De novo sequencing and comparative transcriptome analysis of white petals and red labella in *Phalaenopsis* for discovery of genes related to flower color and floral differentation

Yuxia Yang[1], Jingjing Wang[1,2], Zhihu Ma[3], Guosheng Sun[3], Changwei Zhang[1]*

[1] College of Horticulture, Nanjing Agricultural University, 1# Weigang, Nanjing, Jiangsu Province, 210095, P.R. China

[2] Dongxin Department, Jiansu Provincial Agricultural Reclamation and Development Corporation, 26# Dongfang Middle Road, Lianyungang, Jiangsu Province, 222248, P.R. China

[3] Zhenjiang Institute of Agricultural Sciences in Hilly Area of Jiangsu Province, 112# Ninghang Road, Jurong, Jiangsu, 212400, P.R. China

Abstract

Phalaenopsis is one of the world's most popular and important epiphytic monopodial orchids. The extraordinary floral diversity of *Phalaenopsis* is a reflection of its evolutionary success. As a consequence of this diversity, and of the complexity of flower color development in *Phalaenopsis*, this species is a valuable research material for developmental biology studies. Nevertheless, research on the molecular mechanisms underlying flower color and floral organ formation in *Phalaenopsis* is still in the early phases. In this study, we generated large amounts of data from *Phalaenopsis* flowers by combining Illumina sequencing with differentially expressed gene (DEG) analysis. We obtained 37 723 and 34 020 unigenes from petals and labella, respectively. A total of 2736 DEGs were identified, and the functions of many DEGs were annotated by BLAST-searching against several public databases. We mapped 837 up-regulated DEGs (432 from petals and 405 from labella) to 102 Kyoto Encyclopedia of Genes and Genomes pathways. Almost all pathways were represented in both petals (102 pathways) and labella (99 pathways). DEGs involved in energy metabolism were significantly differentially distributed between labella and petals, and various DEGs related to flower color and floral differentiation were found in the two organs. Interestingly, we also identified genes encoding several key enzymes involved in carotenoid synthesis. These genes were differentially expressed between petals and labella, suggesting that carotenoids may influence *Phalaenopsis* flower color. We thus conclude that a combination of anthocyanins and/or carotenoids determine flower color formation in *Phalaenopsis*. These results broaden our understanding of the mechanisms controlling flower color and floral organ differentiation in *Phalaenopsis* and other orchids.

Keywords: *Phalaenopsis*; RNA-seq; transcriptome; flower color; diversity

Introduction

Phalaenopsis amabilis, an epiphytic monopodial orchid, is an important ornamental species with a huge global market, especially in western countries. Ornamental quality in *Phalaenopsis* is influenced by many factors, including flower color, fragrance and shape, cut-flower longevity, flowering control, and abiotic stress tolerance [1]. Among these, flower color and diversity are the two main factors visually impacting ornamental and commercial values of *Phalaenopsis*.

Phalaenopsis flowers are zygomorphic and include three outer tepals (T1–T3, also known as sepals) in the first whorl, two lateral inner tepals (t1 and t2, the petals), and a median inner tepal (t3, the lip or labella) [2,3]. Floral initiation and development are regulated by complex networks of exogenous and endogenous signals, including light stimuli, hormones, ligand–receptor interactions, signal transduction pathways, and transcription factor cascades [4]. A recent theory known as "the orchid code" proposes a reasonable model in which four different class B *AP3/DEF*-like MADS-box genes have played a vital role in the evolution of the orchid perianth by their combinatorial interaction. Orchid floral diversity is an evolutionary consequence of two duplication events and associated changes that occurred in the regulatory regions of the class B *AP3/DEF*-like MADS-box genes, which was followed by sub- and neo-functionalization. In orchids, class B *AP3/DEF*-like MADS-box genes are divided into four distinct clades: *PeMADS2*-like (clade 1), *OMADS3*-like (clade 2), *PeMADS3*-like (clade 3), and *PeMADS4*-like (clade 4), each having its own specific expression pattern. The combined expressions of clade 1 and clade 2 genes mediate the development of the three outer tepals, while the combination of clade 1, clade 2, and clade 3 genes leads to the development of the lateral inner tepals. Labella development is determined by a combination of genes

---

* Corresponding author. Email: changweizh@njau.edu.cn

Handling Editor: Przemysław Wojtaszek

from all four clades. Differential expression of clade 3 genes is obviously responsible for differences between inner and outer tepals, whereas differential expression of clade 4 genes differentiates the lateral inner tepals from the labella [2,3]. Orchid *PI/GLO*-like genes, found to be present in the *AP3/DEF*-like gene copies, are also necessary for current floral tissue development [4]. Despite this knowledge, however, the regulatory network controlling orchid floral development remains unclear.

Flower color is derived from the three major classes of plant pigments: anthocyanins, betalains, and carotenoids [5]. Of these, anthocyanins are the major contributors to flower color [6]. A class of water-soluble flavonoids, anthocyanins are synthesized in the cytosol and localized in vacuoles. Through the phenylpropanoid pathway, they provide a wide range of colors, ranging from orange-red to violet-blue in dark-colored flowers [5]. Despite their structural variety, anthocyanins are only categorized into six chromophore forms: pelargonidin, cyanidin, peonidin, delphinidin, petunidin, and malvidin [7,8]. The anthocyanin biosynthetic pathway has been well elaborated [9].

In orchids, the primary anthocyanin in red flowers is a cyanidin derivative that is typically modified by glycosylation and acylation [10]. The glycosylation-related gene *PeU-FGT3* plays a critical role in red color formation in *Phalaenopsis* [8]. Several important enzymes, such as chalcone synthase (CHS), chalcone isomerase (CHI), dihydroflavonol 4-reductase (DFR), and anthocyanidin synthase (ANS), are involved in the formation of colored anthocyanidins [5].

Compared with studies in other flowering plants, the molecular basis of floral color development has not been well characterized in orchids. A better understanding of the molecular mechanisms underlying orchid flower color and floral organ formation is thus needed. Furthermore, few transcriptomic-based investigations of the functions of genes related to flower color and floral differentiation have been reported for *Phalaenopsis*. To expand knowledge regarding flower color and diversity for *Phalaenopsis* breeding, in this study we analyzed differential gene expression between petals and labella using the Illumina RNA-Seq method. Our study generated a huge number of *Phalaenopsis* transcript sequences during floral formation that can be used to discover putative genes related to flower color and floral differentiation. By comparing relative gene expression levels between petals and labella, novel insights can be gleaned into orchid floral development. Our study therefore provides a foundation for future research on mechanisms underlying floral development in *Phalaenopsis* and other orchids.

Material and methods

Plant material and sample collection

Phalaenopsis plants with white petals and red labella (Fig. 1) were grown in greenhouses at Nanjing Agriculture University under natural light conditions and a controlled temperature of 22–27°C. Petals and labella were collected at the full-bloom stage. The two samples were immersed in liquid nitrogen and stored at −80°C until subjected to total RNA extraction.

Total RNA extraction, cDNA library construction, and Illumina deep sequencing

Total RNA was extracted from petal and labella samples using a Trizol kit (Takara, Japan). Total RNA quality and quantity were analyzed using a Nanodrop 2000 instrument (Thermo Scientific) and a ChipRNA 7500 Series II Bioanalyzer (Agilent). The two total RNA samples were delivered to Beijing Biomarker Biotechnology Co. (Beijing, China) for the construction of cDNA libraries using an mRNA-Seq Sample Preparation kit (Illumina) according to the manufacturer's instructions. The sequencing of the two samples was performed on an Illumina HiSeq 2000 system.

Sequence assembly and annotation

The raw image data produced from sequencing was transformed by base calling into raw reads. Transcriptome de novo assembly was carried out with the Trinity short-read assembly program, which generated in turn contigs, transcripts, and unigenes. To identify unigene putative functions, their sequences were aligned using BLASTX (*E*-value ≤10⁻⁵) against the following public protein databases: National Center for Biotechnology Information non-redundant (Nr) and nucleotide (Nt) databases and SwissProt, TrEMBL, Clusters of Orthologous Groups (COG), Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. The Blast2GO software package was used to compare and determine unigene GO annotations. Finally, WEGO software was used to obtain GO functional classifications for all annotated unigenes.

Identification of differentially expressed genes (DEGs)

To identify DEGs between the two samples, the following formula was used to calculate significance (*P*) of differences in transcript accumulation for each gene:

$$P(y/x) = \left(\frac{N2}{N1}\right)^y \frac{(x+y)!}{x!\,(y)!\,(1+\frac{N2}{N1})^{(x+y+1)}}$$

where *N1* and *N2* represent the total number of clean reads from petals and labella, respectively, and *x* and *y* represent the number of reads mapping to the given gene. We then
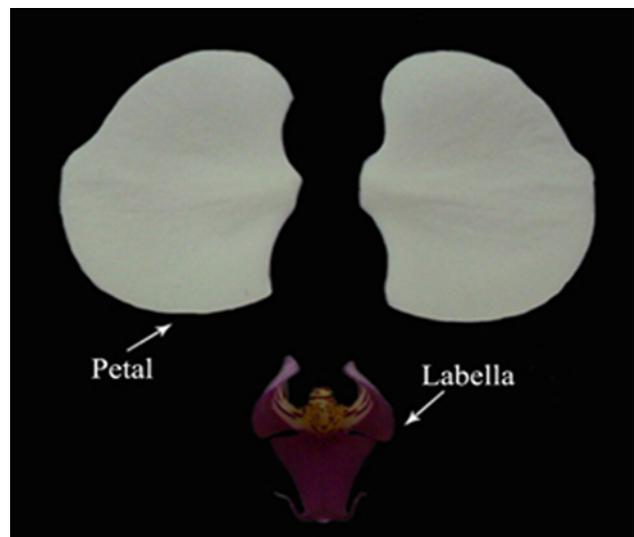

Fig. 1 Photographic image of selected flower materials

used false discovery rate (FDR) to determine the threshold of the *P*-value in multiple tests. To identify DEGs, we used a combination of two criteria: expression fold-change |log$_2$Ratio| ≥ 1 and FDR-adjusted *P*-value < 0.001.

We performed GO and KEGG functional enrichment analysis to determine which DEGs were significantly enriched in GO terms and metabolic pathways ($P \leq 0.05$) compared with the selected transcriptome background. Significance was calculated according to the formula:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{N}}$$

where *N* is the number of genes with GO/KEGG annotations, *n* is the number of DEGs in the set of *N* genes, *M* is the number of genes mapped to a certain GO/KEGG term, and m is the number of DEGs in the set of *M* genes.

Results

Short-read de novo sequencing and assembly

We pooled equal amounts of RNA from the two samples (petals and labella) to construct cDNA libraries for transcriptome sequencing and analysis on a Genome Analyzer IIx platform using Illumina sequencing technology. Using this sequencing approach, raw reads were generated from both ends of the cDNA fragments. After data filtering using stringent quality criteria (i.e., removal of sequences shorter than 65 bp or with CycleQ20 values less than 100%), we obtained 10 734 813 clean reads comprising 2 134 924 797 nucleotides (nt) from the petal library and 16 224 038 clean reads comprising 3 276 861 889 nt from the labella library. The clean reads were assembled de novo into contigs using the Trinity software package. The results of the sequencing assembly are presented in Tab. 1. A total of 1 057 077 and 2 105 819 contigs were assembled from petals and labella, respectively, with corresponding mean lengths of 75 bp and 67 bp and N50s of 99 bp and 65 bp. Of these, 16 999 (1.61%) contigs from petals and 17 334 (0.82%) contigs from labella were longer than 500 bp. The contigs from petals and labella were further assembled into 55 101 and 67 026 scaffolds with mean lengths of 811 bp and 948 bp, respectively. Scaffold N50 lengths were 1280 bp for petals and 1168 bp for labella. In total, 15 207 scaffolds assembled from petals and 24 853 scaffolds from labella coded for transcripts longer than 1 kb. Finally, 37 723 unigenes were obtained from the petals, with a final unigene N50 length of 1 125 bp and a total length of 26 656 163 nt; similarly, 34 020 unigenes were generated from the labella data, with a final unigene N50 length of 1398 bp and a total length of 27 636 110 nt. These unigenes were then organized into a transcriptome database for the identification of putative genes related to flower color and floral differentiation.

DEG analysis between white petal and labella libraries

A primary goal of transcriptome sequencing is comparison of gene expression levels in different samples. In this study, a large number of DEGs were estimated by

Tab. 1 Summary of sequencing and assembly results.

Assembled data	Statistics of data production	White petals	Red labella
Reads			
	No. of reads	10 734 813	16 224 038
	Total nucleotides (nt)	2 134 924 797	3 276 861 889
	GC percentage (%)	46.80	45.08
	Q20 percentage (%)	100.00	100.00
Contigs			
	No. of contigs	1 057 077	2 105 819
	Total nucleotides (nt) in contigs	79 618 808	141 300 083
	Length of N50 (bp)	99	65
	Mean length of contigs (bp)	75	67
	No. of contigs above 500 bp	16 999	17 334
Scaffolds			
	No. of scaffolds	55 101	67 026
	Total nucleotides (nt) in scaffolds	44 692 377	63 535 969
	Length of N50 (bp)	1280	1443
	Mean length of scaffolds (bp)	811	948
	No. of scaffolds above 500 bp	23 777	40 620
Unigenes			
	No. of unigenes	37 723	34 020
	Total nucleotides (nt) in unigenes	26 656 163	27 636 110
	Length of N50 (bp)	1125	1398
	Average length of unigenes (bp)	707	812
	No. of unigenes above 500 bp	15 903	15 984

comparing gene expression between the petal and labella libraries. Unigene expression was calculated using the RPKM method [11]. DEGs detected with at least two-fold differences (FDR < 0.001 and |log$_2$Ratio| ≥ 1) between petal and labella libraries are shown in Fig. 2. Using these criteria, we identified 2736 DEGs between petals and labella. Of these, 1277 were up-regulated and 1459 were down-regulated in petals. A greater number of genes were expressed only in the petals (243) than in the labella (62). This result suggests that petal and labella development may involve totally different processes.

DEG functional annotation

To validate and annotate the assembled DEGs, the 2736 DEGs were subjected to BLASTX comparisons (*E*-value ≤ 1×10^{-5}) against several public protein databases to identify putative functions of the unigene sequences. As a result, 2698, 2183, 2319, 2706, 2446, 837, and 1077 DEGs were found to have homologous sequences in Nr, Nt, SwissProt, TrEMBL, GO, KEGG, and COG databases, respectively.
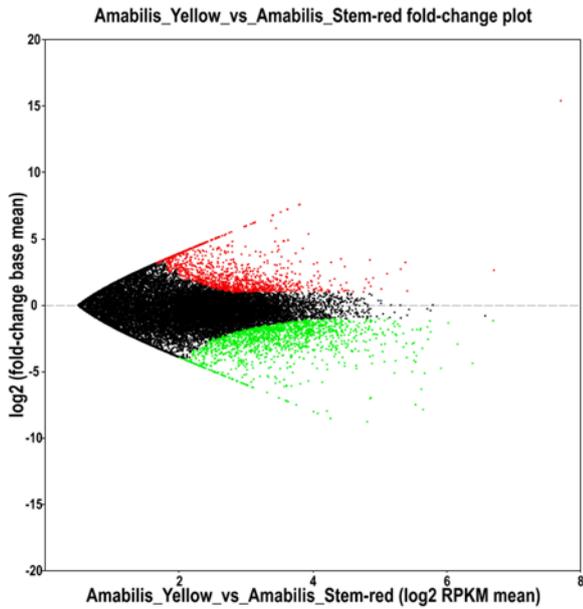
Fig. 2 Scatter plot of relative gene expression in petals vs. labella. Red dots above zero on the y-axis indicate genes having a higher expression level in petals, green dots below zero correspond to genes having a higher expression level in labella, and black dots represent genes that were similar in both libraries. False discovery rate < 0.001 and |log$_2$Ratio| ≥ 1 were used as thresholds to judge the significance of gene expression differences.

COG ANALYSIS OF DEG PATTERNS. COG analysis is used to classify orthologous gene products, with all proteins in an orthologous gene cluster considered to have evolved from the same ancestral protein. The COG database, which contains putative protein sequences encoded by all genes from bacteria, archaea, and eukaryotes that have complete genome sequences, facilitates exploration via sequence analysis of evolutionary relationships among these groups [12].

The newly identified sequences were aligned against the known COG database to predict and classify their possible functions. In total, 1077 unigenes were assigned to at least one COG classification. Among the 25 COG categories (Fig. 3), the cluster for "General function prediction only" represented the largest group (262 unigenes; 24.3%), followed by "Posttranslational modification, protein turnover, chaperones" (146; 13.6%), "Transcription" (137; 12.7%) and "Translation, ribosomal structure and biogenesis" (118; 10.9%). The categories of "Cell motility" (3; 0.28%), "Nuclear structure" (1; 0.09%) and "Extracellular structures" (0; 0%) contained the smallest numbers of unigenes. This information should prove to be a valuable resource for studying specific processes and functions in *Phalaenopsis*.

GO FUNCTIONAL ENRICHMENT. GO is an international standardized gene functional classification system that describes properties of genes and their products in any organism [13]. Based on sequence homology, 2446 DEGs were categorized into three main categories (biological process, cellular component, and molecular function) and 62 subcategories (Fig. 4). Multiple terms were frequently assigned to the same transcript; thus, 15 159 GO term annotations were assigned to at least one GO category under biological
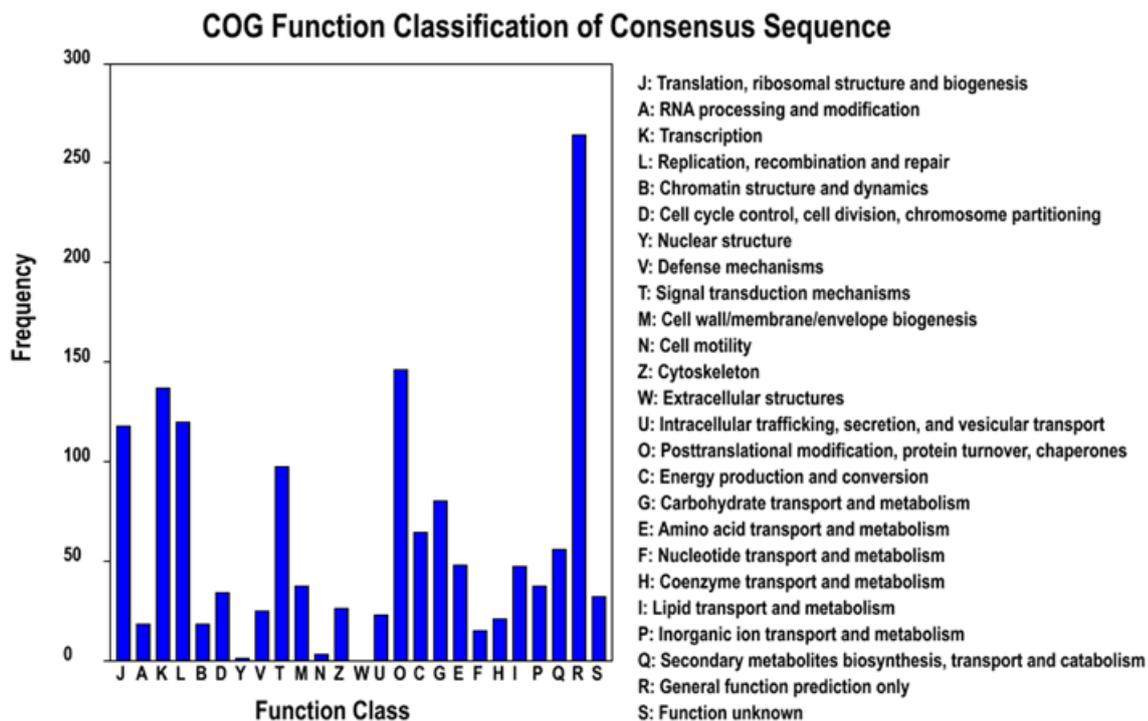

Fig. 3 Clusters of orthologous genes (COG) annotations of putative proteins. All putative proteins were aligned to the COG database and functionally classified into one or more of 25 molecular families. The capital letters on the x-axis correspond to the COG categories listed to the right of the histogram, and the y-axis indicates the numbers of DEGs assigned to the corresponding COG category.
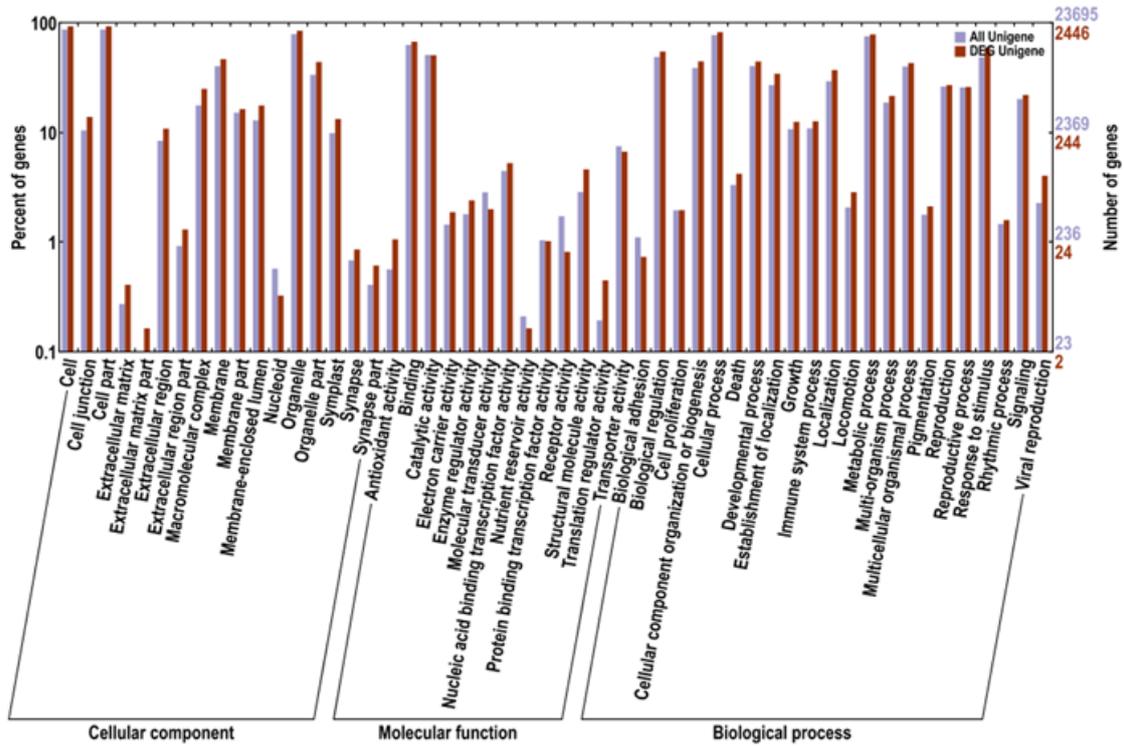
Fig. 4 Gene ontology (GO) classification of all annotated unigenes and DEGs. GO classification according to biological process, cellular component, and molecular function categories of all annotated 23 695 unigenes and 2446 differentially expressed unigenes. Subcategories are indicated on the x-axis. The right y-axis indicates the number of genes in each category, while the left y-axis represents the percentage of genes within each GO category falling into a specific subcategory. Gray bars correspond to all annotated unigenes and red bars indicate differentially expressed unigenes (DEGs).

processes, 11 346 under cellular components, and 3552 under molecular function. Within the molecular function category, the majority of DEGs were related to binding (1655; 67.6%). With respect to cellular components, most assignments were to cell (2283; 93.2%) and cell part (2283; 93.2%) subcategories. Among biological processes, the cellular subcategory (2016; 82.3%) was the most highly represented. These results suggest that the flowering stage in *Phalaenopsis* is predominantly controlled by genes related to cellular structure and molecular interactions. In contrast, almost no DEGs were assigned to virion (2; 0.8%) and virion part (2; 0.8%) under cellular components, cell killing (1; 0.4%) under biological processes, and chemoattractant activity (1; 0.4%) and receptor regulator activity (1; 0.4%) under molecular function.

KEGG PATHWAY ENRICHMENT ANALYSIS. Pathway enrichment analysis identifies significantly enriched metabolic pathways or signal transduction pathways in DEGs by comparing them with the whole genome background. Using pathway enrichment analysis, DEG-associated metabolic and signal transduction pathways can be identified. On the basis on sequence homology, 837 up-regulated DEGs (432 in petals and 405 in labella) were mapped to 102 KEGG pathways. This result suggests that the development and formation of petals and labella involve two totally different processes. Almost all KEGG pathways were found in the two floral parts: 102 pathways in petals and 99 in labella (Tab. 2).

Among DEGs, 441 could be mapped to 10 secondary metabolic sub-pathways. We identified many gene families that encode key enzymes involved in major secondary metabolic pathways, such as those associated with betalain, flavone and flavonol, phenylpropanoids, flavonoids, terpenoids, and carotenoids. A total of 2822 genes involved in the 10 sub-pathways were detected in our study. The broad coverage obtained for these secondary metabolic genes provides more abundant information for examining secondary metabolite biosynthesis in *Phalaenopsis*.

Expression levels of these genes exhibited great differences. The 432 DEGs up-regulated in petals mapped to 102 KEGG pathways distributed into the categories of metabolism (292 unigenes), genetic information processing (122 unigenes), organism systems (12 unigenes), cellular processes (4 unigenes), environmental information processing (9 unigenes), and human diseases (2 unigenes). The distribution of the 405 DEGs upregulated in labella into 94 KEGG pathways was as follows: genetic information processing (201 unigenes), metabolism (146 unigenes), cellular processes (38 unigenes), environmental information processing (19 unigenes), organism systems (15 unigenes), and human diseases (50 unigenes). The broad distribution of both petal and labella unigenes in these essential processes suggests that the associated biochemical pathway networks are complicated.

Tab. 2 Some unigenes mapped to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

KEGG pathways	Sub-pathways of KEGG pathway	Branches of sub-pathway	PN	PNU	PL	PLU
Metabolism	Energy metabolism	Photosynthesis–antenna proteins	11	0	11	11
		Photosynthesis	12	2	12	10
		Carbon fixation in photosynthetic organisms	14	0	14	14
		Oxidative phosphorylation	32	3	32	29
	Biosynthesis of other secondary metabolites	Tropane, piperidine and pyridine alkaloid biosynthesis	1	0	1	1
		Flavonoid biosynthesis	3	2	6	4
		Flavone and flavonol biosynthesis	1	0	2	2
		Stilbenoid, diarylheptanoid and gingerol biosynthesis	2	1	2	1
		Betalain biosynthesis	1	1	0	0
		Isoquinoline alkaloid biosynthesis	2	1	1	1
	Lipid metabolism	Glycerolipid metabolism	3	1	2	2
		Glycerophospholipid metabolism	4	1	3	3

PN – number of mapped unigenes from petals; PNU – number of mapped up-regulated unigenes from petals; PL – number of mapped unigenes from labella; PLU – number of mapped up-regulated unigenes from labella.

With respect to metabolic processes (Tab. 2), the distributions of petal and labella unigenes involved in energy metabolism were significantly different. As 67 unigenes from labella were participants in this pathway compared with 10 unigenes from petals, more DEGs in labella were accordingly related to this KEGG sub-pathway. These results suggest that the labella is associated with a more complex energy metabolism mechanism than are petals; the labella may thus not only be an energy sink, but may also serve as an energy source for its own function or supply energy to other floral organs.

Genes related to floral differentiation and color

Transcription factors control genomic expression and play vital roles in all aspects of higher plant life cycles. The distribution of transcription factors associated with floral differentiation – including MADS (16 DEGs), MYB, ARF, NAC, and TCP – was similar between petals and labella, but their expression levels were significantly different (Tab. 3).

Orchids produce flowers of a variety of colors ranging from greenish, white, and yellow to intense red, blue, and brown. Several anthocyanin biosynthetic DEGs, such as *CHI*, *CHS*, UDP-glucose: flavonoid-3-O-glucosyltransferase (*UFGT3*), *DFR*, and *ANS*, were uncovered in this study (Tab. 4).

Discussion

In this study, we obtained 37 723 and 34 020 unigenes from petals and labella, respectively. A total of 2736 DEGs were identified, many of which were functionally annotated by BLAST searches against several public databases. We mapped 837 DEGs (432 up-regulated in petals and 405 up-regulated in labella) to 102 KEGG pathways. Almost all of these pathways were found in both flower parts: 102 pathways in petals and 99 in labella. DEGs involved in energy metabolism were significantly differentially distributed

between labella and petals. In addition, several DEGs related to flower color and floral differentiation were found in labella and petals.

Genes related to floral differentiation

MADS-box genes have long been noted for their roles in floral organ identity. In this study, we detected 16 genes differentially expressed between petals and labella that mapped to 29 MADS-like genes. All 16 of these DEGs were expressed in petals, with 9 found to be up-regulated. Only 12 of the 16 DEGs were expressed in labella, with 7 up-regulated in that floral organ. One of the latter seven DEGs was a *MADS4*-like unigene, which is specifically expressed in the labella and can be ectopically expressed in the labella-like petals of a peloric mutant [14]. Another identified unigene was homologous to AGL9-like genes, which may function at the top of a proposed regulatory hierarchy of MADS-box genes controlling floral development [15]. The TCP family is an ancient group of plant developmental transcription factors that regulate cell division in vegetative and reproductive structures and that are essential to the establishment of flower zygomorphy [16]. Three unigenes associated with TCP were identified as expressed in both petals and labella, although all three had much higher expression levels in labella than in petals. Studies have shown that TCP may have played a role in the diversification and radiation of Orchidaceae [17]. The elevated expression of these genes in labella imply that TCP has a more prominent role in patterning of labella than of petals. MYB transcription factors, which can affect overall petal morphology by influencing cell shape [18], were more numerous and more highly expressed in labella (9) than in petals (3). These genes may therefore also contribute to differences in differentiation between the labella and petals. We detected 10 unigenes that encode auxin response factor (ARF); although all of them were expressed in both labella and petals, more were up-regulated in labella (7 unigenes) than in petals (3 unigenes). Auxin signaling plays key roles in many plant growth and developmental processes from

Tab. 3 Some genes related to floral differentiation and their expressions in petals and labella.

Genes related to floral differentiation	Total No. of DEGs	No. of DEGs in petals	No. of up-regulated DEGs in petals	No. of DEGs in labella	No. of up-regulated DEGs in labella
MADS	16	16	9	12	7
MYB	12	9	3	12	9
ARF	10	10	3	10	7
NAC	6	6	1	6	5
bZIP	5	5	2	5	3
AP2	4	4	2	4	2
TCP	3	3	0	3	3
WRKY	3	3	1	3	2
YABBY	1	1	1	1	0
HD-ZIP	1	1	0	1	1
C3H	1	1	1	1	0

Tab. 4 Some genes related to flower color and their expressions in petals and labella.

Classification of flower color genes	Genes related to flower color	Total No. of DEGs	No. of DEGs in petals	No. of up-regulated DEGs in petals	No. of DEGs in labella	No. of up-regulated DEGs in labella
Genes related to anthocyanin synthesis	*CHI*	3	3	3	3	0
	CHS	1	0	0	1	1
	DFR	1	0	0	1	1
	UFGT3	1	1	0	1	1
	ANS(AS)	1	0	0	1	1
	F3'H	1	0	0	1	1
Genes related to carotenoid synthesis	*PSY*	1	1	1	1	0
	PAL	2	2	0	2	2
	FLS	1	1	0	1	1
	PDS	1	1	1	1	0
	DXS	1	1	1	1	0
	IPI	1	1	0	0	0
	GGPS	1	1	1	1	0
Some other related genes	*MYB*	12	9	3	12	9
	bZIP	5	5	2	5	3

embryogenesis to senescence [19]. Because more DEGs are up-regulated in labella, auxin may have a more complex role in labella patterning than in petals. We also found NAC transcription factors in both labella and petals, with a higher number of up-regulated DEGs identified in labella (5) than in petals (1). A previous study has shown that NAC transcription factors may be involved in pathogen infection responses in rice [20]. The much higher expression observed in labella compared with petals may serve as a defensive weapon in the labella, which functions as a landing platform and nectar guide for pollinators. WRKY transcription factors are key regulators of many plant processes, including biotic and abiotic stress response, senescence, seed dormancy, and seed germination [21]. We found WRKY transcription factors in both labella and petals, with a greater number of up-regulated DEGs identified in labella (2 DEGs) than in petals (1 DEGs). WRKY1 regulates sesquiterpene synthase gene expression and the accumulation of sesquiterpene induced by bacterial and fungal infection [22]; consequently, the observed coexpression of *WRKY* genes and monoterpene and sesquiterpene synthesis genes in labella and petals suggests that the function of these genes may be related to defense against biotic stress. The much higher expression in labella than in petals also suggests a defensive role for WRKY transcription factors in the labella, a landing platform for pollinators. We also identified genes that affect floral organ development [23–25], namely *YABBY*, basic region/leucine zipper motif (*bZIP*) transcription factor, and *C3H*-type zinc finger protein genes. These developmental genes were found to have different expression levels between labella and petals, suggesting that they are associated with labella and petal development but may have different functional mechanisms.

Genes related to flower color

A single *PeUFGT3* gene was detected in our study, and was found to be more highly expressed in red labella than

in white petals. This observation is consistent with a report that *PeUFGT3* plays a critical role in red color formation in *Phalaenopsis* [8]. On the basis of these findings, we propose the following mechanism to explain red color formation in the labella (Fig. 5). First, CHS, the key enzyme in the flavonoid biosynthesis pathway, catalyzes the condensation of three acetate residues from malonyl-CoA with 4-coumaroyl-CoA to form tetrahydroxychalcone. Next, tetrahydroxychalcone is stereospecifically isomerized by CHI to colorless flavanone, which is catalyzed further to colorless dihydroquercetin by flavanone 3-hydroxylase (F3H) and flavonoid 3′-hydroxylase (F3′H). Dihydroquercetin is catalyzed in turn by DFR, ANS, and UFGT3 to form an anthocyanin (cyanidin-3-glucoside), thereby generating the red color of the labella [5,6]. In contrast to a previous study [8], however, we did not identify any DEGs encoding CHS, ANS, DFR, or F3′H in white petals. To explain this discrepancy, we suggest two possibilities. First, no method is perfect; we may thus have failed to detect the expression of these genes in white petals because of technical limitations. Alternatively, because different *Phalaenopsis* plants have inherited different characteristics, an anthocyanin synthesis pathway may not have existed in our chosen white petal samples (possibly the main reason for their white color) [26]. We also identified several DEGs that encode key enzymes regulating carotenoid synthesis in petals and labella, such as phytoene synthase (PSY), phenylalanine ammonialyase (PAL), and phytoene desaturase (PDS). In composites, floral coloration in the orange-red to red range is determined by a combination of anthocyanins and/or carotenoids [27]; the same observation has been made for orchids such as *Disa* hybrids [28]. The anthocyanin pathway may thus not be the only pathway controlling flower color formation in Phalaenopsis, especially red color. Pigment biosynthesis is
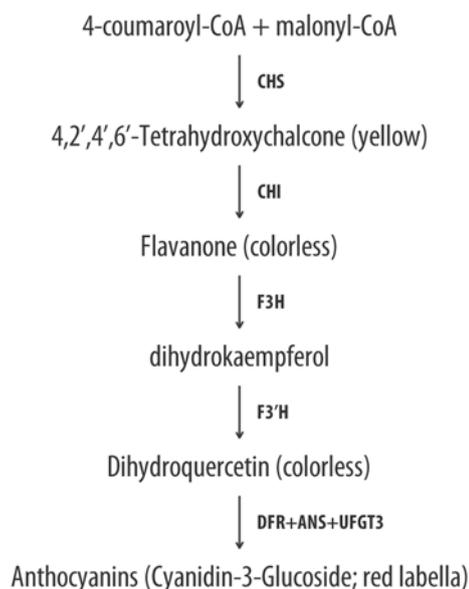

Fig. 5 Proposed pathway for synthesis of red coloration in labella.

regulated by a set of WD40, MYB, and bHLH transcription factors similar to those involved in trichome and cell shape development [9,29,30]. In this study, we detected 12 *MYB* genes in petals and labella, with more DEGs present and more strongly expressed in labella (9) than in petals (3). This differential expression may be responsible for the contrasting coloration of labella and petals.

The results of our study should serve as a valuable genetic resource for future investigations of the molecular basis of flower color and floral organ formation in *Phalaenopsis*, and inform *Phalaenopsis* flower breeding.

Authors' contributions
The following declarations about authors' contributions to the research have been made: the experimental plan was conceived and designed by CZ; participated in sample collection and RNA preparation and performed the experiments: JW, YY; contributed to data analysis, bioinformatics analysis, and manuscript preparation: ZM, GS; drafted and revised the manuscript: JW, YY. JW and YY contributed equally to this work.

Supplementary material
The following supplementary material for this article is available online at http://pbsociety.org.pl/journals/index.php/asbp/rt/suppFiles/asbp.2014.023/0:
1. Tab S1: unigenes mapped to KEGG pathways.

References
1. Teixeira da Silva JA, Chin DP, Van PT, Mii M. Transgenic orchids. Sci Hortic. 2011;130(4):673–680. http://dx.doi.org/10.1016/j.scienta.2011.08.025
2. Mondragón-Palomino M, Theißen G. MADS about the evolution of orchid flowers. Trends Plant Sci. 2008;13(2):51–59. http://dx.doi.org/10.1016/j.tplants.2007.11.007
3. Mondragon-Palomino M, Theissen G. Why are orchid flowers so diverse? Reduction of evolutionary constraints by paralogues of class B floral homeotic genes. Ann Bot. 2009;104(3):583–594. http://dx.doi.org/10.1093/aob/mcn258
4. Salemme M, Sica M, Gaudio L, Aceto S. Expression pattern of two paralogs of the PI/GLO-like locus during *Orchis italica* (Orchidaceae, Orchidinae) flower development. Dev Genes Evol. 2011;221(4):241–246. http://dx.doi.org/10.1007/s00427-011-0372-6
5. Tanaka Y, Sasaki N, Ohmiya A. Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. Plant J. 2008;54(4):733–749. http://dx.doi.org/10.1111/j.1365-313X.2008.03447.x
6. Tanaka Y, Brugliera F, Kalc G, Senior M, Dyson B, Nakamura N, et al. Flower color modification by engineering of the flavonoid biosynthetic pathway: practical perspectives. Biosci Biotechnol Biochem. 2010;74(9):1760–1769. http://dx.doi.org/10.1271/bbb.100358
7. Fukuchi-Mizutani M. Biochemical and molecular characterization of a novel UDP-glucose:anthocyanin 3'-O-glucosyltransferase, a key enzyme for blue anthocyanin biosynthesis, from gentian. Plant Physiol. 2003;132(3):1652–1663. http://dx.doi.org/10.1104/pp.102.018242
8. Chen WH, Hsu CY, Cheng HY, Chang H, Chen HH, Ger MJ. Down-regulation of putative UDP-glucose: flavonoid 3-O-glucosyltransferase gene alters flower coloring in *Phalaenopsis*. Plant Cell Rep. 2011;30(6):1007–1017. http://dx.doi.org/10.1007/s00299-011-1006-1
9. Koes R, Verweij W, Quattrocchio F. Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci. 2005;10(5):236–242. http://dx.doi.org/10.1016/j.tplants.2005.03.002
10. Harborne JB, Williams CA. Anthocyanins and other flavonoids

(January 1998 to December 2000). Nat Prod Rep. 2001;18(3):310–333. http://dx.doi.org/10.1039/b006257j

11. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5(7):621–628. http://dx.doi.org/10.1038/nmeth.1226

12. Xu P, Liu Z, Fan X, Gao J, Zhang X, Zhang X, et al. De novo transcriptome sequencing and comparative analysis of differentially expressed genes in *Gossypium aridum* under salt stress. Gene. 2013;525(1):26–34. http://dx.doi.org/10.1016/j.gene.2013.04.066

13. Hao QN, Zhou XA, Sha AH, Wang C, Zhou R, Chen SL. Identification of genes associated with nitrogen-use efficiency by genome-wide transcriptional analysis of two soybean genotypes. BMC Genomics. 2011;12(1):525. http://dx.doi.org/10.1186/1471-2164-12-525

14. Tsai WC, Hsiao YY, Pan ZJ, Hsu CC, Yang YP, Chen WH, et al. Molecular biology of orchid flowers. Adv Bot Res. 2008;47:99–145. http://dx.doi.org/10.1016/s0065-2296(08)00003-7

15. Yu H. Identification and characterization of three orchid MADS-box genes of the AP1/AGL9 subfamily during floral transition. Plant Physiol. 2000;123(4):1325–1336. http://dx.doi.org/10.1104/pp.123.4.1325

16. Mondragon-Palomino M, Trontin C. High time for a roll call: gene duplication and phylogenetic relationships of TCP-like genes in monocots. Ann Bot. 2011;107(9):1533–1544. http://dx.doi.org/10.1093/aob/mcr059

17. Sargent RD. Floral symmetry affects speciation rates in angiosperms. Proc R Soc Lond B Biol Sci. 2004;271(1539):603–608. http://dx.doi.org/10.1098/rspb.2003.2644

18. Baumann K, Perez-Rodriguez M, Bradley D, Venail J, Bailey P, Jin H, et al. Control of cell and petal morphogenesis by R2R3 MYB transcription factors. Development. 2007;134(9):1691–1701. http://dx.doi.org/10.1242/dev.02836

19. Guilfoyle TJ, Hagen G. Auxin response factors. Curr Opin Plant Biol. 2007;10(5):453–460. http://dx.doi.org/10.1016/j.pbi.2007.08.014

20. Nakashima K, Tran LSP, van Nguyen D, Fujita M, Maruyama K, Todaka D, et al. Functional analysis of a NAC-type transcription factor OsNAC6 involved in abiotic and biotic stress-responsive gene expression in rice: rice OsNAC6 functions in stress responses. Plant J. 2007;51(4):617–630. http://dx.doi.org/10.1111/j.1365-313X.2007.03168.x

21. Rushton DL, Tripathi P, Rabara RC, Lin J, Ringler P, Boken AK, et al. WRKY transcription factors: key components in abscisic acid signalling. Plant Biotechnol J. 2012;10(1):2–11. http://dx.doi.org/10.1111/j.1467-7652.2011.00634.x

22. Rushton PJ, Bokowiec MT, Han S, Zhang H, Brannock JF, Chen X, et al. Tobacco transcription factors: novel insights into transcriptional regulation in the Solanaceae. Plant Physiol. 2008;147(1):280–295. http://dx.doi.org/10.1104/pp.107.114041

23. Golz JF, Hudson A. Plant development: YABBYs claw to the fore. Curr Biol. 1999;9(22):R861–R863. http://dx.doi.org/10.1016/S0960-9822(00)80047-0

24. Jakoby M, Weisshaar B, Dröge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, et al. bZIP transcription factors in *Arabidopsis*. Trends Plant Sci. 2002;7(3):106–111. http://dx.doi.org/10.1016/S1360-1385(01)02223-3

25. Huang P, Ju HW, Min JH, Zhang X, Chung JS, Cheong HS, et al. Molecular and physiological characterization of the *Arabidopsis* thaliana oxidation-related zinc finger 2, a plasma membrane protein involved in ABA and salt stress response through the ABI2-mediated signaling pathway. Plant Cell Physiol. 2012;53(1):193–203. http://dx.doi.org/10.1093/pcp/pcr162

26. van der Krol AR. Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. Plant Cell. 1990;2(4):291–299. http://dx.doi.org/10.1105/tpc.2.4.291

27. Kishimoto S, Sumitomo K, Yagi M, Nakayama M, Ohmiya A. Three routes to orange petal color via carotenoid components in 9 compositae species. J Jpn Soc Hortic Sci. 2007;76(3):250–257. http://dx.doi.org/10.2503/jjshs.76.250

28. Matsui S. Varietal differences in flower colors of *Cattleya aurantiaca*. Res Bull Fac Agric Gifu Univ Jpn. 1992;57:181–185.

29. Tatsuzawa F, Ichihara K, Shinoda K, Miyoshi K. Flower colours and pigments in *Disa* hybrid (Orchidaceae). South Afr J Bot. 2010;76(1):49–53. http://dx.doi.org/10.1016/j.sajb.2009.06.017

30. Grotewold E. The genetics and biochemistry of floral pigments. Ann Rev Plant Biol. 2006;57(1):761–780. http://dx.doi.org/10.1146/annurev.arplant.57.032905.105248