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 differentiation

Yuxia Yang¹, Jingjing Wang¹,², Zhihu Ma³, Guosheng Sun³, Changwei Zhang¹*

¹ College of Horticulture, Nanjing Agricultural University, 1# Weigang, Nanjing, Jiangsu Province, 210095, P.R. China
² Dongxin Department, Jiangsu Provincial Agricultural Reclamation and Development Corporation, 26# Dongfang Middle Road, Lianyungang, Jiangsu Province, 222248, P.R. China
³ Zhejiang 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...
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 anthocyanins [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:

\[
P(y|x) = \frac{N_2}{N_1^2} \frac{(x + y)!}{x! \cdot y!} \left(1 + \frac{N_2}{N_1} \frac{x+y+1}{x+y} \right)
\]

where \(N_1\) and \(N_2\) 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...
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 \([\log \text{Ratio}] \geq 1\) 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:

\[
P = 1 - \sum_{i=0}^{n-1} \binom{M}{i} \frac{(N-M)}{(n-i)} \frac{(n-M)}{(n-i)}
\]

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 948 bp, respectively. Scaffold N50 lengths were 1280 bp for petals and 1168 bp for labella.

<table>
<thead>
<tr>
<th>No. of unigenes</th>
<th>37 723</th>
<th>34 020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total nucleotides (nt)</td>
<td>26 656 163</td>
<td>27 636 110</td>
</tr>
<tr>
<td>Length of N50 (bp)</td>
<td>1125</td>
<td>1398</td>
</tr>
<tr>
<td>Average length of unigenes (bp)</td>
<td>707</td>
<td>812</td>
</tr>
<tr>
<td>No. of unigenes above 500 bp</td>
<td>15 903</td>
<td>15 984</td>
</tr>
</tbody>
</table>

#### 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 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 \(|\log \text{Ratio}| \geq 1\)) 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 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 \((E\text{-value} \leq 1 \times 10^{-5})\) 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.
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...
processes, 11,346 under cellular components, and 3,552 under molecular function. Within the molecular function category, the majority of DEGs were related to binding (1,635; 67.6%). With respect to cellular components, most assignments were to cell (2,283; 93.2%) and cell part (2,283; 93.2%) subcategories. Among biological processes, the cellular subcategory (2,016; 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 of 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 2,822 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.
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...
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

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

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

<table>
<thead>
<tr>
<th>Classification of flower color genes</th>
<th>Genes related to flower color</th>
<th>Total No. of DEGs</th>
<th>No. of DEGs in petals</th>
<th>No. of up-regulated DEGs in petals</th>
<th>No. of DEGs in labella</th>
<th>No. of up-regulated DEGs in labella</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genes related to anthocyanin synthesis</td>
<td>CHI</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CHS</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>DFR</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>UFGT3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ANS(AS)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>FYH</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Genes related to carotenoid synthesis</td>
<td>PSY</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>PAL</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>FLS</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>PDS</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>DXS</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>IPI</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Some other related genes</td>
<td>MYB</td>
<td>12</td>
<td>9</td>
<td>3</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>bZIP</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

**Tab. 4** Some genes related to flower color and their expressions in petals and labella.
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-coumaryl-CoA to form tetrahydroxylchalcone. Next, tetrahydroxylchalcone is stereospecifically isomerized by CHI to colorless flavanone, which is catalyzed further to colorless dihydroquercetin by flavanone 3-hydroxylase (F3H) and flavonoid 3’-hydroxylase (F3H1). 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 F3H 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 ammonia lyase (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 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.

References

10. Harborne JB, Williams CA. Anthocyanins and other flavonoids
Yang et al. / Comparative transcriptome analysis of Phalaenopsis flowers 199


