

RESEARCH ARTICLE

Digital Gene Expression Analysis Based on *De Novo* Transcriptome Assembly Reveals New Genes Associated with Floral Organ Differentiation of the Orchid Plant *Cymbidium ensifolium*

Fengxi Yang, Genfa Zhu*

Guangdong Key Laboratory of Ornamental Plant Germplasm Innovation and Utilization, Environmental Horticulture Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, 510640, P. R. China

* genfazhu@163.com



OPEN ACCESS

Citation: Yang F, Zhu G (2015) Digital Gene Expression Analysis Based on *De Novo* Transcriptome Assembly Reveals New Genes Associated with Floral Organ Differentiation of the Orchid Plant *Cymbidium ensifolium*. PLoS ONE 10 (11): e0142434. doi:10.1371/journal.pone.0142434

Editor: Serena Aceto, University of Naples Federico II, ITALY

Received: June 29, 2015

Accepted: October 21, 2015

Published: November 18, 2015

Copyright: © 2015 Yang, Zhu. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This research was supported by grants from the Strategic Emerging Industries of Guangdong Province (2012A020800003), the National Natural Science Foundation of China (31471915), and Agriculture common key technology research (2011A020102007).

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Cymbidium ensifolium belongs to the genus *Cymbidium* of the orchid family. Owing to its spectacular flower morphology, *C. ensifolium* has considerable ecological and cultural value. However, limited genetic data is available for this non-model plant, and the molecular mechanism underlying floral organ identity is still poorly understood. In this study, we characterize the floral transcriptome of *C. ensifolium* and present, for the first time, extensive sequence and transcript abundance data of individual floral organs. After sequencing, over 10 Gb clean sequence data were generated and assembled into 111,892 unigenes with an average length of 932.03 base pairs, including 1,227 clusters and 110,665 singletons. Assembled sequences were annotated with gene descriptions, gene ontology, clusters of orthologous group terms, the Kyoto Encyclopedia of Genes and Genomes, and the plant transcription factor database. From these annotations, 131 flowering-associated unigenes, 61 *CONSTANS-LIKE* (*COL*) unigenes and 90 floral homeotic genes were identified. In addition, four digital gene expression libraries were constructed for the sepal, petal, labellum and gynostemium, and 1,058 genes corresponding to individual floral organ development were identified. Among them, eight MADS-box genes were further investigated by full-length cDNA sequence analysis and expression validation, which revealed two *APETALA1/AGL9*-like MADS-box genes preferentially expressed in the sepal and petal, two *AGA-MOUS*-like genes particularly restricted to the gynostemium, and four *DEF*-like genes distinctively expressed in different floral organs. The spatial expression of these genes varied distinctly in different floral mutant corresponding to different floral morphogenesis, which validated the specialized roles of them in floral patterning and further supported the effectiveness of our *in silico* analysis. This dataset generated in our study provides new insights into the molecular mechanisms underlying floral patterning of *Cymbidium* and supports a valuable resource for molecular breeding of the orchid plant.

Introduction

Cymbidium is a very economically important flowering genus of the orchid family (Orchidaceae) in China, Japan, Korea, and Southeast Asia [1, 2]. *Cymbidium ensifolium*, which belongs to the subgenus *Jensoa* of genus *Cymbidium*, is a prevalently potted plant and cut flower cultivated in China [3, 4]. Its unique floral organ includes three sepals in the first whorl, three petals in the second whorl and productive parts in the center of the flower. The sepals and petals together are called the tepals. Among them, two of the petals are similar to each other and resemble unmodified sepals, while the third is highly modified and is called the labellum (or lip). Its shape is ovate to triangular and often has a kinked to undulated margin. The male and female productive organs are highly fused to form a gynostemium (or column), which evolved through complete fusion of the style, stigma and staminal filament, and has four pollinia on a semi-circular viscidium [5, 6]. Despite this spectacular flower morphology, various mutations occur frequently in the orchid family and greatly diversify the floral morphology [7–12]. For example, the gynostemium is replaced by a new emerged flower in the multi-tepals mutant, which continues to produce sepals, petals, and gynostemium centripetally, whereas in the non-gynostemium mutant, the gynostemium is absent and replaced by new sepals in the center part of the flower. Owing to their highly specialized and diversified floral morphology, these mutants are much more valuable and attract more horticultural and commercial attention. However, the functional genomic studies and the gene discovery associated with floral patterning regulation remains greatly limited in the *Cymbidium* orchid due to the polyploidy genomes and long juvenile phases [13, 14]. Despite their apparent importance, molecular and genetic approaches to orchid flower development and evolution are still in their infancy.

In recent years, next-generation sequencing (NGS) technologies have provided powerful tools for high-throughput sequence determination. Various NGS-based RNA-sequencing modes have made obtaining massive sequences with enormous depth and coverage possible, thus enabling the discovery of novel genes [15–17]. An increasing number of studies have highlighted the utility of cDNA sequencing for discovering candidate genes associated with floral development, floral scent production, or flowering time in various orchid species (such as genera *Phalaenopsis*, *Oncidium*, *Ericina*, *Dendrobium*) in the absence of a genomic sequence [18–21]. For example, using genetic recourses, four *APETALA3*-like genes, *PeMADS2*, *PeMADS3*, *PeMADS4*, and *PeMADS5*, were found to display distinct floral morphogenetic roles in various floral organs of *Phalaenopsis*. Additionally, *PeMADS6* was found to be involved in petaloid formation and correlated with ovary development [22, 23]. In *Oncidium Gower Ramsey*, *OMADS1* was isolated and characterized as a functional gene promoting flower transition and formation by acting as an activator of *FT* and *SOC1*, and was able to strongly interact with *OMADS3*, which influenced flower formation and floral initiation [24]. In *Dendrobium crumenatum*, B function genes *DcOAP3A/B* and *DcOPI* could form heterodimers and further interact with *DcOSEP* to form higher protein complexes and mediate flower formation and development [25]. However, similar studies in the genera *Cymbidium* are limited, and the number of genes in current databases is still not enough to support further functional genomics studies when compared with other plant species [26, 27]. A very recent result from next-generation sequencing generated 51,696 genes for *C. ensifolium* [26]. Nevertheless, little functional genomics research has been performed, and the molecular aspects of flowering control and floral organ development of *C. ensifolium* remain largely unknown.

To obtain an overview of the gene expression information and provide plenty of functionally characterized candidate genes directly associated with floral development in *Cymbidium*, we sequenced the floral transcriptome of *C. ensifolium* ‘tiansesu’, which is a typical *C. ensifolium* cultivar, well-known for its valuable horticultural traits, including large flowers and varied

floral morphology [1, 3]. A total of 10.9 Gb clean sequence data and 111,892 unigenes were generated, and an informative EST dataset was obtained, which can be used as an important resource for molecular breeding, marketable traits, and investigating various biological processes in *Cymbidium*. To facilitate identifying sets of genes involved in individual floral organ development, we developed DGE tags derived from different floral organs. More than 1,000 genes with highly differential expression levels were identified as candidates connected with floral pattern regulation. Among them, we concentrated on the MIKC-type transcription factors and found 16 MADS-box genes expressed preferentially among the sepal, petal, labellum and gynostemium. Phylogenetic analysis demonstrated that 14 of these MADS-box genes clustered well with the floral homeotic genes isolated from other orchid species. Eight of them were further characterized by gene cloning and expression analysis. The result from quantitative RT-PCR suggested distinct roles of them on floral organ identity, which indicated that we developed a useful transcriptome database for identifying the candidate genes responding to the individual floral organ development. This study provides new insights into the molecular mechanisms underlying floral patterning and broadens our understanding of flower development in *Cymbidium*.

Materials and Methods

Plant materials and growth conditions

The plants of *C. ensifolium* 'tiansesu' used in this study were artificially cultivated and collected from the cultivation base of Environmental Horticulture Research Institute, Guangdong Academy of Agricultural Sciences, China. All of the plants were grown and maintained in pots in a greenhouse at day/night temperatures of 26/23°C under a 16-h light /8-h dark photoperiod.

cDNA preparation and Illumina sequencing

For transcriptome sequencing, the cDNA library was prepared from an equal mixture of RNAs isolated from flower buds and mature flowers of *C. ensifolium* 'tiansesu'. For the Digital Gene Expression (DGE) analysis, we constructed four independent cDNA libraries for the sepal, petal, labellum, and gynostemium separately. Total RNA was extracted from approximately 0.5 g of each tissue using the TRIzol reagent (Invitrogen). Individual mRNAs were purified from total RNA using the Oligotex mRNA Midi Kit (QIAGEN) and quantified using a NanoDrop 2000 spectrophotometer (Thermo Scientific) to generate the cDNA library according to the Illumina manufacturer's instructions. Briefly, fragmentation buffer was added to interrupt mRNA to short fragments. Random hexamer primers were added to these short fragments to synthesize the first-strand cDNA. The second-strand cDNA was synthesized using the SuperScript double-stranded cDNA synthesis kit (Invitrogen) and purified with a QiaQuick PCR extraction kit (QIAGEN). The double-stranded cDNA was sequenced using Illumina HiSeq 2500 platform.

Sequence assembly and annotation

The resulting raw sequence reads with weak signal or low quality were screened and trimmed by GS FLX pyrosequencing software, to yield a final dataset comprised of high-quality (HQ) (>99.5% accuracy of single-base reads) sequences. Prior to assembly, primer and adapter sequences were trimmed from the HQ dataset, and sequences shorter than 50 bp were removed. The remaining data were assembled into unique sequences (including contigs) using trinityrn (http://trinityrnaseq.sourceforge.net/analysis/extract_proteins_from_trinity_transcripts.html).

For unigene annotations, the genes in which could find protein coding sequences were searched against the NCBI non-redundant protein (Nr) database (<http://www.ncbi.nlm.nih.gov>) and the SwissProt protein database (<http://www.expasy.ch/prot>) using the BLASTP algorithm with an E-value cut-off of 10^{-5} . The remaining sequences were searched against public databases, using the BLASTX algorithm with an E-value cut-off of 10^{-5} [28].

Gene ontology (GO) classification was obtained by GO terms in the database (<http://www.geneontology.org/>). The unigene sequences were also aligned to the COG database to predict and classify functions [29]. Pathway assignments were made according to the Kyoto Encyclopedia of Genes and Genome (KEGG) mapping project, (<http://www.genome.jp/>) [30]. Enzyme commission (EC) numbers were assigned to unique sequences that had BLASTX scores with an E-value cut-off of 10^{-5} as determined by searching the KEGG database. The unique sequences were allocated to specific biochemical pathways according to the corresponding EC distribution in the KEGG database.

For identification of transcription factor-related unigenes in our *C. ensifolium* dataset, the protein sequences of predicted transcription factors for rice were downloaded from the Plant Transcription Factor Database (PTFDB; <http://plantfdb.cbi.pku.edu.cn/>). These protein sequence set of each transcription factor family from rice was BLAST against the clusters and singletons in our dataset with use of the TBLASTN programs. Sequence similarity was considered significant at E-value $< 10^{-5}$.

Expression analysis of DGE tags

The gene expression was calculated by the numbers of reads mapped to the reference sequence. To map the DGE tags, the sequenced raw data were filtered to remove low quality tags (including tags with an unknown nucleotide “N”, empty tags, and tags with only one copy number). Clean tags were mapped to our transcriptome reference database using TopHat [31]. No more than two read-mismatches and read-gap-length were allowed in the alignment.

The gene expression level was calculated by the value of FPKM (Fragments Per Kilobase of transcript per Million mapped reads). $RPKM = \frac{\text{total exon Fragments}}{\text{mapped Fragments (millions)} \times \text{exon length (KB)}}$. To compare the differences in gene expression among different floral organs, the tag frequency in each DGE libraries was statistically analyzed according to the method described by Audic and Claverie. We used $FDR < 0.001$ and an absolute value of $|\log_2(\text{Fold change})| > 1$, $q_value < 0.05$ as the threshold to determine the significant difference in gene expression [32].

Gene ontology and pathway enrichment analysis of DEG

The differentially expressed genes were annotated using GO and KEGG enrichment analyses according to a method similar to that described by Wang [32], which firstly mapped all DEGs to GO terms (or KEGG pathways) in the database (<http://www.geneontology.org/>, <http://www.genome.ad.jp/>), calculating gene numbers for every term (or pathway). Followed by a hypergeometric test to find the significantly enriched terms in DEGs compared with the genome background. We take the corrected P-value ≤ 0.05 or Q-value ≤ 0.05 as a threshold for the significantly enriched GO terms or KEGG pathways, respectively, in DEGs.

Real-time quantitative RT-PCR (qRT-PCR) analysis

When determining the circadian rhythm of the *COL* genes, leaves from one-year-old *C. ensifolium* plants were collected in 4-h intervals for 24 h after the start of light exposure in LD conditions (16-h light /8-h dark). Each sample was collected from three plants. For the expression of MADS-box genes, mature flowers were dissected into their individual floral organs (sepals,

petals, labellums and gynostemiums). Total RNAs were extracted using TRIzol RNA extraction method (Tiangen, China) and reverse transcribed into cDNA using PrimeScript[®] RT reagent kit with gDNA Eraser (Perfect Real Time) (Takara, China). The qRT-PCR was performed with a Bio-Rad CFX-96 RealTime PCR System (Bio-Rad, US) in a final volume of 20 μ l containing 2 μ l cDNA, 10 μ l SYBR premix Ex taq[™] (Takara, Japan), 0.4 μ l 10 mM the forward and reverse primers each, and 7.2 μ l RNase-free water. The thermal cycling conditions were as follows: 95°C 5 min, 40 cycles at 95°C for 5 s for denaturation and 59.8 C for 25s for annealing and extension. Ubiquitin was used as an internal control for normalization to make a comparison of gene expression level between the accessions. The primers designed with the software Primer 5.0 were listed in [S1 Table](#).

Results and Discussion

Transcriptome sequencing and assembly

We pooled the cDNA library from an equal mixture of RNAs isolated from flower buds and mature flowers. After sequencing, 109,366,378 reads (89.92% of the raw data) that passed quality control were entered into the Trinity assembly process and yielded 136,553 contigs, with 148,968,762 total residues. These contigs were further assembled into 111,892 unigenes, including 1,227 clusters and 110,665 singletons, with a minimum unigene size of 201 bp, a maximum size of 16,814 bp, an N50 of 1,697 bp and a mean length of 932.03 bp ([Table 1](#)). As shown in [Fig 1](#), the assembly produced a substantial number of large unigenes: 14,154 unigenes were > 2,000 bp in length and 34,450 unigenes were > 1,000 bp, although most unigenes were between 300 and 500 bp in length. To validate the effectiveness of our de novo assembly, the unigenes generated in this study were BLAST against the previously available ESTs derived from *C. ensifolium*. Results showed that among 111,892 unigenes, 101,486 (90.7%) matched the previously published ESTs. And the remaining 10,406 unigenes that did not match probably represent the sequences newly identified in *C. ensifolium*. These massive sequences within our dataset provide more comprehensive gene expression information that could be used as a valuable resource for exploring the genetic diversity of *C. ensifolium* and for comparative genomic studies among orchid plants.

Functional annotation and classification of the transcriptome

To functionally annotate the *C. ensifolium* floral transcriptome, the unigene sequences were used as query in a BLAST search against the NCBI non-redundant (Nr) protein, gene ontology (GO), clusters of orthologous group (COG), and the Kyoto Encyclopedia of Genes and

Table 1. Summary of sequencing and de novo assembling of the floral transcriptome in *C. ensifolium*.

Summary of <i>C. ensifolium</i> Floral Transcriptome	
Raw reads	121,626,996
Clean reads	109,366,378
Total clean nucleotides	10,924,593,174 bp
Total contigs	136,553
Total residues	148,968,762 bp
Total unigenes	111,892
Smallest unigene	201 bp
Largest unigene	16814 bp
Average length	932.03 bp

doi:10.1371/journal.pone.0142434.t001

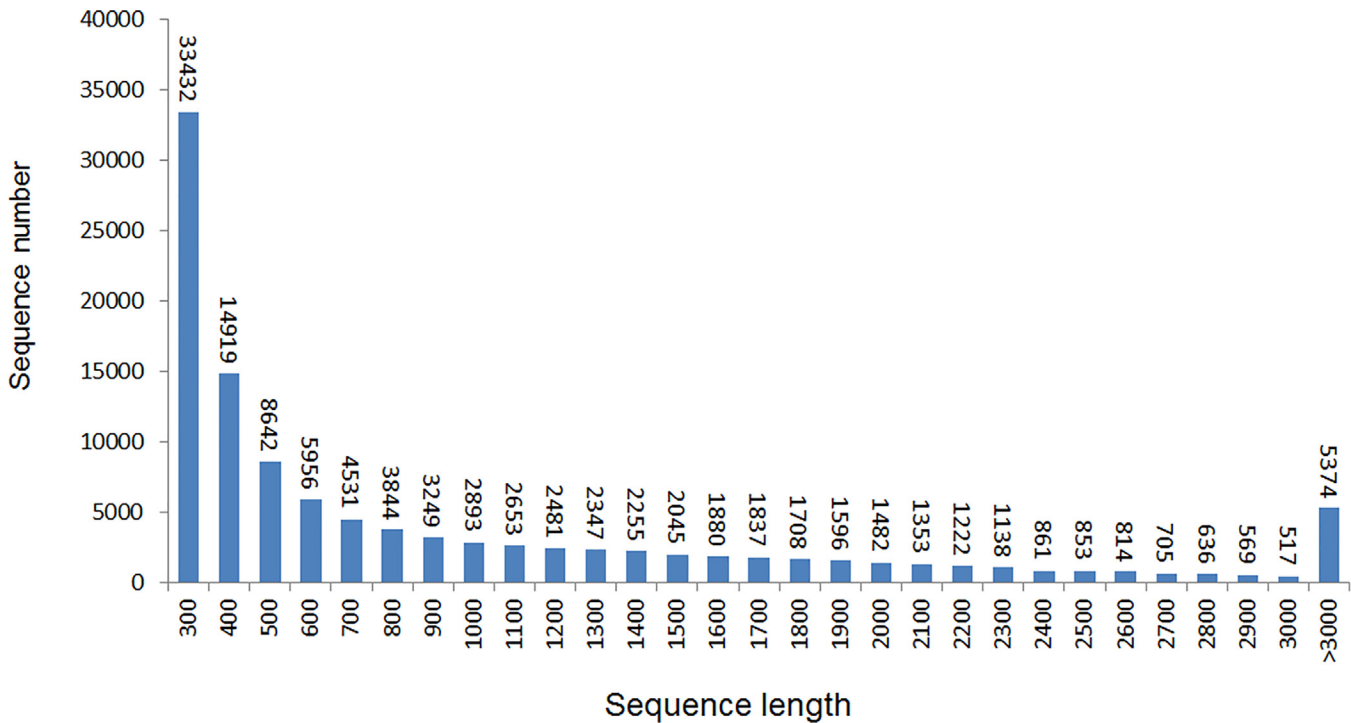


Fig 1. The length distribution of assembled unigenes. The x-axis represents the sequence length in base pairs. The y-axis represents the unigenes number.

doi:10.1371/journal.pone.0142434.g001

Genomes (KEGG) databases using a cut off E-value of e^{-5} . Consequently, a total of 62,139 unigenes (55.53% of all the assembled unigenes) provided a significant BLAST result (Table 2) and most matches were found in the NCBI-Nr database (47,285 or 42.26%). According to the E-value distribution of the top hits in the databases, 51.75% of the matched sequences showed strong homology ($< 1.0e^{-50}$), while 48.25% of the matched sequences showed moderate homology (between $1.0e^{-5}$ and $1.0e^{-50}$) (Fig 2A). The identity distribution pattern showed that 11.28% of these alignments had a similarity higher than 80%, 37.67% between 60% and 80%, and 51.05% lower than 60% (Fig 2B).

In the COG classification, 43,018 unigenes (38.45%) were classified into 25 functional classifications (Fig 3). ‘General function prediction’ was dominant (11,569), followed by ‘Posttranslational modification’ (6,795), ‘Signal transduction’ (5,266) and ‘Secondary metabolites biosynthesis’ categories (2,199). ‘RNA processing and modification’, ‘Carbohydrate transport and metabolism’, and ‘Transcription’ also shared a relatively high-percentage of genes among

Table 2. Summary of the blast hits against the known protein database.

	Num	Percent
Total	62,139	55.53%
Nr	47,285	42.26%
ORF	30,091	26.89%
COG	43,018	38.45%
GO	32,632	29.16%
KEGG	21,461	19.18%

doi:10.1371/journal.pone.0142434.t002

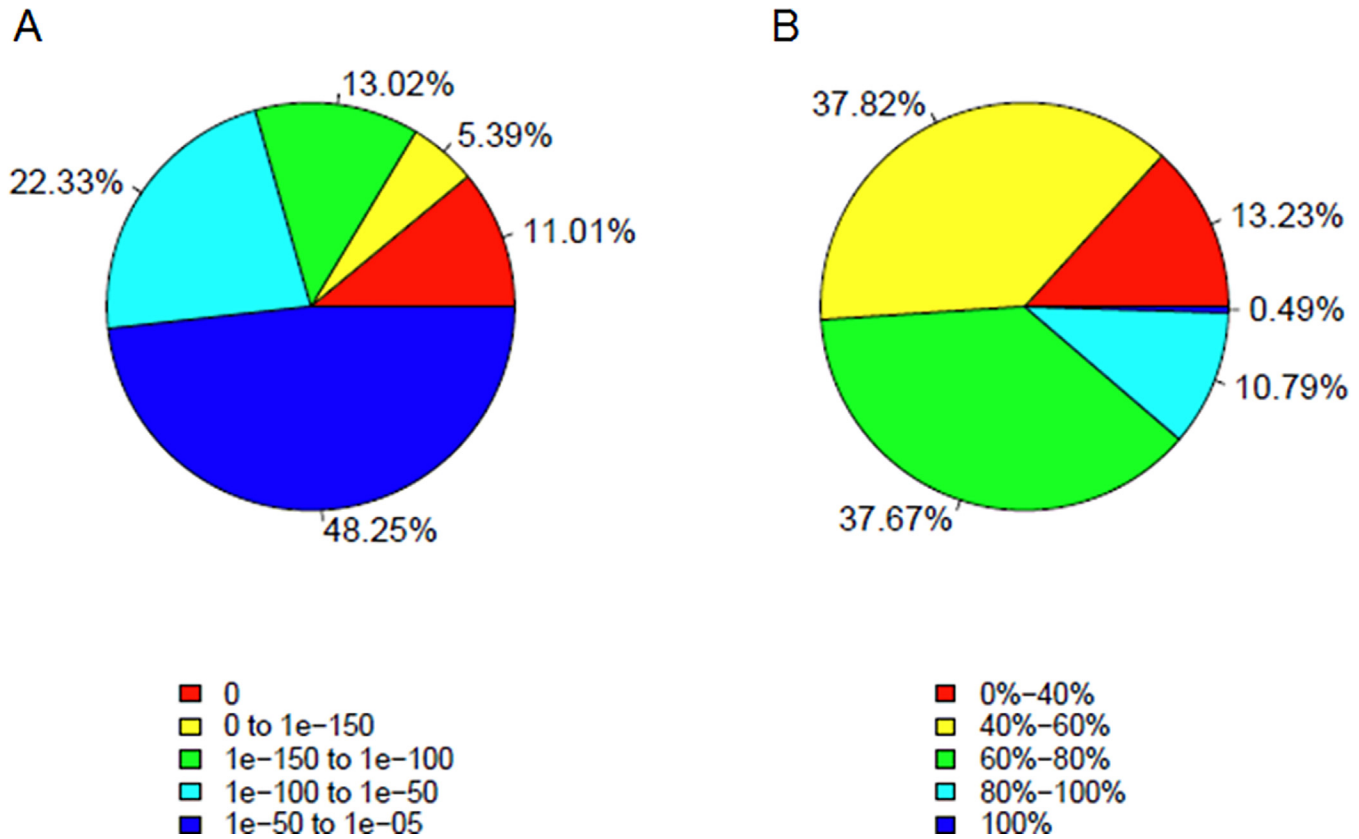


Fig 2. Characteristics of homology search of the unigenes in Nr database.

doi:10.1371/journal.pone.0142434.g002

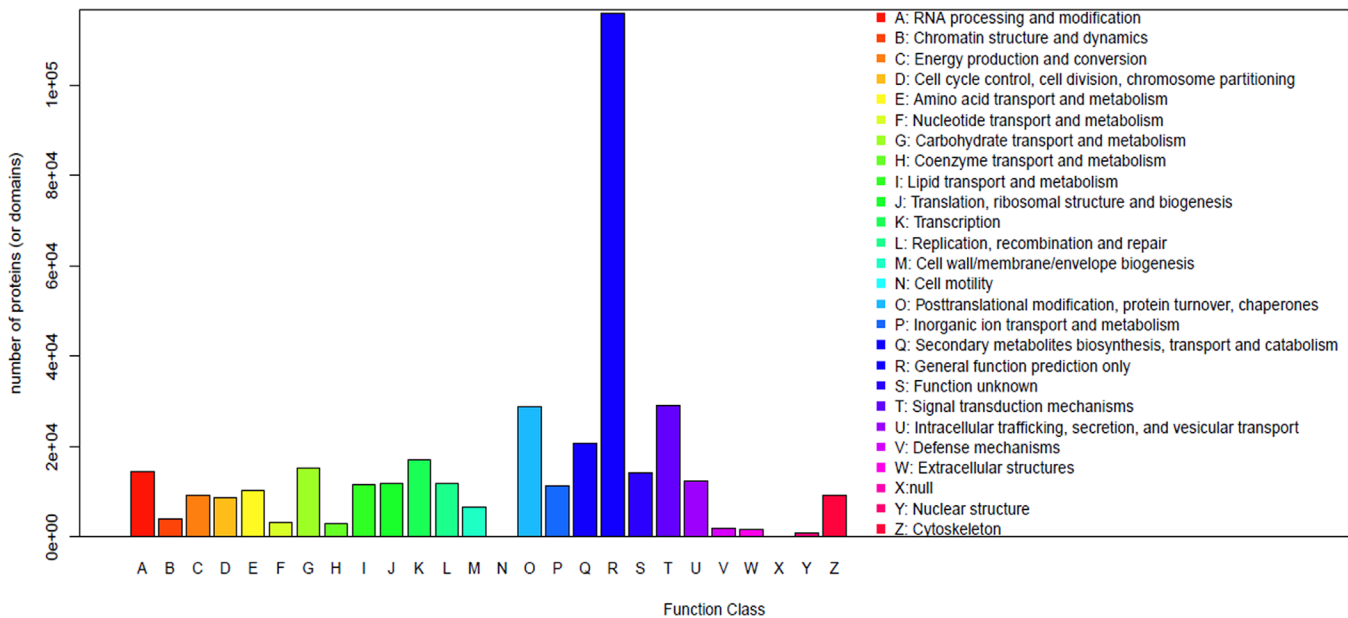


Fig 3. COG function classification of assembled unigenes.

doi:10.1371/journal.pone.0142434.g003

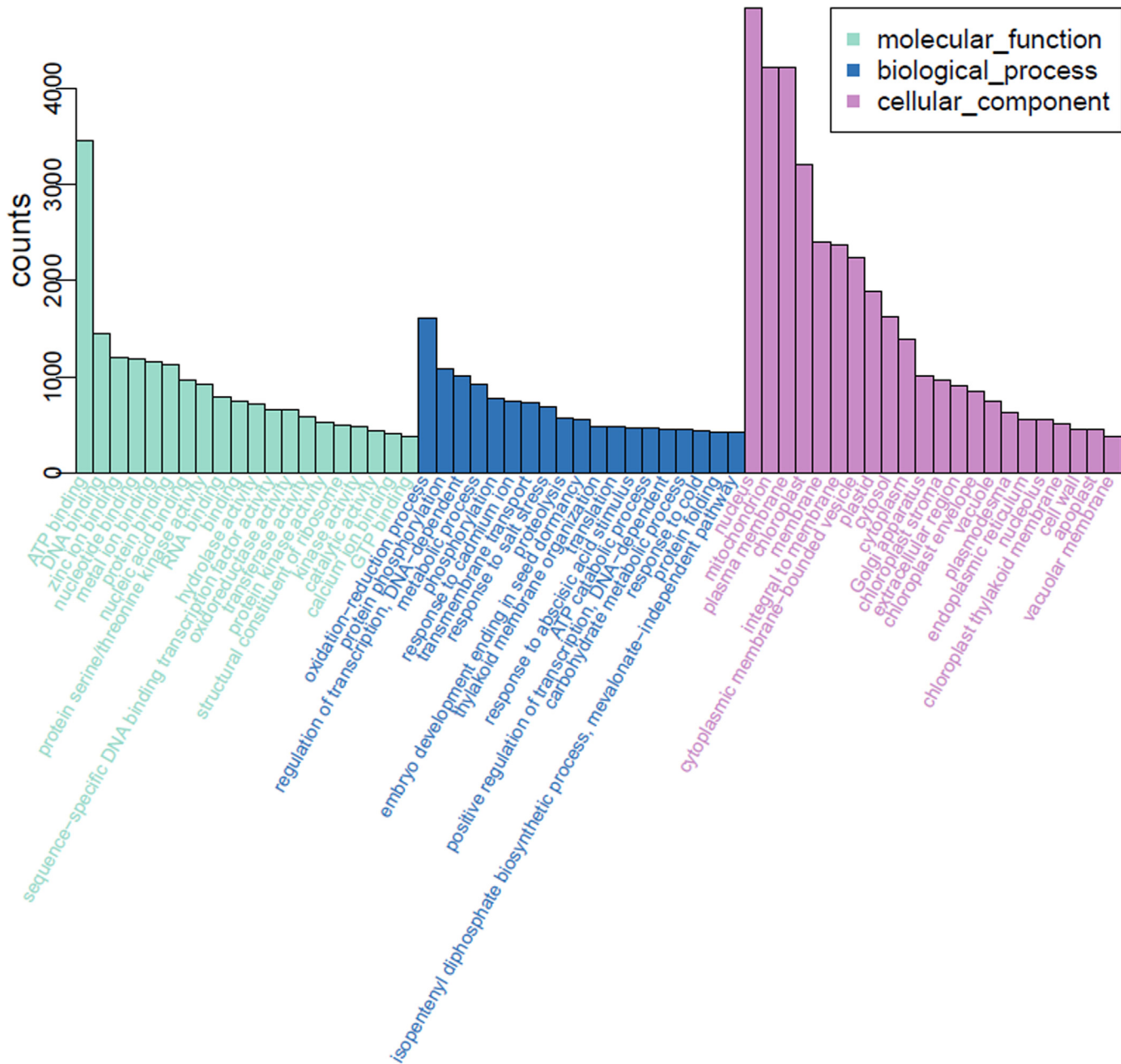


Fig 4. GO classification of unigenes of assembled unigenes.

doi:10.1371/journal.pone.0142434.g004

the categories. Only a few genes matched the terms ‘Nuclear structure’ and ‘Extracellular structures’.

When aligned to the GO system, 32,632 (29.16%) unigenes were categorized into 61 functional groups (Fig 4). Of which, 20 GO terms were related to ‘molecular functions’, 19 were related to ‘biological process’, and 22 were related to ‘cellular component’. The three largest percentages of genes within these three function categories were ‘ATP binding’ (9,585, 48.3%), ‘oxidation-reduction process’ (14,119, 71.2%) and ‘nucleus’ (10,809, 54.5%). In addition, we noticed that a high-percentage of genes came from the ‘mitochondrion’ and ‘plasma

membrane' groups in the 'cellular component' category and a relative high percentage of genes came from the 'binding' and 'protein phosphorylation' groups in the 'molecular functions' and 'biological process' categories, respectively. However, only few genes came from the terms 'extracellular region part' and no gene came from the terms 'metallochaperone', 'nutrient reservoir', or 'translation regulator'.

We also mapped the assembled unigenes to the reference canonical pathways in KEGG, including metabolism, genetic information processing, environmental information processing, cellular processes, and organism systems (<http://www.genome.jp/kegg/pathway.html>). In total, 21,461 unigenes (37,583,629 reads) were mapped into 329 KEGG pathways representing 19.18% of the assembled unigenes. Among these, 3,776 unigenes were related to metabolism, which represented the most mapped sequences. Additionally, 2,688 unigenes corresponded to genetic information processing, 1,238 unigenes belonged to organism systems, 1,094 unigenes were classified as cellular processes, and 823 unigenes mapped to environmental information processing (Table 3). The presence of genes for all of these essential biological processes suggests that these sequences account for most of the comprehensive transcriptome. These pathways provide valuable resources for investigating specific biological processes during *C. ensifolium* floral development. Among these, the circadian clock is an important part of the photoperiod pathway controlling flower initiation. A circadian rhythm was found in the KEGG pathway and its detailed metabolic pathway is shown in S1 Fig. The 66 unigenes involved (S2 Table) will facilitate further studies on the effects of the photoperiod pathway on flowering control in *C. ensifolium*.

Identification of transcription factors

Transcriptional regulation of floral development has been well documented in *Arabidopsis* and many other non-model plants, but little has been known in *C. ensifolium*. In this research, we identified the transcription factors from the transcriptome of *C. ensifolium* by aligning rice transcription factor sequences (downloaded from <http://plantfdb.cbi.pku.edu.cn/>) against the EST dataset. In a total, 4,632 unigenes were identified, belonging to 56 putative transcription factor families using Blastx with a cut off E-value of below 10^{-5} (Fig 5).

Among the all putative transcription factor encoding genes, the Nin-like family, which is identified as core symbiotic genes required for establishing symbiosis between legumes and nitrogen fixing bacteria, collectively called the Rhizobium, are the most abundantly expressed transcription factor genes, accounting for 17.5% of the overall transcription factors. In contrast, expression studies in *Arabidopsis* and rice genomes found barely nine and three Nin-like proteins, respectively, suggesting that Nin homologs constitute small gene families in these model plants [33]. The huge number of Nin-like homologs found in *Cymbidium*, which is greatly different from in *Arabidopsis* and rice, might be related to the importance of their functions in symbiosis and nitrogen metabolism in orchidaceae [33, 34]. In addition, 387 and 361 genes were identified as FAR1 and bZIP respectively. C3H (237), MYB (220), C2H2 (218), bHLH (218), ERF (205), G2-like (167), WRKY (177), NAC (148), and MYB-related (134) families were also found. These nine families formed ~30% of the expressed transcription factors, and the number in each family was more than in *Arabidopsis* and the rice genomes. Whole-genome duplication occurred in early monocots resulting in polyploidy, gene duplication, gene loss and rearrangements. This may also be the reason why greater numbers of transcription factors are found in the orchid genome, as suggested by the presence of the four AP3-like paralogs that form the basis for the complicated floral morphologies of *Phalaenopsis* [18]. Paralogous genes are present in at least four out of the five subfamilies of the orchidaceae, suggesting that some genes may have been undergone gene duplication. However, no more than five members of the

Table 3. Summary of the KEGG pathway and their corresponding gene number.

KEGG Pathways	Sub-pathways of KEGG Pathway	Number of unigenes	Number of reads
Cellular Processes		1,094	1,693,012
	Transport and catabolism	595	1,187,266
	Cell motility	113	123,121
	Cell growth and death	414	417,722
	Cell communication	161	268,922
Genetic Information Processing		2,688	4,615,076
	Transcription	436	618,315
	Translation	1,103	2,358,454
	Folding, sorting and degradation	845	1,804,603
	Replication and repair	484	203,113
Environmental Information Processing		823	986,367
	Membrane transport	149	96,975
	Signal transduction	663	874,444
	Signaling molecules and interaction	13	14,951
Organismal Systems		1,238	1,772,131
	Immune system	334	598,690
	Endocrine system	269	512,321
	Circulatory system	99	103,034
	Digestive system	245	154,968
	Excretory system	140	177,267
	Nervous system	283	429,307
	Sensory system	46	33,599
	Development	46	50,614
	Environmental adaptation	309	461,208
Metabolism		3,776	7,328,381
	Carbohydrate metabolism	907	2,176,302
	Energy metabolism	840	2,277,509
	Lipid metabolism	574	1,700,740
	Nucleotide metabolism	459	354,067
	Amino acid metabolism	644	1,498,932
	Metabolism of other amino acids	250	682,150
	Glycan biosynthesis and metabolism	310	152,222
	Metabolism of cofactors and vitamins	512	473,178
	Metabolism of terpenoids and polyketides	272	419,236
	Biosynthesis of other secondary metabolites	22	566,666
	Xenobiotics biodegradation and metabolism	104	398,766
	Total	21,461	37,583,629

doi:10.1371/journal.pone.0142434.t003

transcription factor families Whirly, LFY, RAV, VOZ, HRT-like, and HB-PHD were detected. This result is consistent with the previous report that few members of these six families were found in *Phalaenopsis* [35]. This might result from the rare expression of these genes. Large-scale deeper transcriptome sampling and sequencing efforts will help identify genes related to transcription factor families in orchids.

Table 4. Representatives of putative flowering-time genes in *C.ensifolium*.

	Gene ID	Homologous gene	Gene Nr-ID	Identity Nr_top (%)	
Photoperiod pathway	Comp58987	PHYA	XP_002278610.1	68.4	
	Comp54211	PHYB	XP_003558068.1	71.7	
	Comp5851		XP_004295077.1	70.8	
	Comp35281	PIF3	XP_002276198.2	38.6	
	Comp53073		XP_002276198.2	37.4	
	Comp52955	EIF3	ABL11477.1	38	
	Comp50981	LHY	NP_001131529.2	34	
	Comp52658		BAC99516.1	76.5	
	Comp57367		BAC99516.1	37.8	
	Comp57910	CCA1	XP_004306608.1	80.6	
	Comp35965	COP1	BAA94422.1	60.2	
	Comp54784	TOC1	ADO51646.1	62.5	
	Comp48883	GI	AEB35426.1	72.4	
	Comp56138		ADP92454.1	80.3	
	Comp27950	CHS	AGJ50587.1	83	
	Comp48799		NP_001064831	59.3	
	Comp37213	CO	NP_001047975.1	87.2	
	Comp51149		NP_001057441.1	86.7	
	Floral integrator pathway	Comp26556	FT	AFS17371.1	54.5
		Comp39786		ADP89905.1	78.3
Comp41343			ADW76861.1	98.5	
Comp44994			AFU08240.1	73.8	
Comp48314			ADP89470.1	63.8	
Comp301941			BAD01612.1	86.3	
Comp36081		AP2	AGI62045.1	72.8	
CL2711Contig1			ABU68665.1	48.9	
CL2228Contig1			AGI62044.1	72	
Comp43524			AGI62041.1	82.1	
Comp51324			AGI62047.1	80.2	
Comp26637			AGK07583.1	68.6	
Comp47321			AGK07583.1	84.8	
Comp55126			NP_001236377.1	84	
Floral meristem identity		Comp45071	SOC1	AFQ31623.1	66.4
	Comp47662		AFQ31623.1	74.8	
	Comp58027	AP1	AFQ31623.1	99.6	
	Comp51302	LFY	AGE45851.1	97.8	
Vernalization pathway	Comp45077	VRN1	AEV22381.1	58.9	
	Comp59230	FLC	ACZ26524.1	26.1	
	CL346Contig1	FLD	AAX51267	58.7	
	Comp58957		NP_001148070.1	50	

doi:10.1371/journal.pone.0142434.t004

CONSTANS-LIKE (COL) genes has been identified and plays a central role in the photoperiod response pathway by mediating between the circadian clock and the floral integrators [36–38]. To identify the *COL* genes in *C. ensifolium*, the *Arabidopsis* CO and COL amino acid sequences were used to screen for homologs in our dataset using tBLASTN. In total, 61 putative COL homologs were identified and are listed in S5 Table. After translation, sets of these proteins grouped together and showed homology with the COL proteins from *Arabidopsis*. As shown in

[S2 Fig](#), 16 COL homologs of *C. ensifolium* were clustered well with 17 *Arabidopsis* COL proteins and can be divided into three subgroups as previously defined in *Arabidopsis* [36, 37].

The COL genes belonging to different subgroups are expected to perform distinct biological roles, but only a few of them can be regulated by the circadian clock and light, and plays a central regulatory role in the photoperiod pathway [36]. Here, we present a further determination of six COL genes as shown in Figs 6 & 7. When compared to other COL genes from *Arabidopsis* and other orchid plants including *Phalaenopsis hybrid*, *Cymbidium sinense* and *Erycina pusilla*, these six selected genes can be divided into three subgroups: unigene comp51149 in group I was clustered with CsCOL, and PhalCOL, as well as *Arabidopsis* AtCO/AtCOL1/AtCOL2. Group II included unigenes comp56744, comp44679, comp59285, and comp51441. They are closely related to *Erycina pusilla* EpCOL8/EpCOL10 and *Arabidopsis* AtCOL10/AtCOL13/AtCOL14/AtCOL15. Comp47321 was assigned to group III and clustered well with AtCO6/AtCOL7/AtCOL8. The temporal expression of them also indicated three patterns of light regulation (Fig 7). Genes that belong to the Type I expression pattern, including unigenes comp56744, comp44679, and comp51441, were abundant in the light. Type II genes, which included comp47321 and comp51149 were expressed constitutively in the light or dark. Contrarily, the expression of Type III gene comp59285 was increased in the dark, but repressed in the light. These expression patterns may correlate with the regulatory regions or biological functions of them, and further investigations using these genomic sequences will provide more insight into the photoperiodic control of flowering in the orchid plants.

DGE sequencing and analysis

To identify the transcriptional responses in individual flower organs, we constructed independent sepal, petal, labellum, and gynostemium DGE libraries, and sequenced them using the Illumina HiSeq 2500 sequencing platform. For each library, the total number of counts for each read was determined, and all the reads were mapped back to the reference transcriptome. The DGE sequencing quality evaluation and alignment statistics are shown in [Table 5](#), which revealed that the clean reads in each sample ranged from 26 to 30 million. Among the clean tags, 28,314,897, 28,885,091, 26,626,616, and 20,985,335 sequences were corresponding to 91.15%, 92.47%, 90.35%, and 79.82%, respectively, of the sepal, petal, labellum, and gynostemium transcriptome libraries, respectively. Thus, good coverage of the transcript profiles was provided.

Gene expression profiling among different floral organs

To characterize the gene expression profile of each floral organ, we determined the differentially expressed unigenes between every two samples using the RPKM method with an algorithm developed by Audic et al [39]. From the six pairwise comparisons among sepals, petals, labellum and gynostemium, we found a total of 3,994 unigenes had significant changes in expression, ranged from 77 to 1,298 in different pairwise comparisons. The largest differences occurred between the petals and the gynostemium, and there were 982 and 316 transcripts up- and down-regulated in the gynostemium, respectively (Fig 8). The smallest difference occurred between sepals and petals, in which only 77 differentially expressed tags were identified. The differences among the other comparisons ranged from 399 to 1,255 transcripts.

To determine the putative biological functions of these differentially expressed unigenes, we mapped them to every term in the GO database (<http://www.geneontology.org/>) and 3,292 unigenes were annotated. This analysis revealed that 41 GO terms showed significantly different expression levels among different floral organs. These GO terms included 'structural constituents of ribosome' and 'ATP binding' in the 'molecular function' category, 'ribosome

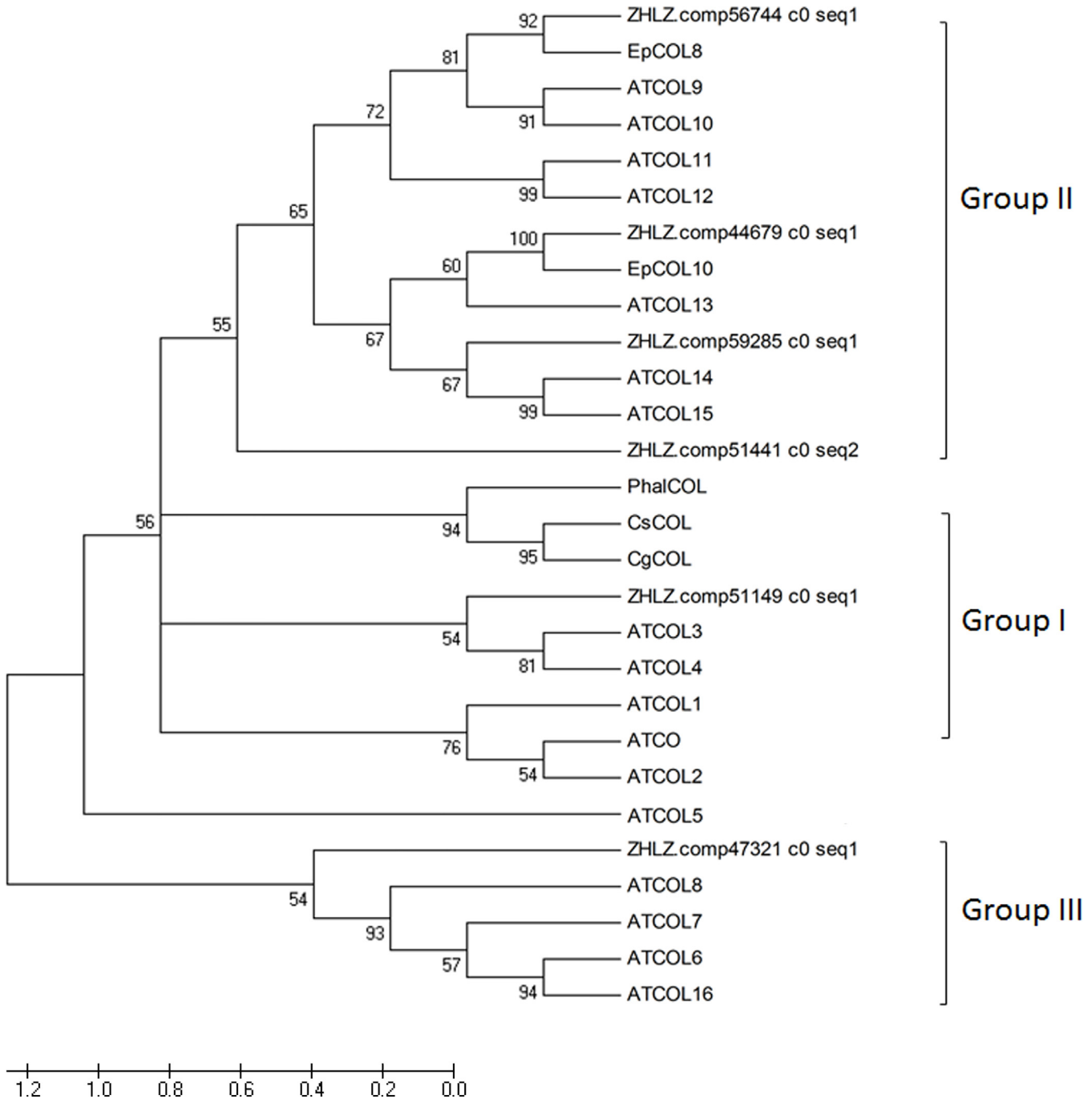


Fig 6. Phylogenetic analysis of the CONSTANS-like proteins from different plant species. Amino acid sequences were aligned by the ClustalW 2.0, and phylogenetic relationships were reconstructed using a maximum-likelihood (ML) method in PHYML software with JTT amino acid substitution model. Bootstrap values for 1,000 replicates were used to assess the robustness of the trees. Previously published plant MADS-box protein sequences were retrieved from GenBank database. (AtCO: NP_197088, AtCOL1: NP_197089, AtCOL2: NP_186887, AtCOL3: Q9SK53, AtCOL4: Q940T9.2, AtCOL5: Q9FHH8, AtCOL6: Q8LG76, AtCOL7: Q9C9A9, AtCOL8: Q9M9B3, AtCOL9: NP_001118599, AtCOL10: Q9LUA9, AtCOL11: O23379, AtCOL12: Q9LJ44, AtCOL13: O82256, AtCOL14: O22800, AtCOL15: Q9C7E8, AtCOL16: Q8RWD0, PhalCOL: FJ469986, CsCO: GU168786. EpCOL8: KC836891.1, EpCOL10: KC836893.1)

doi:10.1371/journal.pone.0142434.g006

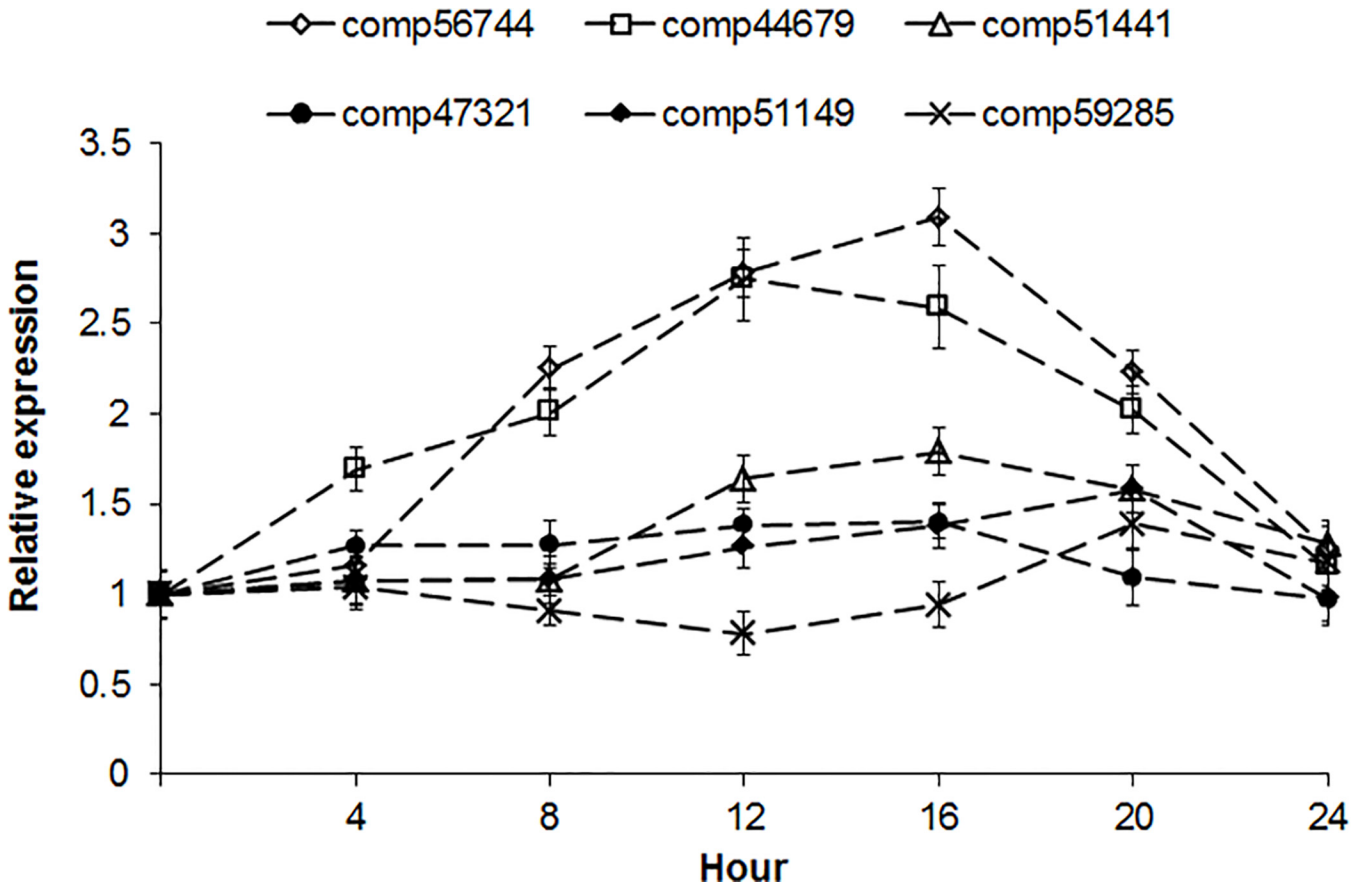


Fig 7. Accumulation of *Cymbidium ensifolium* COL genes in LD conditions. Leaves from one-year-old *C. ensifolium* plants were collected in 4-h intervals for 24 h after the start of light exposure in LD conditions (16-h light /8-h dark).

doi:10.1371/journal.pone.0142434.g007

biogenesis', 'translation', and the 'oxidation-reduction process' in the 'biological process' category, and 'membrane-bounded vesicle integral to membrane', and 'ribosome' in the 'cellular component' category (S6 Table). Most of the enrichment occurred in the gynostemium as opposed to the other three libraries, and these genes were mainly correlated to 'binding', 'membrane-bounded organelles', and 'organelle parts', which correlated well with the genes identified in the stamen or pollen transcriptomes of other plants.

Table 5. Summary of DGE sequencing quality and alignment.

	Sepal	Petal	Labellum	Gynostemium
Total reads	31,063,650(100%)	31,237,880(100%)	29,464,502(100%)	26,291,414(100%)
Total mapped	28,314,897(91.15%)	28,885,091(92.47%)	26,626,616(90.35%)	20,985,335(79.82%)
Multiple mapped	3,882,139(12.50%)	3,057,936(9.79%)	3,595,240(12.20%)	5,328,168(20.27%)
Uniquely mapped	24,432,758(78.65%)	25,827,155(82.68%)	23,031,376(78.15%)	15,657,167(59.55%)
Read1 mapped	14,166,443(45.60%)	14,454,262(46.27%)	13,321,458(45.20%)	10,499,498(39.94%)
Read2 mapped	14,148,454(45.55%)	14,430,829(46.20%)	13,305,158(45.15%)	10,485,837(39.88%)
Reads map to '+'	14,170,934(45.62%)	14,437,837(46.22%)	13,337,405(45.26%)	10,524,193(40.03%)
Reads map to '-'	14,143,963(45.53%)	14,447,254(46.25%)	13,289,211(45.09%)	10,461,142(39.79%)
Duplication	12,403,029(39.93%)	15,231,879(48.76%)	11,439,220(38.82%)	2,605,567(9.91%)

doi:10.1371/journal.pone.0142434.t005

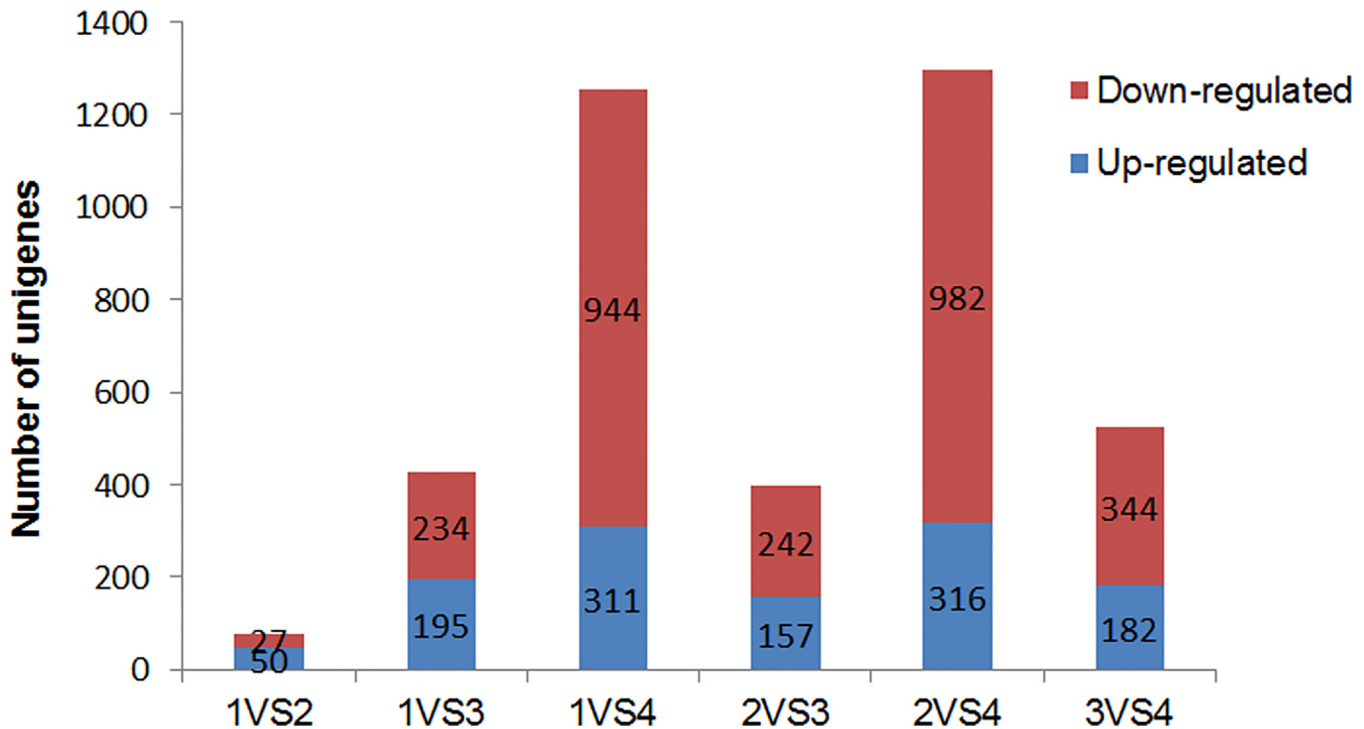


Fig 8. Transcripts differentially expressed between different floral organs. Up- and down-regulated transcripts were quantified. The results of six comparisons between each two samples are shown. 1, sepal; 2, petal; 3, Labellum; 4, gynostemium.

doi:10.1371/journal.pone.0142434.g008

For the pathway enrichment analysis, we mapped those differentially expressed unigenes to terms in the KEGG database and searched for KEGG terms that were significantly enriched compared with the transcriptome background. In total, 1,734 unigenes were assigned to 176 KEGG pathways. The pathways that were most represented by the unigenes were ‘ribosome’ (72), followed by ‘oxidative phosphorylation’ (29), ‘Phenylpropanoid biosynthesis’ (9), ‘Glucconeogenesis’ (8), ‘protein processing in endoplasmic reticulum’(8), and ‘plant-pathogen interaction’(7) (S7 Table). Consistent with the result from the GO annotation, most differences occurred between the gynostemium and the petal, and the functional transcripts involved in the ‘ribosome and oxidative phosphorylation’ pathway accounted for a large proportion. Thus, further studies on the precise roles of these genes will provide new insights into the flower patterning mechanism of *C. ensifolium*.

Differentially expressed genes specific for each floral organ

To further identify the genes specifically responding to individual floral organ development, we selected genes that showed at least a fourfold higher expression after filtering statistically significant expression differences using the p-value threshold of 0.05. Moreover, genes that were up-regulated in more than one of the organs were not included. In total, we identified 10 petal-, 5 sepal-, 279 labellum- and 764 gynostemium-specific genes. Most of the enriched transcripts were connected to cell structures, plant metabolic process and the oxidation-reduction process, including cDNAs encoding homologs of cell wall-related genes, such as cellulose synthase-like proteins and cell wall-associated hydrolase family members, membrane associated proteins, such as membrane protein, photosystem II thylakoid membrane protein, and MATE efflux membrane protein, and a number of genes involved in secondary metabolism

Table 6. Tissue-specific genes revealed from DGE expression of various flower organs.

4-fold up in sepals	
Sequence ID	Putative function
Metabolism	
Comp49718	NADP-dependent oxidoreductase P2 [<i>Zea mays</i>]
Comp60267	O-acyltransferase WSD1-like [<i>Glycine max</i>]
Comp59584	endo-beta-1,4-glucanase [<i>Pyrus communis</i>]
Comp45151	O-methyltransferase [<i>Ricinus communis</i>]
Comp28605	Putative phospholipase a2 precursor [<i>Arabidopsis thaliana</i>]
Cell structure	
Comp55252	myosin light chain 2 [<i>Lonomia obliqua</i>]
Comp53878	Alpha-expansin 8 precursor [<i>Oryza sativa Japonica Group</i>]
4-fold up in petals	
Sequence ID	Putative function
Metabolism	
Comp58735	transketolase, chloroplastic-like [<i>Oryza sativa Japonica Group</i>]
Comp58022	1-D-desoxyxylulose 5-phosphate synthase (DXS) [<i>Vitis vinifera</i>]
Comp56799	transketolase [<i>Camellia sinensis</i>]
Comp53861	transketolase isoform 1 [<i>Camellia sinensis</i>]
Comp52466	sphinganine C(4)-monooxygenase 2-like isoform 1 [<i>Solanum lycopersicum</i>]
4-fold up in the labellum	
Sequence ID	Putative function
Cell structure	
Comp64311	plasma membrane associated protein [<i>Populus trichocarpa</i>]
Comp57083	leucine-rich repeat transmembrane protein kinase [<i>Ricinus communis</i>]
Comp8922	putative apolipoprotein A-I precursor [<i>Meleagris gallopavo</i>]
Comp8510	40S ribosomal protein S10 [<i>Vitis vinifera</i>]
Comp58037	Cell wall-associated hydrolase [<i>Solanum lycopersicum</i>]
Metabolic process	
Comp56608	3-hydroxyacyl-CoA dehydrogenase [<i>Arabidopsis thaliana</i>]
Comp58442	transketolase [<i>Camellia sinensis</i>]
Comp48310	alcohol dehydrogenase [<i>Solanum lycopersicum</i>]
Signal transduction	
Comp56125	NBS-LRR disease resistance protein precursor [<i>Meleagris gallopavo</i>]
Comp33694	ADP-ribosylation factor, arf [<i>Riptortus pedestris</i>]
4-fold up in the gynostemium	
Sequence ID	Putative function
Cell Structure	
Comp58037	cell wall-associated hydrolase [<i>Veillonella sp. 6_1_27</i>]
Comp57393	cell wall-associated hydrolase [<i>Neisseria meningitidis</i>]
Comp55155	cell wall-associated hydrolase [<i>Neisseria meningitidis</i>]
Comp54344	cellulose synthase-like protein D4-like [<i>Arabidopsis thaliana</i>]
Comp51052	expansin-A4-like [<i>Ricinus communis</i>]
Comp44366	YABBY domain transcription factor family protein [<i>Populus tomentosa</i>]
Comp36904	Protein HIS-71 [<i>Caenorhabditis elegans</i>]
Comp49964	UPF0497 membrane protein [<i>Arabidopsis thaliana</i>]
Comp49767	major protein body membrane protein MP27 like [<i>Capsella rubella</i>]
Comp44488	ABA-induced plasma membrane protein protein [<i>Arabidopsis thaliana</i>]

(Continued)

Table 6. (Continued)

Comp106967	putative MATE efflux membrane protein [<i>Solanum lycopersicum</i>]
Comp17934	60S ribosomal protein L30-like isoform 1 [<i>Vitis vinifera</i>]
Metabolic process	
Comp52552	polygalacturonase precursor-like [<i>Ricinus communis</i>]
Comp7369	serine acetyltransferase [<i>Medicago truncatula</i>]
Comp57200	sesquiterpene synthase [<i>Oryza sativa Indica</i>]
Comp36904	Histone H3.3 [<i>Oryza sativa Indica</i>]
Comp43882	tryptophan synthase alpha subunit, putative [<i>Ricinus communis</i>]
Comp54620	cytochrome P450 A [<i>Capsicum annuum</i>]
Comp35538	metallothionein-like protein [<i>Sorghum bicolor</i>]
Comp59258	heavy metal ATPase [<i>Populus trichocarpa</i>]
Comp102609	ATP carrier protein 1-like [<i>Glycine max</i>]
Oxidation–reduction process	
Comp53444	stearoyl-ACP desaturase-like protein [<i>Glycine max</i>]
Comp8922	apolipoprotein A-I [<i>Lagopus lagopus</i>]
Comp48818	myo-inositol oxygenase [<i>Solanum lycopersicum</i>]
Comp54494	L-ascorbate oxidase-like protein [<i>Aegilops tauschii</i>]
Comp47117	seq polyphenol oxidase [<i>Ananas comosus</i>]
Comp49915	ACC oxidase [<i>Cymbidium hybrid cultivar</i>]
Comp59496	alcohol dehydrogenase, putative [<i>Ricinus communis</i>]
Comp27255	AMP dependent CoA ligase, putative [<i>Ricinus communis</i>]
Comp27226	cytokinin hydroxylase-like [<i>Cicer arietinum</i>]
Transcription	
Comp58618	MADS-box transcription factor [<i>Oryza sativa Indica</i>]
Comp58442	MADS-box transcription factor [<i>Oryza sativa Indica</i>]
Comp52003	MADS-box transcription factor [<i>Oryza sativa Indica</i>]
Comp50822	MADS-box transcription factor [<i>Oryza sativa Indica</i>]
Comp60086	R2R3-MYB transcription factor [<i>Oryza sativa Indica</i>]
Comp52709	probable WRKY transcription factor [<i>Oryza sativa Indica</i>]
Comp1478	heat shock 70 kDa protein cognate [<i>Glycine max</i>]
Signal transduction	
Comp44110	phytochrome C [<i>Oryza sativa Indica</i>]
Comp54413	ras-related protein RABA1f-like [<i>Linum usitatissimum</i>]
Comp56125	NBS-LRR disease resistance protein precursor [<i>Glycine max</i>]
Comp45941	CBL-interacting protein kinase [<i>Oryza sativa Indica</i>]
Comp43483	ralf-like 22 protein [<i>Vitis vinifera</i>]

doi:10.1371/journal.pone.0142434.t006

and oxidation-reduction pathways, such as the acetyltransferase, transketolase, and NADP-dependent oxidoreductase (Table 6). The specifically up-regulated expression of these genes strongly suggests putative functions for them in flower organ development and/or differentiation.

To validate the discovery methods, we randomly chose 10 unigenes showing differential expression among these four floral organs and designed primers for quantitative RT-PCR. An EST-encoding ubiquitin was used as an internal control, to which gene expression was normalized. The same mRNA flower samples used in the microarray analysis served as templates. The results exhibited differential expression among the four libraries and were identical to those obtained by DGE expression profiling (Fig 9).

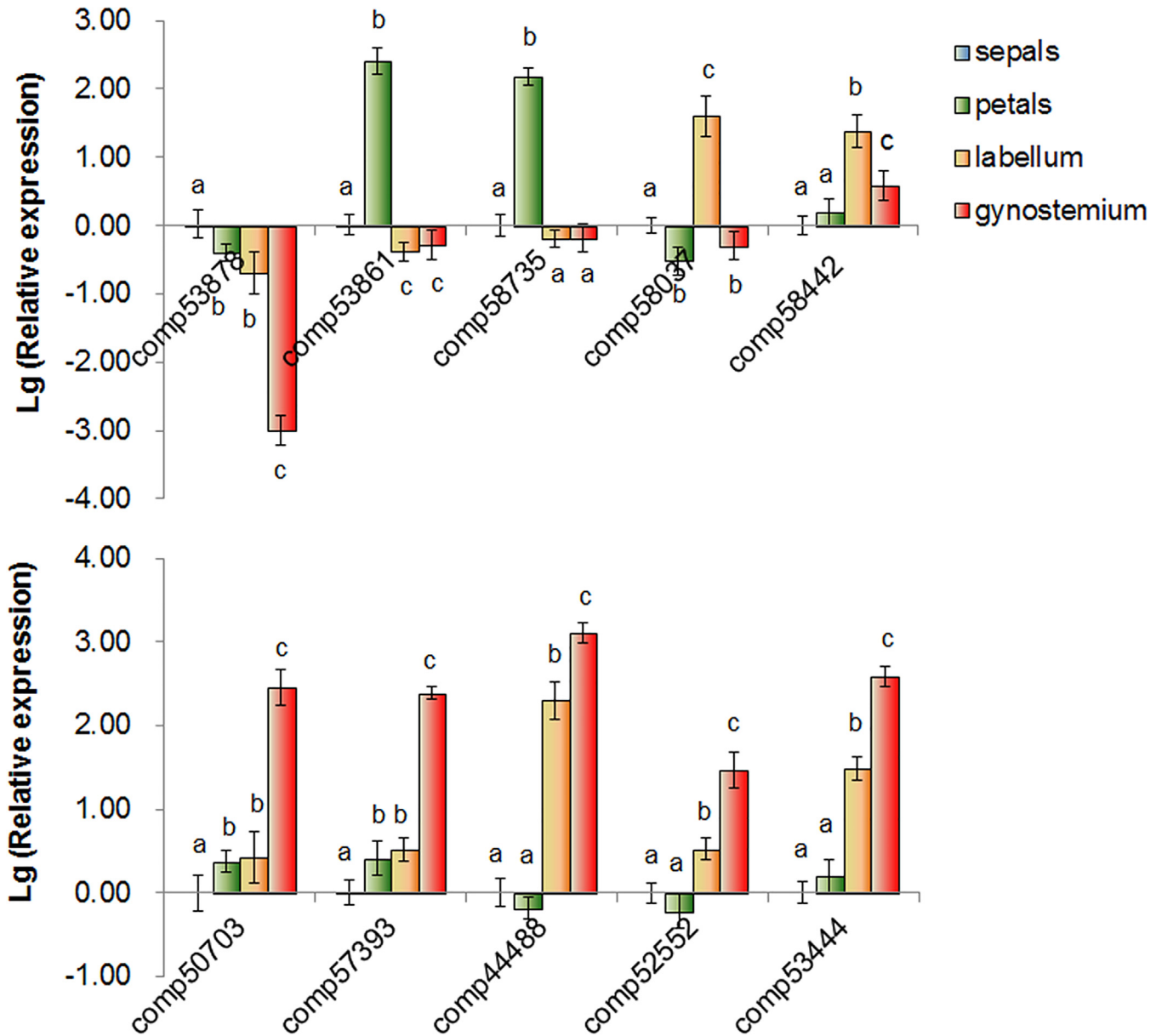


Fig 9. The quantitative RT-PCR analysis of gene expression in sepals, petals, labellums and gynostemiums. The y-axis indicates fold change in expression among the samples. The Lg (Relative Quantitation) of the genes in the sepals was calibrated as zero. Error bars indicate the standard deviation of the mean (SD) (n = 3). Three replicates were analyzed, with similar results. a, b, and c, d, one way ANOVA with Bonferroni multiple comparison test significant at P<0.05 between two of the individual floral organs sepal, petal, Labellum, gynostemium.

doi:10.1371/journal.pone.0142434.g009

Differentially expression of transcription factors

The transcription factors play a critical role in the control of plant reproductive development. In our study, a total of 4,632 putative transcription factors were detected in the floral transcriptome of *C. ensifolium*. DGE expression analysis revealed 126 of them were differently expressed in different floral organs, suggesting a potential role of them in the regulation of different floral organ development. As shown in Fig 10, there are 2 to 88 transcription factors had significant changes in expression in the pairwise comparisons among sepals, petals, labellum and

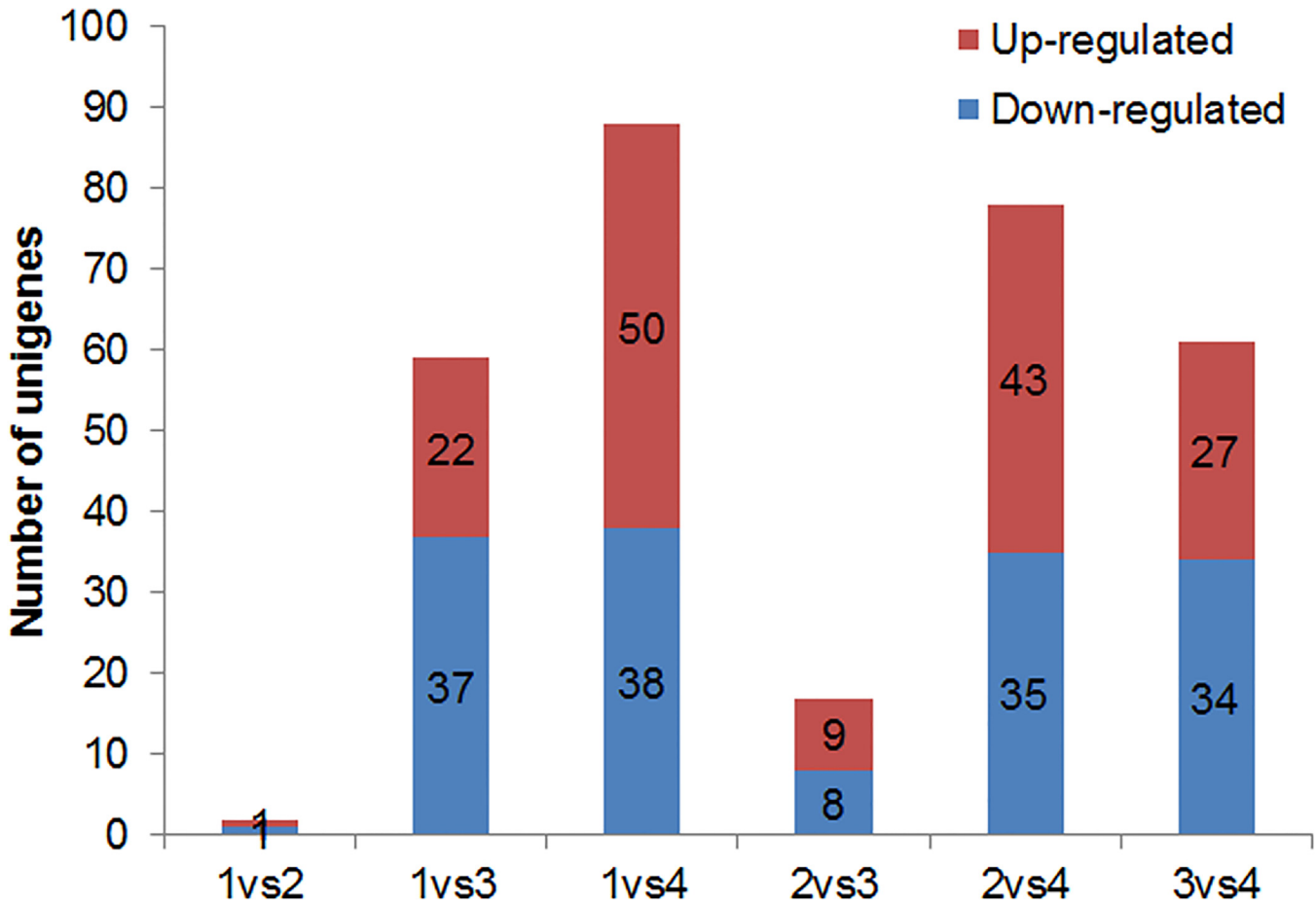


Fig 10. Transcription factors differentially expressed among different floral organs. Up- and down-regulated transcripts were quantified. The results of six comparisons between each two samples are shown. 1, sepal; 2, petal; 3, Labellum; 4, gynostemium.

doi:10.1371/journal.pone.0142434.g010

gynostemium. Typically, the gynostemium contains the highest number of differentially expressed genes, representing the most specialized reproductive organ of the orchid flower.

We next concentrated on the transcription factors responding to the gynostemium development and found 30 and 28 transcription factors specifically up- and down-regulated in the gynostemium, respectively, as opposed to the other three libraries. Among these genes, the expression of 5 NF-YC homologs, 4 MIKC-type MADS-box genes, 2 YABBY and 2 MYB transcription factors were increased by more than 4-fold; whereas 10 unigenes with bZIP-like sequences, 4 unigenes with FAR1-like sequences, 2 unigenes with LBD sequences, and 2 unigenes with WRKY-like sequences showed significantly lower abundance in the gynostemium (Fig 11 and S8 Table), suggesting a crucial role of them for floral development. Our work represents the first identification of TFs responding to individual floral organ development, and should be useful for understanding the floral patterning regulation of *C. ensifolium*.

Differentially expression of floral homeotic genes

Floral homeotic genes dominantly control the organ identity in developing flowers [18–24]. Most of these genes belong to the MIKC-type MADS-box gene family and can be categorized into several distinct gene groups: APETALA1/AGL9, APETALA3/PISTILLATA, and the AGAMOUS groups [18–24]. From the digital gene expression profiling, we found 16

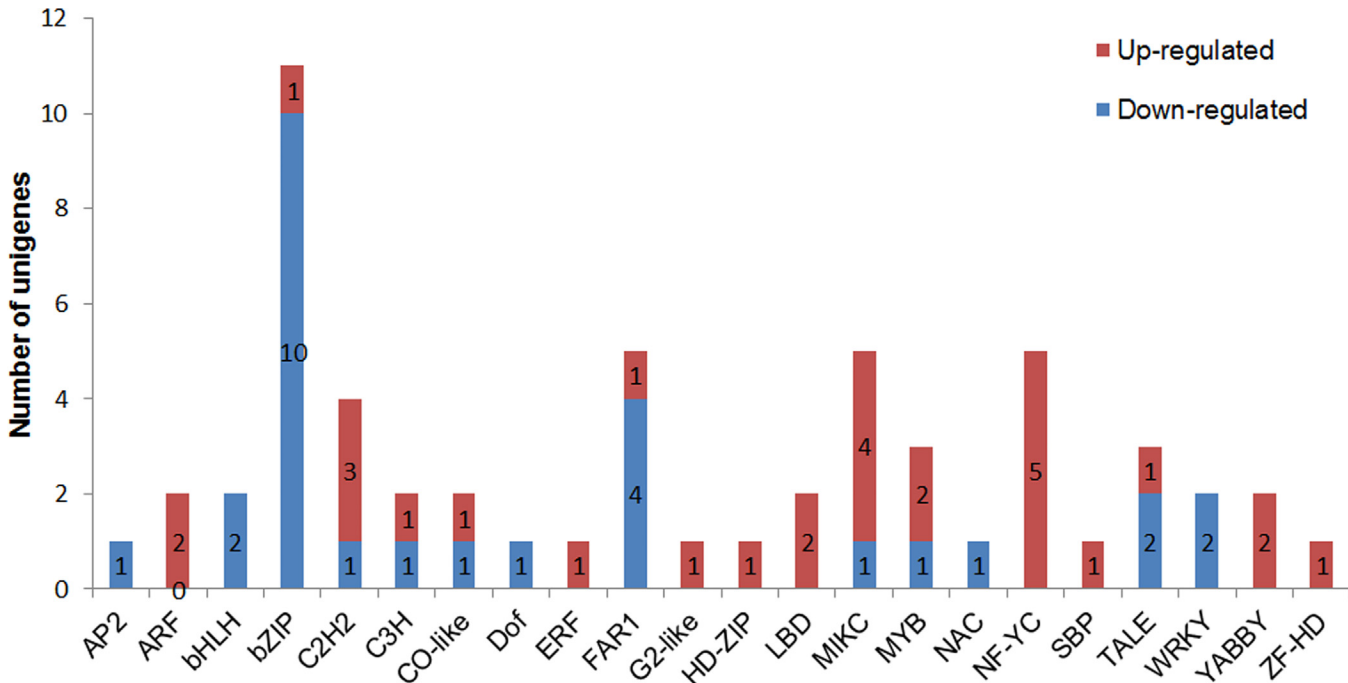


Fig 11. Transcription factors correlated with the development of the gynostemium. Up-(red) and down-regulated (green) in the gynostemium as opposed to other DGEs were quantified and annotated with the plant transcription factor database.

doi:10.1371/journal.pone.0142434.g011

MADS-box genes differentially expressed among the sepals, petals, labellum and gynostemium (S9 Table). Phylogenetic analysis demonstrated 14 of them clustered well with the floral homeotic genes isolated from other orchid species. Among them, five genes fall well into the APE-TALA3/PISTILLATA group, four genes belong to the AGAMOUS group and other five were classified to the APETALA1/AGL9 gene group (Fig 12).

The topology of the phylogenetic tree indicated that four of the MADS-box paralogs, comp26961, comp35346, comp53790 and comp55590 constituted a well-supported subclade of B-class DEF-like gene of *C. ensifolium*, as were evident in *Phalaenopsis* [18]. To further characterize these four DEF-like paralogs in *C. ensifolium*, complete sequences of them were identified by 5'RACE, and named as CeDEF-like1 (referring to comp26961), CeDEF-like2 (comp35346), CeDEF-like3 (comp53790), and CeDEF-like4 (comp55590). Conceptual translation of the ORFs yielded proteins of 227 aa, 222 aa, 225 aa, and 218 aa respectively. Multiple sequence alignments with the DEF-like subclade proteins in *Phalaenopsis* demonstrated that all of the four CeDEF-like proteins have a typical MIKC-type domain (Fig 13), and shared 62–79% similarity.

In addition to the four DEF-like genes, two AGOUGOUS-like C-class genes and two APE-TALA-like E-class genes were also identified in our study. Among them, the full-length cDNAs of C-class gene comp50822 and comp52003 were identical to CeMADS1 and CeMADS2 respectively, consisting well with the previous reported that showed CeMADS1 and CeMADS2 function in the regulation of reproductive organ development [27]. The E-class gene comp57688 and comp58442 were indicated as the orthologs of PeSEP1 and PeSEP3 respectively (Fig 12 and S3 Fig). Amino acid sequences of their proteins shared 67–83% identity and 79–87% similarity with PeSEP1 and PeSEP2, respectively, and possessed the conserved MIK domain and a divergent C-terminal domain with the conserved SEP I and SEP II motifs (S3 Fig).

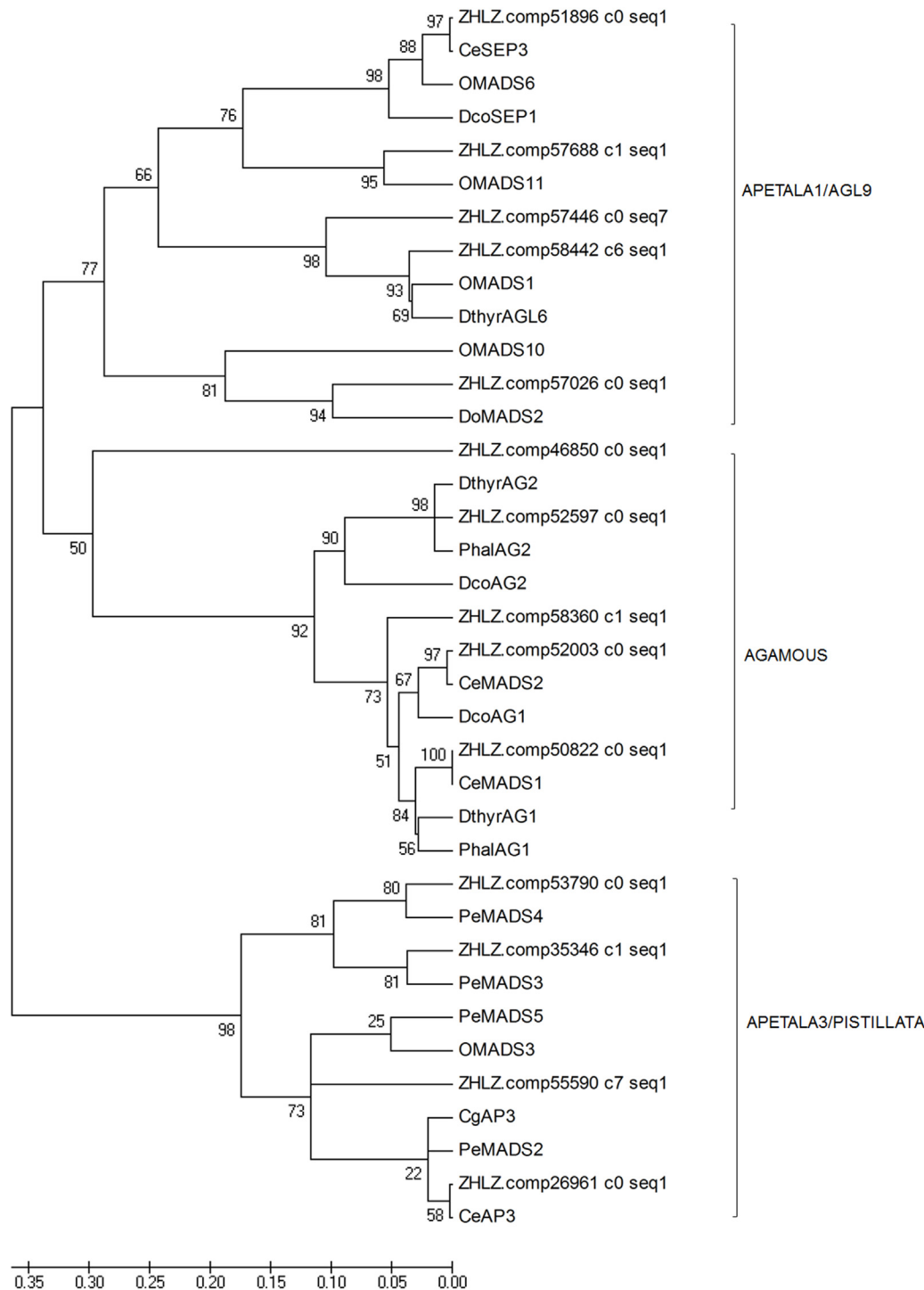


Fig 12. Phylogenetic analysis of the MADS-box genes differentially expressed among different floral organs. Amino acid sequences were aligned by the ClustalW 2.0, and phylogenetic relationships were reconstructed using a maximum-likelihood (ML) method in PHYML software with JTT amino acid substitution model. Previously published plant MADS-box protein sequences were retrieved from GenBank database (PeMADS2: AAR26628, CeAP3-like1: AFH66788, CeAP3-like2: AFH66787, CgAP3:ADI58460, CeMADS2: ADP00516, CeMADS1: ADP00515, CePI-like: AFH66786, OMADS11: ADJ6724, OMADS6: ADJ67238, OMADS1: ADJ67237, OMADS10: ADJ67240, OMADS2: AIJ29175, OMADS3: AAO45824, OMADS4: AIJ29176, DthyrAG1: AAY86364, DthyrAG2: AAY86365, PhalAG1: BAE80120, PhalAG2: BAE80121, PeMADS2: AY378149, PeMADS3: AY378150, PeMADS4: AY378151, PeMADS5: AY378148).

doi:10.1371/journal.pone.0142434.g012

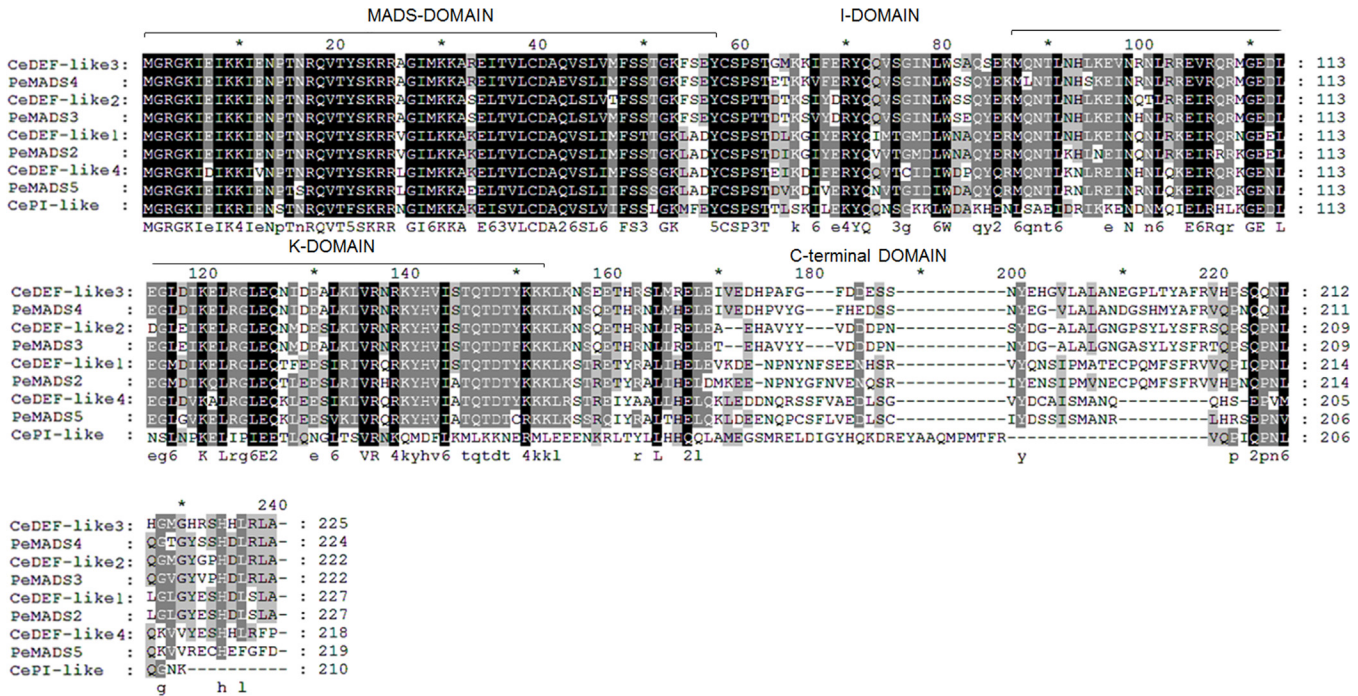


Fig 13. Sequence comparison of the DEF-like genes of *C. ensifolium* and *P. equestris*. Amino acid sequences were aligned by the ClustalW 2.0. The MADS-, I-, K-, and C-domains are indicated on top of the column.

doi:10.1371/journal.pone.0142434.g013

Confirmation of differential expression of floral homeotic genes

To validate the results obtained from the digital gene expression and determine the potential roles of the floral homeotic genes referred above, we confirmed their expressions among different floral organs by quantitative RT-PCR. As shown in Fig 14, four paralogs of *DEF-like* MADS-box genes displayed distinctive spatial expression patterns in various floral organs: *DEF-like1* was predominantly expressed in sepals and petals, but showed up weakly in the gynostemium and hardly detected in the labellum, whereas, *DEF-like2* was expressed strongly in the labellum and petals, but to a lesser extent in the gynostemium and not expressed in sepals. By contrast, expression of *DEF-like3* was restricted to the gynostemium and labellum. *DEF-like4* was predominantly expressed in petals, sepals, and hardly detected in the gynostemium. Comp57688 and comp58442, which belong to *API/AGL9*-like gene group, were preferentially expressed in sepals, petals and extremely lower in the labellum and gynostemium. On the contrary, the *AGAMOUS*-like genes comp50822 and comp52003 were found to be restricted to the gynostemium (Fig 14A & 14B), consisting well with the previous work that showed C-class gene *CeMADS1/2* were specific to reproductive organ development in *C. ensifolium* [27].

Furtherly, we validated the substantially differential expressions of these MADS-box genes in different floral mutant to confirm their functions in floral morphogenesis. Our result indicated that the B-class gene *DEF-like1* was expressed much higher in the labellum of the multi-tepala mutant, whereas expression of *DEF-like4* was extremely eliminated in petals and sepals compared with those in the wild type. These results were in agreement with the findings from other orchid species that showed the petaloid character of the outer tepals is due to heterotopic expression of class B genes such as *HrDEF*, *DcOAP3A*, *PeMADS2* and *PeMADS5*, and *OMADS3* in *Habenaria radiata*, *Dendrobium crumenatum*, *Phalaenopsis equestris* and the

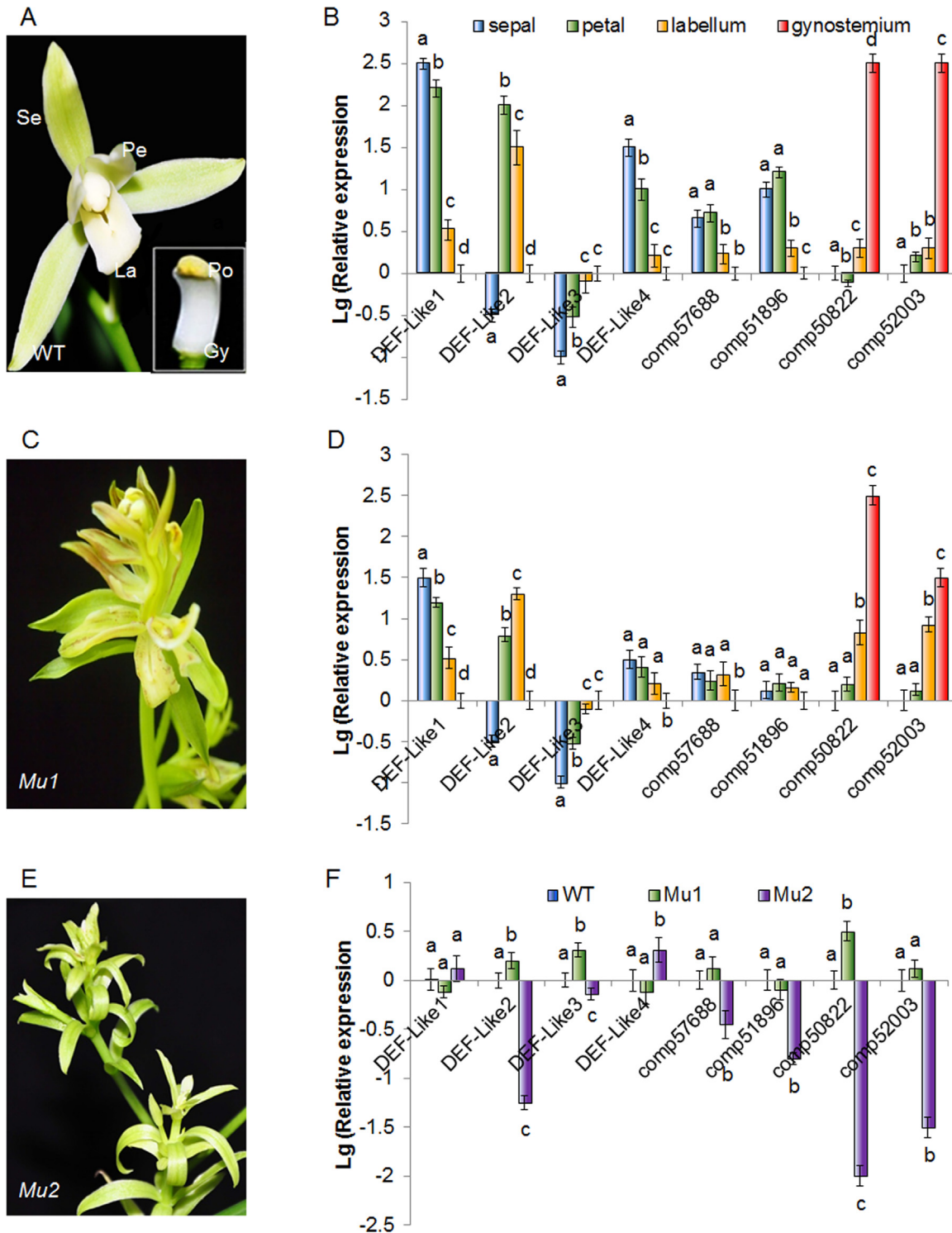


Fig 14. Flower morphology and expression levels of the MADS-box genes in *Cymbidium ensifolium* ‘tiansesu’ (WT, A&B), multi-lobed mutant (*Mu1*, C&D) and non-gynostemium mutant (*Mu2*, E&F) Se, sepal primordium; Pe, petal primordium; La, Labellum; Gy, gynostemium; Po, pollinia. The y-axis indicates fold change in expression among the samples. The ubiquitin gene served as the internal control. Error bars indicate the standard deviation of the mean (SD) (n = 3). Three replicates were analyzed, with similar results. a, b, c, and d, one way ANOVA with Bonferroni multiple comparison test significant at P<0.05 between two of the individual floral organs sepal, petal, Labellum, gynostemium.

doi:10.1371/journal.pone.0142434.g014

hybrid *Oncidium* "Gower Ramsey", respectively [18, 19, 24, and 40]. In addition, the preferential expression of *SEP-like* genes in the tepals were eliminated in the multi-tepal mutant compared to those in the wild type, and the expression of C-class genes comp50822 and comp52003 were also extended to the labellum of multi-tepal mutant (Fig 14C & 14D). By contrast, we found the overall expression level of the genes which preferentially expressed in reproductive organs, including comp53790, comp50822 and comp52003, were extremely lower in the flower of the non-gynostemium mutant that exhibited petal-to-sepal transformations (Fig 14E & 14F), likely due to the absence of normal petals and gynostemium in the center of the flower. The distinctive expression patterns of these MADS-box genes indicated their divergent functions in controlling floral organ development in different whorls, which have evolved into petaloid perianth (sepals and petals) and reproductive-related organs (labellum and gynostemium). Thus, the data generated in this study is sufficient to be used as a tool to investigate the role of some specific floral homeotic genes that show comparative expression levels among different floral organs.

Conclusion

High-throughput sequencing and *de novo* assembly were performed for *C. ensifolium*. As a result, more than 10 Gb data were obtained and assembled into 111,892 unigenes with an average length of 932.03 base pairs, which represent orthologs of known plant genes, as well as potential new genes. This systems bioinformatics survey, combined with molecular biology analyses, provide a more comprehensive gene information of *C. ensifolium* than was previously available, including more functional genes and transcription factors as well as detailed information on the relevant genes involved in flowering and floral development. In this work, we also investigated the gene expression profiling of the individual floral organs through four independent digital gene expression libraries prepared from the sepal, petal, labellum and gynostemium, which revealed 3,994 differently expressed gene tags. Further functional annotation of these tags provided a comprehensive understanding of the transcriptome complexity of floral structures and organ identities. To the best of our knowledge, this study provides a more comprehensive genomic resource and is the first informative EST dataset for individual floral organs of *Cymbidium*. This information broadens our understanding of the mechanisms of floral patterning and contributes to molecular and genetic research in the orchid plants.

Supporting Information

S1 Fig. Metabolic pathway of the circadian rhythm for unigenes by KEGG
(JPG)

S2 Fig. Phylogenetic analysis of the *COL* genes with its homologues
(TIF)

S3 Fig. Sequence alignment of *SEP-like* MADS box proteins with its homologues
(TIF)

S1 Table. The primers used for Realtime RT-PCR.
(TXT)

S2 Table. The unigenes involved in the circadian rhythm annotated by KEGG pathway.
(TXT)

S3 Table. Representative floral homeotic gene including the ABCDE classes.
(XLS)

S4 Table. A total of 129 MADS-box genes in *C. ensifolium*.
(TXT)

S5 Table. The homology genes of *CONSTANS*-like family in *C. ensifolium*.
(TXT)

S6 Table. Gene set enrichment analysis by GO annotation.
(XLS)

S7 Table. Gene set enrichment analysis by KEGG metabolic pathway classification.
(XLS)

S8 Table. Unigenes encoding putative transcription factors associated with the gynostemium development.
(XLS)

S9 Table. Sequences of the MADS-box genes differentially expressed among different floral organs.
(TXT)

Acknowledgments

We acknowledge the Guangzhou Ribobio Corporation for technical assistance in data processing and bioinformatics analysis, and the Yuandong Chinese Orchid Corporation for plant cultivation.

Author Contributions

Conceived and designed the experiments: GZ FY. Performed the experiments: FY. Analyzed the data: FY GZ. Contributed reagents/materials/analysis tools: FY. Wrote the paper: FY GZ.

References

1. Liu ZJ, Chen SC, Ru ZZ, Chen LJ (2006) Chinese *Cymbidium* plants. Beijing: Science Press.
2. Du Puy DJ, Cribb P (1988) The Genus *Cymbidium*. Helm Christopher, London, Timber Press.
3. Huang J and Dai S (1998) The numerical taxonomy of Chinese *Cymbidium*. Journal of Beijing Forestry University 20: 38–43.
4. Yukawa T, Stern WL (2002) Comparative vegetative anatomy and systematics of *Cymbidium* (Cymbidieae: Orchidaceae). Bot J Linn Soc 138:383–419.
5. Rajkumari JD, Longjam RS (2005) Orchid flower evolution. J Genet 84:81–84. PMID: [15876589](#)
6. Aceto S, Gaudio L (2011) The MADS and the Beauty: Genes Involved in the Development of Orchid Flowers. Curr Genomics 12:342–356. doi: [10.2174/138920211796429754](#) PMID: [22294877](#)
7. Wang HZ, Lu JJ, Hu X, Liu JJ (2011) Genetic variation and cultivar identification in *Cymbidium ensifolium*. Plant Syst Evol 293:101–110.
8. Schiestl FP, Ayasse M, Paulus HF, Lofstedt C, Hansson BS, Ibarra F, et al. (1999) Orchid pollination by sexual swindle. Nature 399:421–421.
9. Chang YY, Kao NH, Li JY, Hsu WH, Liang YL, Wu JW, et al. (2010) Characterization of the possible roles for B class MADS box genes in regulation of perianth formation in orchid. Plant Physiol 152:837–53. doi: [10.1104/pp.109.147116](#) PMID: [20018605](#)
10. Cozzolino S, Widmer A (2005) Orchid diversity: an evolutionary consequence of deception? Trends Ecol Evol 20:487–494. PMID: [16701425](#)
11. Gravendeel B, Smithson A, Slik FJ, Schuiteman A (2004) Epiphytism and pollinator specialization: drivers for orchid diversity? Philos Trans R Soc Lond B Biol Sci 359:1523–1535. PMID: [15519970](#)
12. Tsai WC, Hsiao YY, Pan ZJ, Chen HH (2010) Molecular mechanisms underlying orchid floral morphogenesis. Acta Hort (ISHS) 878:115–123.

13. Tsai WC, Pan ZJ, Hsiao YY, Chen LJ, Liu ZJ (2014) Evolution and function of MADS-box genes involved in orchid floral development. *J Syst Evol* 52:397–410.
14. Cai J, Liu X, Vanneste K, Proost S, Tsai WC, Liu KW, et al. (2015) Corrigendum: The genome sequence of the orchid *Phalaenopsis equestris*. *Nat Genet* 47:304. doi: [10.1038/ng0315-304a](https://doi.org/10.1038/ng0315-304a) PMID: [25711871](https://pubmed.ncbi.nlm.nih.gov/25711871/)
15. Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28:1086–1092. doi: [10.1093/bioinformatics/bts094](https://doi.org/10.1093/bioinformatics/bts094) PMID: [22368243](https://pubmed.ncbi.nlm.nih.gov/22368243/)
16. Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, et al. (2010) De novo assembly and analysis of RNA-seq data. *Nat Methods* 7:909–912. doi: [10.1038/nmeth.1517](https://doi.org/10.1038/nmeth.1517) PMID: [20935650](https://pubmed.ncbi.nlm.nih.gov/20935650/)
17. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. (2012) Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 2012:251364. doi: [10.1155/2012/251364](https://doi.org/10.1155/2012/251364) PMID: [22829749](https://pubmed.ncbi.nlm.nih.gov/22829749/)
18. Tsai WC, Kuoh CS, Chuang MH, Chen WH, Chen HH (2004) Four DEF-like MADS box genes displayed distinct floral morphogenetic roles in *Phalaenopsis* orchid. *Plant Cell Physiol* 45:831–844. PMID: [15295066](https://pubmed.ncbi.nlm.nih.gov/15295066/)
19. Hsu HF, Yang CH (2002) An orchid (*Oncidium* Gower Ramsey) AP3-like MADS gene regulates floral formation and initiation. *Plant Cell Physiol* 43:1198–1209. PMID: [12407200](https://pubmed.ncbi.nlm.nih.gov/12407200/)
20. De Paolo S, Salvemini M, Gaudio L, Aceto S (2014) De novo transcriptome assembly from inflorescence of *Orchis italica*: analysis of coding and non-coding transcripts. *PLoS One* 9(7):e102155. doi: [10.1371/journal.pone.0102155](https://doi.org/10.1371/journal.pone.0102155) PMID: [25025767](https://pubmed.ncbi.nlm.nih.gov/25025767/)
21. Teixeira da Silva JA, Aceto S, Liu W, Yu H, Kanno A (2014) Genetic control of flower development, color and senescence of *Dendrobium* orchids. *Scientia Horticulturae* 175:74–86.
22. Pan ZJ, Chen YY, Du JS, Chung MC, Tsai WC, Wang CN, et al. (2014) Flower development of *Phalaenopsis* orchid involves functionally divergent SEPALLATA-like genes. *New Phytol* 202:1024–1042. doi: [10.1111/nph.12723](https://doi.org/10.1111/nph.12723) PMID: [24571782](https://pubmed.ncbi.nlm.nih.gov/24571782/)
23. Tsai WC, Lee PF, Chen HI, Hsiao YY, Wei WJ, Pan ZJ, et al. (2005) PeMADS6, a GLOBOSA/PISTILLATA-like gene in *Phalaenopsis equestris* involved in petaloid formation, and correlated with flower longevity and ovary development. *Plant Cell Physiol* 46:1125–1139. PMID: [15890679](https://pubmed.ncbi.nlm.nih.gov/15890679/)
24. Kim SY, Yun PY, Fukuda T, Ochiai T, Yokoyama J, Kameya T, Kanno A (2007) Expression of a DEFICIENS-like gene correlates with the differentiation between sepal and petal in the orchid, *Habenaria radiata* (Orchidaceae). *Plant Sci* 172:319–326.
25. Zhang XM, Zhao L, Larson-Rabin Z, Li DZ, Guo ZH (2012) De novo sequencing and characterization of the floral transcriptome of *Dendrocalamus latiflorus* (Poaceae: Bambusoideae). *PLoS One* 7(8): e42082. doi: [10.1371/journal.pone.0042082](https://doi.org/10.1371/journal.pone.0042082) PMID: [22916120](https://pubmed.ncbi.nlm.nih.gov/22916120/)
26. Li X, Luo J, Yan T, Xiang L, Jin F, Qin D, Sun C, Xie M (2013) Deep sequencing-based analysis of the *Cymbidium ensifolium* floral transcriptome. *PLoS One* 8(12):e85480. doi: [10.1371/journal.pone.0085480](https://doi.org/10.1371/journal.pone.0085480) PMID: [24392013](https://pubmed.ncbi.nlm.nih.gov/24392013/)
27. Wang SY, Lee PF, Lee YI, Hsiao YY, Chen YY, Pan ZJ, Liu ZJ, Tsai WC (2011) Duplicated C-class MADS-box genes reveal distinct roles in gynostemium development in *Cymbidium ensifolium* (Orchidaceae). *Plant Cell Physiol* 52(3):563–577. doi: [10.1093/pcp/pcr015](https://doi.org/10.1093/pcp/pcr015) PMID: [21278368](https://pubmed.ncbi.nlm.nih.gov/21278368/)
28. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676. PMID: [16081474](https://pubmed.ncbi.nlm.nih.gov/16081474/)
29. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. PMID: [12969510](https://pubmed.ncbi.nlm.nih.gov/12969510/)
30. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27–30. PMID: [10592173](https://pubmed.ncbi.nlm.nih.gov/10592173/)
31. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25:1105–1111. doi: [10.1093/bioinformatics/btp120](https://doi.org/10.1093/bioinformatics/btp120) PMID: [19289445](https://pubmed.ncbi.nlm.nih.gov/19289445/)
32. Wang Y, Huang H, Ma Y, Fu J, Wang L, Dai S (2014) Construction and de novo characterization of a transcriptome of *Chrysanthemum lavandulifolium*: analysis of gene expression patterns in floral bud emergence. *Plant Cell Tiss Organ Cult* 116:297–309.
33. Schauser L, Wieloch W, Stougaard J (2005) Evolution of NIN-like proteins in *Arabidopsis*, rice, and *Lotus japonicus*. *J Mol Evol* 60:229–237. PMID: [15785851](https://pubmed.ncbi.nlm.nih.gov/15785851/)
34. Konishi M, Yanagisawa S (2013) Arabidopsis NIN-like transcription factors have a central role in nitrate signalling. *Nat Commun* 4:1617. doi: [10.1038/ncomms2621](https://doi.org/10.1038/ncomms2621) PMID: [23511481](https://pubmed.ncbi.nlm.nih.gov/23511481/)

35. Hsiao YY, Chen YW, Huang SC, Pan ZJ, Fu CH, Chen WH, et al. (2011) Gene discovery using next-generation pyrosequencing to develop ESTs for *Phalaenopsis* orchids. *BMC Genomics* 12:360. doi: [10.1186/1471-2164-12-360](https://doi.org/10.1186/1471-2164-12-360) PMID: [21749684](https://pubmed.ncbi.nlm.nih.gov/21749684/)
36. Robson F, Costa MM, Hepworth SR, Vizir I, Pineiro M, Reeves PH, et al. (2001) Functional importance of conserved domains in the flowering-time gene *CONSTANS* demonstrated by analysis of mutant alleles and transgenic plants. *Plant J* 28:619–631. PMID: [11851908](https://pubmed.ncbi.nlm.nih.gov/11851908/)
37. Suarez-Lopez P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G (2001) *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature* 410:1116–1120. PMID: [11323677](https://pubmed.ncbi.nlm.nih.gov/11323677/)
38. Chou ML, Shih MC, Chan MT, Liao S Y, Hsu CT, Haung YT, et al. (2013) Global transcriptome analysis and identification of a *CONSTANS*-like gene family in the orchid *Erycina pusilla*. *Planta* 237: 1425–1441. doi: [10.1007/s00425-013-1850-z](https://doi.org/10.1007/s00425-013-1850-z) PMID: [23417646](https://pubmed.ncbi.nlm.nih.gov/23417646/)
39. Audic S, Claverie JM (1997) The significance of digital gene expression profiles. *Genome Res* 7:986–995. PMID: [9331369](https://pubmed.ncbi.nlm.nih.gov/9331369/)
40. Mondragón-Palomino M, Hiese L, Härter A, Koch MA, Theißen G (2009) Positive selection and ancient duplications in the evolution of class B floral homeotic genes of orchids and grasses. *BMC Evol Biol* 9: 81. doi: [10.1186/1471-2148-9-81](https://doi.org/10.1186/1471-2148-9-81) PMID: [19383167](https://pubmed.ncbi.nlm.nih.gov/19383167/)