Genome-wide transcriptome analysis of genes involved in flavonoid biosynthesis between red and white strains of *Magnolia sprengeri* pamp

Shou-Guo Shi¹², Mei Yang¹, Min Zhang¹, Ping Wang¹, Yong-Xiang Kang¹ and Jian-Jun Liu*¹

**Abstract**

**Background:** *Magnolia sprengeri* Pamp is one of the most highly valuable medicinal and ornamental plants of the Magnolia Family. The natural color of *M. sprengeri* is variable. The complete genome sequence of *M. sprengeri* is not available; therefore we sequenced the transcriptome of white and red petals of *M. sprengeri* using Illumina technology. We focused on the identity of structural and regulatory genes encoding the enzymes involved in the determination of flower color.

**Results:** We sequenced and annotated a reference transcriptome for *M. sprengeri*, and aimed to capture the transcriptional determinants of flower color. We sequenced a normalized cDNA library of white and red petals using Illumina technology. The resulting reads were assembled into 77,048 unique sequences, of which 28,243 could be annotated by Gene Ontology (GO) analysis, while 48,805 transcripts lacked GO annotation. The main enzymes involved in the flavonoid biosynthesis, such as phenylalanine ammonia-Lyase, cinnamat-4-Hydroxylase, dihydroflavonol-4-reductase, flavanone 3-hydroxylase, flavonoid-3′-hydroxylase, flavonol synthase, chalcone synthase and anthocyanidin synthase, were identified in the transcriptome. A total of 270 transcription factors were sorted into three families, including MYB, bHLH and WD40 types. Among these transcription factors, eight showed 4-fold or greater changes in transcript abundance in red petals compared with white petals. High-performance liquid chromatography analysis of anthocyanin compositions showed that the main anthocyanin in the petals of *M. sprengeri* is cyanidin-3-O-glucoside chloride and its content in red petals was 26-fold higher than that in white petals.

**Conclusion:** This study presents the first next-generation sequencing effort and transcriptome analysis of a non-model plant from the Family Magnoliaceae. Genes encoding key enzymes were identified and the metabolic pathways involved in biosynthesis and catabolism of *M. sprengeri* flavonoids were reconstructed. Identification of these genes and pathways adds to the current knowledge of the molecular biology and biochemistry of their production in plant. Such insights into the mechanisms supporting metabolic processes could be used to genetically enhance flower color among members of the Magnoliaceae.

**Keywords:** Transcriptome, Flavonoid biosynthesis, *Magnolia sprengeri*, Flower color

**Background**

*Magnolia sprengeri* Pamp is one of the most valuable medicinal and ornamental plants of the Magnolia Family, which is native to the Qinling Mountains of Shaanxi Province and the Daba Mountains of Hubei Province, China [1]. The flower color (i.e., color of the petals) of *M. sprengeri* varies widely from white to red. Flower color has evolved via interaction with evolving pollinating insects. The extreme color variation ranges from pure white color, with only the faintest pale purple stripe on the base of the petals abaxially, to an intense red color on both sides. Previous studies showed that red and white were the two main types of abaxial color patterns. Anthocyanin is the primary flower pigment in higher plants, and its accumulation is tightly linked with flower development and color changes in most cases [2]. Natural phenotypic variations offer an opportunity to elucidate the role of anthocyanin genes that lead to extreme colors of *M. sprengeri*. Anthocyanin
biosynthesis via the flavonoid metabolism pathway has also been well studied in flower development because of its high anti-oxidation properties and ultraviolet protection [3-7]. Anthocyanin is synthesized from phenylalanine, and catalyzed by phenylalanine ammonia-lyase (PAL), which is controlled by two groups of genes. The first group consists of the structural genes, including PAL, chalcone synthase (CHS), flavanone 3-hydroxylase (F3H), dihydroflavonol-4-reductase (DFR), anthocyanidin synthase (ANS), and UDP-glucose flavonoid 3-O-glucosyltransferase (UFGT), which represent the enzymes responsible for the biochemical reactions of anthocyanin synthesis [8-13] (Figure 1). The second group involves regulatory genes or transcription factors (TFs), which regulate the structural genes expressions during anthocyanin biosynthesis. The important TFs for anthocyanin biosynthesis belong to the MYB, bHLH, and WD40 families [14-21]. The coordinated expressions of the two categories of genes lead to anthocyanin accumulation during the color development process.

Transcriptome analysis of an organism is a particularly effective method for gene discovery, especially in non-model plants for which no reference genome sequences are available [22]. At the same time, it may provide powerful tools to identify differentially expressed genes, and its possible use in modern plant breeding continues to attract the attention of many plant biologists [23-26].

Sequencing technologies have dramatically accelerated genome-wide studies of transcriptomes and have been widely used to explore gene structure and gene expression, even in plants without a genome reference [27-29]. Illumina sequencing technology has been applied recently to transcriptome analyses of plant and animals, and can generate large amounts of sequence data cheaply and quickly [30-33].

In this study, we first sequenced the transcriptomes of white and red petals of M. sprengeri using Illumina technology (Figure 2). We focused on the discovery of structural genes and regulatory genes encoding enzymes involved in the anthocyanin biosynthetic pathway. We obtained sets of upregulated and downregulated genes from red and white flowers, and identified some candidate genes related to anthocyanin synthesis in M. sprengeri. The assembled annotated transcriptome sequences provide a valuable genomic resource to further understand the molecular basis of variations of flower color in M. sprengeri.

Results and discussion

Sequencing and sequence assembly

A cDNA library from red and white petals was sequenced using Illumina sequencing in a single run which generated 39,652,898 sequences with 4,004,942,698 nucleotides (bp) from red petals and 68,698,774 sequences with 4,004,942,698 nucleotides (bp) from red petals and 68,698,774 sequences with 4,004,942,698 nucleotides (bp).
6,938,576,174 bp from white petals. After the removal of the low-quality raw reads (too short, empty, too many Ns), we obtained 39,652,898 (red) and 68,698,774 (white) high-quality sequences (Table 1). These short reads were assembled into 39,990 total genes, 77,048 total isogenes, and 91,433,742 total resides with an average length of 1,186.71 bp. The largest isogene was 15,888 bp and the smallest was 351 bp (Table 2). The sequence length distribution is shown in Figure 3.

In this study, we obtained 35,642,032 sequences and 62,964,028 high-quality sequences in red and white petals of *M. sprengeri*, respectively (Table 2). By comparison, the assembly of 39,990 *M. sprengeri* sequences from GenBank (using the GS De novo Assembler) led to only 77,048 unique sequences. The unique sequences derived from GenBank sequences and Illumina sequences were compared by a BLAST search, where matches were defined as having an identity >90% and an overlap >100 bp. Our Illumina sequencing efforts produced 77,048 unique sequences. Unique sequences that were not present in GenBank were considered as the novel transcripts of *M. sprengeri*. The large quantity of unique sequences should cover the vast majority of genes from *M. sprengeri* petals, providing, for the first time, a powerful gene resource for this medicinal and ornamental plant.

**Gene ontology (GO) annotation**

GO annotation provides a description of gene products in terms of their associated molecular functions, cellular components, and biological processes [34]. GO functional interpretations for plants are primarily based on the *Arabidopsis thaliana* genome. GO terms were assigned to 28,243 *M. sprengeri* sequences based on sequence similarities with known proteins and annotated using The Arabidopsis Information Resource (TAIR) using GO slim. Among the 77,048 spliced transcripts, 28,243 had GO annotated transcripts while 48,805 had no annotated transcripts (Figure 4). The GO annotations of the unique sequences were most frequently related to molecular function (23,227 unique sequences), followed by biological processes (20,540 unique sequences) and cellular components (16,690 unique sequences). For each sequence, the specifically annotated GO terms provide a broad overview of the groups of genes cataloged in the transcriptome. Finally, the functions of the identified transcripts were determined to be involved different biological processes.

**Clustering of orthologous group (COG) and eukaryote clusters of orthologous groups (KOG) classification**

The GO analysis identified well-represented categories within the cellular component group, including sequences related to the chloroplast, the plasma membrane, and the ribosome. Additionally, the sequences encoded a broad set of transcripts that could be assigned to molecular function categories. To further examine the integrity of our transcriptome library and the effectiveness of the annotation process, we identified the unigene numbers with COG and KOG classification. Altogether, there were 25,626 unigenes identified from all unigenes with COG (Figure 5) and KOG classifications (Figure 6). Among the 24 COG categories, the cluster of “General function prediction” accounted for the largest proportion (3,823, 14.9%) followed by “Replication, recombination and repair” (2,120, 8.3%), “Transcription” (2,102, 8.2%) and “Signal transduction mechanisms” (1,792, 7.0%). The categories of “Nuclear structure” (7) accounted for 0.03%, “Secondary metabolites biosynthesis, transport and catabolism” (437) accounted for 1.7%, and “RNA processing and modification” (254) accounted for 1.0%. Among the 25 KOG categories, the cluster of “General function prediction” accounted for the largest proportion (3,682, 14.4%), followed by “Signal transduction mechanisms” (2,346, 9.2%), “Posttranslational modification, protein turnover, chaperones” (1,971, 7.7%) and “Transcription” (1,388, 5.4%). The categories of “Nuclear structure” (90, 0.35%), “Secondary metabolites biosynthesis, transport and

### Table 1 Summary of sequencing for *M. sprengeri*

| Sequencing     | No. of sequences | No. of bases  | No. of high-quality reads | No. of bases  |
|----------------|------------------|--------------|---------------------------|--------------|
| Red petal      | 39,652,898       | 4,004,942,698| 35,642,032                | 30,265,311,514|
| White petal    | 68,698,774       | 6,938,576,174| 62,964,028                | 5,246,524,133|

### Table 2 Splicing results for *M. sprengeri*

| Type                  | Sum   |
|-----------------------|-------|
| Total genes (n)       | 39,990|
| Total isogenes (n)    | 77,048|
| Total resides (bp)    | 91,433,742|
| Average length (bp)   | 1,186.71|
| Largest isogene (bp)  | 15,888|
| Smallest isogene (bp) | 351    |

The best represented groups were protein metabolism, developmental processes, response to abiotic or biotic stimuli, and response to stress and transport. These GO annotations provided valuable clues to investigate the specific processes, molecular functions, and cellular structures of the *M. sprengeri* transcriptome.
catabolism (519, 2.0%), “RNA processing and modification” (1,121, 4.4%), and “Extracellular structures” (62, 0.24%) had the least proportions.

Pathway assignment based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) classification system

The KEGG classification system provides an alternative functional annotation of genes according to their associated biochemical pathways [35]. KEGG annotations for M. sprengeri transcripts were based on sequence similarity searches against the KEGG database, and matches were assigned the corresponding enzyme commission (EC) number. Overall, 12,082 M. sprengeri unique sequences were assigned KEGG annotations, of which only 1,696 unique sequences were assigned to the biosynthesis of secondary metabolites pathways.

Metabolic pathways were well represented among M. sprengeri unique sequences, most of which were associated with amino acid metabolism, galactose metabolism, biosynthesis of secondary metabolites, aminoacyl-tRNA biosynthesis, and flavonoid biosynthesis. Notably, the transcripts encoding all the enzymes involved in the

Figure 3 Length distribution of sequencing reads and contigs of M. sprengeri.

Shi et al. BMC Genomics 2014, 15:706
http://www.biomedcentral.com/1471-2164/15/706

Figure 4 Histogram representing Gene Ontology (GO) classification. GO categories, shown on the x-axis, were grouped into three main ontologies: biological process, cellular component, and molecular function. The right y-axis indicates the number of genes in each category, while the left y-axis indicates the percentage of total genes in that category. The ‘all gene’ indicates that the unigenes were those assembled from reads from the red and white sample.
Figure 5 Histogram representing clusters of orthologous groups (COG) classification. A total of 25,626 unigenes were assigned to 24 categories in the COG classification. The y-axis ‘Frequency’ indicates the number of genes in a specific functional cluster. The right side legend shows a description of the 24 functional categories.

Figure 6 Histogram representing clusters of orthologous groups (KOG) classification. 25,626 unigenes were assigned to 25 categories by KOG classification. The y-axis ‘Frequency’ indicates the number of genes in a specific function cluster. The legend on the right shows a description of the 25 functional categories.
flavonoid biosynthesis pathway were present in our Illumina sequences dataset (Table 3). The expression of all candidate genes in the petals (Additional file 1: Figure S2) was confirmed by quantitative real-time PCR (qPCR) (Figure 7).

**Candidate genes encoding enzymes involved in the biosynthesis of flavonoids**

Flavonoid biosynthesis is an integral part of secondary metabolism; therefore, it should be considered within the context of cellular metabolism. The color of the petal is affected by the flavonoids metabolism pathways. Changes in the transcript abundances of the genes encoding enzymes in these pathways are listed in Tables 3 and 4.

The levels of transcripts encoding the first enzymes in the flavonoids biosynthesis, such as PAL (EC.4.3.1.5), were markedly higher in red petals than in white petals. The transcript abundance of the flavonoids biosynthesis enzymes, including C4H (EC.1.14.13.11), DFR (EC.1.1.1.219), F3H (EC.1.14.11.9), flavonoid-3′-hydroxylase (F3′H), flavonol synthase (FLS; EC.1.14.11.23), CHS (EC.2.3.1.74), and ANS (EC.1.14.11.19), were also higher in red petals.

The Plant Transcription Factor Database was used to search the *M. sprengeri* transcripts dataset to identify the genes encoding putative TFs or transcriptional regulators [36]. A total of 270 transcripts were predicted to be TFs and were sorted into three families (data not shown). Of these genes, the expression of eight MYB genes showed 4-fold or greater changes in red petals compared with white petals. Further studies are needed to determine whether the changes in transcript abundance of these putative TFs could be related to the regulation of flavonoid metabolism.

Flavonoids are a large group of polyphenolic compounds and are a structurally diverse class of plant secondary metabolites. They are important for defense against pathogens and herbivores, protection from harmful ultraviolet radiation, and flower pigmentation for attracting pollinators [37-39]. In addition to their physiological functions in plants, flavonoids display a wide range of anti-oxidant, anti-microbial, anti-inflammatory, and anti-cancer activities [39]. As a dietary component, flavonoids are considered to have health-promoting and disease-preventing properties. Recently, flavonoids have been intensively investigated as potent pharmaceuticals for treating chronic human pathological conditions [40-44].

According to the present transcriptomic analysis, the expression of eight genes was altered in flavonoid biosynthesis, including PAL, C4H, F3H, FLS, CHS, DFR, and ANS. It has been known for decades that flavonoids are synthesized from phenylalanine. This process involves three steps: first, PAL, cinnamic acid 4-hydroxylase, and 4-coumarate coenzyme A ligase catalyze the conversion of phenylalanine to; second, CHS, CHI, F3H, F3′H, and DFR catalyze the conversion of *p*-coumaroylCoA to leucoanthocyanidins; and third, ANS catalyzes the conversion of leucoanthocyanidins to form anthocyanidins [9]. In this process, any changes in the expression of the genes encoding these enzymes can lead to the production of different anthocyanidin species [45-47], and such changes may lead to the production of the red petal phenotype of *M. sprengeri*. Our study also identified certain genes encoding TFs related to flavonoid biosynthesis, such as MYBs, bHLHs, and WD40s. The differentially expressed structural genes and differentially expressed TF genes that may be associated with flavonoid biosynthesis are shown in Tables 4 and 5. Such information would help provide a deeper understanding of how changes in gene expression are related to the changes in the color of *M. sprengeri* flowers.

Based on this comparison, almost all of the candidate genes involved in the flavonoid biosynthesis were present in the transcriptome datasets of *M. sprengeri* in this study. These results highlight the immense capacity of high-throughput sequencing to discover genes in metabolic pathways.

**Anthocyanin accumulation**

Anthocyanin accumulation is tightly linked with flower development and color changes in most cases [2]. To

---

**Table 3 Genes encoding enzymes involved in flavonoid biosynthesis in M. sprengeri**

| Name | Description                  | Number of transcripts KO no. | EC no.   |
|------|------------------------------|------------------------------|----------|
| PAL  | PhenylalanineAmmonia-Lyase   | K10775                       | EC.4.3.1.5|
| C4H  | Cinnamat-4-Hydroxylase       | K13065                       | EC.1.14.13.11|
| DFR  | DiHydroflavonol-4-Reductase  | K13082                       | EC.1.1.1.219|
| F3H  | Flavanone 3-Hydroxylase      | K00475                       | EC.1.14.11.9|
| F3′H | Flavonoid-3′-Hydroxylase     | K05280                       | EC.1.14.13.21|
| CHI  | Chalcone Isomerase           | K01859                       | EC.5.5.1.6|
| FLS  | Flavonol Synthase            | K05278                       | EC.1.14.11.23|
| CHS  | Chalcone Synthase            | K00660                       | EC.2.3.1.74|
| ANS  | Anthocyanidin Synthase       | K05277                       | EC.1.14.11.19|
examine the accumulation of anthocyanin in the petals of red and white _M. sprengeri_, the flower extracts were subjected to high-performance liquid chromatography (HPLC) analysis. The HPLC data showed that the main anthocyanin in the petals of _M. sprengeri_ is cyanidin-3-O-glucoside chloride. The accumulation of cyanidin-3-O-glucoside chloride in petals of red and white _M. sprengeri_ was 3.421 and 0.132 mg per 100 g samples, respectively (Additional file 2: Figure S1). Our results indicate that red petals accumulate 26-fold more cyanidin-3-O-glucoside chloride than white petals (Figure 8).

**Conclusions**

Illumina next-generation sequencing technology was used for sequencing and transcriptome analysis of the non-model plant _M. sprengeri_ pamp. We identified the

| Transcript       | Annotation                  | FPKM-R | FPKM-W |
|------------------|-----------------------------|--------|--------|
| comp55739_c0_seq2 | Phenylalanine Ammonia-Lyase | 967.65 | 454.39 |
| comp53876_c0_seq1 | Cinnamat-4-Hydroxylase     | 447.91 | 143.63 |
| comp43573_c0_seq2 | Dihydroflavonol-4-Reductase | 264.97 | 7.46   |
| comp47103_c0_seq1 | Flavanone-3-Hydroxylase    | 420.34 | 389.55 |
| comp49650_c0_seq1 | Flavonol synthase          | 786.51 | 299.47 |
| comp50438_c0_seq1 | Chalcone synthase          | 705.94 | 225.14 |
| comp51086_c0_seq1 | Anthocyanidin reductase    | 344.97 | 203.23 |

FPKM-R, fragments per kilobase of transcript per million fragments mapped red; FPKM-W, fragments per kilobase of transcript per million fragments mapped white.
genes encoding key enzymes and reconstructed the metabolic pathways involved in biosynthesis and catabolism of flavonoid of M. sprengeri. Our results promote understanding of the mechanisms underlying various metabolic processes, and will enable the genetic manipulation of flower color in M. sprengeri.

The accumulation of flavonoids and the discovery of genes associated with their biosynthesis and metabolism in M. sprengeri are intriguing and worthy of further investigation. The sequences and pathways identified here represent the genetic framework required for further studies. Quantitative transcriptomics in concert with physiological and biochemical analysis in M. sprengeri under conditions that stimulate production and accumulation of flavonoids could help provide insights into the regulation of, and links between, these pathways.

### Methods

#### Plant materials

The petals of red and white M. sprengeri were harvested from approximately 50-year-old trees in March 2012 from Wufeng County, Hubei Province, China (Figure 2). We selected 10 trees with red flowers and 10 trees with white flowers for petal collection. Nine petals of each color were selected for RNA-sequencing experiments while three petals (around 0.5 g) of each color were taken for the HPLC experiments. For qPCR, we used additional 5–10 petals to isolate total RNA. After cleaning, the petals were cut into small pieces, immediately frozen in liquid nitrogen, and stored at −80°C until further processing.

#### RNA preparation

The TRizol® reagent (Invitrogen) was used to extract total RNA from the petals of red and white M. sprengeri according to the manufacturer’s instructions (Invitrogen, USA). The purity of all RNA samples was assessed at an absorbance ratio of OD260/280 and the RNA quality was tested using a 1% ethidium bromide-stained (EtBr-stained) agarose gel. A GeneQuant100 spectrophotometer (GE Healthcare, UK) assessed the RNA concentration before processing.

#### cDNA synthesis and Illumina sequencing

Clontech’s SMART cDNA synthesis kit (Clontech, USA), was used to produce first-strand cDNA from 5 μg of total RNA extracted from the petals of M. sprengeri, according to the manufacturer’s instructions. The samples were treated with RNase-free DNase I (Takara Biotechnology, China). To construct a cDNA library, oligo (dT) magnetic beads were used to purify poly (A) mRNA from total RNA. The RNA was then fragmented into small pieces by the addition of fragmentation buffer. These short fragments served as templates to synthesize first-strand cDNA using random hexamer primers. Second-strand cDNA was synthesized using buffer, dNTPs, RNaseH, and DNA polymerase I. A QiaQuick PCR extraction kit purified the

### Table 5 Changes in transcript abundance of predicted transcription factors and regulators about flavonoid biosynthesis in M. sprengeri red and white

| Transcript | Annotation | FPKM-R | FPKM-W |
|------------|------------|--------|--------|
| comp25552_c0_seq1 | MYB domain protein 20 | 2.54 | 0.7 |
| comp45264_c0_seq2 | R2R3 MYB transcription factor | 13.27 | 0.35 |
| comp45298_c0_seq2 | R2R3-MYB transcription factor MYB9 | 9.24 | 1.02 |
| comp46821_c0_seq4 | MYB6 | 39.37 | 9.38 |
| comp47225_c0_seq1 | MYB-related protein 306 | 38.92 | 7.95 |
| comp47225_c1_seq1 | MYB-related protein 306 isoform 1 | 42.62 | 13.5 |
| comp47747_c0_seq2 | MYB transcription factor | 89.68 | 17.14 |
| comp50161_c1_seq1 | MYB-related protein 306 isoform 2 | 288.96 | 72.35 |

FPKM-R, fragments per kilobase of transcript per million fragments mapped red; FPKM-W, fragments per kilobase of transcript per million fragments mapped white.
short fragments. These fragments were washed with elution buffer for end repair and poly (A) addition and then ligated to sequencing adapters. Suitable fragments, as judged by agarose gel electrophoresis, were selected for use as templates for PCR amplification. An Illumina HiSeq™2000 sequenced the cDNA library using paired-end technology in a single run.

Transcriptome assembly and annotation
The Solexa GA pipeline 1.6 generated the transcriptome de novo assembly. After the removal of low-quality reads, the Trinity de novo assembler (http://trinityrnaseq.sourceforge.net/) [48,49] assembled processed reads with an identity value of 95% and a coverage length of 100 bp [48,49]. First, the overlap information in the short reads was used to construct high-coverage contigs, and then the short reads were assembled into contigs. We then realigned the short reads onto the contigs and estimated the distance and relation of the two contigs using the pair-end linkage and insert size information. Unreliable linkages between the two contigs were filtered and the remaining contigs with compatible connections were linked to each other, and had at least three read-pairs. The last step was to close gaps in the scaffolds. We gathered the paired-end reads with one end mapped to the contigs and another end located in the gaps and performed local assembly with the unmapped end to extend the contig sequence into the small gaps in the scaffolds. CAP3 [50] was used (with default parameters) to reduce redundancy and to combine scaffolds and single-end contigs in the separate assemblies.

To annotate the *M. sprengeri* transcriptome, we performed a BLAST search against the non-redundant (NR) database in NCB!@, SWISS-PROT, KEGG, and COG with a cut-off E-value of \( \leq 10^{-5} \). The Blast2GO software (http://www.blast2go.com/b2ghome) obtained the GO annotations and the corresponding EC numbers of the sequences.

Pathway assignment with KEGG
Pathway assignments were mapped according to the KEGG database (http://www.genome.ad.jp/kegg/kegg2.html) (version KEGG) [51]. EC numbers were assigned to unique sequences that had BLASTX scores with an E value cut-off of \( 10^{-5} \) after searching the KEGG protein databases. The unique sequences were mapped to specific biochemical pathways according to the corresponding EC distribution in the KEGG database.

Quantitative Real-time PCR (qPCR) Analyses
To remove any contaminating genomic DNA before cDNA synthesis, we treated the total RNA with RNase-free DNase I (Invitrogen, USA) according to the manufacturer’s instructions. A NanoDrop™ 1000 spectrophotometer was used to quantify the RNA before and after this DNase I treatment, and RNA quality and integrity were checked by electrophoresis using agarose gels stained with ethidium bromide. For qPCR, first-strand cDNA was synthesized with 2 \( \mu \)g of total RNA in a volume of 20 \( \mu \)L, using a SYBR®PrimeScript™ RT-PCR Kit II (Takara, China) plus random hexamers and oligo(dT) primers. After reverse-transcription, the reaction product was diluted 10-fold with sterile water. Real-time PCR was performed on an iQ5 instrument (Bio-Rad, USA) using SYBR Green qPCR kits (Takara, China) according to the manufacturer’s instructions. Primer sequences are listed in Table 6. Real-time PCR was performed on an iQ5 instrument (Bio-Rad, USA) using SYBR Green qPCR kits (Takara, China) according to the manufacturer’s instructions. Primer sequences are listed in Table 6. Real-time PCR was performed on an iQ5 instrument (Bio-Rad, USA) using SYBR Green qPCR kits (Takara, China) according to the manufacturer’s instructions. Primer sequences are listed in Table 6.

Table 6 Primers used in quantitative real-time PCR

| Gene ID     | Primer sequence(forward) 5'-3' | Primer sequence(reverse) 5'-3' |
|-------------|---------------------------------|--------------------------------|
| 55739       | TAACGGAGCCGGAACAGGAA            | GAGAAATTGGCCGCAATCA            |
| 53876       | ACGCACCCTGCCGGCAGT              | ATTTCAACCTTGTCATCTT            |
| 43573       | TCGTGGAGCGTCCGGAGGAC            | AGCGTGATGGTCGCCAGGTC           |
| 47103       | GGCGTCGTTGGCTGGATGT            | TCCCGCTAGTGGATTTGGGAGG         |
| 49650       | GCAGAACAATCATCCACCTCCTCAA       | CAGGACGACGAATAAACAAGGAGA       |
| 50438       | GCAGCCATCAGCAACTAC             | AATAATCTCCCTCCCTACCA           |
| 51086       | GGGGACTCTACACACAGGAA           | CTAACCGGAGAGATATTGAC           |
| 25552       | GCCTCACTAAAACCTCTACTTTCC       | TGGACCTTCTGACGCTTA             |
| 45264       | CAATCCGCTCCGTAAGGAC            | CCCCGTGTAATGGAGAA              |
| 45298       | TTTATTTAGTGGCAGAA              | ATTACGGATGTGGCAGGG             |
| 46821       | GCCGAATCATACTCTTGAGCAT          | TCTGGCCTTGCAGCTG              |
| 47225       | ATGCCTAGTGGATTGTTG             | CCATGGTACCTGAGGAGAGAGG         |
| 47747       | CCGAAGAAGATGGCAGAAA            | GAGGACGACGAGGATTCA             |
| 50161       | GCTGTTGGTGGGAGAGGA             | TACGCAGATGGTGCGGAG             |
| 53953(Actin)| GGGCTGATTGTTGCTGGAGAG          | GTGGTGCTGGTCGGTGGAG            |
time PCR reactions were carried out in 20-μL volumes containing a 10-μM concentration of each primer, 40 ng of cDNA, and 10 μL of SYBR Premix Ex Taq™ II. Thermal-cycling conditions included an initial heat-denaturing step at 95°C for 3 min; then 40 cycles of 95°C for 20 s, 58°C for 20 s, and 72°C for 20 s. Fluorescence was measured at the end of each cycle. A melting-curve analysis was performed by heating the PCR product from 58°C to 95°C. Expression data were presented as relative units after normalization to the act control, using the 2^△△CT method. Values for mean expression and SD were calculated from the results of three independent experiments.

HPLC analysis of anthocyanin
Magnolia sprengeri petals (0.5 g) were ground in 1.5 mL of 70% methanol containing 2% formic acid at 4°C, then centrifuged at 10,000 g for 10 min at 4°C. The supernatant was passed through a 0.22-μm syringe filter before HPLC analysis. Anthocyanins were investigated on an Agilent 1100 HPLC equipped with a diode array detector (Agilent Technology), as described by Zhang et al. [52]. The total anthocyanin concentration was calculated based on a cyanidin-3-O-glucoside standard (Sigma-Aldrich, St. Louis, MO, USA).

Additional files

Additional file 1: Full-length cDNA sequences of genes used in qPCR assay.
Additional file 2: Anthocyanidin analysis of petals in red and white flower color. Peaks of HPLC were identified by retention time compared with cyaniding-3-O-glucoside chloride standards. (A: Red petals; B: White petals).

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
SGS and MY were responsible for collecting tissue samples and the HPLC analysis. SGS, MZ, and PW performed the RNA extractions and preparation, SGS and MY were responsible for collecting tissue samples and the HPLC analysis. SGS, JJL, and YXK contributed to the analysis. SGS, MZ, and PW performed the RNA extractions and preparation, SGS and MY were responsible for collecting tissue samples and the HPLC analysis. SGS, MZ, and PW performed the RNA extractions and preparation, SGS and MY were responsible for collecting tissue samples and the HPLC analysis. SGS, MZ, and PW performed the RNA extractions and preparation, SGS and MY were responsible for collecting tissue samples and the HPLC analysis.

Acknowledgments
This work was supported by the National Forestry Research and Special Public Service Sectors (No. 200904004).

Author details
1 College of Forestry, Northwest A&F University, Yangling, Shaanxi 712100, China. 2Life Sciences Department, Yuncheng University, Yuncheng, Shanxi 044000, China.

Received: 22 March 2014 Accepted: 21 August 2014 Published: 23 August 2014

References
1. Kang YX, Ejder E: Magnolia sprengeri Pamp.: morphological variation and geographical distribution. Plant Biosyst 2011, 145:906–923.
2. Weiss D: Regulation of flower pigmentation and growth: multiple signaling pathways control anthocyanin synthesis in expanding petals. Physiol Plant 2000, 110:152–157.
3. thal 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, 166:1500–1513.
4. Iriani N, Grotewold E: Light-induced morphological alteration in anthocyanin-accumulating vacuoles of maize cells. BMC Plant Biol 2005, 5:7.
5. Pasco P, Barton H, Zagrobski P, Gorinstein S, Foltz 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.
6. Ordidge M, García-Macías P, Battey NH, Gordon MH, John P, Lovegrove JA, Wysini 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, 91:1597–1604.
7. 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 2007, 45:1271–1276.
8. Koes RE, Quattrocchio F, Mol JNM: The flavonoid biosynthetic pathway in plants: function and evolution. Biosess 1994, 16:123–132.
9. Horation TA, Cornish EC: Genetics and biochemistry of anthocyanin biosynthesis. Plant Cell 1995, 7:1071–1083.
10. Winkel-Shirley B: Flavonoid biosynthesis: a colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol 2001, 126:485–493.
11. Tan JF, Wang MJ, Tu LL, Nie YC, Lin YJ, Zhang XL: The Flavonoid Pathway Regulates the Petal Colors of Cotton Flower. Plo S One 2013, 8:e27364.
12. Chen SM, Li CH, Zhu XR, Deng YM, Sun W, Wang LS, Chen FD, Zhang Z: The identification of flavonoids and the expression of genes of anthocyanin biosynthesis in the chrysanthemum flowers. Biol Plant 2012, 56:458–464.
13. Huang W, Sun W, Wang Y: Isolation and molecular characterisation of flavonoid 3′-hydroxylase and flavonoid 3′-hydroxylase genes from a traditional Chinese medicinal plant, Epimedium sagittatum. Gene 2012, 497:125–30.
14. Ramsay NA, Walker AR, Mooney M, Gray JC: Two basic-helix-loop-helix genes (MYC-146 and GL3) from Arabidopsis can activate anthocyanin biosynthesis in a white-flowered Matthiola incana mutant. Plant Mol Biol 2003, 52:679–688.
15. Bogs J, Jaffe FW, Takos AM, Walker AR, Robinson SP: The grapevine transcription factor VvMYBPA1 regulates proanthocyanidin synthesis during fruit development. Plant Physiol 2007, 143:1347–1363.
16. Dubos C, Le Gourrierec J, Baudry A, Hueso F, Lepiniec L, Routaboul JM, Alboresi A, Weisshaar B, Lepiniec L: MYBL2 is a new regulator of flavonoid biosynthesis in Arabidopsis thaliana. Plant J 2008, 55:940–953.
17. Gonzalez A, Zhao M, Leavitt JM, Lloyd AM: Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. Plant J 2008, 53:814–827.
18. Palapol Y, Ketsa S, Lin-Wang K, Ferguson I, Allan AC: A MYB transcription factor regulates anthocyanin biosynthesis in mangosteen (Garcinia mangostana L) fruit during ripening. Planta 2009, 229:1233–34.
19. Zhou LL, Shi MZ, Xie DF: Regulation of anthocyanin biosynthesis by nitrogen in TG1–GL3/TT8–PAP1-programmed red cells of Arabidopsis thaliana. Planta 2012, 236:825–837.
20. Huang YJ, Song S, Allan AC, Liu XF, Yin XR, Xu CJ, Chen KS: Differential activation of anthocyanin biosynthesis in Arabidopsis and tobacco over-expressing an R2R3 MYB from Chinese bayberry. Plant Cell Tiss Organ Cult 2013, 113:491–499.
21. Li Y, Zhang T, Shen ZW, Xu Y, Li JY: Overexpression of maize anthocyanin regulatory gene Lc affects rice fertility. Biotechnol Lett 2013, 35:115–119.
22. Parkinson JJ, Blaxter M: Expressed sequence tags: an overview. Methods Mol Biol 2009, 533:1–12.
23. Shadeo A, Chari R, Vatcher G, Campbell J, Lonergan KM, Matisic J, van Nierkerk D, Ehlen T, Miller D, Follen M, Lam WL, Macaulay C: Comprehensive serial analysis of gene expression of the cervical transcriptome. BMC Genomics 2007, 8:142.
24. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009, 10:57–63.

25. Tao X, Gu YH, Wang HY, Zheng W, Li X, Zhao CW, Zhang Z. Digital gene expression analysis based on integrated de novo transcriptome assembly of sweet potato (Ipomoea batatas (L.) Lam). PLoS One 2012, 7:e26334.

26. Wu P, Qin Z, Zhao W, Zhou X, Wu T, Xin M, Guo Q. Transcriptome analysis reveals differentially expressed genes associated with propamocarb response in cucumber (Cucumis sativus L) fruit. Acta Physiol Plant 2013, 35:2393–2406.

27. Chen S, Luo H, Li Y, Sun Y, Wu Q, Niu Y, Song J, Lv A, Zhu Y, Sun C, Steinmetz A, Qian Z. 454 EST analysis detects genes putatively involved in ginsenoside biosynthesis in Panax ginseng. Plant Cell Rep 2011, 30:1593–1601.

28. Shen GM, Dou W, Niuz JZ, Jiang HB, Yang WJ, Jia FX, Hu F, Cong L, Wang JJ. Transcriptome analysis of the oriental fruit fly (Bactrocera dorsalis). Plos One 2011, 6:e29127.

29. Liu B, Jiang G, Zhang Y, Li J, Li X, Yue J, Chen F, Liu H, Li H, Zhu S, Wang J. Ran C. Analysis of transcriptome differences between resistant and susceptible strains of the citrus red mite panonychus citri(acaec. tenuicaudatae). Plos One 2012, 7:e36234.

30. Deluc LG, Grimplet J, Wheatley MD, Tillett RL, Qullici DR, Osborne C, Schooley DA, Schlauch KA, Cushman JC, Cramer GR. Transcriptomic and metabolite analyses of Cabernet Sauvignon grape berry development. BMC Genomics 2007, 8:29.

31. Zenoni S, Ferrarini A, Giacomelli E, Yurmeck L, Fasoli M, Malerba G, Bellini D, Pezzotti M, Delledonne M. Characterization of transcriptomic complexity during berry development in Vitis vinifera using RNA-Seq. Plant Physiol 2010, 152:1787–95.

32. Sweetman C, Wong DC, Ford CM, Drew DP. Transcriptome analysis at four developmental stages of grape berry (Vitis vinifera cv. Shiraz) provides insights into regulated and coordinated gene expression. BMC Genomics 2012, 13:691.

33. Fan XD, Wang JQ, Yang N, Dong YZ, Liu L, Wang FW, Wang N, Chen H, Liu WC, Sun YP, Wu JY, Li HY. Gene expression profiling of soybean leaves and roots under salt, saline-alkali and drought stress by high-throughput Illumina sequencing. Gene 2013, 512:392–402.

34. Berardini TZ, Mundodi S, Reiner L, Huala E, Garcia-Hernandez M, Zhang P, Mueller LA, Yoon J, Doyle A, Lander G, Moseyko N, Yoo D, Xu J, Ziockier B, Montoya M, Miller N, Weems D, Rhee SY. Functional annotation of the Arabidopsis genomeseing controlled vocabularies. Plant Physiol 2004, 135:745–755.

35. Kanekia M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic Acids Res 2004, 32:D227–D230.

36. Perez-Rodriguez P, Riaho-Pachon DM, Caneda LG, Reising SA, Kersten B, Mueller-Rebour B, Pf FTDB: updated content and new features of the plant transcription factor database. Nucleic Acids Res 2010, 38(suppl 1):D822–D827.

37. Winkel-Shirley B. Biosynthesis of flavonoids and effects of stress. Curr Opin Plant Biol 2002, 5:18–223.

38. Williams CA, Grayer RJ. Anthocyanins and other flavonoids. Nat Prod Rep 2004, 21:539–73.

39. Buer CS, Imin N, Djordjevic MA. Flavonoids: new roles for old molecules. J Integr Plant Biol 2010, 52:98–111.

40. Greenwald P. Clinical trials in cancer prevention: current results and perspectives for the future. J Nutr 2004, 134:3507S–3512S.

41. Hou DX, Fuji M, Tahrara N, Yoshimoto M. Molecular mechanisms behind the chemopreventive effects of anthocyanidins. J Nutr Biochem 2004, 15:321–325.

42. Allison EM, Borradaile NM, Edwards JY, Huy MF. Inhibition of microsomal triglyceride transfer protein expression and apolipoprotein B100 secretion by the citrus flavonoid naringenin and by insulin involves activation of the mitogen-activated protein kinase pathway in hepatocytes. Diabetes 2005, 54:1676–83.

43. Pospolítková J, Polkowský K, Skierská JS, Mazurek AP. In vitro toxicity evaluation in the development of new anticaner drugs-genistein glycosides. Cancer Lett 2005, 229:67–75.

44. Woo Y, Chiu SY, Hyun J, Lee SD, Lee YH, Lim Y. Flavonones inhibit the clonogenicity of HCT116 colorectal cancer cells. Int J Mol Med 2012, 29:403–408.

45. Reddy VS, Dash S, Reddy AR. Anthocyanin pathway in rice (Oryza sativa L): identification of a mutant showing dominant inhibition of anthocyanins in leaf and accumulation of proanthocyanidins in pericarp. Theor Appl Genet 1995, 91:301–312.

46. Ronald K, Verweij W, Qustracchio F. Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci 2005, 10:236–242.

47. Huang W, Sun W, Lv H, Luo M, Zeng S, Pattanakis S, Yuan L, Wang Y. A R2R3-MYB transcription factor from Epimedium sagittatum regulates the flavonoid biosynthetic pathway. Plos One 2013, 8:e70778.

48. Price DP, Nagarajan V, Churbanov A, Houdé P, Milligan B, Drake L, Gustafsson JE, Hansen IA. The fat body transcriptomes of the yellow fever mosquito Aedes aegypti, pre- and post- blood meal. Plos One 2013, 8:e22573.

49. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Ratakta R, Zeng QD, Chen ZH, Mauceli E, Hacochen N, Gnirke A, Rhind N, Palma FD, Birnir BW, Nussbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnology 2011, 29:444–452.

50. Huang X, Madan A. CAP3: A DNA sequence assembly program. Genome Res 1999, 9:686–877.

51. Kanekia M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Kataba T, Kadowa S, Okuda S, Tokumatsu T, Yamashita Y. KEGG for linking genomes to life and the environment. Nucleic Acids Res 2008, 36:D480–484.

52. Zhang Y, Li P, Cheng L. Developmental changes of carbohydrates, organic acids, amino acids, and phenolic compounds in ‘Honeicip’ apple flesh. Food Chem 2010, 123:1013–1018.