



# Transcriptome analysis reveals anthocyanin acts as a protectant in *Begonia semperflorens* under low temperature

Huitao Bi<sup>1</sup> · Meili Guo<sup>1</sup> · Jiawan Wang<sup>1</sup> · Ying Qu<sup>1</sup> · Wenli Du<sup>2</sup> · Kaiming Zhang<sup>1</sup>

Received: 3 October 2016 / Revised: 12 July 2017 / Accepted: 18 November 2017 / Published online: 12 December 2017  
© Franciszek Górski Institute of Plant Physiology, Polish Academy of Sciences, Kraków 2017

## Abstract

Anthocyanins are natural bioactive pigments in plants that play important roles in many physiological functions. They are found in various tissues and can protect plants against different stress conditions. Anthocyanins are synthesized and accumulate in nutritional organs, which is crucial for plants to adapt to and resist adverse environmental conditions, including high exposure to light, ultraviolet light, low temperatures, drought, pests and disease. Some progress has been made in understanding the adaptability of anthocyanin to the external environment. *Begonia semperflorens* is an excellent model for studying the function and regulation of anthocyanin synthesis. To investigate the biosynthesis and regulation of anthocyanins, RNA sequencing techniques were employed to investigate anthocyanin biosynthesis induced by low temperature in *B. semperflorens* leaves. A total of 74,779 unigenes with a mean length of 1249 bp were assembled. Functional annotations were implemented using five protein databases. Differentially expressed genes involved in the process of anthocyanin biosynthesis were identified. This study represents the first report of a broad-scale gene expression study on *B. semperflorens*.

**Keywords** Anthocyanins · Transcriptome · *B. semperflorens*

## Abbreviations

WBM	WD40-bHLH-MYB
UFGT	Flavonoid 3- <i>O</i> -glucosyltransferase
LT	Low temperature
CK	Control
COG	Clusters of orthologous groups
GO	Gene ontology
KEGG	The Kyoto Encyclopedia of Genes and Genomes
DEGs	Differentially expressed genes
Nr	Non-redundant
FDR	False discovery rate
4CL	4-Coumarate: Coenzyme A Ligase

ANS	Anthocyanidin synthase
CHI	Chalcone isomerase
CHS	Chalcone synthase
DFR	Dihydroflavonol 4-reductase
GST	Glutathione <i>S</i> -transferase
PAL	Phenylalanine ammonia-lyase
ROS	Active oxygen free radicals

## Introduction

Anthocyanins are a group of colourful bioactive pigments that contribute to the display of red-to-purple colours in various plant tissues, including leaves, stems, roots, fruit and tissues relating to the flower (Lingua et al. 2013). Anthocyanins also play important physiological and ecological roles in plants by attracting pollinators and seed dispersers, protecting plants from irradiation due to intense light exposure and scavenging free radicals produced in cells under stressful conditions (Zhang et al. 2011a, b). Anthocyanins are usually biosynthesized in specific developmental stages (e.g. the juvenile or senescent phase) and in specific plant tissues (e.g. in flowers, fruits and seeds). Anthocyanin biosynthesis is induced by many biotic and abiotic stressors, including nutrient deficiency, low temperature, intense light

Communicated by R. Baczek-Kwinta.

Huitao Bi and Jiawan Wang contributed equally to this work.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11738-017-2578-3>) contains supplementary material, which is available to authorized users.

✉ Kaiming Zhang  
miss1199@126.com

<sup>1</sup> College of Forestry, Henan Agricultural University, Nongye Road 63, Zhengzhou 450002, China

<sup>2</sup> Fuzhou Institute of Vegetable Sciences, Fuzhou, China

exposure, wounding and pathogen infection (Zhang et al. 2010a, b; Xie et al. 2012; Van den Ende and El-Esawe 2014; Fogelman et al. 2015; Sivankalyani et al. 2016).

During recent decades, numerous studies have focused on the biosynthesis of anthocyanin in various plants. To date, all known pathways and genes involved in anthocyanin biosynthesis have been characterized in the model plant *Arabidopsis thaliana* using molecular, genetic and biochemical data. Researchers have revealed that the WD40-bHLH-MYB (WBM) transcription factor complexes regulate the biosynthesis of anthocyanin under light conditions in this model (Stracke et al. 2010; Petroni and Tonelli 2011).

*Begonia semperflorens* is native to South America and has naturalized elsewhere in the tropics and subtropics (Golding and Wasshausen 2002). Since it is a perpetual perennial plant, it is widely used in gardening. The leaves of *B. semperflorens* vary between different varieties, from copper and green to variegated colours. The copper leaf phenotype has been found to be rich in anthocyanin and thrives with bright colour in sunny conditions. The leaf edge of the green leaf phenotype has been observed to turn red in the sun and is thought to easily accumulate anthocyanin under circumstantial conditions such as low temperature, high temperature and nitrogen deficiency. With such properties, *B. semperflorens* is an excellent model to study the function and regulation of anthocyanin synthesis (Zhang et al. 2013, 2016).

Until now, few studies have focused on *B. semperflorens*, especially at the molecular level. In earlier research, we found that surplus photosynthate in *B. semperflorens* could induce up-regulation of several critical enzymes (e.g. phenylalanine ammonialyase PAL, chalcone isomerase CHI, dihydroflavonol 4-reductase DFR, and flavonoid 3-*O*-glucosyltransferase UFGT) in the anthocyanin biosynthesis pathway and as a result could induce anthocyanin biosynthesis. Excess carbohydrate was also found to trigger the bioprocess (Zhang et al. 2013). In our later research, we observed that, during low-temperature conditions, the colours of *B. semperflorens* cultivars turn red, indicating that anthocyanin synthesis occurs in the plant in such circumstances (Zhang et al. 2013). We predict that low temperature is closely related to anthocyanin synthesis and the induction of anthocyanin synthesis can protect plant photosynthetic organs chloroplast from the damage of the relative excess light energy under low temperature. To understand the molecular mechanisms of this biosynthesis process in *B. semperflorens*, in this study the transcriptomes of *B. semperflorens* under low temperature (LT) and normal growing temperature (CK) were analysed. To the best of our knowledge, this represents the first report of the broad-scale transcriptome of *B. semperflorens*.

*Begonia semperflorens* belongs to the Family: Begoniaceae, Order: Violales. There are many ornamental plants in Begoniaceae, and related research about species belonging

to the Begoniaceae is rare. Studying the transcriptome of *B. semperflorens* will contribute to our current understanding of the genetic background of many plants in the Violales taxonomic order.

## Materials and methods

### Plant material

Seeds of *B. semperflorens* ‘Super Olympia’ were germinated in a mixture of peat and vermiculite (7:3, v/v) in a growth chamber (10-h day length, 100 mol m<sup>-2</sup> s<sup>-1</sup> photosynthetic photon flux density (PPFD), 25/15 °C day/night temperatures and 100% relative humidity). The seedlings were watered daily and fertilized weekly with Hoagland’s nutrient solution. Approximately 30 days after germination, seedlings were selected and transferred to pots with a growth medium of peat, vermiculite and perlite (6:3:1, v/v/v). Normal control and low-temperature-treated samples were named CK and LT, respectively. CK plants were cultured under 25/15 °C (day/night) and 300 μmol m<sup>-2</sup> s<sup>-1</sup> intensity light for 10 h with 95% relative humidity every day. LT plants were cultured under 15/5 °C (day/night) and 300 μmol m<sup>-2</sup> s<sup>-1</sup> intensity light for 10 h with 95% relative humidity every day. After 12 days, the second pair of leaves growing from the apex of the plant of each sample was picked. Four parallel samples from four separate plants for each condition were prepared. All samples were harvested at the same time. At least 0.3 g of leaves was prepared for each sample and immediately transferred and snap-frozen in liquid nitrogen. After that, they were stored at – 80 °C for RNA extraction.

### Anthocyanin quantification

The anthocyanin content was determined according to the methods described by Mita et al. (1997) with minor modifications. Frozen leaf tissue (300 mg) was homogenized in 3 ml of 1% (v/v) HCl:MeOH at 4 °C for 1 day. Following centrifugation at 3500×g for 15 min, the supernatant was measured spectrophotometrically at 530 and 657 nm. One unit of anthocyanins equals one absorbance unit (A<sub>530</sub>–A<sub>657</sub> × 1/4) per ml of extraction solution.

### RNA isolation, library preparation and sequencing

Total RNA was isolated from plant cells with a Quick RNA isolation kit (Biotek Corporation, Beijing, China) according to the manufacturer’s protocol, followed by RNA purification using RNase-free DNase I (Takara, Dalian, China) in order to remove the genomic DNA contaminant. Thereafter, RNA degradation and contamination were monitored through

electrophoresis on a 1% agarose gel. The NanoPhotometer<sup>®</sup> spectrophotometer (IMPLEN, Westlake Village, CA, USA) was used to check RNA purity. RNA concentration was measured using Qubit<sup>®</sup> RNA Assay Kit in Qubit<sup>®</sup> 2.0 Fluorometer (Life Technologies, CA, USA). RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, Palo Alto, CA, USA). Three micrograms of total RNA per sample was pooled and used for subsequent RNA-Seq. Sequencing libraries were generated using the NEBNext<sup>®</sup> Ultra<sup>™</sup> RNA Library Prep Kit for Illumina<sup>®</sup> (NEB, Ipswich, MA, USA) following the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. We purified mRNA from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEB Next First Strand Synthesis Reaction Buffer (5×). First strand cDNA was synthesized using a random hexamer primer and M-MuLV Reverse Transcriptase (RNase H<sup>-</sup>). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase digestion. After adenylation of 3' ends of DNA fragments, NEB Next Adaptor with hairpin loop structure was ligated to prepare for hybridization. To select cDNA fragments preferentially 150–200 bp in length, the library fragments were purified using the AMPure XP system (Beckman Coulter, Beverly, MA, USA). Then, 3 µl of USER Enzyme (NEB) was used with the size-selected adaptor-ligated cDNA at 37 °C for 15 min followed by 5 min at 95 °C before PCR. Then, PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. At last, the PCR products were purified using the AMPure XP system, and library quality was assessed on the Agilent Bioanalyzer 2100 system.

The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina HiSeq 2000 platform and paired-end reads were generated.

### Quality control, transcriptome assembly and gene functional annotation

Raw data (raw reads) in fastq format were first processed through in-house Perl scripts. In this step, clean data (clean reads) were obtained by removing reads containing adapter, reads containing poly-N and low-quality reads from raw data. At the same time, Q20, Q30, GC-content and sequence duplication level of the clean data were calculated. All the downstream analyses were based on clean data with high quality. The left files (read1 files)

from all libraries/samples were pooled into one left.fq file, and the right files (read2 files) into one right.fq file. Transcriptome assembly was accomplished based on the left.fq and right.fq using Trinity (Grabherr et al. 2011) with *min\_kmer\_cov* set to 2 by default and all other parameters set to default. Contig is the seq extended by *k*-mer with *k* – 1 overlap. The *k* is 25. The filtered data were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under the Accession Numbers SRS1381787 and SRS1396330. BLASTX alignment (Altschul et al. 1997) with an *E* value  $\leq 10^{-5}$  was used to implement functional annotation between uni-genes and protein databases. Protein databases, including the Nr (<ftp://ncbi.nih.gov/blast/db/nr>), Swiss-Prot (<http://www.ebi.ac.uk/swissprot/>), KEGG (<http://www.genome.jp/kegg/>), COG (<http://www.ncbi.nlm.nih.gov/COG/>) and GO (<http://geneontology.org/>) databases, were used for BLAST search and annotation.

### Expression annotation

Gene expression levels were determined by RSEM (RNA-Seq by Expectation Maximization) (Li and Dewey 2011) for each sample: clean data were mapped back onto the assembled transcriptome by bowtie software with parameters of “-q -phred33-quals -n 2-a -m 300 s”, and the read count for each gene was obtained from the mapping results. Prior to differential gene expression analysis, for each sequenced library, the read counts were adjusted by the edgeR program package through one scaling normalized factor. Differential expression analysis of two samples was performed using the DEGseq (Anders and Huber 2010) R package. P-values were adjusted using *qvalue* (Broberg 2005). The false discovery rate (FDR)  $\leq 0.01$  coupled with the absolute value of  $\log_2$  (fold change)  $\geq 1$  was set as the threshold for differential expression, and FDR  $\leq 0.01$  coupled with the absolute value of  $\log_2$  (fold change)  $\geq 2$  was set as the threshold for significantly differential expression.

GO enrichment analysis of the differentially expressed genes (DEGs) was implemented by the GOrse R packages-based Wallenius non-central hyper-geometric distribution (Young et al. 2010), which can adjust for gene length bias in DEGs. KEGG (Kanehisa et al. 2008) is a resource database that utilizes molecular information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies, for understanding high-level functions and utilities of biological systems, such as the cell, the organism and the ecosystem (<http://www.genome.jp/kegg/>). We used KOBAS (Mao et al. 2005) software to test the statistical enrichment of differentially expressed genes in KEGG pathways.

## Quantitative real-time reverse transcription PCR (qRT-PCR)

Using qRT-PCR, we selected three anthocyanin-related genes (*BsF3H*, *BsPAL* and *BsANS*) from the DEGs to convince our analysis on the DEGs. The primers were shown in Table A1.

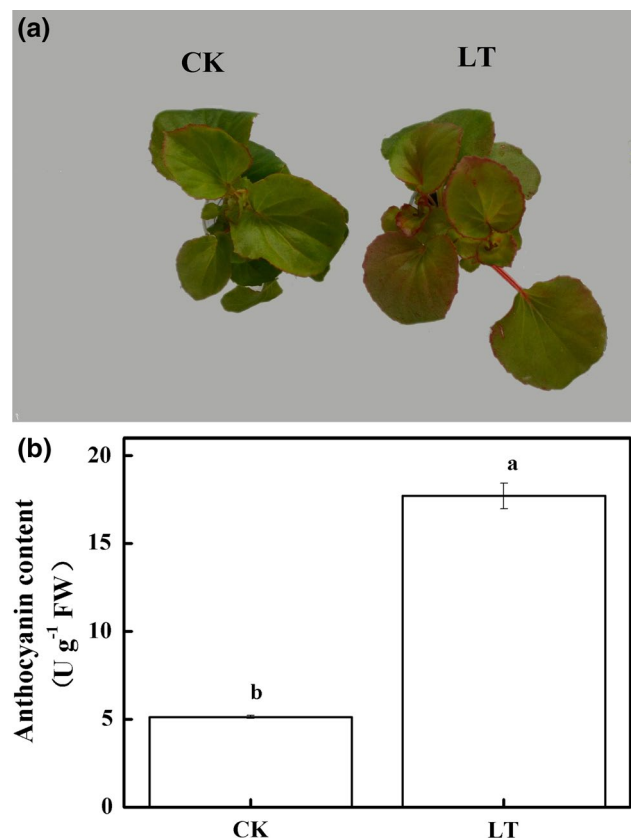
To verify the effect of low temperature on anthocyanin biosynthesis, we treated *B. semperflorens* seedlings with low temperature and dichlorophenyl dimethylurea (DCMU), a diuron, significantly inhibits light-dependent anthocyanin accumulation in *Arabidopsis*) based on the previous studies.

*B. semperflorens* seedlings with four leaves were divided into three groups: CK (normal temperature and light; 25/15 °C, 10 h, 300  $\mu\text{mol m}^{-2} \text{s}^{-1}$ ); LT (low temperature and light; 16/6 °C, 10 h, 300  $\mu\text{mol m}^{-2} \text{s}^{-1}$ ); LT + DCMU (10  $\mu\text{M}$  DCMU was applied to leaves 6 h before low-temperature treatment). Leaves were harvested at 0, 9 h, 1, 4 and 6 days after the start of each treatment, and then used for performing QRT-PCR. Gene expression was measured under different conditions with the StepOne and StepOnePlus Real-Time PCR Systems (Applied Biosystems, Foster City, CA, USA) using SYBR® Premix Ex Taq™ II (TaKaRa). The *B. semperflorens* 18S rRNA gene (Genebank Accession No. KJ959633) was used as an internal reference to identify differences in each cDNA templates. The QRT-PCR cycling conditions were as follows: 95 °C for 10 s, followed by 40 cycles of 95 °C for 10 s and 60 °C for 45 s. We employed the  $2^{-\Delta\Delta C_t}$  method (Livak and Schmittgen 2001) to analyse relative gene expression based on three biological replicates.

## Results and discussion

### Anthocyanin was triggered by low temperature

Many adverse environmental factors, such as ultraviolet radiation, high or low temperature, drought and pest invasion, can induce anthocyanin biosynthesis. Low temperature can induce anthocyanin biosynthesis in many plants (Xie et al. 2012; Huang et al. 2012; Schulz et al. 2015) by decreasing nutrient absorption and photosynthetic rates, two important triggers of anthocyanin biosynthesis (Peng et al. 2008; Zhang et al. 2010a, b). Anthocyanin biosynthesis is known to be triggered by excess sucrose (Zhang et al. 2013; Solfanelli et al. 2006), which usually occurs in plants at low temperatures. In this study, two different conditions, low temperature (LT) and normal control temperature (CK), were used to build libraries using Illumina sequencing technology (Fig. 1a). As shown in Fig. 1b, LT-seedlings showed significant higher content of anthocyanin, which is consistent with our observation that leaves of LT-seedlings turned red, and the colour deepened with time. Based on the



**Fig. 1** Tissues of *B. semperflorens* (a) and anthocyanin content (b). LT low temperature, CK control. The data are the means of 4–6 independent replicates,  $\pm$  SE. Means denoted by the same letter do not significantly differ at  $P < 0.05$  in a single group according to Tukey's test

induction of anthocyanin biosynthesis by low temperature (Zhang et al. 2013), this results proved the casual correlation between anthocyanin biosynthesis and low temperature from transcriptome.

### RNA-Seq and sequence assembly

After quality checks, data with low quality were filtered out, and 25,022,374 and 22,506,321 clean reads were placed in the LT and CK libraries, respectively. The phred quality score  $> 30$  (Q30) was 92.10 and 94.07% for the two samples, respectively. The GC content of the LT and CK libraries was 46.81 and 47.49%, respectively (Table 1).

Short clean reads were then assembled into 9,984,407 contigs in the total library, which were then further assembled into 174,352 transcripts and subsequently 74,779 unigenes. N50 of the transcripts and unigenes was 2200 and 1249 bp, respectively. The mean length of the transcripts and unigenes was 1312 and 684 bp, respectively (Table 2).

Another CK seedling was used as a biological replicate with the former one. RNA-seq was done, and their



**Table 1** Summary of the sequencing and assembly of the transcriptome of *B. semperflorens*

	LT	CK
Read number	25,022,374	22,506,321
Base number	7,487,611,916	6,725,691,034
GC content (%)	46.81	47.49
Q30 (%)	92.10	94.07

**Table 2** Summary statistics of the sequence assembly

Length range	Contig	Transcript	Unigene
200–300	9,930,952 (99.46%)	36,727 (21.06%)	31,238 (41.77%)
300–500	23,582 (0.24%)	26,007 (14.92%)	17,829 (23.84%)
500–1000	15,426 (0.15%)	29,376 (16.85%)	11,895 (15.91%)
1000–2000	9198 (0.09%)	41,263 (23.67%)	8361 (11.18%)
2000 +	5249 (0.05%)	40,979 (23.50%)	5456 (7.30%)
Total number	9,984,407	174,352	74,779
Total length	433,487,072	228,915,199	51,193,959
N50 length	45	2200	1249
Mean length	43.42	1312.95	684.60

correlation was analysed. The correlation coefficients of the expression of the two duplicate samples are shown in Fig. A1. The duplicate showed linear correlation with its corresponding sample. The  $r^2$  value ( $r$  represents Pearson  $r$  values) was 0.9779.

### Transcriptome functional annotation

Since there is no genome of a closely related species available, we employed the transcriptome of *B. semperflorens* to study the molecular mechanisms of the anthocyanin biosynthesis process. All unigenes were aligned using databases including the non-redundant (Nr) protein database, Swiss-Prot database, Kyoto Encyclopedia of Genes and Genomes (KEGG) database, Cluster of Orthologous Groups (COG) database and Gene Ontology (GO) protein database. A total

of 42,027 unigenes were identified and annotated; 41,161, 28,154, 17,165, 12,577 and 25,415 unigenes were identified in the five databases, respectively (Table 3).

### Differentially expressed genes

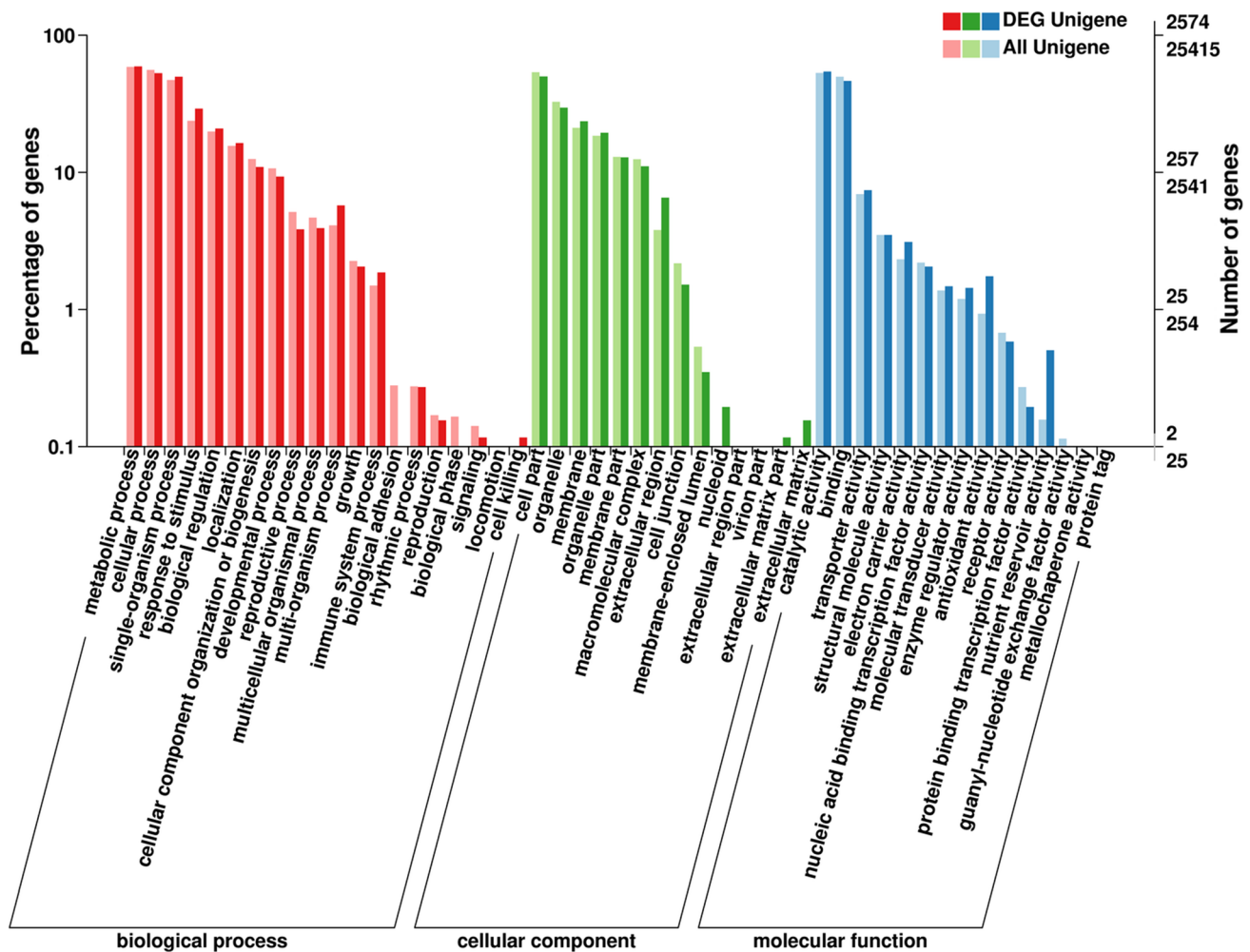
Differentially expressed genes (DEGs) were screened with the criteria of false discovery rate (FDR)  $\leq 0.01$  coupled with absolute value of  $\log_2$  ratio  $\geq 2$ . A total of 4507 DEGs (6.03% of the all unigenes) were identified between the two samples, among which 2333 were up-regulated and 2174 were down-regulated in the comparison of LT and CK (Table A2). Of the 4507 DEGs, 4032 could be identified and annotated in the five protein databases (Table A3). A total of 2574 DEGs were annotated by GO classification, and they were assigned to 49 subgroups: 14 in the “cellular component” category, 15 in “molecular function” category and 20 in “biological process” category. The subgroups containing the highest numbers of DEGs were “metabolic process”, “cellular process”, “cell part”, “cell” and “catalytic activity” (Fig. 2). In the COG annotation, 1980 DEGs were identified as being involved in 25 categories (Fig. 3). In the KEGG annotation, 1531 DEGs were identified as being involved in 113 pathways (Fig. 4). “Carbon metabolism”, “biosynthesis of amino acids” and “ribosome” were the highest three levels with the most unigenes (Table A4), which are consistent with our previous results that excess carbohydrate contents accumulated under low temperature triggered anthocyanin biosynthesis (Zhang et al. 2013).

### Manual identification of anthocyanin synthesis and anthocyanin transporting genes

To identify anthocyanin synthesis and anthocyanin transporting genes in *B. semperflorens*, we used the unigenes in BLAST searches of five protein libraries. Differentially expressed genes (DEGs) were screened with the criteria of false discovery rate (FDR)  $\leq 0.01$  coupled with absolute value of  $\log_2$  ratio  $\geq 1$ . A total of 85 unigenes had sequence similarities to known anthocyanin synthesis genes from other plants, including twenty (4-Coumarate: Coenzyme A

**Table 3** Annotation of the unigenes against different databases

Database	Annotated number	$300 \leq \text{length} < 1000$	Length $\geq 1000$
COG	12,577	3990	5284
GO	25,415	9527	7712
KEGG	17,165	6489	5012
KOG	23,379	8295	8029
Pfam	26,483	9156	11,219
Swissprot	28,154	10,506	9663
nr	41,161	15,646	13,048
All	42,027	15,840	13,141



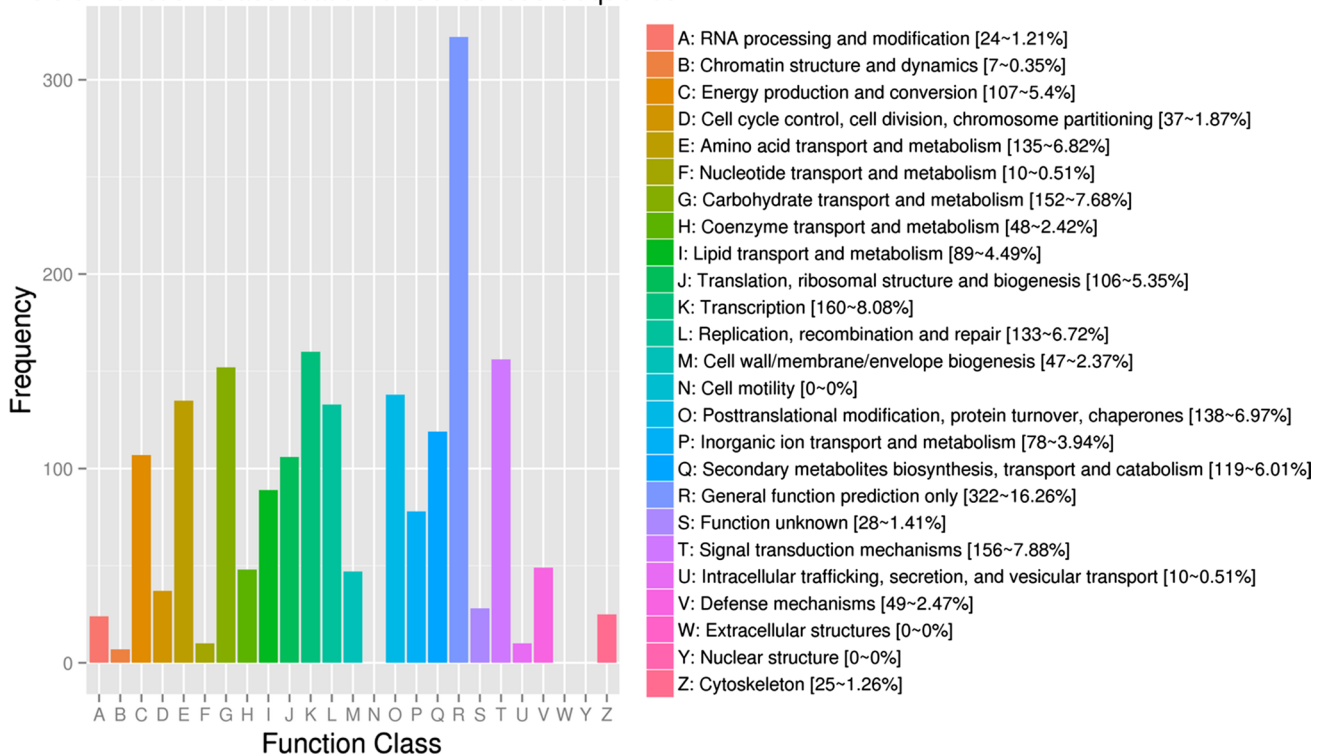
**Fig. 2** Gene ontology classification of distinct proteins that were detected in *B. semperflorens* leaves. 2574 DEGs were annotated by GO classification and they were assigned to 49 subgroups

Ligase) 4CL encoding genes, five anthocyanidin synthase (ANS) encoding genes, nine CHI encoding genes, nine F3H encoding genes, 22 encoding CHS genes, eight encoding DFR genes, nine glutathione S-transferase (GST) encoding genes and six PAL encoding genes (Table A5). Of all the DEGs, five 4CL encoding genes, three ANS encoding genes, three CHI encoding genes, one F3H encoding gene, 14 encoded chalcone synthase (CHS) genes, one DFR encoding gene, four GST encoding genes, three PAL encoding genes and another 26 anthocyanin-related genes were found (Table 4). Unigenes involved in anthocyanin metabolism (141 unigenes), ascorbic acid (45 unigenes), glutathione (184 unigenes), NADP<sup>+</sup> (107 unigenes), NADPH (168 unigenes) and oxidoreduction (20 unigenes) were also retrieved (Table A6). Of these, 11 ascorbic acid encoded genes, 38 glutathione encoded genes, 17 NADP<sup>+</sup> encoded genes, 19 NADPH encoded genes and three oxidoreduction encoded genes were found (Table A7). Sixty-seven DEPs in the

phenyl propionate synthesis pathway and 31 DEPs in the flavonoid biosynthetic pathway were identified in the comparison of LT and CK. The process of anthocyanin synthesis is implemented by these two pathways. Fifty-eight DEPs participating in starch and sucrose metabolism were also identified in the KEGG analysis.

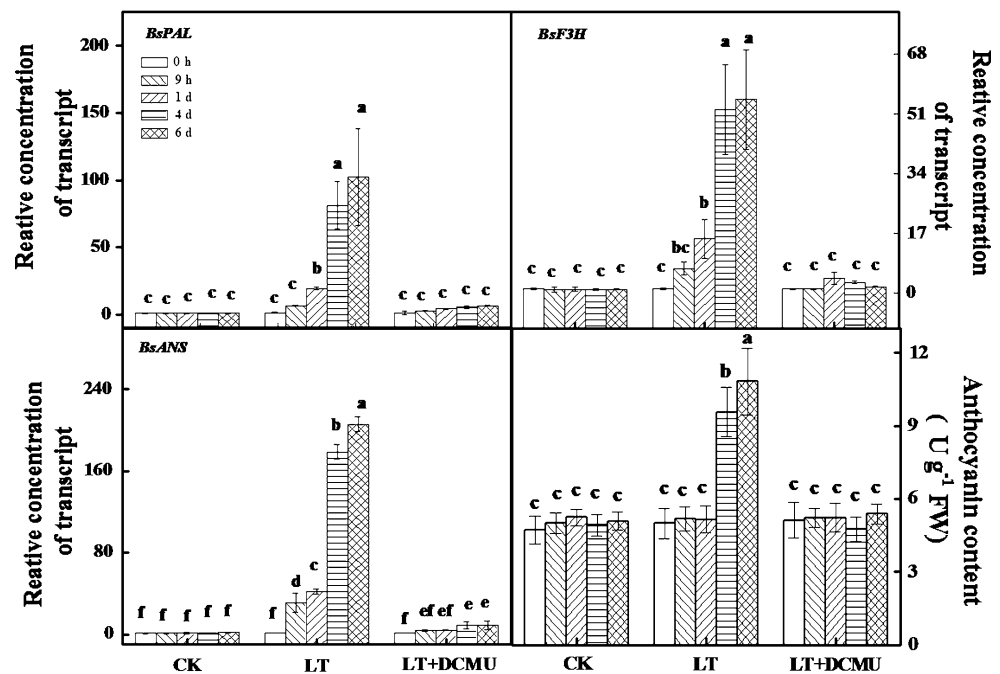
DEGs of the biosynthesis of anthocyanin in this study can be divided into two categories: (1) structure genes, which directly encode enzymes in anthocyanin synthesis and exist in different plants, and (2) regulating genes, which regulate the expression patterns of structure genes, spatially and temporally affecting pigment accumulation. That is, low temperature may not only directly trigger the anthocyanin biosynthesis genes to accumulate anthocyanins in plants; it may also regulate the process using transcription factors. Moreover, low-temperature-induced anthocyanin synthesis enhanced the related enzymes in anthocyanin biosynthesis pathways and GST enzymes in anthocyanin transport

## COG Function Classification of Consensus Sequence



**Fig. 3** Clusters of orthologous groups function classification of distinct proteins that were detected in *B. semperflorens* leaves. Protein or domains were annotated and divided into 23 specific categories

**Fig. 4** Changes in anthocyanin contents, relative transcript levels for four key genes in anthocyanin biosynthesis [*BsPAL* (a), *BsCHS* (b), *BsF3H* (c) and *BsANS* (d)] in leaves of *B. semperflorens*. Values represent mean  $\pm$  SD of three replicates



pathways, resulting in the accumulation of anthocyanin in plant leaves and stems. At the same time, more callose was accumulated in the phloem of leaf and stem, resulting in a

blockage of carbohydrates, one of the substrates for anthocyanin biosynthesis. Two methods, cutting off downward transported carbohydrates with ring cutting and adding

**Table 4** Manual identification of anthocyanin synthesis and anthocyanin transporting genes in the comparison of LT vs. CK in *B. semperflorens* leaves

Gene name	ID	log2FC	Regulated	nr_annotation
4CL	c29652.graph_c0	2.810433	Up	4-Coumarate-CoA ligase 1
	c11199.graph_c0	3.083223	Up	4-Coumarate:CoA ligase 2
	c21305.graph_c0	1.826176	Up	Hypothetical protein JCGZ_08679
	c12222.graph_c0	4.132124	Up	4-Coumarate:CoA ligase
	c17546.graph_c0	1.68307	Up	4-Hydroxycinnamoyl-CoA ligase 3
ANS	c37577.graph_c1	3.116364	Up	Leucoanthocyanidin dioxygenase isoform 2
	c23900.graph_c0	5.24669	Up	Hypothetical protein PRUPE_ppa007856 mg
	c15537.graph_c0	-4.07543	Down	Leucoanthocyanidin dioxygenase
CHI	c33286.graph_c0	4.753767	Up	Class IV chitinase
	c35985.graph_c0	-1.35128	Down	Hypothetical protein Csa_5G523170
	c39856.graph_c0	1.887581	Up	Chalcone isomerase
CHS	c20218.graph_c0	2.265481	Up	Chalcone synthase
	c24763.graph_c0	3.600508	Up	Chalcone synthase
	c30348.graph_c0	4.823641	Up	Hypothetical protein PHAVU_002G038700 g
	c32174.graph_c1	5.510193	Up	Chalcone synthase, partial
	c40536.graph_c0	4.362687	Up	Chalcone synthase 3
	c11536.graph_c0	-2.50278	Down	Chalcone synthase
	c32174.graph_c2	-5.97191	Down	Chalcone synthase
	c44830.graph_c0	-5.49946	Down	Os07g0411300
	c36292.graph_c0	1.691505	Up	ATP-dependent RNA helicase
	c32496.graph_c0	4.718337	Up	Chalcone synthase
	c24533.graph_c0	3.544999	Up	Chalcone synthase 1
	c18612.graph_c0	6.367716	Up	PREDICTED: chalcone synthase G-like
	c27045.graph_c0	5.440108	Up	PREDICTED: chalcone synthase G-like
	c31732.graph_c0	3.199831	Up	Chalcone synthase 1
	DFR	c17369.graph_c0	-2.15027	Down
F3H	c30323.graph_c0	-1.51578	Down	PREDICTED: transcription factor bHLH123-like isoform X1
GST	c15438.graph_c0	2.79215	Up	Glutathione transferase GST 23
	c9340.graph_c0	1.911664	Up	Glutathione transferase
	c28093.graph_c0	-4.65567	Down	PREDICTED: probable glutathione S-transferase
	c16028.graph_c0	3.130772	Up	Unknown
PAL	c20551.graph_c0	3.019905	Up	Phenylalanine ammonia lyase 2
	c26186.graph_c0	3.228863	Up	Phenylalanine ammonia-lyase
	c30218.graph_c0	2.918035	Up	Phenylalanine ammonia-lyase, partial
Anthocyanin	c37577.graph_c1	3.116364	Up	Leucoanthocyanidin dioxygenase isoform 2
	c46979.graph_c0	-4.21165	Down	PREDICTED: probable carotenoid cleavage dioxygenase 4, chloroplastic-like
	c35972.graph_c0	-2.09802	Down	PREDICTED: uncharacterized protein LOC103484635
	c38357.graph_c0	-2.08063	Down	PREDICTED: root phototropism protein 3-like isoform X1
	c16870.graph_c0	6.211788	Up	PREDICTED: flavonoid 3'-monooxygenase
	c30191.graph_c0	3.547008	Up	Anthocyanidin 3-O-glucosyltransferase;
	c32174.graph_c2	-5.97191	Down	Chalcone synthase
	c39187.graph_c1	2.0409	Up	Unknown
	c45798.graph_c0	-4.39453	Down	PREDICTED: ribonuclease 1-like
	c32931.graph_c0	3.755843	Up	Flavonoid 3'-hydroxylase
	c41751.graph_c0	-2.33939	Down	Phototropic-responsive NPH3 family protein
	c32765.graph_c0	2.277611	Up	Chalcone isomerase
	c41793.graph_c0	-2.80397	Down	Hypothetical protein CISIN_1g005842 mg
	c45656.graph_c0	-4.07543	Down	PREDICTED: fructose-bisphosphate aldolase



**Table 4** (continued)

Gene name	ID	log2FC	Regulated	nr_annotation
	c40229.graph_c0	2.604837	Up	4-Coumarate:coenzyme A ligase 4
	c32496.graph_c0	4.718337	Up	Chalcone synthase
	c15926.graph_c0	-2.9363	Down	DNA damage repair protein/transcription factor
	c40170.graph_c0	3.336283	Up	Hypothetical protein M569_15523, partial
	c27045.graph_c0	5.440108	Up	PREDICTED: chalcone synthase G-like
	c29239.graph_c0	-3.06297	Down	PREDICTED: chalcone synthase C2-like
	c37256.graph_c0	2.209363	Up	PREDICTED: ATP-citrate synthase alpha chain protein 2
	c25089.graph_c0	-2.99727	Down	Hypothetical protein JCGZ_07595
	c39187.graph_c0	2.944699	Up	Thermospermine synthase ACAULIS5
	c36865.graph_c1	1.413623	Up	Leucine-rich receptor-like protein kinase family protein
	c31732.graph_c0	3.199831	Up	Chalcone synthase 1
	c34516.graph_c0	3.268737	Up	4-Coumarate:CoA ligase 3

exogenous sucrose to promote the accumulation of carbohydrates in plants, are used to verify the hypothesis (Zhang et al. 2013). The results support the conclusion that carbohydrates directly trigger anthocyanin synthesis under low temperatures.

### Analysis of anthocyanin concentration under low temperature

To further convince the DEGs, we selected three core anthocyanin biosynthesis genes (*BsPAL*, *BsF3H* and *BsANS*) to verify the relationship between anthocyanin biosynthesis and low temperature using qRT-PCR.

As shown in Fig. 4, three genes expression were significantly up-regulated in seedlings after treatment with low temperature (16/5 °C) for 9 h or 1 day, which resulted in a significant increase in anthocyanin level in seedlings after treatment with low temperature for 4 days. Treatment of DCMU significantly decreased all the three genes expression, which resulted in no changes in anthocyanin level in turn.

### Anthocyanin as a protectant in low-temperature environments

Anthocyanin synthesis induced by low temperatures may be a protective mechanism for plants. According to the cellular component of the GO analysis of the DEGs, the synthesis of anthocyanin in cells is mainly concentrated in the cytoplasm around the vacuole. Based on the KEGG pathway analysis, we can conclude that the process is implemented by the phenyl propionate synthesis pathway and flavonoid biosynthetic pathway and ends with transport into the vacuole.

Based on the KEGG enrichment analysis (Fig. 5 and Table A4), we can infer that anthocyanins play two roles under low temperature, first by functioning as

antioxidants. The most different expression unigenes over-presented in photosynthesis and photosynthesis-antenna proteins. Low temperature inactivates enzymes and organisms activities, and plants respond to low temperature by reducing photosynthetic rate, which produce relative excess light energy produced in plants. Excess light energy will produce a large number of active oxygen free radicals (ROS), such as  $\cdot\text{O}_2^-$ ,  $\text{H}_2\text{O}_2$  and  $\cdot\text{OH}$ . These free radicals can directly oxidize and damage DNA, proteins, lipids and other biological macromolecules. As a result, the light mechanism of a plant leaf can be destroyed, leading to photooxidation and even light damage, thereby hindering the normal physiological metabolism of plants. This is consistent with the study of Cheng et al. (2016). From the KEGG enrichment analysis (Fig. 5 and Table A4), LT-seedlings also showed higher Enrichment Factor in phenylpropanoid biosynthesis, phenylalanine metabolism and flavonoid biosynthesis. Anthocyanins are flavonoids, and they therefore have the physiological function of the antioxidant activity of flavonoids. On the one hand, the molecular structure of anthocyanins has a number of phenolic hydroxyl groups and can release electrons through their oxidation, directly removing various free radicals and inhibiting oxidation. Such functions were also reported by Amin et al. (2015). On the other hand, anthocyanins can also protect lipids from over-oxidation and other oxidative modification through chelating metal ions such as iron and copper, which act as catalysts in the free radical reactions, as verified in another study (Zhang and Chen 2000). In addition, anthocyanins can reduce the oxidation of low-density lipoprotein (LDL) by increasing the activities of superoxide dismutases and glutathione in cells. Second, anthocyanins accumulation under low temperature may function as a carbon sink to consume relative excess carbohydrate contents. As shown in Fig. 5 and Table A4, the

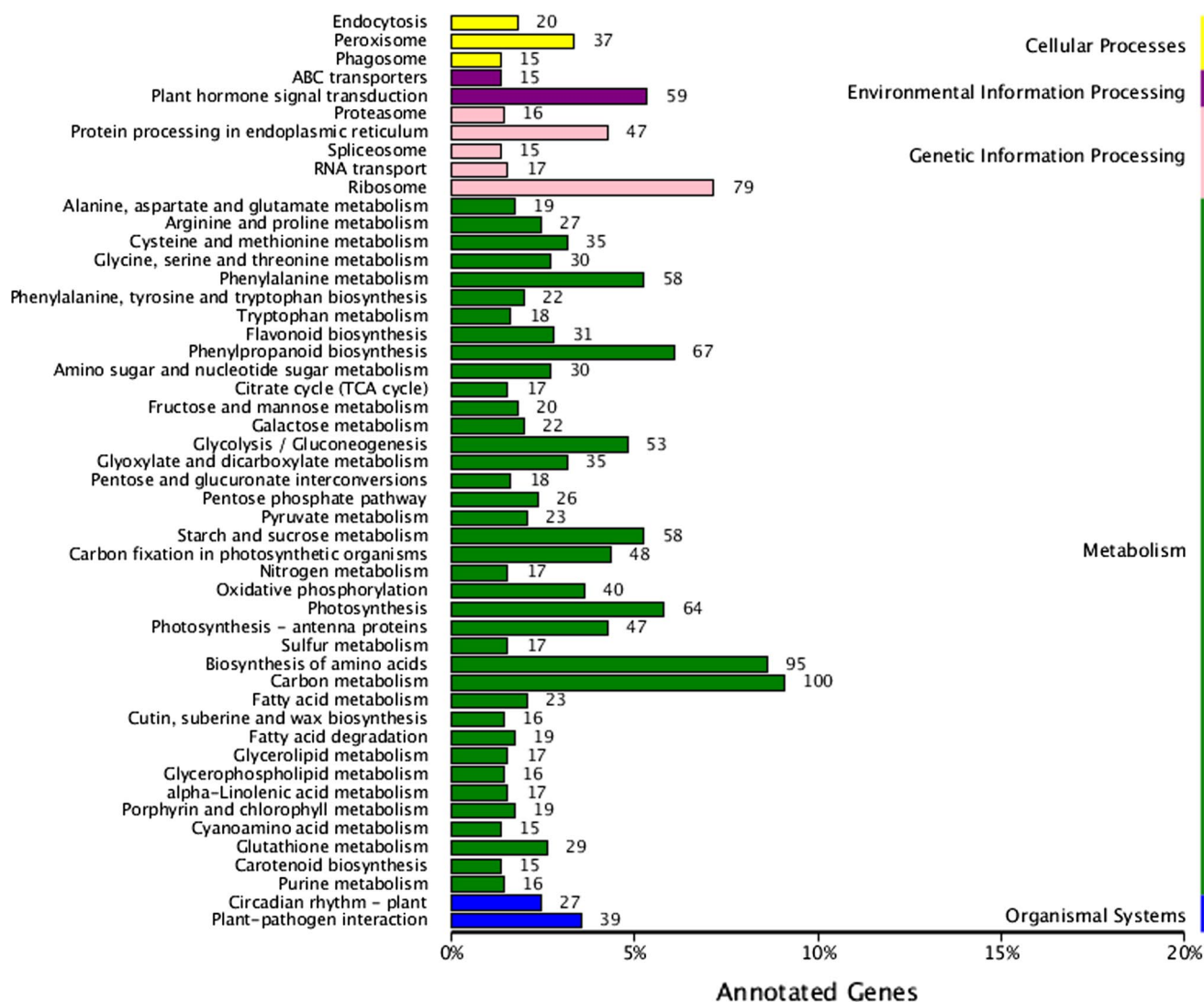


Fig. 5 Kyoto Encyclopedia of Genes and Genomes analysis of the identified differentially expressed genes (DEGs) in the comparison of LT vs. CK. 1,531 DEGs were identified involving in 113 pathways

most unigenes over-presented were in carbon metabolism. Meanwhile, increases in secondary metabolism, particularly flavonoid biosynthesis, indicated that anthocyanins are associated with carbon metabolism under low temperature. Under low temperature, relative excess carbohydrates accumulate in plants as a result of reduced respiratory rate and growth rate. Higher carbohydrate content has been found in many plants as a trigger of anthocyanin biosynthesis (Zhang et al. 2013; Scafidi et al. 2015). Anthocyanins are produced from a C6–C3–C6 skeleton using sugars as one substrate. Therefore, the induction of anthocyanin biosynthesis in *Begonia* can help consume excess carbohydrate, which otherwise will inhibit photosynthesis.

### Conclusions and future perspectives

To understand the molecular mechanisms of the biosynthesis and regulation process in *B. semperflorens*, the transcriptomes of *B. semperflorens* leaves under normal and low-temperature conditions were analysed. To the best of our knowledge, this represents the first report of the broad-scale transcriptome of *B. semperflorens* and will also contribute to our current understanding of the genetic background of many plants in the *Violales* taxonomic order.

Further research on the mechanism of the protection function of anthocyanins is needed to determine how environmental signals influence anthocyanin biosynthesis, how

the signals are transduced, how anthocyanin functions in plant protection, the relationship between anthocyanin and phytochrome and the effects of anthocyanin on plant photosynthesis. These functions of anthocyanin are important for plants adapting to the environment. In addition, there are a variety of anthocyanins, and since different plant species contain different anthocyanins, their physiological functions to adapt to the environment are also distinctly different. In conclusion, anthocyanins play an important role in plant defence and adaptation to adverse environments, and further research is needed to understand and maximize their function.

**Author contribution statement** KZ and BH conceived and designed the experiments; MG performed the experiments; WD analysed the data; JW contributed reagents/materials/analysis tools; KZ and BH wrote the paper.

**Acknowledgements** This research was supported by the National Science Foundation of China (Grant No. 31100562) and The Key Scientific Research Project of High Education in Henan Province (Grant No. 15A220006).

### Compliance with ethical standards

**Conflict of interest** The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

### References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Amin HP, Czank C, Raheem S, Zhang Q, Botting NP, Cassidy A, Kay CD (2015) Anthocyanins and their physiologically relevant metabolites alter the expression of IL-6 and VCAM-1 in CD40L and oxidized LDL challenged vascular endothelial cells. *Mol Nutr Food Res* 59(6):1095–1106
- Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11:R106
- Broberg P (2005) A comparative review of estimates of the proportion unchanged genes and the false discovery rate. *BMC Bioinform* 6:199
- Cheng DD, Zhang ZS, Sun XB, Zhao M, Sun GY, Chow WS (2016) Photoinhibition and photoinhibition-like damage to the photosynthetic apparatus in tobacco leaves induced by *Pseudomonas syringae* pv. *Tabaci* under light and dark conditions. *BMC Plant Biol* 16(1):29–39
- Fogelman E, Tanami S, Ginzberg I (2015) Anthocyanin synthesis in native and wound periderms of potato. *Physiologia plantarum* 153(4):616–626
- Golding J, Wasshausen DC (2002) Begoniaceae, edition 2: part I: annotated species list: Part ii: Illustrated key, abridgment and supplement. *Contrib U S Natl Herb* 43:1–289
- Grabherr MG, Haas BJ, Yassou M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mucelli E, Hacohen N, Gnirke A, Rhind N, Palma F, Birren W, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652
- Huang ZA, Zhao T, Fan HJ, Wang N, Zheng SS, Ling HQ (2012) The upregulation of NtAN2 expression at low temperature is required for anthocyanin accumulation in juvenile leaves of Lc-transgenic tobacco (*Nicotiana tabacum* L.). *J Genet Genom* 39:149–156
- Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y (2008) Kegg for linking genomes to life and the environment. *Nucleic Acids Res* 36:480–484
- Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform* 12:323
- Lingua G, Bona E, Manassero P, Marsano F, Todeschini V, Cantamessa S, Copetta A, D'Agostino G, Gamalero E, Berta G (2013) Arbuscular mycorrhizal fungi and plant growth-promoting pseudomonads increases anthocyanin concentration in strawberry fruits (*Fragaria × ananassa* var. *Selva*) in conditions of reduced fertilization. *Int J Mol Sci* 14(8):16207–16225
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-ΔΔC<sub>T</sub></sup> Method. *Methods* 25:402–408
- Mao X, Cai T, Olyarchuk JG, Wei L (2005) Automated genome annotation and pathway identification using the keggorthology (KO) as a controlled vocabulary. *Bioinformatics* 21:3787–3793
- Mita S, Murano N, Akaike M, Nakamura K (1997) Mutants of *Arabidopsis thaliana* with pleiotropic effects on the expression of the gene for β-amylase and on the accumulation of anthocyanin that are inducible by sugars. *Plant J* 11(4):841–851
- Peng MS, Hudson D, Schofield A, Tsao R, Yang R, Gu HL et al (2008) Adaptation of Arabidopsis to nitrogen limitation involves induction of anthocyanin synthesis which is controlled by the NLA gene. *J Exp Bot* 59:2933–2944
- Petroni K, Tonelli C (2011) Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Sci* 181:219–229
- Scafidi P, Maria G, Barbagallo, Downey MO (2015) Influence of light exclusion on anthocyanin composition in 'cabernet sauvignon'. *Front Plant Sci* 06:107–110
- Schulz E, Tohge T, Zuther E, Fernie AR, Hinch DK (2015) Natural variation in flavonol and anthocyanin metabolism during cold acclimation in *Arabidopsis thaliana* accessions. *Plant Cell Environ* 38:1658–1672
- Sivankalyani V, Feygenberg O, Diskin S, Wright B, Alkan N (2016) Increased anthocyanin and flavonoids in mango fruit peel are associated with cold and pathogen resistance. *Postharvest Biol Technol* 111:132–139
- Solfanelli C, Poggi A, Loreti E, Alpi A, Perata P (2006) Sucrose-specific induction of the anthocyanin biosynthetic pathway in Arabidopsis. *Plant Physiol* 140(2):637–646
- Stracke R, Jahns O, Keck M, Tohge T, Niehaus K, Fernie AR, Weishaar B (2010) Analysis of production of flavonol glycosides-dependent flavonol glycoside accumulation in *Arabidopsis thaliana* plants reveals myb11-, myb12- and myb111-independent flavonol glycoside accumulation. *New Phytol* 188:985–1000
- Van den Ende W, El-Esawe SK (2014) Sucrose signaling pathways leading to fructan and anthocyanin accumulation: a dual function in abiotic and biotic stress responses? *Environ Exp Bot* 108:4–13
- Xie XB, Li S, Zhang RF, Zhao J, Chen YC, Zhao Q, Yao YX, You CX, Zhang XS, Hao YJ (2012) The bHLH transcription factor MdbHLH3 promotes anthocyanin accumulation and fruit colouration in response to low temperature in apples. *Plant Cell Environ* 35(11):1884–1897

- Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-Seq: accounting for selection bias. *Genome Biol* 11:R14
- Zhang HY, Chen DZ (2000) Theoretical characterization and application of free radical scavenging activity of phenolic antioxidants. *Acta Biophys Sin* 16:1–9
- Zhang KM, Yu HJ, Shi K, Zhou YH, Yu JQ, Xia XJ (2010a) Photoprotective roles of anthocyanins in *Begonia semperflorens*. *Plant Sci* 179:202–208
- Zhang KM, Yu HJ, Shi K, Zhou YH, Yu JQ, Xia XJ (2010b) Photoprotective roles of anthocyanins in *Begonia semperflorens*. *Plant Sci* 179:202–208
- Zhang Y, Zheng S, Liu Z, Wang L, Bi Y (2011a) Both HY5 and HYH are necessary regulators for low temperature-induced anthocyanin accumulation in *Arabidopsis* seedlings. *J Plant Physiol* 168(4):367–374
- Zhang YQ, Zheng SH, Liu ZJ, Wang LG, Bi YR (2011b) Both hy5 and hyh are necessary regulators for lowtemperature-induced anthocyanin accumulation in *arabidopsis* seedlings. *J Plant Physiol* 168:367–374
- Zhang KM, Li Z, Li Y, Li YH, Kong DZ, Wu RH (2013) Carbohydrate accumulation may be the proximate trigger of anthocyanin biosynthesis under autumn conditions in *Begonia semperflorens*. *Plant Biol* 15:991–1000
- Zhang KM, Guo ML, He D, Wu RH, Li YH (2016) The inhibition effect and excessive carbon flux resulting from blocking anthocyanin biosynthesis under darkness in *Begonia semperflorens*. *J Plant Growth Regul* 35:22–30