not have enzymatic activity and fails to catalyze the formation of naringein, which results in a redirection of the flux of metabolites. Delphinidin was detected in the white perianths, indicating that the presence of eriodictyol in *I. laevigatas* var. *alba*, while naringin, as one of the substrates of eriodictyol, was not detected. Therefore, eriodictyol can only be produced with 2'3'4,4',6'-Pentahydroxychalcone as substrates, which fully demonstrates that the metabolic pathway of flavonoids in *I. laevigatas* var. *alba* is from the initial compound ρ-Courmaroyl-CoA flowing through Caffeoyl-CoA to eriodictyol, and finally through to the delphinidin pathway (Fig.1-F).

A structural or regulatory blockage in the ABP would decrease the amount of flavonoid intermediates below the blockage, but would increase the amount of intermediates in upstream side branches (depending on the dynamics of metabolite flux through the pathway) [37]. Consistent results were found in *I. lutescens*, where the colorless anthocyanins (including chalcones, flavones, and flavonols) is higher than that of purple [38]. Lou et al. [7] concluded that the limitation of flux in upstream reactions and the multishunt process in downstream reactions led to the process of elimination of red pigmentation in the white flowers of *M. armeniacum*. They argue that the proportion of corresponding products (flavonols and anthocyanins) changed due to the competition between FLS and DFR on the basis of consistency of substrates, caused a redirection of the flux of metabolites through a side-chain to other products [7]. Similar arguments may exist between *I. laevigatas* and *I. laevigatas* var. *alba*.

Previous studies have shown that inactivation of CHI was an enhancer accompanying with one or more

flavonoid pathway genes, including flavonol synthase (FLS) and ANR [39-44]. The genetic and biochemical effects of chalcone isomerase-like (CHIL) are still unclear, but CHIL overexpression can increase the accumulation of wild-type Arabidopsis proanthocyanidins and flavonols [45]. When FLS competes for substrates with DFR, excessive CHI genes enhance FLS's competitiveness to substrates, especially when DFR is down-regulated. Previous research have shown that FLS can promote the conversion of dihydroflavonols to flavonols, resulting in the relatively higher accumulation of anthoxanthins (flavones and flavonols) in the nearly white flowers than in deeper color flowers [46]. The competition between FLS and DFR for substrates seems to exist, controlling metabolic flux to anthoxanthins (flavones and flavonols) or anthocyanins branches respectively, thus affecting the variation of flower color. Furthermore, studies show that Ipomoea CHI mutants have been shown to produce pale yellow flowers that accumulate chalcone 2'-O-glucoside rather than lacking flavonoid [47-49]. Chalcone 2'-O-glucoside is one of the naringenin chalcone by-products. Therefore, naringenin chalcone is more easily glucosidated to form chalcone 2'-O-glucoside. Similary, RT results showed that CHS was significantly up-regulated in white perianths. We speculated that CHI was also co-expressed with the CHS gene as an enhancer, which resulted in the tranformationsfer of naringenin chalcone into chalcone 2'-O-glucoside. Of course, this hypothesis needs to be further validated by detecting the related flavonones and flavonols.

As discussed earlier, Cy is the most important anthocyanin for the blue color in *I. laevigatas*. The question is why was Del synthesized instead of Cy in the presence of a consistent background of intermediate compounds (eriodictyol). We assume that Cy was present for a very short time or in a very small amount in white flowers. There may be multiple reasons for this phenomenon. Hydroxylation of flavanones to dihydroflavonols are mediated by F3H, F3'H and F3'5'H [50]. F3'5'H tend to be

produce delphinidin-based anthocyanins and F3’ H tend to produce predominantly cyanidin-based anthocyanins [51]. Due to the transfer of the upstream flux pathway, with the decrease of F3'H catalytic substrates, the flux of cyanidin-based anthocyanins also greatly decreases. Relatively, F3'5'H occupies more flux, therefore, there is a greater etabolic flow into the delphinidin pathway. The content of dihydronquercetin then decreased further. At the same time, the down-regulation of the expression of DFR and ANS resulted in the extremely small cyaniding content, which could not be detected or be catalyzed into the colorless proanthocyanidins under the action of ANR. In the delphinidin pathway, when the substrate background was the same (eriodictyol), the down-regulation of F3H makes more substrates catalyzed by F3'5'H, resulting in reduced metabolic flux. Therefore, the Del content was greater than Cy. In summary, the transfer of the upstream flux pathwat and the multi-stage split in the downstream reaction are responsible for the inability of white *I. laevigatas* to accumulate Cy.

In addition, peonidin, which is the methylated derivative of Cy. is detected in blue perianths. This indicates that the anthocyanins synthesis in the late stage of Cy is accomplished by methylation Additional research is needed to better understand and detect the related metabolites.

Conclusions

Although anthocyanins have been elaborated in some *Iris* species, the current information is far from complete in relation to the mutation mechanism behind the phenotype of *I. laevigata* var. *alba*. The transcriptomics data combined with liquid chromatography provides an effective method for identifying the key genes responsible for the absence of flower color. One hundred and forty-three (143) unigenes were identified as putative homologues of color-related genes, including 1 up-regulated and 12 down-regulated unigenes. Anthocyanin components associated with blue formation have been

identified. These anthocyanin components include Pel, Del and Cy. Among these, Cy is the primary pigment. We combined the metabolites and transcriptome data of the anthocyanin synthesis pathway to simulate the expression patterns of related structural genes in the flavonoid metabolic pathway. Through differential comparison, reverse metastasis of corresponding metabolites, we speculated that the upstream flux is redistributed by up-regulated CHI, the downstream gene (F3H, DFR, ANS) is down-regulated, and the competition between DFR and FLS for substrates make the color change from blue to white. This indicates that the absence of flower color is not determined by a single factor, but by the interaction and coordination of genes. This provided molecular biological references for the study of *I. laevigata*, especially variation in flower color.

Methods

Plant materials

I. laevigata(purple perianths) and *I. laevigata* var. *alb*a(white perianths) were used in this study. *I. laevigata* referenced the flora of China. *I. laevigata* var. *alba* was published by authors Ling Wang et al. in 2016. *I. laevigata* and *I. laevigata* var. *alb*a were cultivated in the campus nursery of Northeast Forestry University of China in 2014. The perianths of *I. laevigata* and *I. laevigata* var. *alba* were collected at the flowering stage (Fig. 1.A-D) on May 20, 2017, at 8:00 AM. Thereafter, fresh samples were immediately cryopreserved in liquid nitrogen and stored in a -80°C freezer for subsequent transcriptome sequencing, RNA extraction, and flavonoid analysis.

Pigment extraction and HPLC analysis

Anthocyanins were detected by HPLC. The freshly ground petals weighed 3.0 g and were extracted in the 2:1:1 v/v/v mixture (absolute ethanol: water: HCl) and homogenized. The material was then placed

in a constant volume 50 ml colorimetric tube ultrasonic extraction performed for 30 min, followed by shaking for 1 min., with hydrolysis in a boiling water bath for 1 h and then cooled. The volume was adjusted to a volume of 50 ml twice with the above mixed solution, and then allowed to stand. The supernatant was collected using a 0.45 μm reinforced nylon membrane filter for HPLC analysis. Standards for anthocyanins were provided by Technology Innovation Co., Ltd. (Qingdao, China).

Liquid chromatography was performed using a ThermoFisher U3000 HPLC. The HPLC included a C18 column (250×4.6 mm, 5 μm), a flow rate 0.8 mL/min, injection volume 10 μL, column temperature 35°C, and anthocyanidin chromatography. The anthocyanidin was measured at 530 nm, and anthocyanin determination content was based 100 g fresh weight in milligrams, and was repeated three times to record the peak area.

RNA extraction, cDNA library construction and RNA sequencing

Total RNA extraction from *I. laevigata* was extracted using a Qubit® RNA kit, and RNA purity was checked by 1% agarose gel electrophoresis and a Nanophotometer® spectrophotometer (IMPLEN, CA, USA). RNA integrity was assessed using the RNA Nano 6000 Assay Kit from the Agilent Bioanalyzer 2100 System. (Agilent Technologies, CA, USA). In order to increase the understanding of the molecular basis of the flower color variation of *I. laevigata*, two different flower blooms were used to construct cDNA libraries (Fig.1B, D). RNA sequencing was commissioned by Beijing Baimai Biotechnology Co., Ltd. The eukaryotic mRNA was enriched with magnetic beads with Oligo (dT). The mRNA was randomly interrupted by the addition of Fragmentation Buffer. The first cDNA strand was synthesized using random hexamers using mRNA as a template, and then the second cDNA strand was synthesized by adding buffer, dNTPs, RNase H and DNA polymerase I. The cDNA was purified

by AMPure XP beads. The purified double-stranded cDNA was subjected to end repair. A tail was added and the sequencing linker was ligated, and then AMPure XP beads were used for fragment size selection.Finally, a cDNA library was obtained by PCR enrichment. After the library was constructed, the concentration and insert size of the library were detected using Qubit2.0 and Agilent 2100, respectively, and the effective concentration of the library was accurately quantified by Q-PCR method to ensure the library quality. High-throughput sequencing was performed with HiSeq2500 after passing the library.

Transcription assembly and gene function annotation

Data filtering was performed on raw data, and the joint sequences and low quality reads were removed to obtain high quality clean data. The cDNA library was sequenced based on Sequencing By Synthesis (SBS) technology. After sequencing quality control, 14.81 Gb clean data was obtained. The Q30 base percentage of each sample was not less than 91.70%. After obtaining high-quality sequencing data, the sequencing reads were broken into shorter fragments (K-mer) using the Trinity platform, and then these small fragments were extended into longer fragments (Contig) and the overlap between the fragments was utilized. The fragment was obtained, and finally the transcript sequence was identified in each fragment set by using the method of De Bruijn graph and sequencing read information. Unigene sequences were aligned with NR, Swiss-Prot, GO, COG, KOG, eggNOG4.5, and KEGG databases using BLAST software. KEGG Orthology results of unigene in KEGG were obtained using KOBAS 2.0, and the amino acid sequence of unigene was predicted to be used. HMMER software was compared with the Pfam database to obtain unigene's annotation information. In order to obtain as much of the *I. laevigata* anthocyanin gene sequence as possible ,, KEGG was used to establish a reference map of the secondary metabolic pathway related to the *I. laevigata* flower color The

corresponding reference sequence was downloaded from the public database based on the nodal enzyme name and EC identification code provided by the KEGG database. All flower color related candidate genes annotated in the *I. laevigata* database and the corresponding reference sequences were subjected to BLASTx alignment.

Gene expression and co-expression analysis

The reads obtained by sequencing were compared with the unigene library by Bowtie, and the expression level was estimated according to the comparison result and RSEM. The expression abundance of the corresponding unigene is expressed by the FPKM value. In the process of differential expression analysis, the well-established Benjamini-Hochberg method was used to correct the p-value of the original hypothesis test, and finally the corrected p-value. FDR (False Discovery Rate) was a key indicator for differentially expressed gene screening to reduce false positives caused by independent statistical hypothesis testing of expression values of a large number of genes. In the screening process, the FDR is less than 0.01 and the difference fold FC (Fold Change) is greater than or equal to 2 as a screening criterion. FC represents the ratio of the expression levels between the two samples (groups). The hierarchically clustered analysis of the differentially expressed genes was performed, and the genes with the same or similar expression behavior were clustered to display the differential expression patterns of the gene sets under different experimental conditions.

Real-time quantitative PCR

Real-time quantitative PCR (q-PCR) specific primers were designed using Primer 5 software and the detailed information is shown in Additional file 4. Total RNA was extracted and the synthesis of the first cDNA strand was performed as described in the PrimeScriptTM RT reagent Kit (TaKaRa)

instructions. The cDNA obtained by reverse transcription was used as a template to perform q-PCR. The q-PCR analysis was run on an ABI 7500 real-time PCR detection system (using BiesysTEMS) using the SyBR® Premix ExtAGTM kit (TakaaBio Inc.). The PCR reaction was at 35 cycles (95 °C for 5 min, 95°C for 15 s; 60°C for 30 s; 72°C for 30 s; 72°C for 10 min). After RT-qPCR, a melting curve was generated to test the specificity of the product. The Actin gene was selected as an internal reference gene. To ensure the authenticity and reproducibility of the results, at least two independent biological replicates and three technical replicates were performed for each treatment, and differential analysis of the two conditions/groups was performed using the DESeq R package (1.10 1). Relative expression levels apply to the $2^{-\Delta\Delta Ct}$ method.

Abbreviations

ANR: Anthocyanidin reductase; ANS: Anthocyanidin synthase; ABP: anthocyanin biosynthesis pathway; CHS: Chalcone synthase; CHI: Chalcone isomerase; Cy: Cyanidin; COG: Clusters of Orthologous Groups of proteins; DFR: Dihydroflavonol 4-reductase; Dp: Delphinidin; DEGs: Differentially expressed genes; F3H: Flavanone 3-hydroxylase; F3'H: Flavonoid 3'-hydroxylase; F3'5'H: Flavonoid 3'5'-hydroxylase; FLS: Flavonol synthase; GO: Gene Ontology; HPLC: High liquid chromatography; KEGG: Kyoto Encyclopedia of Genes and Genomes; KOG: Eukaryotic Orthologous Groups of proteins; NR: Non-redundant; Pfam: Protein family; N50: Covering 50 % of all the nucleotide sequences of the largest unigene length; Pel：Pelargonidin; Q30 percentage: Percentage of bases with sequencing error rate lower than 1‰; RNA-Seq: RNA sequencing; RT-qPCR: Real-time quantitative polymerase chain reaction; UFGT: Anthocyanidin 3-O-glucosyltransferase;

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its additional files.

Competing interests

The authors declare that they have no competing interests.

Funding

This project is supported by grants from the National Natural Science Foundation of China (Grant No.31670344) and the Fundamental Research Funds for the Central Universities (Grant No.2572019DF08) and Excellent Youth Foundation of Heilongjiang Province of China (Grant No.YQ2019C018), which provide support for data collection and the design of the study, the National Key Technologies R&D Program (Grant No. 2016YFC0500308-02), which provide support for the analysis and interpretation of data.

Authors' contributions

DW carried out sequence data analysis and drafted the manuscript. SY contributed to manuscript revision. LW organized the manuscript. JY assisted with the sample collection and RNA extraction. All authors read and approved the final version of the manuscript.

Acknowledgements

The authors are grateful to the National Natural Science Foundation of China(Grant No.31670344), the Fundamental Research Funds for the Central Universities(Grant No.2572019DF08) and Excellent Youth Foundation of Heilongjiang Province of China(Grant No.YQ2019C018) for supporting this research.

Author details

[1] School of Landscape Architecture,Northeast Forestry University,Harbin,150040,China

[2] Forestry Research Institute of Heilongjiang Province,Harbin,150081,China

[3] State Key Laboratory of Tree Genetics and Breeding, School of Forestry, Northeast Forestry University, Harbin 150040, China

References

1. Wilson CA. Phylogenetic relationships among the recognized series in Iris section Limniris. Systematic Botany. 2009; 34: 277–284.
2. Wang L, Su L, Fan L-J, Shang F-J, Chen H-L. A new infraspecific taxon of Iris laevigata (Iridaceae) in northeast China. A Journal for Botanical Nomenclature. 2016; 25:111–113.
3. Zhao Y., Noltie HJ, Mathew B. Flora of China, Iridaceae in Z. Y. Wu and P.H. Raven [eds.]; Science Press: Beijing, China, and Missouri Botanical Garden Press: St. Louis, Missouri, USA. 2000; 24: 297–313.
4. Yabuya T. High-performance liquid chromatographic analysis of anthocyanins in induced amphidiploids of *Iris laevigata* fisch. x *I. ensata* thumb. Euphytica. 1987; 36: 381–387.
5. Yabuya T. High-performance liquid chromatographic analysis of anthocyanins in Japanese garden Iris and its wild forms. Euphytica. 1991; 52:215–219.

6. Sangaalofa T, Clark WSV. A systems approach to identifying correlated gene targets for the loss of colour pigmentation in plants. BMC Bioinformatics. 2011; 12(343): 1–18.

7. Lou Q, Liu, Y Qi Y, Jiao S, Tian F, Jiang L, Wang Y. Transcriptome sequencing and metabolite analysis reveals the role of delphinidin metabolism in flower colour in grape hyacinth. Journal of Experimental Botany. 2014; 65(12): 3157–3164.

8. Ithal N, Reddy AR. Rice flavonoid pathway genes, OsDfr and OsAns, are induced by dehydration, high salt and ABA, and contain stress responsive promoter elements that interact with the transcription activator, OsC1-MYB. Plant Sci. 2004; 166: 1505–1513.

9. Irani N, Grotewold E. Light-induced morphological alteration in anthocyanin-accumulating vacuoles of maize cells. BMC Plant Biol. 2005; 5:7.

10. Pasko P, Barton H, Zagrodzki P, Gorinstein S, Folta M, Zachwieja Z. Anthocyanins, total polyphenols and antioxidant activity in amaranth and quinoa seeds and sprouts during their growth. Food Chem. 2009; 115: 994–998.

11. Ordidge M, García-Macías P, Battey NH, Gordon MH, John P, Lovegrove JA, Vysini E, Wagstaffe A, Hadley P. Development of color and firmness in strawberry crops is UV light sensitive, but color is not a good predictor of several quality parameters. J Sci Food Agric. 2011; 92: 1597–1604.

12. Fengjuan F, Minjun L, Fengwang M, Lailaing C. Phenylpropanoid metabolites and expression of key genes involved in anthocyanin biosynthesis in the shaded peel of apple fruit in response to sun exposure. Plant Physiol Bioch. 2013; 69: 54–61.

13. Zoran Jeknić SJ, Slađana Jevremović, Angelina Subotić, Tony H. H.Chen. Alteration of flower color in Iris germanica L. 'Fire Bride' through ectopic expression of phytoene synthase gene (crtB) from Pantoea agglomerans. Plant Cell Reports. 2014; 33(8): 1307–1321.

14. Jin XH, Huang H, Wang L, Sun Y and Dai SL. Transcriptomics and Metabolite Analysis Reveals the Molecular Mechanism of Anthocyanin Biosynthesis Branch Pathway in Different Senecio cruentus Cultivars. Frontiers in plant science. 2016, 7: 1–14.

15. Sasaki N, Nakayama T. Achievements and perspectives in biochemistry concerning anthocyanin modification for blue flower coloration. Plant Cell Physiol. 2015; 56: 28–40.

16. Zhu ML, Zheng XC, Shu QY, Li H, Zhong PX, Zhang HJ, Xu YJ, Wang LJ, Wang LS. Relationship between the composition of flavonoids and flower colors variation in tropical waterlily (Nymphaea) cultivars. PLoS One. 2012; 7: 334–335.

17. Yabuya T, Yamaguchi MA, Fukui YK, Katoh KJ. Characterization of anthocyanin p-coumaroyltransferase in flowers of Iris ensata. Plant Science. 2001; 16: 499–503.

18. Boase, M. R., Lewis, D. H., Davies, K. M., Marshall, G. B., Patel, D., Schwinn, K. E., et al.. Isolation and antisense suppression of flavonoid 3'5-hydroxylase modifies flower pigments and colour in cyclamen. BMC Plant. Biol. 2010; 10: 107.

19. Tanaka Y, Tsuda S, Kusumi T. Metabolic engineering to modify flower color. Plant Cell Physiol. 1998; 39: 19–26.

20. IWashina T. The structure and distribution of the flavonoids in plants. J Plant Res. 2000; 113: 287–99.

21. Yabuya T, Noda T. The characterization of autoallotetraploid hybrids between *Iris ensata* Thunb. and *I. laevigata* Fisch. Euphytica. 1998; 103: 325–328.

22. Hopkins R, Rausher MD. Identification of two genes causing reinforcement in the Texas wildflower Phlox drummondii. Nature. 2011; 469: 411–414.

23. Lister, Carolyn. Flavonoid functions in plants; Flavonoids: Chemistry, Biochemistry and Applications. 2006.

24. Thill J, Miosic S, Ahmed R, Schlangen K, Muster G, Stich K, Halbwirth H: 'Le Rouge et le Noir'. A decline in flavone formation correlates with the rare color of black dahlia (Dahlia variabilis hort.) flowers. BMC Plant Biology. 2012; 12(225): 1–13.

25. Tomoko Mitsunami MN, Ivan Galis, Kabir Md Alamgir, Yuko Hojo, Kohei Fujita, Nobuhiro Sasaki, Keichiro Nemoto, Tatsuya Sawasaki, Gen-ichiro Arimura. Overexpression of the PAP1 Transcription Factor Reveals a Complex Regulation of Flavonoid and Phenylpropanoid Metabolism in Nictiana tabacum Plants Attacked by Spodoptera litura. PLOS ONE. 2014; 9 (9): 9.

26. Mizuno T, Uehara A, Mizuta D, Yabuya T, Iwashina T. Contribution of anthocyanin–flavone copigmentation to grayed violet flower color of Dutch iris cultivar ‘Tiger's Eye’ under the presence of carotenoids. Scientia Horticulturae. 2015; 186: 201–206.

27. Mizuno T, Yabuya T, Kitajima J, Iwashina T. Identification of novel C-glycosylflavones and their contribution to flower colour of the Dutch iris cultivars. Plant Physiology and Biochemistry. 2013; 72: 116–124.

28. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 1999; 27: 29–34.

29. Jia N, Shu Q, Wang L, Du H, Xu Y, Liu Z. Analysis of petal anthocyanins to investigate coloration mechanism in herbaceous peony cultivars. Sci Hortic. 2008; 117: 167–173.

30. Jia N, Shu Q, Wang D, Wang L, Liu Z, Ren H. Identification and characterization of anthocyanins by high-performance liquid chromatography-electrospray ionization-mass spectrometry in herbaceous peony species. J Am Soc Hortic Sci. 2008; 133: 418–426.

31. Zhong P, Wang L, Li S, Xu Y, Zhu M. The changes of floral color and pigments composition during the flowering period in Paeonia lactiflora Pallas. Acta Horticulturae Sinica. 2012; 39: 2271–2282.

32. Castellarin,S.D., Di Gaspero, G. Transcriptional control of anthocyanin biosynthetic genes in extreme phenotypes for berry pigmentation of naturally occurring grapevines. BMC Plant Biology.

2007; 7(1): 46.

33. Lin-Wang K, Bolitho K, Grafton K, Kortstee A, Karunairetnam S, K McGhie T, V Espley R, P Hellens R, C Allan A. An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. BMC Plant Biology. 2010; 10(1): 50.

34. Feng C, Chen M, Xu C J, Bai L, Yin XR, Li X, Allan CA, Ferguson IB, Chen KS. Transcriptomic analysis of Chinese bayberry (Myrica rubra) fruit development and ripening using RNA-Seq. BMC Genomics. 2012; 13(1): 19.

35. Yuan Y, Ma X, Shi Y, Tang D. Isolation and expression analysis of six putative structural genes involved in anthocyanin biosynthesis in Tulipa fosteriana. Scientia Horticulturae. 2013; 153(3):93–102.

36. Johnson ET, Ryu S, Yi HK, Shin B, Cheong H, Choi G. Alteration of a single amino acid changes the substrate specificity of dihydroflavonol 4-reductase. Plant J. 2001; 25: 325–33.

37. Casimiro-Soriguer I, Narbona E, Buide ML, Del Valle JC, Whittall JB: Transcriptome and Biochemical Analysis of a Flower Color Polymorphism in Silene littorea (Caryophyllaceae). Frontiers in Plant Science. 2016; 7: 204.

38. Wang H, Conchou L, Bessière, Jean-Msrie, Cazals G, Schatz B, Imbert E. Flower color polymorphism in Iris lutescens (Iridaceae): biochemical analyses in light of plant-insect interactions. Phytochemistry. 2013; 94: 123–134.

39. Ma L. Genomic evidence for COP1 as a repressor of light-regulated gene expression and development in Arabidopsis. The Plant Cell. 2002; 14: 2383–2398.

40. Oravecz A, Baumann A, Mate Z, Brzezinska A, Molinier J, Oakeley EJ, Adam E, Schafer E, Nagy F, Ulm R. Constitutively photomorphogenic1 is required for the UV-B response in Arabidopsis. The Plant Cell. 2006; 18:1975–1990.

41. Pan Y, Michael TP, Hudson ME, Kay SA, Chory J, Schuler MA. Cytochrome P450 monooxygenases as reporters for circadian-regulated pathways. Plant Physiology. 2009; 150: 858–878.

42. Morita Y, Takagi K, Fukuchi-Mizutani M, Ishiguro K, Tanaka Y, Nitasaka E, Nakayama M, Saito N, Kagami T, Hoshino A, Iida S. A chalcone isomerase-like protein enhances flavonoid production and flower pigmentation. The Plant Journal. 2014; 78(2): 294–304.

43. Yonekura-Sakakibara K, Tohge T, Matsuda F, Nakabayashi R, Takayama H, Niida R, Watanabe-Takahashi A, Inoue E, Saito K. Comprehensive Flavonol Profiling and Transcriptome Coexpression Analysis Leading to Decoding Gene–Metabolite Correlations in Arabidopsis. The Plant Cell. 2008; 20(8): 2160–2176.

44. Xie DY, Sharma SB, Paiva NL, Ferreira D, Dixon RA. Role of Anthocyanidin Reductase, Encoded by BANYULS in Plant Flavonoid Biosynthesis. Science Reprint. 2003; 299: 396–399.

45. Jiang W, Yin Q, Wu R, Zheng G, Liu J, Dixon RA, Pang Y. Role of a chalcone isomerase-like

protein in flavonoid biosynthesis in Arabidopsis thaliana. Experimental Botany. 2015; 66(22): 7165–7179.

46. Gao L, Yang H, Liu H, Yang J, Hu Y. Extensive Transcriptome Changes Underlying the Flower Color Intensity Variation in Paeonia ostii. Frontiers in Plant Science. 2016; 6: 1205.

47. Morita Y, Takagi K, Fukuchi-Mizutani M, Ishiguro K, Tanaka Y, Nitasaka E, Nakayama M, Saito N, Kagami T, Hoshino A, Iida S. A chalcone isomerase-like protein enhances flavonoid production and flower pigmentation. Plant J. 2014, 78: 294–304.

48. Iida S, Morita Y, Choi JD, Park KI, Hoshino A. Genetics and epigenetics in flower pigmentation associated with transposable element in morning glories. Adv. Biophys. 2004; 38: 141–159.

49. Saito N, Tatsuzawa F, Hoshino A, Abe Y, Ichimura M, Yokoi M, Toki K, Morita Y, Iida S, Honda T. Anthocyanin pigmentation controlled by the speckled and c–1 mutations of the Japanese morning glory. Journal of the Japanese Society for Horticultural Science. 2011; 80(4): 452–460.

50. Wu Q, Wu J, Li SS, Zhang HJ, Feng CY, Yin DD, Wu RY, Wang LS. Transcriptome sequencing and metabolite analysis for revealing the blue flower formation in waterlily. BMC Genomics. 2016; 17: 897.

51. Tanaka Y, Brugliera F. Flower colour and cytochromes P450. Philosophical transactions of the royal society. 2013; 368:1–14.

Figure Legends

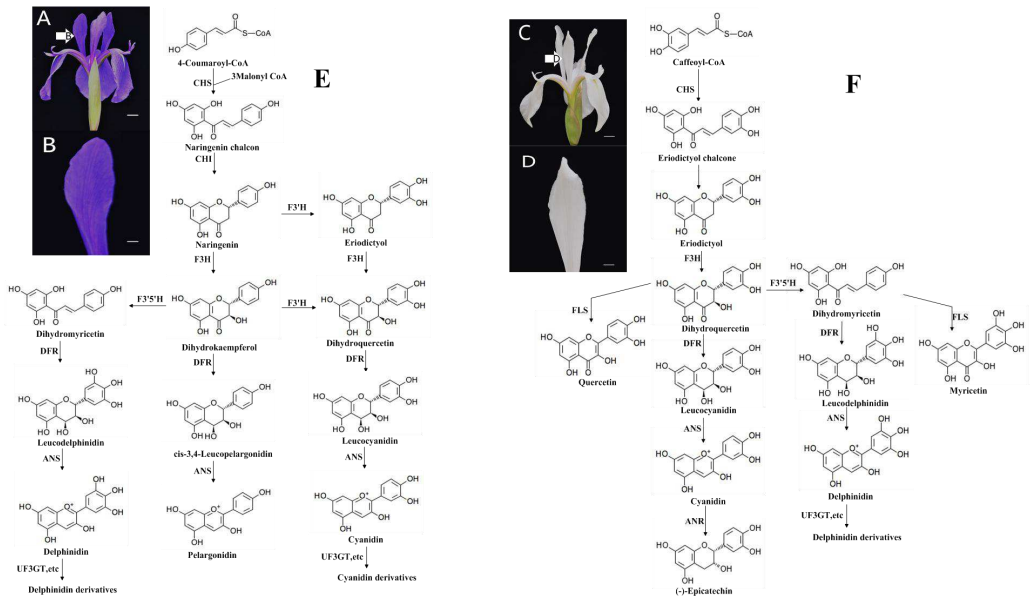


Fig. 1 A diagram of the putative anthocyanin metabolic process in blue or white *I. laevigata* flowers. (A)

Mature inflorescence of *I. laevigata*. Arrows represent the vexilla at the flowering stage. (B) the vexilla at the flowering stage used in deep sequencing. (C) Mature inflorescence of *I. laevigata* (D) The vexilla at the flowering stage used in deep sequencing. The scale bar=5 mm in A and C, and 0.5 mm in B and D. (E) The putative anthocyanin metabolic process in blue *I. laevigata* flowers. (F) The putative anthocyanin metabolic process in white *I. laevigata* flowers.

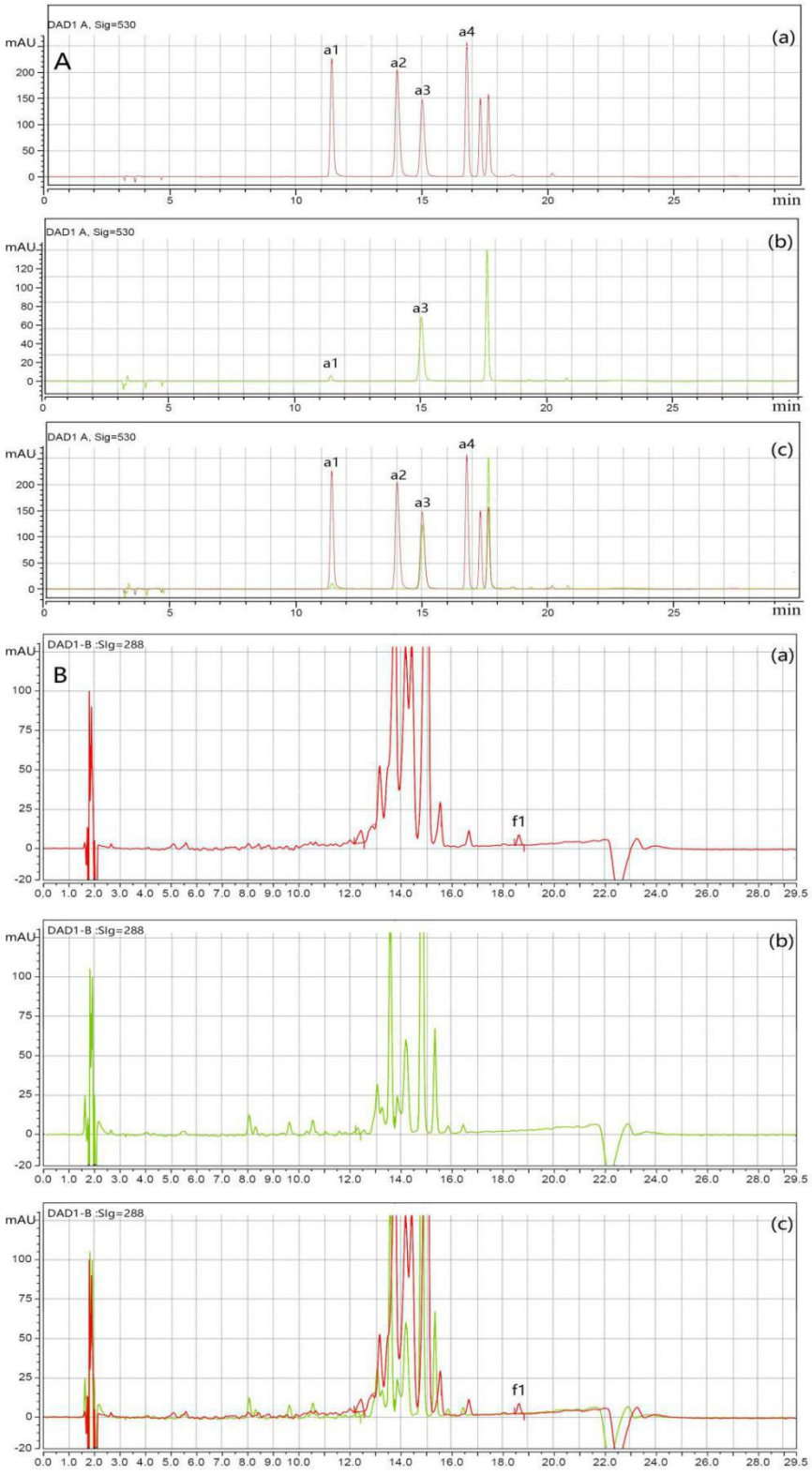


Fig. 2 HPLC of *I.laevigata* and *I.laevigata* var. *alba* about anthocyanins and flavanones.

(a) blue perianths; (b) white perianths; (c) indicates (a) and (b) merge together.

a1-a4 indicates identified anthocyanins; f1 indicates naringenin.

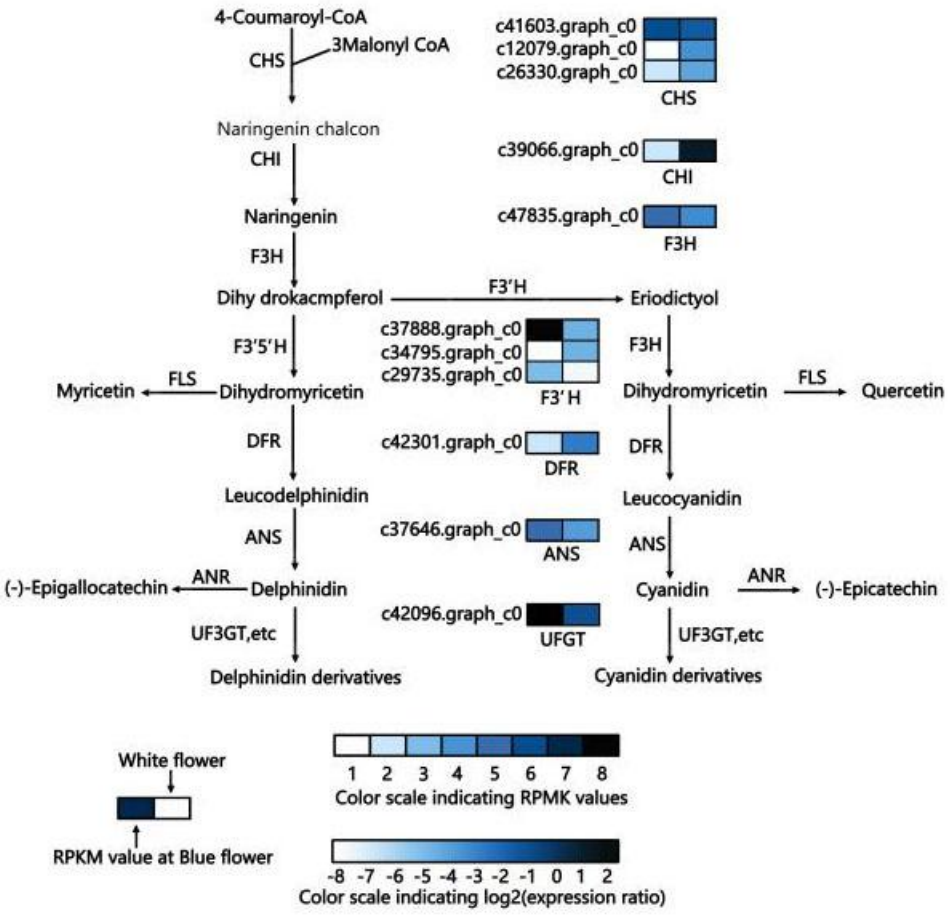


Fig. 3 Schematic of physiological and metabolic data related to flower color development of *I. laevigata*. (A) A detailed part of the Del and Cy metabolic sunetwork showing the subset of nodes or metabolites that constitute the process. Enzyme names and expression patterns are indicated at the side of each step. The expression pattern of each uni-transcript is shown on two grids, with the left one representing the RPKM value of blue flowers, and the right one representing the relative log2 (expression ratio) of blue perianths. The grids with eight different grey scale levels show the absolute, expression magnitude of blue flowers, with the RPKM values 0-8, 8-16, 16-24, 24-32, 32-40, 40-48, 48-56, and 56-64 represented by grey scale levels 1-8, respectively.

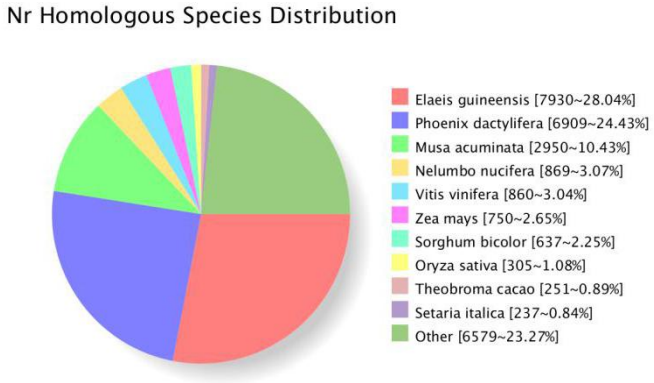


Fig. 4 Numbers and percentages of unigenes matching the eleven top species using BLASTx in the nr database

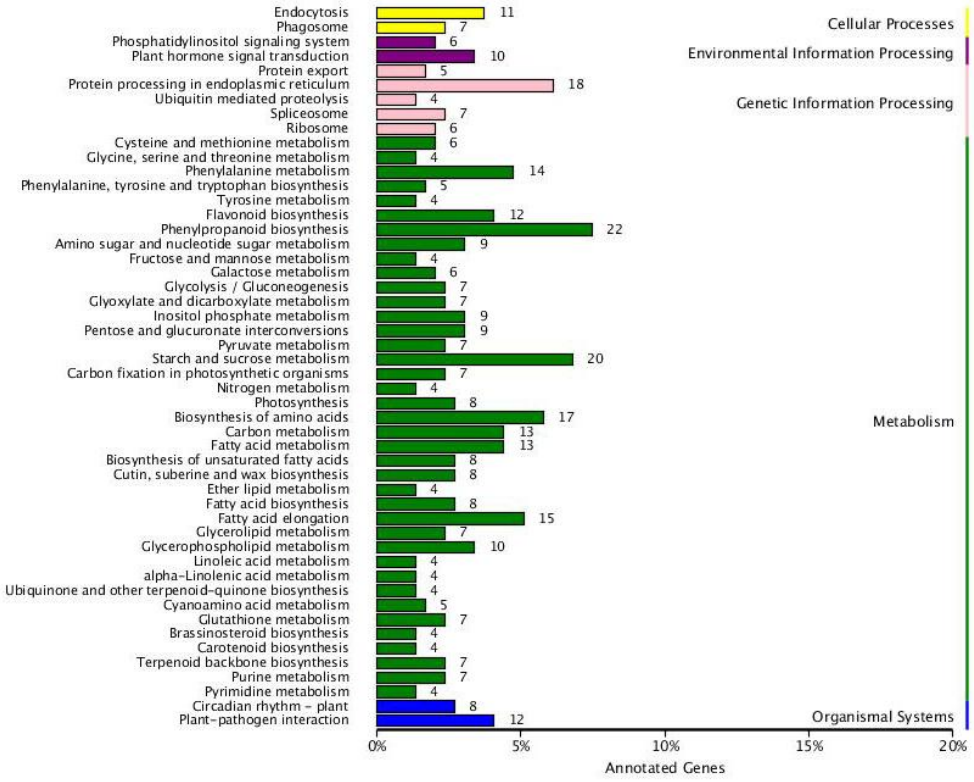


Fig. 5 Classification Map of Differentially Expressed Genes KEGG

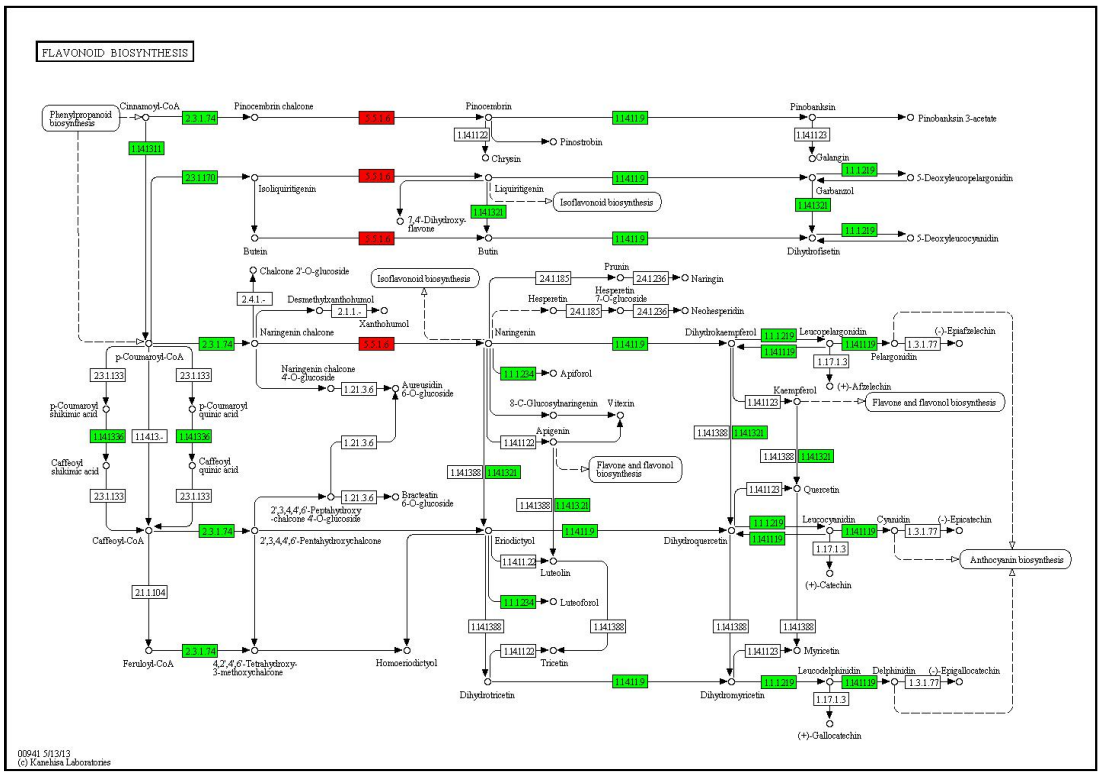


Fig. 6-1 KEGG flavonoid synthesis pathway hypothesis graph

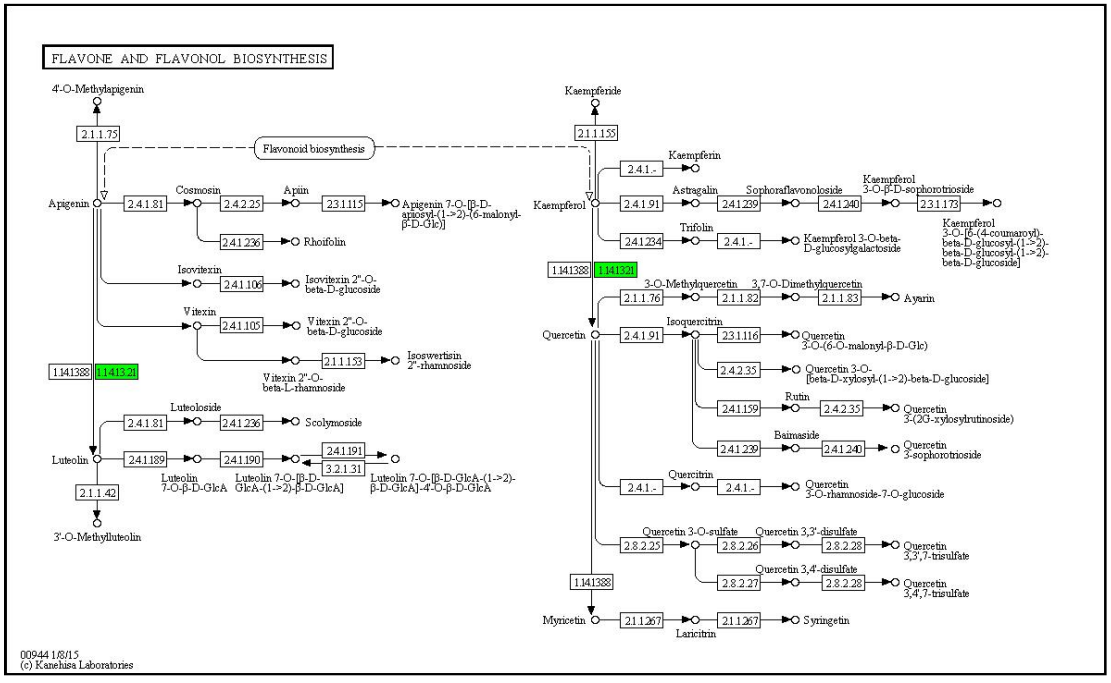


Fig. 6-3 KEGG flavone and flavonol synthesis pathway hypothetical graph

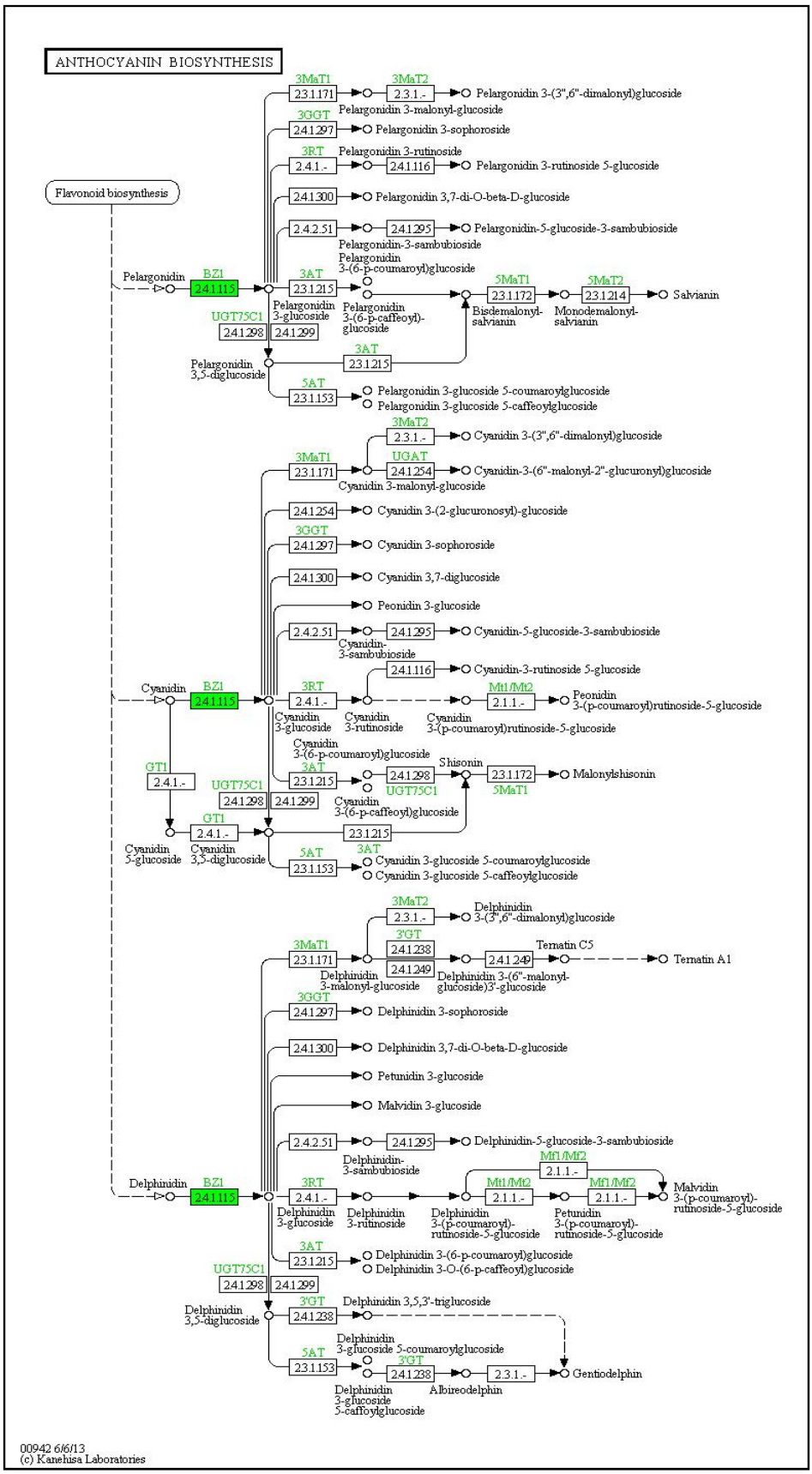


Fig.6-2 KEGG anthocyanin synthesis pathway hypothetical graph

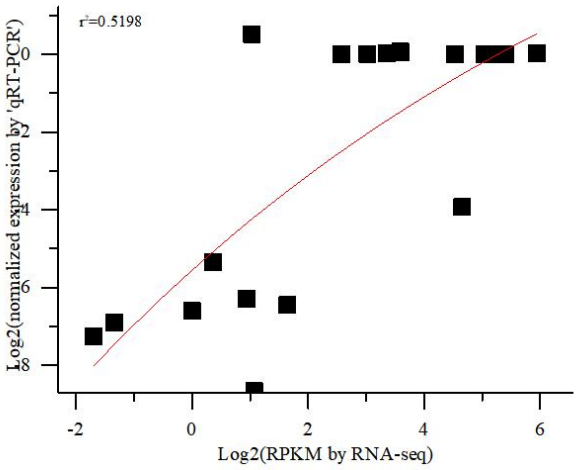
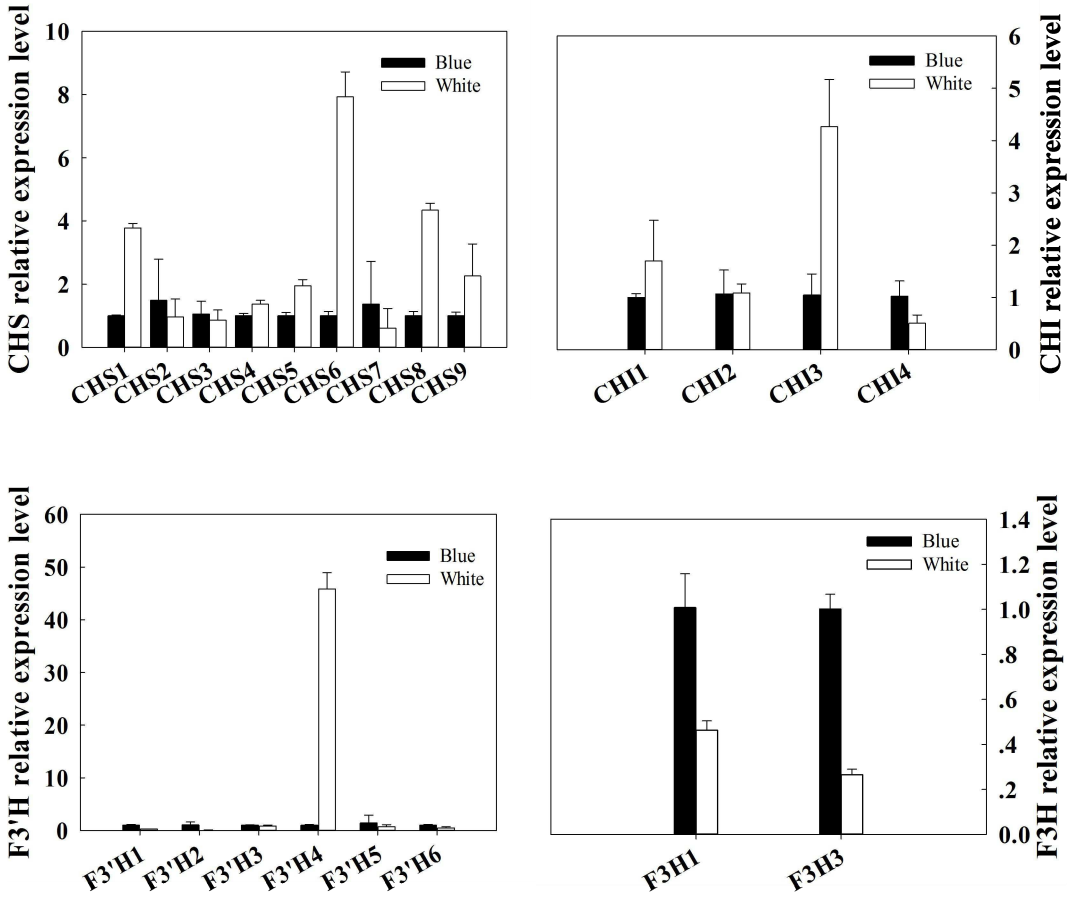


Fig .7 Correlation of gene expression results obtained from q-PCR analysis and RNA-Seq for color-related genes in blue and white flowers.



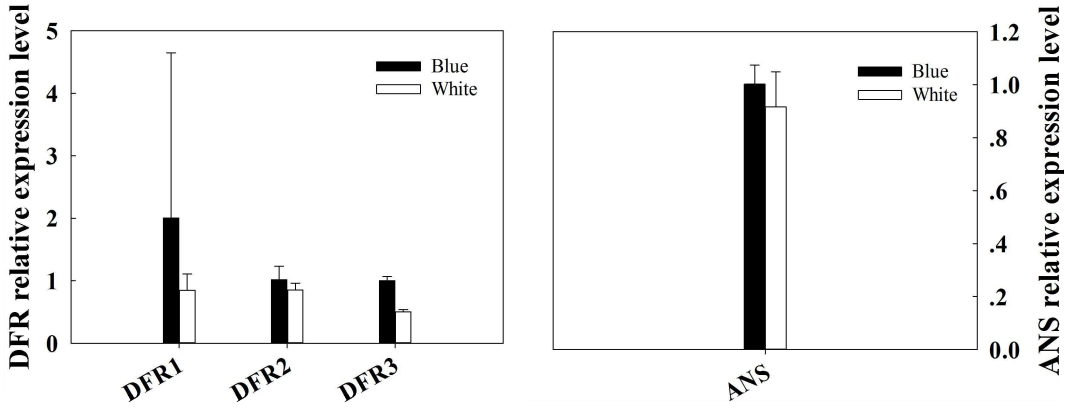


Fig. 8 Expression analysis of 6 candidate DEGs related to color development in *I. laevigata*

Table 1 Contents of flavonoid in perianths of *I. laevigatas* (mg kg^{-1}FW)

Peak	Blue	White
a1	897.28	40.79
a2	1,394.61	--
a3	1,151.41	--
a4	572.49	--
f1	120	--

a1, delphinidin; a2, cyanidin; a3, pelargonin; a4, peonidin; f1, naringenin;

--, not detected.

Table 2 Assembly result statistics

Length Range	Transcript	Unigene
200-300	20,836 (19.16%)	17,395 (26.95%)
300-500	27,220 (25.03%)	20,340 (31.52%)
500-1000	27,276 (25.08%)	14,741 (22.84%)
1000-2000	22,726 (20.89%)	8,283 (12.83%)
2000+	10,710 (9.85%)	3,778 (5.85%)

Total Number	108,768	64,537
Total Length	98,164,896	44,989,208
N50 Length	1,399	1,015
Mean Length	902.52	697.11

Table 3 Candidate genes related to flower pigmentation of *I. laevigata*

Fuction	***Gene***	***Enzyme***	***KO id(EC no.)***	***No.All[a]***	***No.Up[b]***	***No.Down[c]***
Flavonoid biosynthesis	*CHS*	Chalcone synthase	K00660(2.3.1.74)	9	0	3
	CHI	Chalcone isomerase	K01859 (5.5.1.6)	4	1	0
	F3H	Flavanone 3-hydroxylase	K00475 (1.14.11.9)	4	0	1
	F3'H	Flavonoid 3'-hydroxylase	K05280(1.14.13.21)	4	0	3
	F3'5'H	Flavonoid 3',5'-hydroxylase	K13083(1.14.13.88)	2	0	0
	DFR	Dihydroflavonol 4-reductase	K13082 (1.1.1.219)	3	0	1
	ANS	Anthocyanidin synthase	K05277(1.14.11.19)	1	0	1
	C4H	trans-cinnamate 4-monooxygenase	K00487(1.14.13.11)	1	0	1
	C3'H	coumaroylquinate(coumaroylshikimate) 3'-monooxygenase	K09754(1.14.13.36)	3	0	1
	HCT	shikimate O-hydroxycinnamoyltransferase	K13065 (2.3.1.133)	6	0	0
	CCOMT1	coumaroylquinate(coumaroylshikimate) 3'-monooxygenase	K09754(1.14.13.36)	3	0	0
Anthocyanin biosynthesis	*3GT*	anthocyanidin 3-O-glucosyltransferase	K12930(2.4.1.115)	8	0	1
	5AT	Anthocyanin 5-aromatic acyltransferase	K12936 (2.3.1.153)	3	0	0
	GT1	Anthocyanidin 5, 3-O-glucosyltransferase	K12938 (2.4.1.–)	41	0	0

	3MaT2	anthocyanidin 3-O-glucoside-3'',6''-O-dimalonyltransferase	K12932(2.3.1.-)	11	0	0
	Mt1/Mt2、*Mf1/Mf2*	anthocyanin 3'-methyltransferase，anthocyanin 3',5'-methyltransferase	\[d](2.1.1.-)	24	0	0
	5GT	cyanidin 3-O-rutinoside 5-O-glucosyltransferase	2.4.1.116	8	0	0
Flavone and flavonol biosynthesis	*FLS*	flavonol synthase	K05278 (1.14.20.6)	7	0	0
Flavanone biosynthesis	*ANR*	anthocyanidin reductase	K08695 (1.3.1.77)	1	0	0

[a] No. All, the total number of uni-transcripts analysed.

[b] No. Up, the number of uni-transcripts with expression significantly up-regulated in white flowers of *I. laevigata*

[c] No. Down, the number of uni-transcripts with expression significantly down-regulated in white flowers of *I. laevigata* compared with in purple flowers.

[d], omission of numbers for the KO id.

Additional files

Additional file 1: Figure S1. GO classification of all annotated unigenes and DEG unigenes.

Additional file 2: Table S1. Different Expression Genes in *I. laevigata* and *I. laevigata* var. *alba*.

Additional file 3: Table S2. GO functional enrichment of differentially expressed gene.

Additional file 4: Table S3. List of primers for expression analysis of flowering genes.

Statement

The author undertakes that all samples comply with the local legislation and the Convention on trade in endangered species of Wild Fauna and flora: https://www.cites.org/.

Figures

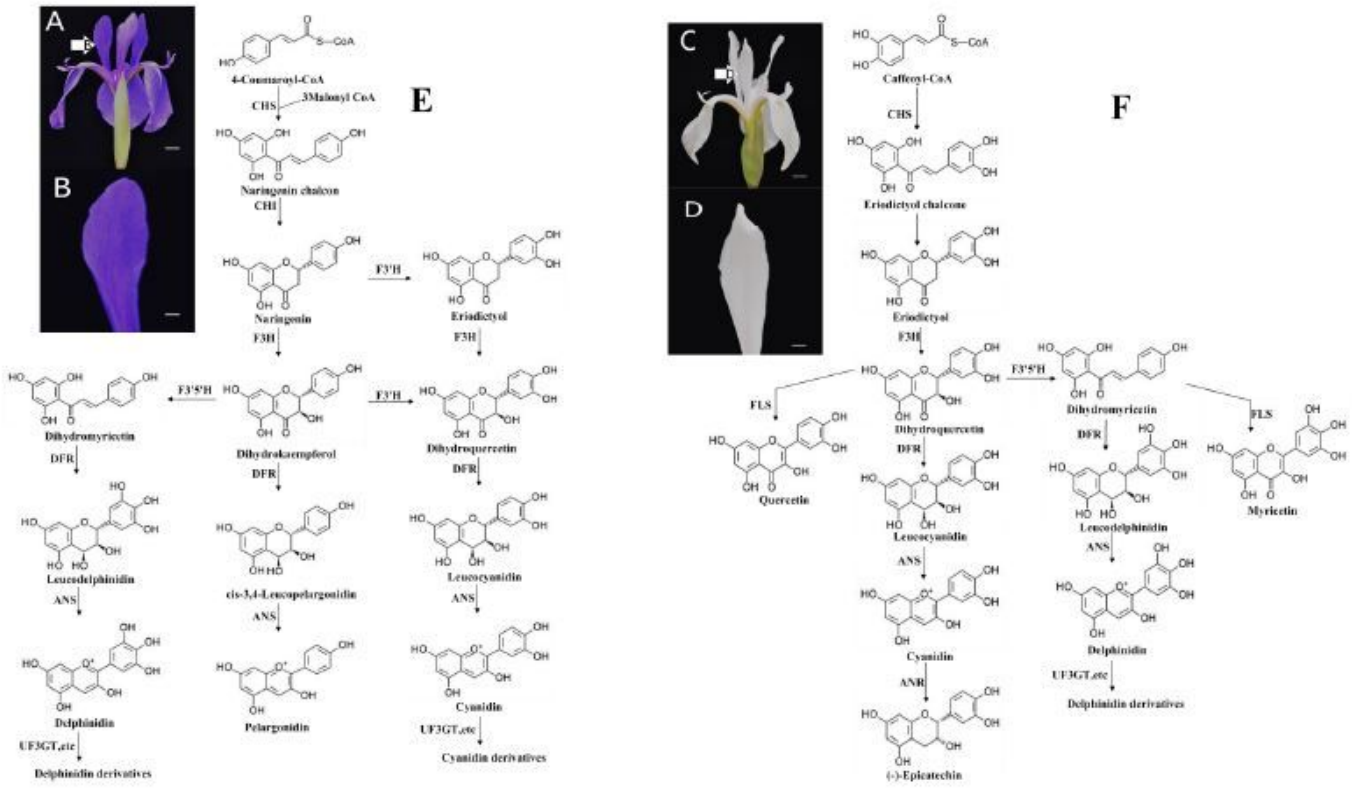


Figure 1

A diagram of the putative anthocyanin metabolic process in blue or white I. laevigata flowers. (A) Mature inflorescence of I. laevigata. Arrows represent the vexilla at the flowering stage. (B) the vexilla at the flowering stage used in deep sequencing. (C) Mature inflorescence of I. laevigata (D) The vexilla at the flowering stage used in deep sequencing. The scale bar=5 mm in A and C, and 0.5 mm in B and D. (E) The putative anthocyanin metabolic process in blue I. laevigata flowers. (F) The putative anthocyanin metabolic process in white I. laevigata flowers.

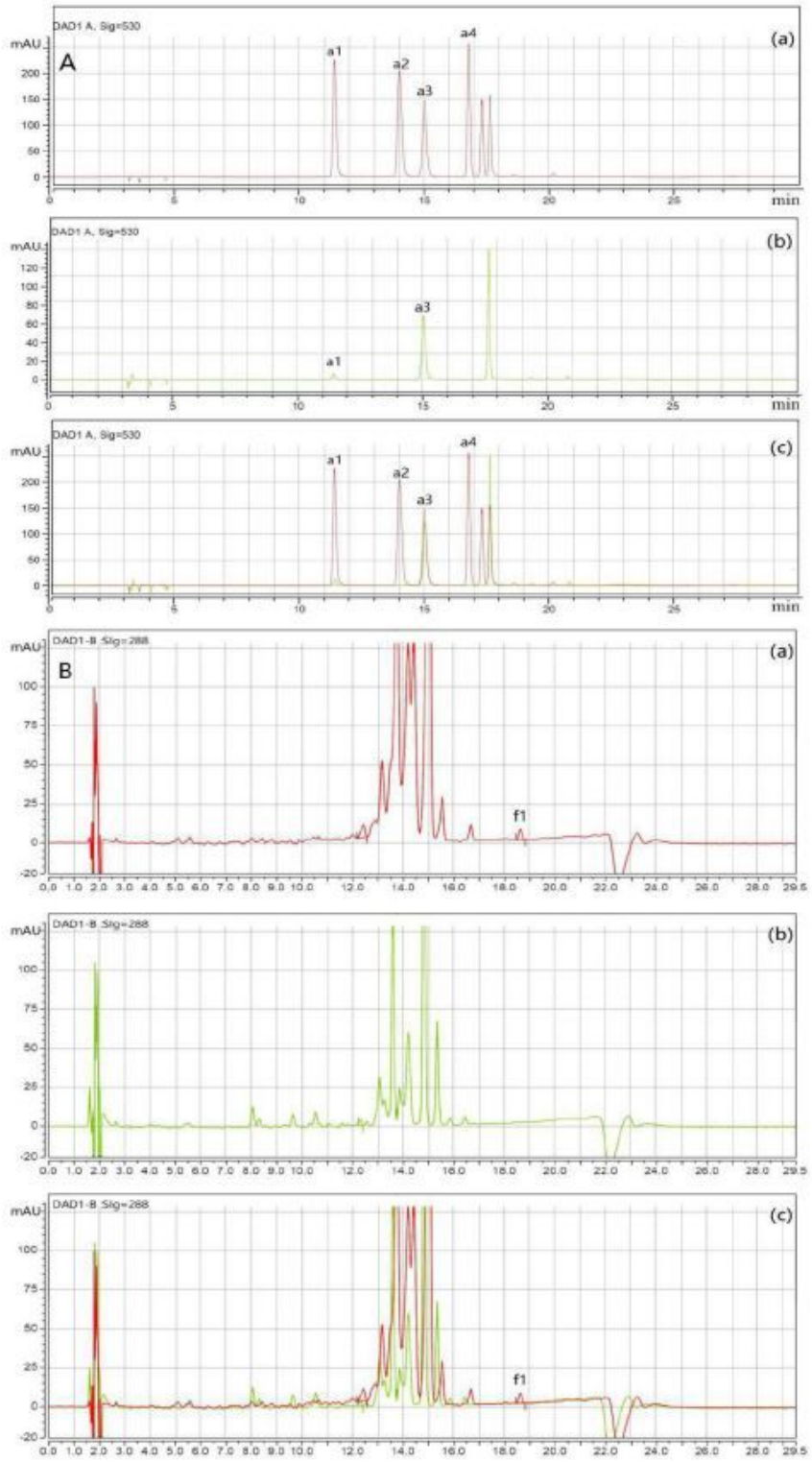


Figure 2

HPLC of I.laevigata and I.laevigata var. alba about anthocyanins and flavanones. (a) blue perianths; (b) white perianths; (c) indicates (a) and (b) merge together. a1-a4 indicates identified anthocyanins; f1 indicates naringenin.

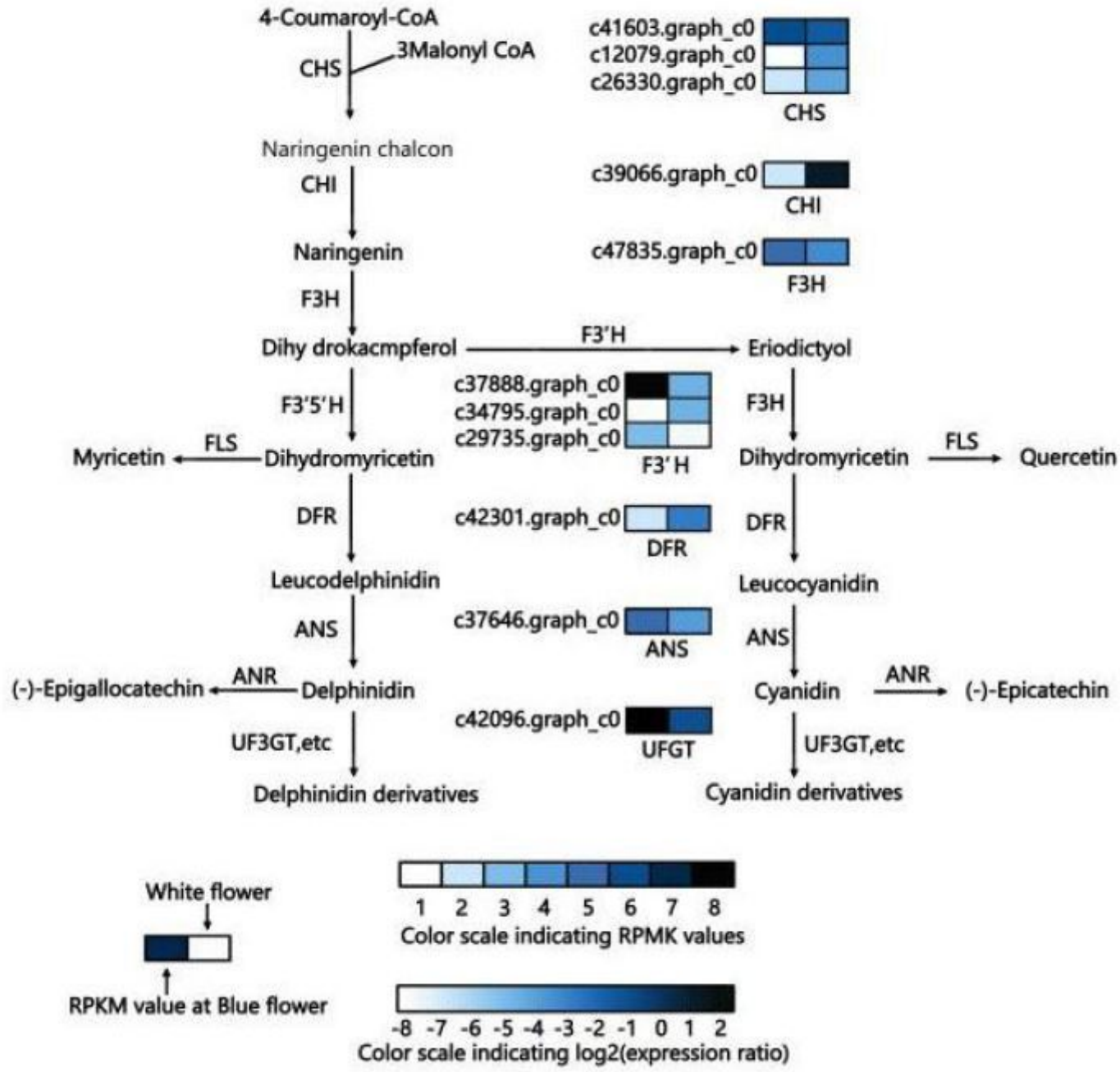


Figure 3

Schematic of physiological and metabolic data related to flower color development of I. laevigata. (A) A detailed part of the Del and Cy metabolic sunetwork showing the subset of nodes or metabolites that constitute the process. Enzyme names and expression patterns are indicated at the side of each step. The expression pattern of each uni-transcript is shown on two grids, with the left one representing the RPKM value of blue flowers, and the right one representing the relative log2 (expression ratio) of blue perianths. The grids with eight different grey scale levels show the absolute, expression magnitude of blue flowers,

with the RPKM values 0-8, 8-16, 16-24, 24-32, 32-40, 40-48, 48-56, and 56-64 represented by grey scale levels 1-8, respectively.

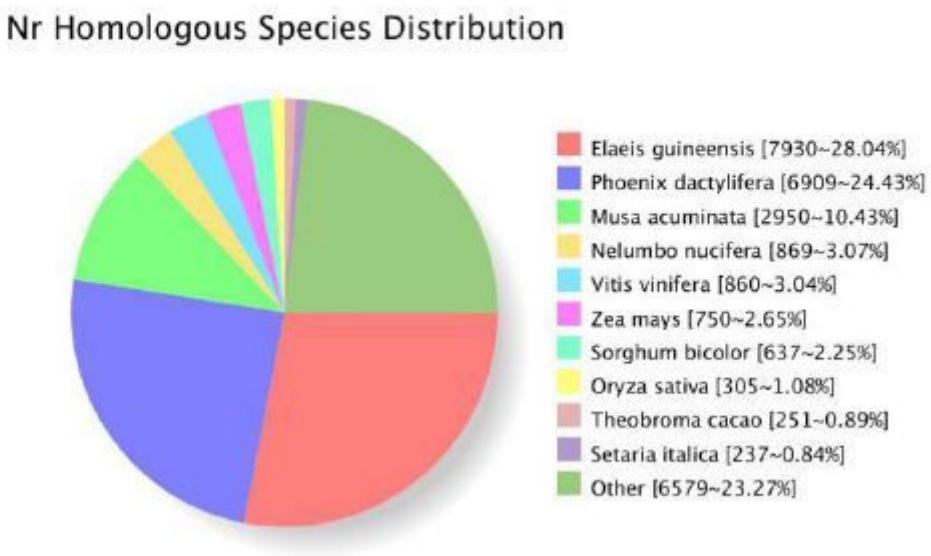


Figure 4

Numbers and percentages of unigenes matching the eleven top species using BLASTx in the nr database

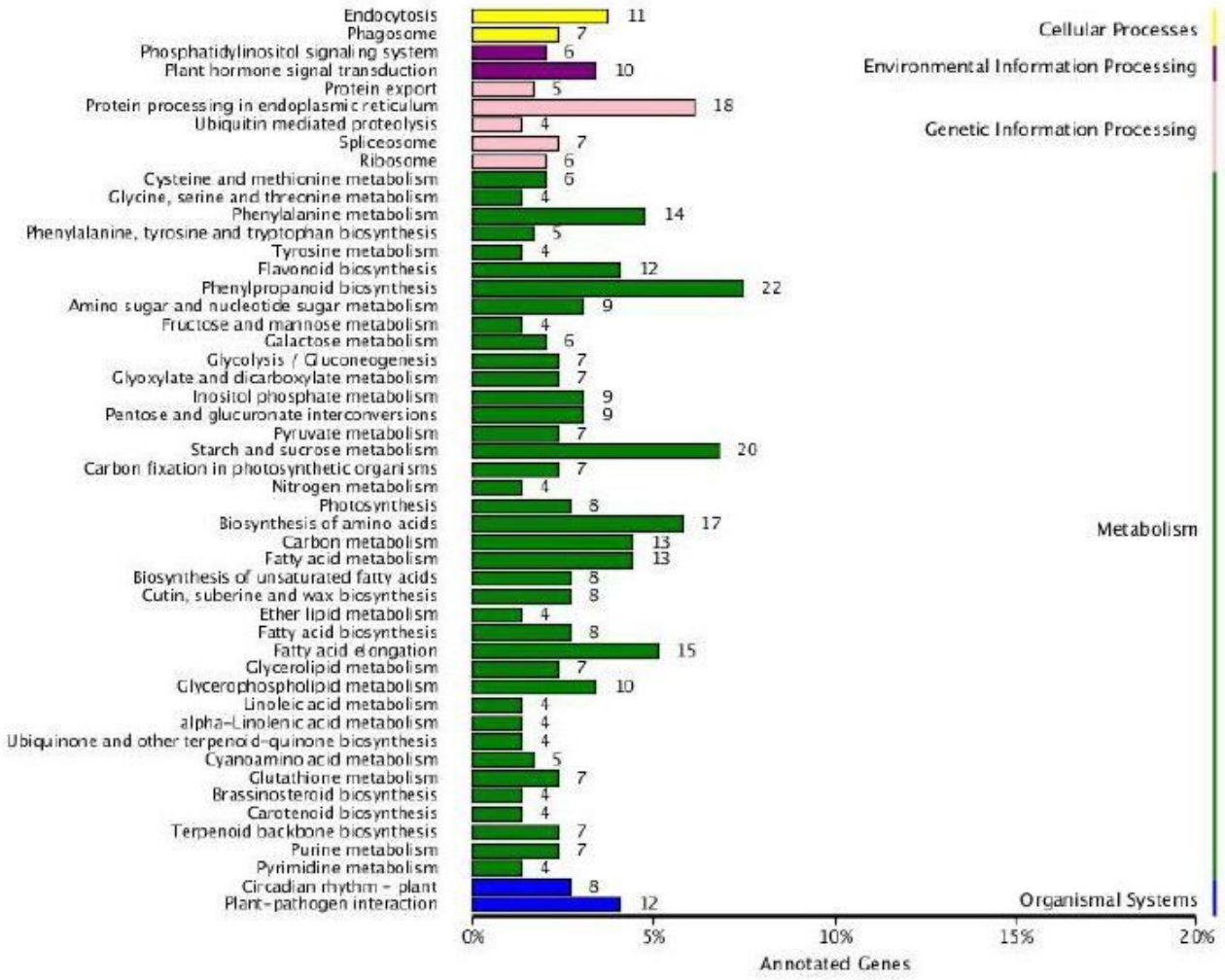


Figure 5

Classification Map of Differentially Expressed Genes KEGG

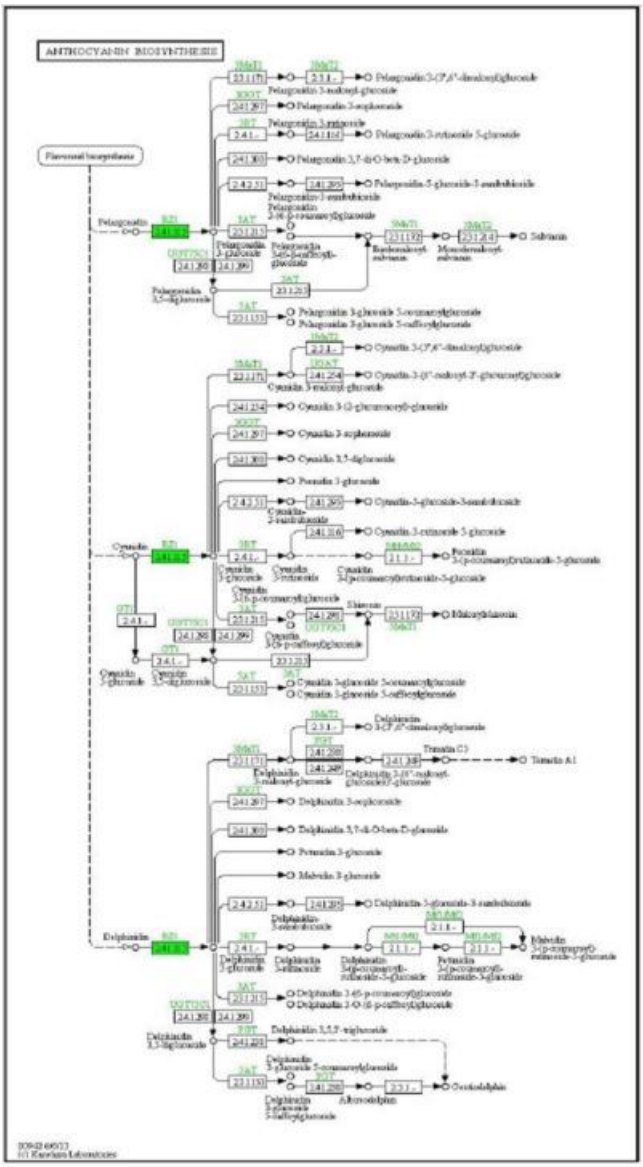


Fig.6-2 KEGG anthocyanin synthesis pathway hypothetical graph

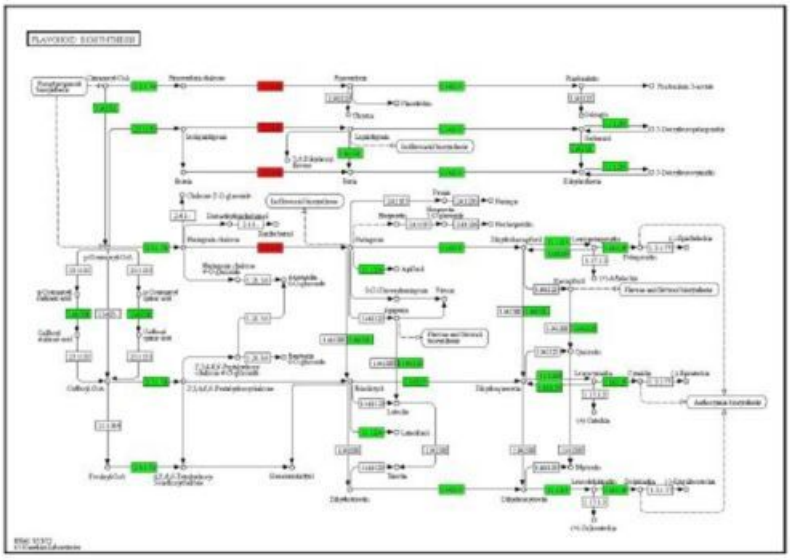


Fig. 6-1 KEGG flavonoid synthesis pathway hypothesis graph

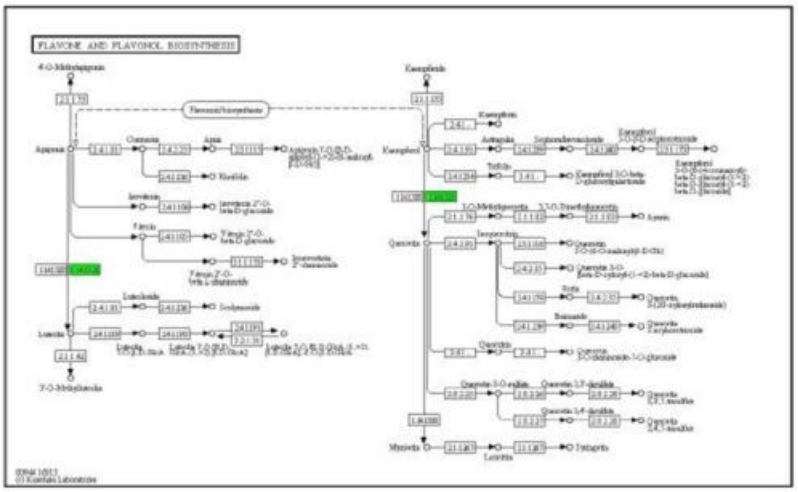


Fig. 6-3 KEGG flavone and flavonol synthesis pathway hypothetical graph

Figure 6

6-1 KEGG flavonoid synthesis pathway hypothesis graph; 6-2 KEGG anthocyanin synthesis pathway hypothetical graph; 6-3 KEGG flavone and flavonol synthesis pathway hypothetical graph

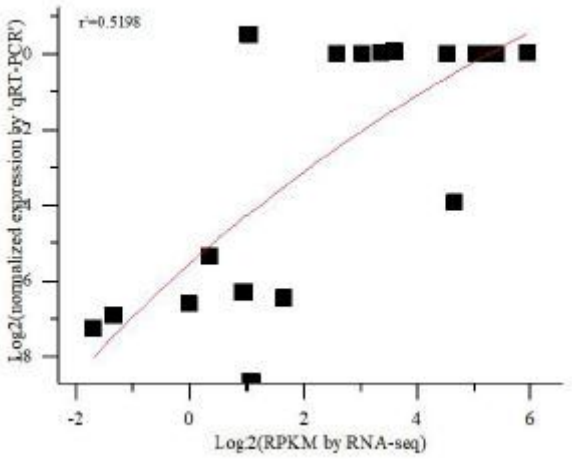


Figure 7

Correlation of gene expression results obtained from q-PCR analysis and RNA-Seq for color-related genes in blue and white flowers.

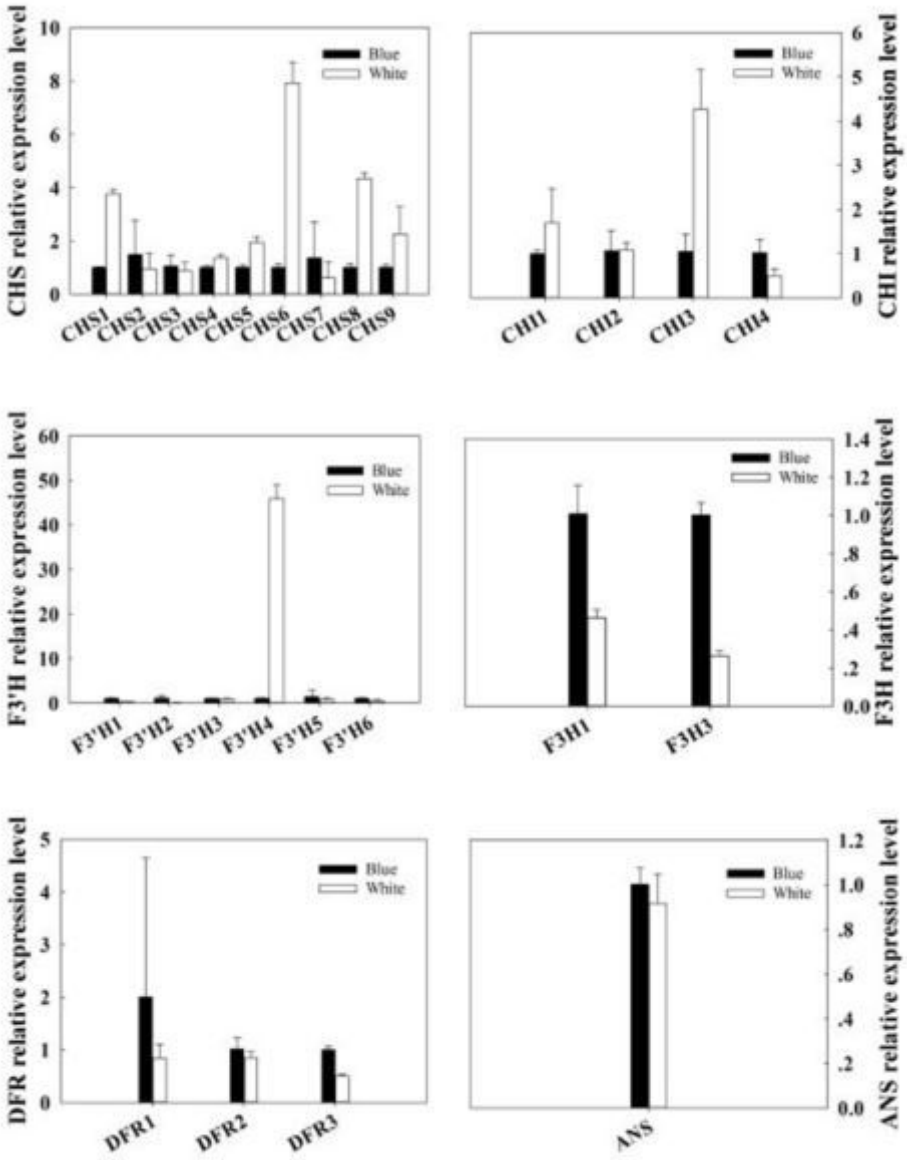


Figure 8

Expression analysis of 6 candidate DEGs related to color development in I. laevigata

Supplementary Files

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

- T03vsT01.DEG.xls
- Book1.xlsx
- TableS3.docx
- T03vsT01.GO.png