



Identification of differentially expressed genes associated with flower color in peach using genome-wide transcriptional analysis

Y. Zhou*, X.X. Wu*, Z. Zhang and Z.H. Gao

College of Horticulture, Nanjing Agricultural University, Nanjing, China

*These authors contributed equally to this study.

Corresponding author: Z.H. Gao

E-mail: gaozhong@njau.edu.cn

Genet. Mol. Res. 14 (2): 4724-4739 (2015)

Received June 11, 2014

Accepted October 27, 2014

Published May 11, 2015

DOI <http://dx.doi.org/10.4238/2015.May.11.5>

ABSTRACT. Flower color is an important trait of the ornamental peach (*Prunus persica* L.). However, the mechanism responsible for the different colors that appear in the same genotype remains unclear. In this study, red samples showed higher anthocyanins content (0.122 ± 0.009 mg/g), which was significantly different from that in white samples (0.066 ± 0.010 mg/g). Similarly to carotenoids content, red extract (0.058 ± 0.004 mg/L) was significantly higher in white extract (0.015 ± 0.004 mg/L). We estimated gene expression using Illumina sequencing technology in libraries from white and red flower buds. A total of 3,599,960 and 3,464,141 tags were sequenced from the 2 libraries, respectively. Moreover, we identified 106 significantly differentially expressed genes between the 2 libraries. Among these, 78 and 28 represented transcripts with a higher or lower abundance of more than 2-fold than in the white flower library, respectively. GO annotation indicated that highly ranked genes were involved in the pigment biosynthetic process. Expression patterns of 11 genes were verified using quantitative reverse transcription-polymerase chain reaction assays. The results suggest that hydroxycinnamoyl-coenzyme A shikimate/quinic

hydroxycinnamoyltransferase, 2-oxoglutarate-dependent dioxygenase, isoflavone reductase, riboflavin kinase, zeta-carotene desaturase, and ATP binding cassette transporter may be associated with the flower color formation. Our results may be useful for scientists focusing on *Prunus persica* floral development and biotechnology.

Key words: Differentially expressed genes; Flower color; *Prunus persica*; Illumina sequencing

INTRODUCTION

Flower color, an important trait of ornamental plants, is determined by plant pigments, such as flavonoids, carotenoids, and betalains. Among these, a colored class of flavonoids, anthocyanins (ACYs), confers a diverse range of colors, including orange, red, violet, and blue. Other flavonoids such as aurones and flavones are pale yellow or invisible to the human eye (Tanaka et al., 2008). Carotenoids are ubiquitously distributed in plants as essential components of photosynthesis and confer a yellow or red color on flowers when they are present. Betalains also yield yellow or red colors, but are found in a limited number of plant lineages (Tanaka et al., 2010). Additionally, the final color of a flower is influenced by factors such as anthocyanin structures, co-pigments (typically flavones and flavonols), metal ions, and vacuolar pH (Yoshida et al., 2009). Modification of flavonoids with hydroxyl, methyl, glycosyl, and acyl groups can produce several thousand different structures (Tanaka et al., 2009).

Cytochrome P450 plays important roles in flavonoid biosynthesis and their colored class of compounds, ACYs, both of which are important floral pigments (Tanaka and Brugliera, 2013). Flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase, belong to CYP75B and CYP75A (Ueyama et al., 2002), respectively. Determining the hydroxylation pattern of the B-ring is essential for the synthesis of cyanidin (typically conferring red/magenta color) and delphinidin-based ACYs, respectively (Tanaka, 2006). With the rapid development of molecular biotechnology and techniques, the regulation mechanism of ACYs has been widely examined. Flavone biosynthesis from flavanones is catalyzed by 2 types of flavone synthases: flavone synthase I (FNS I, a 2OGD) (Martens et al., 2003) and flavone synthase (FNS II, belonging to the CYP93B subfamily of the P450 superfamily). FNS II is more widespread than FNS I, and was first identified in snapdragon.

Next-generation high-throughput sequencing technologies are commonly used in molecular biology analysis as a powerful strategy for identifying genome-wide expression. Recently developed tools allow for the concomitant sequencing of millions of signatures to the genome, as well as identification of specific genes and gene expression levels in a sample tissue (Bentley, 2006). Digital gene expression (DGE) tag profiling is a revolutionary approach for expression analysis (Asmann et al., 2009) and involves a global orthogonal hybridization array validation method, with a nearly unlimited dynamic range. DGE creates genome-wide expression profiles using Solexa/Illumina technology. DGE can be used to analyze the gene expression with very high sensitivity, particularly when comparing 2 very similar samples. The ability to identify, quantify, and annotate expressed genes at the whole genome level without prior sequence knowledge enables large-scale disease classification, leading to higher-confidence target discovery, biological experimentation, and pathway studies. This method combines the sequencing serial analysis of gene expression principle with sequencing technology to generate a digital output that is proportional to the number of

transcripts per mRNA (Anisimov, 2008). In addition, there have been few reports describing the use of Solexa/Illumina technology to analyze gene expression in different floral color tissues. However, the genetics of flower pigment biosynthesis in other species are well understood. Overall, the DGE approach is more valuable for qualitative and quantitative gene expression analysis than traditional microarray-based assays (Hao et al., 2011).

This diploid species *Prunus persica* is naturally self-pollinating, unlike most other cultivated *Prunus* species. *P. persica* belongs to Prunus of Rosaceae and is an important ornamental plant (Sajer et al., 2012) that originated in China and has a cultivation history spanning more than 3000 years. The ornamental peach exhibits immense diversity of flower color, making it an attractive ornamental plant for landscape use. However, very little information is available regarding the molecular mechanism regulating these traits, limiting the scope of rational selection for the improvement of ornamental plants. In this study, we examined ‘Sahong Tao’, which has a color dominated by red and rare in pure white in the genotype.

This is the first genome-wide analysis of gene expression in peach flower buds of different colors. To investigate the differentially expressed genes related to the flower color of *P. persica*, we used high-throughput sequencing technology to survey the gene expression profiles from white and red flower buds. Our data may be useful for studies focusing on *P. persica* floral development and biotechnology.

MATERIAL AND METHODS

Plant materials

In this study, we collected flower buds with white (WF) and red (RF) petals in the pink stage. The 2 types of flower buds were from the same tree of the cv ‘Sahong Tao’ grown on the campus of Nanjing Agricultural University, Nanjing, Jiangsu Province, China (Figure 1). All samples collected were immediately frozen in liquid nitrogen, and stored at -80°C until use. Flower samples were carefully collected and the petals were removed and immediately used for pigment analysis.



Figure 1. Example samples.

ACY and carotenoids determination

ACY contents were determined using the pH differential method (Wrolstad et al., 1982) with a spectrophotometer UV-2450 (Shimadzu Corporation, Japan). Pigments were extracted in 1% HCl. Approximately 0.1 g petal flour was added to 10 mL 1% HCl ultrasonic extraction solution for 0.5 h (30°C, 50 Hz). The homogenate was centrifuged at 15,000 g for 30 min at 4°C. These steps were conducted in the dark and repeated until the petals flour was colorless. Pigment extracts were increased to a volume to 25 mL. Next, 2 mL pigment extract was mixed with 3 mL pH 1.0 (0.4 M KCl-HCl) and pH 5.0 (0.4 M citric acid-sodium hydrogen phosphate) buffers, respectively. The mixture of pH 1.0 buffer and samples were equilibrated for 50 min and pH 5.0 buffer was equilibrated 80 min in the dark. Absorbance was measured at 510 nm. $\Delta OD = OD(\text{pH } 1.0) - OD(\text{pH } 5.0)$. The equation is:

$$ACY = \frac{\Delta OD \times MW \times DF \times 25}{\epsilon \times 0.1} \quad (\text{Equation 1})$$

where MW (molecular weight) = 445.2 g/mol; DF = dilution factor (mL/g); and ϵ (molar extinction coefficient) = 29,600 in $\text{L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$ for cyanidin-3-glucoside. Carotenoid contents were determined based on the Lichtenthaler equation (Wang, 2006). Measurements of ACY and carotenoid in samples were replicated 3 times.

Digital gene expression profiling

Total RNA for Illumina sequencing was isolated from flower bud samples using TRIzol reagent (Invitrogen, Carlsbad, CA, USA), followed by RNase-free DNase treatment (TaKaRa, Dalian, China). Quality and quantity analysis of total RNA, library construction, and sequencing were carried out at the Beijing Genomics Institute (Shenzhen, Guangdong, China). Six micrograms total RNA was extracted and purified using Oligo(dT) magnetic bead adsorption. Next, Oligo(dT) was used as a primer to synthesize cDNA. The 5' ends of tags were generated using 2 endonuclease: *NlaIII* and *DpnII*. All fragments except the 3' cDNA fragments connected to the Oligo(dT) beads were washed away and the Illumina adaptor 1 was ligated to the sticky 5' end of the digested bead-bound cDNA fragments. The junction of Illumina adaptor 1 and the CATG site is the recognition site of *MmeI*, which is a type of endonuclease with different recognition and digestion sites. It cuts 17-bp downstream of the CATG site, producing tags with adaptor 1. After removing the 3' fragments using magnetic bead precipitation, Illumina adaptor 2 was ligated to the 3' ends of tags, creating tags with different adaptors for both ends to form a tag library. After linear polymerase chain reaction (PCR) amplification, fragments were purified by polyacrylamide gel electrophoresis. During quality control steps, the Agilent 2100 Bioanalyzer and ABI StepOnePlus Real-Time PCR System were used to quantify and qualify the sample library. Finally, the library was sequenced using the Illumina HiSeq™ 2000.

Gene expression annotation

The 2 virtual libraries contained all possible CATG+17-bp sequences of the reference gene sequences, which mapped to the GDR (<http://www.rosaceae.org/species/prunus/>)

prunus persica) (Eldem et al., 2012) and NCBI (get the result of blast against nr). All clean tags were mapped to reference sequences and only 1-bp mismatch was considered. Clean tags that mapped to reference sequences from multiple genes were filtered. Any remaining clean tags were designated as unambiguous clean tags. The number of unambiguous clean tags for each gene was calculated and then normalized to number of transcripts per million clean tags (Morrissy et al., 2009). Sense-antisense regulation is very important for gene expression. Sequencing tags mapped to the complementary strand of the sense gene suggest that its antisense strand also produced transcripts and that this gene may use sense-antisense regulation ('t Hoen et al., 2008). Therefore, the clean tags and their antisense sequences were aligned.

Differentially expressed genes and functional analysis

To determine the significance of digital gene expression profiles, a rigorous algorithm has been developed to identify differentially expressed genes (DEGs) between 2 samples (Hao et al., 2011). Significantly differentially expressed genes were designated a P value <0.005, false discovery rate <0.01, and a 2-fold relative change in the threshold of the sequence counts across libraries. We used $FDR \leq 0.001$ and the absolute value of $|\log_2 \text{Ratio}| \geq 1$ as the threshold to determine the significance of differences in gene expression.

To determine their main biological functions, DEGs were mapped to every node of the Gene Ontology (GO) database (<http://www.geneontology.org/>) and pathway enrichment was analyzed. GO is an international standardized gene functional classification system that comprehensively describes the properties of genes and their products in any organism. GO has 3 ontologies: molecular function, cellular component, and biological process. The basic unit of GO is the GO term, and each GO term belongs to a type of ontology.

In gene expression profiling analysis, GO enrichment analysis of functional significance applies a hypergeometric test to map all DEGs to terms in the GO database, looking for significantly enriched GO terms in DEGs compared to the genomic background. The formula is:

$$p = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}}, \quad (\text{Equation 2})$$

where N is the number of all genes with GO annotation; n is the number of DEGs in N ; M is the number of all genes that are annotated to certain GO terms; and m is the number of DEGs in M . Our GO functional enrichment analysis also integrated the clustering analysis of expression patterns. Thus, we can obtain the expression patterns of DEGs annotated to the given GO term.

The P value was subjected to Bonferroni corrected and a corrected P value ≤ 0.05 was chosen as the threshold value. The GO term ($P \leq 0.05$) was defined as a significantly DEG-enriched GO term. We determined the major biological functions of DEGs using this analysis.

Genes typically cooperate with each other to exercise their biological functions. Pathway-based analysis is used to examine the biological functions of genes. KEGG is the major public pathway-related database. Pathway enrichment analysis identifies significantly enriched metabolic pathways or signal transduction pathways in DEGs compared with the whole genome background. The calculating formula is the same as that used in GO analysis. A Q-value of ≤ 0.05 was defined as a significantly DEG (enriched) in these pathways. The most meaningful pathways can be obtained based on the enrichment analysis of DGE pathway significance.

Quantitative real-time PCR (qRT-PCR) analysis

The expression patterns of candidate genes were verified using qRT-PCR. RNA samples used for the qRT-PCR assays were the same as those used for the DGE experiments. Total RNA was isolated using TRIzol reagent according to manufacturer instructions. First-strand cDNA was synthesized according to the method described by Tong et al. (2009). qPCR was performed using a SYBR[®] Green real-time PCR Master Mix (TaKaRa); all primers used and the positional information are listed in [Tables S1](#) and [S2](#). For each reaction, 1 µL diluted cDNA (equivalent to 100 pg total RNA) was mixed with 10 µL 2X SYBR green reaction mix (SYBR[®] Green qRT-PCR Master Mix; TaKaRa). Next, 5 pmol forward and reverse primers were added to prepare the final volume 20 µL. The *ACTIN* gene was used as an endogenous control (forward primer: 5'-ATCATGTTTGAGACCTTCAATG-3', reverse primer 5'-AGAGTCCAGCA-CAATACCAGTT-3') (Nanni et al., 2012). Thermal cycle conditions were as follows: 4 min at 94°C followed by 40 cycles of 20 s at 94°C, 20 s at 60°C, and 43 s at 72°C. Each cDNA was analyzed 3 times, and the average threshold cycle (Ct) was calculated for each sample. The relative expression level of the genes was calculated using the $2^{-\Delta\Delta C_t}$ method.

RESULTS

ACY and carotenoids contents

The pH differential method was employed to determine ACY content between the different petal color samples. The results showed that red petals had the higher ACY content (0.122 mg/g), which was significantly different from the white flowers (0.066 mg/g), as presented in Figure 2. Similarly, for carotenoids content, red extract (0.058 mg/L) was significantly higher than white extract (0.015 mg/L). The content of these pigments directly affects petal color. The results indicated that the color of pigment is related to pigment content and that ACY is the main pigment.

We used Illumina sequencing on DGE to identify DEGs from the WF and RF libraries involved in the flower color in the cv Sahong Tao. A total of 3,599,960 and 3,464,141 tags were obtained from the 2 libraries, respectively (Table 1).

To increase the robustness of the approach, single-copy tags in the 2 libraries (126,894 WF and 118,967 RF) were excluded from further analysis (Figure 3). After discarding low-quality tags (tags containing 'N', adaptor sequences and copy number <2), 3,472,063 and 3,344,217 clean tags remained in the 2 libraries, from which 94,692 (WF) and 90,621 (RF) distinct tags were obtained (Figure 3). There were 4071 additional distinct tags in the WF than in the RF library, possibly representing genes related to floral color. The percentage of distinct tags rapidly declined as copy number increased, indicating that only a small portion of the transcripts were expressed at a high level under the conditions tested.

The distribution of the clean tag copy number is shown in Figure 4. The copy number of distinct clean tags from 2 to 5 was 54,692, from 6 to 10 was 12,784, from 11 to 20 was 9158, from 21 to 50 was 8578, from 51 to 100 was 4256, and >100 was 5,224 in the WF library. In contrast, the copy number of distinct clean tags was mostly distributed from 2 to 5, with 57.76% of the total clean tags in the WF library (Figure 4). In the RF library, the copy number of distinct clean tags from 2 to 5 was 50,539, from 6 to 10 was 12,410, from 11 to 20 was 9177, from 21 to 50 was 8766, from 51 to 100 was 4422,

and >100 was 5307. Similar to the WF library, the copy number of distinct clean tags was mostly distributed from 2 to 5, with 55.77% of the total clean tags in the RF library (Figure 4).

Library saturation was determined by identifying unique tags. Sequencing was considered to have reached saturation when no new unique tags were detected. The results (Figure 5) indicated that the WF and RF libraries were sequenced to saturation, producing a full representation of the transcripts in the conditions tested. In both libraries, fewer unique tags were identified as the number of sequencing tags increased, reaching a plateau shortly after 1 million tags were sequenced. No new unique tags were identified as the total tag number approached 2.7 million in the WF library and 2.5 million in the RF library.

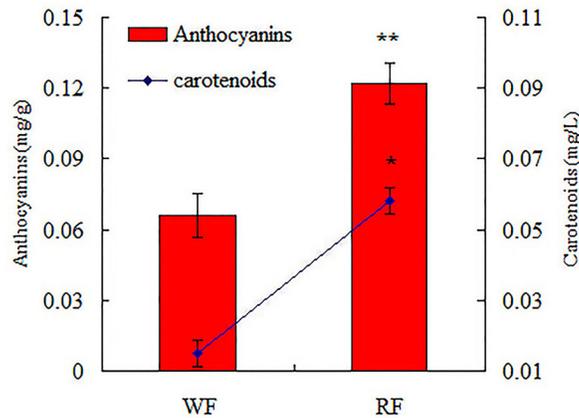


Figure 2. Anthocyanins and carotenoids contents of the petal. * $P < 0.05$, ** $P < 0.01$. Statistics of tag sequencing.

Table 1. Statistics of DGE sequencing.

Summary		WF	RF
Raw Data	Total	3,599,960	3,464,141
	Distinct Tag	222,494	210,441
Clean Tags	Total number	3,472,063	3,344,217
	Distinct Tag number	94,692	90,621
All Tags Mapping to Gene	Total number	2,215,834	2,112,110
	Total % of clean tag	63.82%	63.16%
	Distinct Tag number	42,449	41,225
	Distinct Tag% of clean tag	44.83%	45.49%
Unambiguous Tags Mapping to Gene	Total number	1,796,093	1,724,346
	Total % of clean tag	51.73%	51.56%
	Distinct Tag number	37,968	36,905
	Distinct Tag% of clean tag	40.10%	40.72%
All Tag-mapped Genes	number	16,010	15,676
	% of ref genes	55.78%	54.62%
Unambiguous Tag-mapped Genes	number	12,220	12,011
	% of ref genes	42.58%	41.85%
Mapping to Genome	Total number	916,890	888,158
	Total % of clean tag	26.41%	26.56%
	Distinct Tag number	36,202	34,497
	Distinct Tag% of clean tag	38.23%	38.07%
Unknown Tags	Total number	339,339	343,949
	Total % of clean tag	9.77%	10.28%
	Distinct Tag number	16,041	14,899
	Distinct Tag% of clean tag	16.94%	16.44%

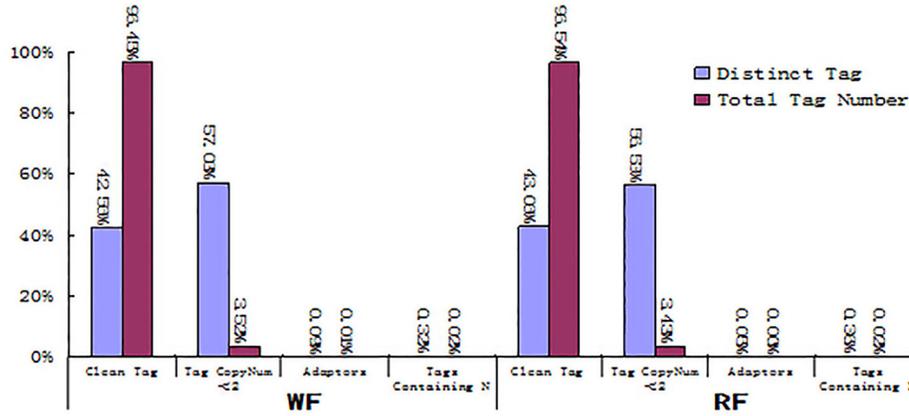


Figure 3. Distribution of tag expression.

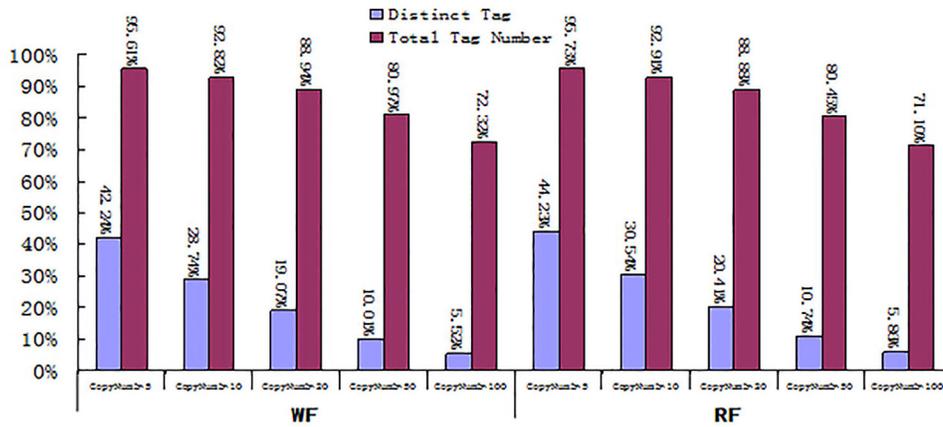


Figure 4. Distribution of clean tag copy number. Total clean tags represent the sum of all clean tag numbers; distinct clean tags represent all types of clean tags.

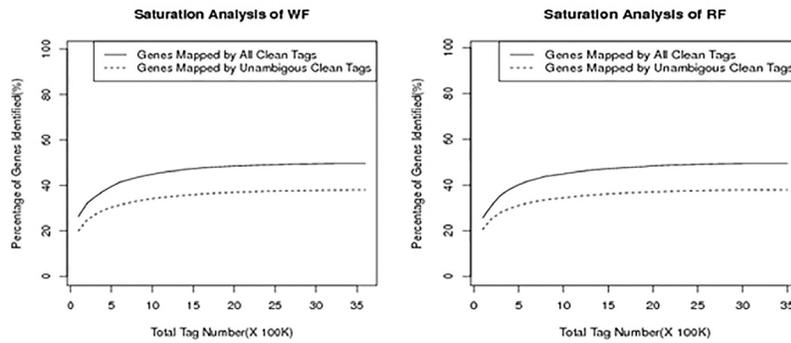


Figure 5. Saturation evaluation of differential expression.

Annotation analysis of unique tags

Distinct tags were compared against the genome and gene sequences of peach using GDR and BLASTn. Tags with a complete match or 1-bp mismatch were further examined. As shown in Table 1, 38.23% of genes in the WF library and 38.07% in the RF library matched the *Prunus persica* genome, but 42,449 (44.83% of unique tags) and 41,225 (45.49% of unique tags) in the WF and RF libraries matched to 16,010 (55.78%) and 15,676 (54.62%) *Prunus persica* gene, respectively.

Further analysis indicated that 37,968 (40.10%) unique tags in the WF library and 36,905 (40.72%) in the RF library matched to only 1 gene sequence in the peach genome (Table 1 and Figure 6), including perfect matching to genes and with a 1-bp mismatch. These data indicate that a larger number of transcripts were present in the RF sample.

Sense regulation, with clean sequencing tags perfectly mapped to the sense gene at a rate of 20.38% (antisense was 16.63%) and 21.15% (antisense was 16.44%) in WF and RF, respectively, was the main regulation model (Figure 6).

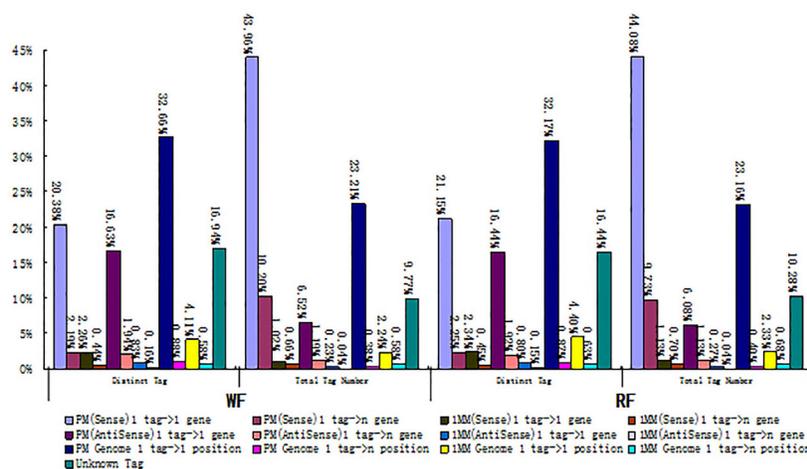


Figure 6. Alignment results of the clean tags. Note: PM (sense or antisense): perfect match to gene (sense or antisense); 1 tag → 1 gene: match to one gene; 1 tag → n gene: match to more than one gene; 1 MM (sense or antisense): match to gene (sense or antisense) with 1-bp mismatch; 1 MM genome: match to genome sequence with 1-bp mismatch; PM genome 1 tag → 1 position: perfect match to genome sequence with one best hit; PM genome 1 tag → n position: perfect match to genome sequence with multiple best hits; unknown tag: no match to gene (sense and antisense) and genome sequence.

qRT-PCR analysis

To confirm the data obtained by Illumina sequencing, a total of 11 candidate genes were randomly selected for qRT-PCR assays. The RNA samples used for these assays were the same as those used for the DEG experiments. Three replicates were used for each sample, after which the average threshold cycle (Ct) was calculated per sample. The relative expression levels were calculated as $2^{-(\Delta\text{Ct of WF} - \Delta\text{Ct of RF})}$. The expression level of each gene in the WF and RF was compared with its abundance from the sequencing data of DGE sequencing (Figure 7). The results of qRT-PCR agreed well with those of DGE analysis.

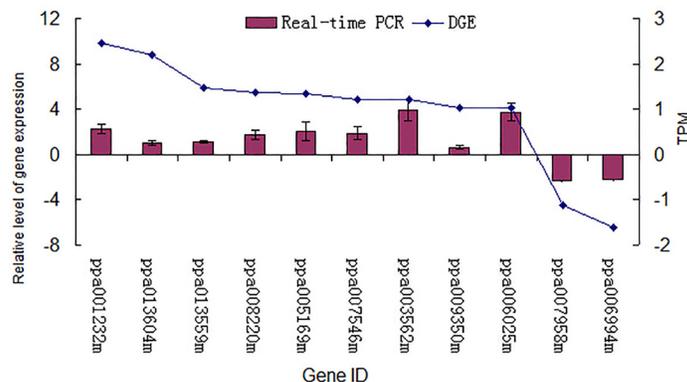


Figure 7. DGE tag profiling and qRT-PCR analysis of the expression of randomly selected genes. All real-time PCR reactions were repeated 3 times and the data are reported as means \pm SD. The x-axis indicates the different genes. The y-axis shows the expression levels: the left shows the relative expression level by qRT-PCR and the right shows tag number per million tags by DGE.

GO functional enrichment analysis of DGEs

As a useful tool for gene functional annotation, WEGO (Web Gene Ontology Annotation Plot) has been widely used in numerous studies. Forty-three DGEs were categorized into 3 ontology groups using $P \leq 0.05$ as the threshold (Figure 8). The cellular components included DNA-directed RNA polymerase II, holoenzyme (2), DNA-directed RNA polymerase complex (2), and nuclear DNA-directed RNA polymerase complex (2). Based on their molecular functions, the genes were classified as follows: carbon-carbon lyase activity (6), carboxy-lyase activity (4), alpha-glucosidase activity (2), lyase activity (6), and glucosidase activity (3). Finally, positive regulation of cell communication (2), carbon fixation (2), cellular metabolic compound salvage (3), glutamine family amino acid metabolic process (3), and cellular response to starvation (3) were identified as biological processes. Pigment biosynthetic process [GO:0046148], the chemical reactions and pathways resulting in the formation of a pigment, any general or particular coloring matter in living organisms, belonged to the term biological processes.

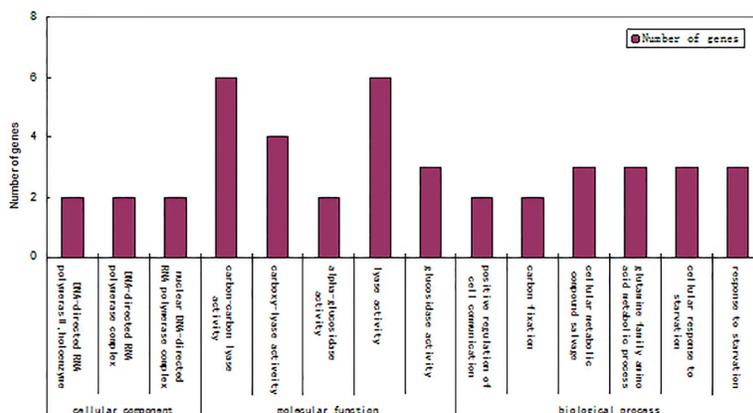


Figure 8. Gene ontology (GO) functional enrichment of differentially expressed genes (DEGs).

Pathway enrichment analysis for DGEs

$Q \leq 0.05$ defined these genes as significantly differentially expressed. There were 58 DEGs associated with 9 pathways (Figure 9). In this study, glyoxylate and dicarboxylate metabolism (ko00630, 4); alanine, aspartate, and glutamate metabolism (ko00250, 4); nitrogen metabolism (ko00910, 3); metabolic pathways (ko01100, 22); biosynthesis of secondary metabolites (ko01110, 15); other glycan degradation (ko00511, 3); *N*-glycan biosynthesis (ko00510, 2); carotenoid biosynthesis (ko00906, 3); and carbon fixation in photosynthetic organisms (ko00710, 2) were all significantly differentially expressed.

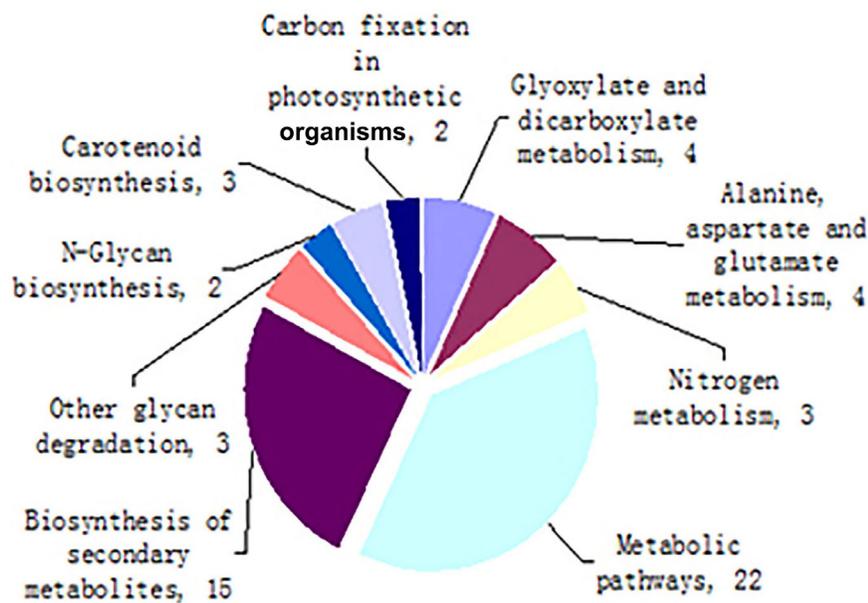


Figure 9. Pathway enrichment analysis.

Comparison of gene expression between the 2 libraries

Differences in tag frequencies in the WF and RF libraries were used to estimate gene expression levels in response to floral color. Transcripts detected with at least a 2-fold difference between the 2 libraries are shown in Figure 10. The red dots (78) and green dots (28) represent transcripts with a higher or lower abundance of more than 2-fold than seen in the WF library, respectively. The blue dots represent transcripts that differed by less than 2-fold between the 2 libraries, which were arbitrarily designated as “no difference in expression.” Among the DGEs, there were 0.07% total unique tags which were decreased by at least 5-fold and 0.07% total unique tags increased by at least 5-fold in the WF library, while the expression level of 99.86% unique tags was within the 5-fold difference between the 2 samples. The DGEs mapped to the GDR are listed in [Table S3](#) and those mapped to Blasts nr are listed in [Table S4](#).

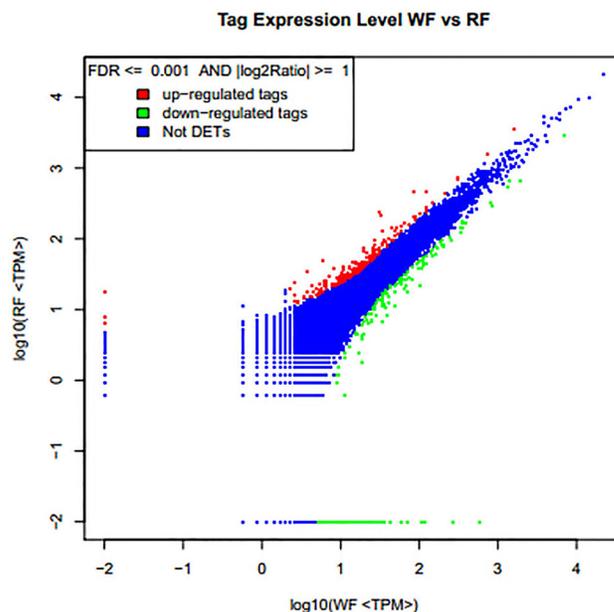


Figure 10. Comparison of gene expression levels between the two libraries. Upregulated genes are shown in red, while downregulated genes are shown in green. The parameters $FDR < 0.001$ and $\log_2 \text{ratio} \geq 1$ were used as the threshold to evaluate the significance of gene expression difference.

DISCUSSION

The evolutionary significance of flower-color polymorphisms found in many flowering plants has been examined in many previous studies, but the mechanisms responsible for their maintenance remain unclear. In this study, we used Illumina sequencing technology to estimate gene expression in libraries from WF and RF buds.

Flavonoid biosynthesis-related genes contribute to flower color

Flavonoids are one of the most abundant and important subgroups of secondary metabolites, with more than 6000 compounds detected in higher plants. They exist in various compositions and concentrations in nearly all plant tissues. In addition to attracting pollinators and dispersers to fruits and flowers, flavonoids also protect against a plethora of stresses (Czemmel et al., 2012). These compounds include 6 major subgroups, and many flavonoid compounds play an important role in flower color (Winkel-Shirley, 2001). Flower color is determined by the proportion of intermediate product. Flavonoid biosynthetic enzymes belong to various enzyme families.

In this study, we identified 15 DGEs involved in the biosynthesis of secondary metabolites. There were 2 upregulated genes and 1 downregulated gene related to flavonoid biosynthesis (ko00941), including hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxylcinnamoyl transferase (Hoffmann et al., 2004) (k13065, -1.38), which plays a critical role in the phenylpropanoid biosynthetic pathway. Notably 2OGD (Kai et al., 2008), (k00475,

8.81) showed the greatest differential expression in the pathway of flavonoid biosynthesis. The 2OGD domain enzymes catalyze the formation of plant hormones and pigments, such as anthocyanidins and flavones. 2OGD may catalyze the conversion of naringenin to dihydrokaempferol. Isoflavone reductase-related protein (k13081, 1.36) plays a major role in the leucocyanidin synthesis of catechin, affecting the entry of leucocyanidin into the anthocyanin biosynthesis pathway, which directly affects differences in peach petal color.

Glutathione *S*-transferase (GST) (Kitamura et al., 2010) is involved in the accumulation of proanthocyanidins. Some of the factors participating in flavonoid transport mechanisms are GST proteins (Tanaka et al., 2010). Flavonoid-related GST-like proteins have been isolated from several plant species, including *Perilla frutescens* (Yamazaki et al., 2008) and *Vitis vinifera*. In this study, the GST gene (ppa011307m) showed a \log_2 ratio >1.97 in the RF compared with the WF. It is involved in anthocyanin accumulation and transport in plants. Riboflavin kinase (Mori and Sakurai, 1995) is an essential enzyme catalyzing the phosphorylation of riboflavin in the presence of ATP and Mg^{2+} to form the active cofactor flavin mononucleotide, which can be further converted to flavin adenine dinucleotide (Karthikeyan et al., 2003). This gene (ppa010350m) is also related to the accumulation of anthocyanins, as riboflavin increases the accumulation of anthocyanins and anthocyanin synthesis is stimulated by riboflavin in plants.

The uridine diphosphate-glucosyltransferase family 1 proteins (family 1 UGTs) (-1.84) are known to glycosylate several classes of plant secondary metabolites (Caputi et al., 2008). Uridine diphosphate-glucosyltransferases catalyze the addition of a glycosyl group from a uridine triphosphate-sugar to a small hydrophobic molecule (Mackenzie et al., 1997). In plants, uridine diphosphate-glucosyltransferase are generally localized in the cytosol and are involved in the biosynthesis of plant natural products such as flavonoids, phenylpropanoids, terpenoids, and steroids, and in the regulation of plant hormones (Bowles et al., 2006). Additionally, these proteins play an important role in the stabilization and enhancement of water solubility and the deactivation/detoxification of natural products, regulating metabolic homeostasis, detoxification of xenobiotics, and the biosynthesis, storage and transport properties of secondary metabolites (Yonekura-Sakakibara and Hanada, 2011).

Zeta-carotene desaturase (ZDS) gene, which is associated with carotenoid biosynthesis

Carotenoids, which are isoprenoids, are notable for their wide distribution, structural diversity, and various functions. More than 700 types of carotenoids have been detected from natural sources (Ye et al., 2008). They are natural pigments synthesized by plants and some microorganisms. Carotenoids exhibit yellow, orange, and red colors, but when they are bound to proteins, they acquire green, purple, or blue colors (Machmudah and Goto, 2013). Lycopene is a red carotenoid which has scavenging reactive oxygen capacity and the strongest anticancer activity (Engelmann et al., 2011). Lycopene is raw material used in the synthesis of other carotenoids.

ZDS (Breitenbach et al., 1998) is one of the rate-limiting enzymes in the carotenoid biosynthetic pathway, and can carry out all 4 desaturation reactions necessary to convert the colorless phytoene into the red colored lycopene, whereas this is achieved by only 1 gene product, CrtI (Bartley et al., 2001). By comparing the 2 libraries, we identified 2 upregulated genes, ppa009815m (1.3) and ppa007546m (1.2), and 1 downregulated gene, ppa007958m (-1.1), all related to carotenoid biosynthesis

Low vacuolar pH may affect color formation

Membranes transporter-mediated and vesicle-mediated are the 2 routes to vacuolar storage. Both types of transporters are primary candidates for catalyzing the vacuolar transport of flavonoids: multidrug and toxic compound extrusion transporters and ATP binding cassette proteins. The membrane transporter-mediated route has been found to consist of GST and ATP binding cassette transporters (Agati et al., 2012). ATP binding cassette-type transporters are much more effective than multidrug and toxic compound extrusion transporters in accumulating flavonoids in the vacuole (Frangne et al., 2002). A recent example of vacuolar pH affecting flower color has been reported. Anthocyanin color depends on vacuolar pH. Intense red rose petals generally have lower vacuolar pH, and turn nearly colorless in the presence of increased pH. In this experiment, the red petals were associated with lower vacuolar pH 4.40, while the white petals were associated with a pH of 4.52 (data not shown). Changing the pH of the pigment extract altered flower color. At pH values below 3, anthocyanins converted from red to pink. At pH values above 3, anthocyanins were typically colorless. This is the hemiacetal form result (Heredia et al., 1998). Thus, pigment color may be pH-dependent. Therefore, regulating vacuolar pH is important for engineering flower color. In this study, the ABCG2 (ppa000185m) gene was downregulated in RF compared to in WF, and the GST (ppa011307m) gene showed inverse expression. This could result in RF accumulating and transporting more anthocyanin to the vacuole compared to WF.

ACKNOWLEDGMENTS

We gratefully acknowledge the Fundamental Research Funds for the Central University (#KYZ201208) and the Jiangsu Province Science Technology Independent Innovation Fund [#CX(12)2011] for providing financial support. This study was sponsored by Qing Lan Project.

[Supplementary material](#)

REFERENCES

- Agati G, Azzarello E, Pollastri S and Tattini M (2012). Flavonoids as antioxidants in plants: location and functional significance. *Plant Sci.* 196: 67-76.
- Anisimov SV (2008). Serial analysis of gene expression (SAGE): 13 years of application in research. *Curr. Pharm. Biotechnol.* 9: 338-350.
- Asmann YW, Klee EW, Thompson EA, Perez EA, et al. (2009). 3' tag digital gene expression profiling of human brain and universal reference RNA using Illumina Genome Analyzer. *BMC Genom.* 10: 531.
- Bartley GE, Scolnik PA and Beyer P (2001). Two *Arabidopsis thaliana* carotene desaturases, phytoene desaturase, and zeta-carotene desaturase, expressed in *Escherichia coli*, catalyze a poly-cis pathway to yield pro-lycopene. *Eur. J. Biochem.* 259: 396-403.
- Bentley DR (2006). Whole-genome re-sequencing. *Curr. Opin. Genet. Dev.* 16: 545-552.
- Bowles D, Lim E-K, Popenberger B and Vaistij FE (2006). Glycosyltransferases of lipophilic small molecules. *Annu. Rev. Plant Biol.* 57: 567-597.
- Breitenbach J, Fernández-González B, Vioque A and Sandmann G (1998). A higher-plant zeta-carotene desaturase in the cyanobacterium *Synechocystis* PCC6803. *Plant Mol. Biol.* 36: 725-732.
- Caputi L, Lim EK and Bowles DJ (2008). Discovery of new biocatalysts for the glycosylation of terpenoid scaffolds. *Chemistry* 14: 6656-6662.
- Czemmel S, Heppel SC and Bogs J (2012). R2R3 MYB transcription factors: key regulators of the flavonoid biosynthetic

- pathway in grapevine. *Protoplasma* 249 (Suppl. 2): S109-S118.
- Eldem V, Celikkol Akcay U, Ozhuner E, Bakir Y, et al. (2012). Genome-wide identification of miRNAs responsive to drought in peach (*Prunus persica*) by high-throughput deep sequencing. *PLoS One* 7: e50298.
- Engelmann NJ, Clinton SK and Erdman JW Jr. (2011). Nutritional aspects of phytoene and phytofluene, carotenoid precursors to lycopene. *Adv. Nutr.* 2: 51-61.
- Frangne N, Eggmann T, Koblischke C, Weissenböck G, et al. (2002). Flavone glucoside uptake into barley mesophyll and *Arabidopsis* cell culture vacuoles. Energization occurs by H⁺-antiport and ATP-binding cassette-type mechanisms. *Plant Physiol.* 128: 726-733.
- Hao QN, Zhou XA, Sha AH, Wang C, et al. (2011). Identification of genes association nitrogen-use efficiency by genome-wide transcriptional analysis of two separate soybean genotypes. *BMC Genom.* 12: 525.
- Heredia F, Francia-Aricha E, Rivas-Gonzalo J, Vicario I, et al. (1998). Chromatic characterization of anthocyanins from red grapes-I. pH effect. *Food Chemistry* 63: 491-498.
- Hoffmann L, Besseau S, Geoffroy P, Ritzenhaller C, et al. (2004). Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinic acid hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *Plant Cell* 16: 1446-1465.
- Kai K, Mizutani M, Kawamura N, Yamamoto R, et al. (2008). Scopoletin is biosynthesized via ortho-hydroxylation of feruloyl CoA by a 2-oxoglutarate-dependent dioxygenase in *Arabidopsis thaliana*. *Plant J.* 55: 989-999.
- Karthikeyan S, Zhou Q, Osterman AL and Zhang H (2003). Ligand binding-induced conformational changes in riboflavin kinase: structural basis for the ordered mechanism. *Biochemistry* 42: 12532-12538.
- Kitamura S, Matsuda F, Tohge T, Yonekura-Sakakibara K, et al. (2010). Metabolic profiling and cytological analysis of proanthocyanidins in immature seeds of *Arabidopsis thaliana* accumulation mutants. *Plant J.* 62: 549-559.
- Machmudah S and Goto M (2013). Methods for extraction and analysis of carotenoids. In: Natural Products (Ramawat KG and Mérillon JM, eds.). Springer, Berlin, pp. 3367-3411.
- Mackenzie PI, Owens IS, Burchell B, Bock K, et al. (1997). The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics* 7: 255-269.
- Martens S, Forkmann G, Britsch L, Wellmann F, et al. (2003). Divergent evolution of flavonoid 2-oxoglutarate-dependent dioxygenases in parsley. *FEBS Lett.* 544: 93-98.
- Mori T and Sakurai M (1995). Effects of riboflavin and increased sucrose on anthocyanin production in suspended strawberry cell cultures. *Plant Sci.* 110: 147-153.
- Morrissy AS, Morin RD, Delaney A, Zeng T, et al. (2009). Next-generation tag sequencing for cancer gene expression profiling. *Genome Res.* 19: 1825-1835.
- Nanni V, Zanetti M, Bellucci M, Moser C, et al. (2012). The peach (*Prunus persica*) defensin PpDFN1 displays antifungal activity through specific interactions with the membrane lipids. *Plant Pathol.* 62: 393-403.
- Sajer O, Scorza R, Dardick C, Zhebentayayeva T, et al. (2012). Development of sequence-tagged site markers linked to the pillar growth type in peach (*Prunus persica*). *Plant Breeding* 131: 186-192.
- 't Hoen PA, Ariyurek Y, Thygesen HH, Vreugdenhil E, et al. (2008). Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Res.* 36: e141.
- Tanaka Y (2006). Flower color and cytochromes P450. *Phytochem. Rev.* 5: 283-291.
- Tanaka Y and Brugliera F (2013). Flower colour and cytochromes P450. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 368: 20120432.
- Tanaka Y, Sasaki N and Ohmiya A (2008). Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J.* 54: 733-749.
- Tanaka Y, Brugliera F and Chandler S (2009). Recent progress of flower color modification by biotechnology. *Int. J. Mol. Sci.* 10: 5350-5369.
- Tanaka Y, Brugliera F, Kalc G, Senior M, et al. (2010). Flower color modification by engineering of the flavonoid biosynthetic pathway: practical perspectives. *Biosci. Biotechnol. Biochem.* 74: 1760-1769.
- Tong Z, Gao Z, Wang F, Zhou J, et al. (2009). Selection of reliable reference genes for gene expression studies in peach using real-time PCR. *BMC Mol. Biol.* 10: 71.
- Ueyama Y, Suzuki KI, Fukuchi-Mizutani M, Fukui Y, et al. (2002). Molecular and biochemical characterization of torenia flavonoid 3'-hydroxylase and flavone synthase II and modification of flower color by modulating expression of these genes. *Plant Sci.* 163: 253-263.
- Wang XG (2006). The Principles and Technology of Experiment in Plant Physiology and Biochemistry. Higher Education Press, Beijing, pp. 134-136.
- Winkel-Shirley B (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* 126: 485-493.
- Wrolstad R, Culbertson J, Cornwell C and Mattick L (1982). Detection of adulteration in blackberry juice concentrates and wines. *J. Assoc. Off. Anal. Chem.* 65: 1417-1423.

- Yamazaki M, Shibata M, Nishiyama Y, Springob K, et al. (2008). Differential gene expression profiles of red and green forms of *Perilla frutescens* leading to comprehensive identification and anthocyanin biosynthetic genes. *FEBS J.* 275: 3494-3502.
- Ye ZW, Jiang JG and Wu GH (2008). Biosynthesis and regulation of carotenoids in *Dunaliella*: progresses and prospects. *Biotechnol. Adv.* 26: 352-360.
- Yonekura-Sakakibara K and Hanada K (2011). An evolutionary view of functional diversity in family 1 glycosyltransferases. *Plant J.* 66: 182-193.
- Yoshida K, Mori M and Kondo T (2009). Blue flower color development by anthocyanins: from chemical structure to cell physiology. *Nat. Prod. Rep.* 26: 884-915.