Transcriptome analysis of flower development and mining of genes related to the flower development in Oncidium

Nengyan Fang, Yuanhua Luo, Ronghui Fan, Xiuxian Ye, Minling Huang*, Huaiqin Zhong*
Institute of Crop Sciences, Fujian Academy of Agricultural Sciences (FAAS), Fujian Engineering Research Center of Characteristic Floriculture, Fuzhou, 350013 China

Received: August 07, 2023
Accepted: September 10, 2023
Published Online: October 15, 2023

Abstract

Oncidium, a kind of orchid plants characterized with unique flower, is one of the four tropical orchids with high ornamental value and favored by consumers. However, our understanding about the molecular basis of its flower development is still limited. Here, we collected Oncidium tissues at different developmental stages for RNA-seq. A total of more than 621 million clean reads were generated. 134,640 unigenes were assembled, and 54,221 unigenes were annotated. The number of DEGs (differentially expressed genes) was the largest in the comparison group F2-vs-F3 (F2 stage compared with F3 stage in flowers). The GO (Gene Ontology) terms and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways enriched for the specific DEGs were diverse at different stages. The common pathways enriched in multiple comparisons were also obtained. 11 key genes were obtained in cutin, suberin and wax biosynthesis pathway. Moreover, 32 candidate genes related to flower development were screened. Most of them presented tissue-specific expressions, especially MADS-box genes described by the ABCDE model. In all, the present data provides a valuable resource for dissecting the molecular mechanism of Oncidium in regulating flower development.

Keywords: Candidate genes, DEGs, Flower development, MADS-box genes, Oncidium, RNA-seq

How to cite this:

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Orchidaceae is the largest and most widely distributed plant family with more than 25,000 species (Leitch et al., 2009). Among them, Oncidium is a large category, widely distributed in the Americas such as Mexico, Brazil, Bolivia and other tropical areas (Hew, 1992). It is one of the most important cut-flowers and potted flowers in the world. Its unique structure and beautiful colors are favored by consumers. The flower of the dicotyledon model plant Arabidopsis has four whorls, and they are four sepals, four petals, six stamens and two carpels. The flower of rice, the model monocotyledon plant, also has the similar structure, and it consists of a pair of bract-like organs (lemma and palea), lodicules (equivalent to
eudicot petals), six stamens, and a carpel. Orchids have evolved into a unique group (Givnish et al., 2015). The structures of orchid flowers have unique diversifications, resulting from the long-term coevolution between insects and plants. The *Oncidium* has three sepals, two petals similar to sepals, a lip and a gynandrium that consists of an anther cap, two pollinium, a column, a carpel and an ovule. The lip, in particular, is an important organ that has evolved and is unique to *Oncidium*. Their unique flowers are so fascinating, and how they form is of great interest to researchers. Also, it is of great significance in *Oncidium* breeding to study the mechanism of the unique flower formation and mining of genes related to the flower development.

In plants, the MADS-box family genes play a critical role in flower developmental process (Callens et al., 2018). These genes can be divided into five classes (A, B, C, D and E), constituting the ABCDE model (Weigel and Meyerowitz, 1994; Theissen, 2001). The ABCDE model is generally conserved (Ferrario et al., 2004). In *Oncidium*, some MADS-box genes have been characterized to regulate the flower development. *OMADS10* is a putative paleoAP1 (*paleoAPETALAI*) ortholog, and overexpression of *OMADS10* only causes moderately early flowering in transgenic *Arabidopsis* (Chang et al., 2009). *OMADS3, OMADS5* and *OMADS9* have been characterized as AP3-like (*APETALAI3*-like) genes (Hsu and Yang, 2002). *OMADS5* can negatively affect lip formation (Chang et al., 2010). *OMADS8* has been characterized as PI (*PISTILLATA*) homolog, and it can cause the sepal into an expanded petal-like structure in transgenic *Arabidopsis* (Chang et al., 2010). *OMADS4* is characterized as AG (*AGAMOUS*) homolog, expressed in the reproductive organ column (Hsu et al., 2010). *OMADS2* is putative STK-like (*SEEDSTICK*-like) gene, expressed in the stigmatic cavity of column and ovary (Hsu et al., 2010). *OMADS6*, characterized as SEP3 (*SEPALATA 3*) homolog, and *OMADS11*, characterized as SEP1/2 (*SEPALATA 1/2*) homolog, regulate flower transition and formation (Hsu et al., 2003; Chang et al., 2009). *OMADS1* and *OMADS7* are characterized as AGL6 (*AGAMOUS*-like 6) homologs, and their ectopic expressions can cause early flowering in *Arabidopsis* (Hsu et al., 2003; Chang et al., 2009).

The next-generation deep-sequencing technologies, such as Illumina RNA-seq provide new approaches to study global transcriptome profiles for specific tissue research materials within a specific period (Marioni et al., 2008). In this study, RNA-Seq was performed on flowers, pseudobulbs, and leaves at different developmental stages using *Oncidium* variety Jihui. DEGs (Differentially expressed genes) were analyzed, GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) enriched were performed for the DEGs, and candidate genes related to the flower development were mined. The purpose was to provide theoretical support for studying the molecular mechanism of *Oncidium* flower regulation.

### Material and Methods

#### Plant material

The *Oncidium* cultivar Jinhui was used in this study, an improved variety obtained from *Oncidium Gower Ramsey* (a hybrid from *Oncidium Goldiana × Oncidium Guiea* Gold). Three-year-old plants were grown in the greenhouse of the Institute of Crop Sciences, Fujian Academy of Agricultural Sciences (FAAS) (Fuzhou, China). Five developmental stages of the flower samples, floral meristem (F1), flower bud when septal and petal were green (F2), flower bud when septal and petal turned yellow (F3), flower bud before opening (F4), and mature flower (F5), were collected (Figure S1E). The leaf and pseudobulb samples were also collected from young, mature and flowering plants (Figure S1A-C). The pseudobulb (PB4) samples from post-flowering plants were also collected (Figure S1D). These samples were used for RNA extraction. Two independent RNA samples were used for each experiment in this study.

#### RNA extraction

Total RNA was extracted from collected samples with TRIzol kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer’s instructions. RNA degradation and contamination were monitored on 1% agarose gels, and purity was checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA). RNA concentration was measured using the Qubit® RNA Assay Kit in the Qubit®2.0 Fluorometer (Life Technologies, CA, USA), and integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

#### cDNA library construction and Illumina sequencing

A total of 3 μg RNA per sample was used for the RNA
sample preparation, and twenty-four libraries were generated using the NEBNext®Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) following the manufacturer’s recommendation. Briefly, after index codes added to attribute sequences to each sample, mRNA purified. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X). First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H Minus). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I. After adenylation of 3’ ends of DNA fragments, NEBNext Adaptor was ligated for hybridization. Then cDNA fragments of 150-200 bp in length were selected for PCR and the products were purified. After library quality assessed, the library preparations were sequenced on an Illumina Hiseq 2000 platform and paired-end reads were generated by Biomarker Inc. (Beijing, China). (NCBI SRA accession PRJNA532798).

Transcriptome assembly and gene annotation
Clean data were obtained by removing reads containing an adapter, reads containing ploy-N, and low-quality reads from raw data. Transcriptome assembly was accomplished using Trinity with min_kmer_cov set to 2, and all other parameters was set to default (Grabherr et al., 2011). The obtained unigenes were aligned against protein databases using blast software (Altschul et al., 1997) for function annotation, including the NCBI non redundant (Nr) protein database, Swiss-Prot (The UniProt Consortium, 2018), GO (Ashburner et al., 2000), COG (Clusters of Orthologous Groups) (Tatusov et al., 2000), KOG (euKaryotic Orthologous Groups) (Koonin et al., 2004), eggNOG4.5 (Huerta-Cepas et al., 2016), and KEGG (Kanehisa et al., 2004). The KEGG Orthology of Unigenes in KEGG was obtained using KOBAS 2.0 program (Xie et al., 2011). The E-value cutoff of Blast was set at 1e-5. After the amino acid sequence of UniGene was predicted, Hmmer software (Eddy, 1988) was used to compare with Pfam (Protein family) (Finn et al., 2014) database, and parameter E-value was set to < 1e-10.

Expression analysis of Unigenes
The sequenced reads were compared with Unigene library using Bowtie (Langmead et al., 2009). According to the comparison results, the expression level was estimated using RSEM (Li and Dewen, 2011). The FPKM (fragments per kilobase of transcript per million mapped reads) value indicated the gene expression level. The average FPKM of unigene <= 1.0 was identified as not expressed. The differential expression analysis was performed using the DESeq (Anders and Huber, 2010). DESeq provided statistical analysis for determining differential expressions in the DGEs, using a model based on the negative binomial distribution. The resulting P-values were adjusted using the Benjami-Hochberg approach to control the false discovery rate (FDR). Genes with an adjusted P-value < 0.01 of FDR and fold change (FC) >= 2 obtained by DESeq were assigned as differentially expressed. FC represented the ratio of expression between two groups.

Phylogenetic analysis of MADS-box genes
The MADS-box family protein sequences of Arabidopsis were obtained from the Arabidopsis Information Resource (TAIR, http://www.Arabidopsis.org/), and the MADS-box family protein sequences of rice (Oryza sativa) were downloaded from the Rice Genome Annotation Project (http://rice.uga.edu/). The MADS domains were analyzed using SMART (http://smart.embl-heidelberg.de/). The MADS-box protein domain sequences of Arabidopsis, rice and Oncidium were aligned using ClustalW program. A neighbor-joining (NJ) phylogenetic tree was constructed using MEGAX64 with default parameters, and the bootstrap analysis was set at 1000 replicates.

Quantitative real-time PCR (qRT-PCR) analysis
To validate the accuracy of the transcriptome profile, total RNA was extracted from samples using the TRIZol kit. The total RNA was reverse transcribed into cDNA using the PrimeScript™ 1st Strand cDNA Synthesis Kit (TaKaRa, Dalian, China), according to the manufacturer’s protocol. Nine unigenes were selected for evaluation by qRT-PCR, using ABI 7500 Real Time System and SYBR GREEN PCR Master Mix (TaKaRa, Dalian, China). The primers were listed in Table - S1. Actin was used as an internal control. Reactions were performed at 94 °C for 2 min, then cycled at 95 °C for 30 s, 48-54 °C for 30 s, 72 °C for 30 s for 35 cycles, and finally, 72 °C for 10 min. Each assay was performed in triplicate. Data analysis was performed using the 2-ΔΔCt method (Livak and Schmittgen, 2001).
Results

Illumina sequencing and sequence assembly
In this study, 621,709,321 clean reads were generated with an average length of 295 bp. The average percentages of GC and Q30 were 46.78% and 91.87%, respectively. After assembly, 134,640 unigenes were obtained with an average length of 954 bp and an N50 of 1,445 bp (Table -1). 31.90% of unigenes were >= 1000 bp in length, and the length distribution of unigenes was shown in Figure S2.

Table-1. Summary for Oncidium transcriptome

<table>
<thead>
<tr>
<th>Items</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of clean reads</td>
<td>621,709,321</td>
</tr>
<tr>
<td>Total clean nucleotides (nt)</td>
<td>183,534,969,996</td>
</tr>
<tr>
<td>Average read length (bp)</td>
<td>295</td>
</tr>
<tr>
<td>GC (%)</td>
<td>46.78</td>
</tr>
<tr>
<td>Q30 (%)</td>
<td>91.87</td>
</tr>
<tr>
<td>Total number of unigenes</td>
<td>134,640</td>
</tr>
<tr>
<td>Mean length of unigenes (bp)</td>
<td>954</td>
</tr>
<tr>
<td>N50 length (bp)</td>
<td>1,445</td>
</tr>
</tbody>
</table>

Notes: Q30 means the probability of base recognition error is 0.1%, or the accuracy rate is 99.9%. N50 means the shortest contig length obtained when the cumulative length of the longest contig length equals 50% of the total assembly length.

Gene annotation and functional classification
In total, 54,221 (40.27%) of the assembled unigenes could be annotated (Table - S2). 53,393 (39.66%) unigenes hit with the Nr databases. 53.38% of the matches presented 60% or more similarity. In a subsequent gene matching analysis, the dominant closest matches included Elaeis guineensis (25.74%), Phoenix dactylifera (22.17%) and Musa acuminate subsp. Malaccensis (9.74%) (Figure S3BC). The Blast Nr E-value distribution was showed in Figure S3A. Totally 32,768 (24.34%) and 35,910 (26.67%) unigenes hit with the Swiss-Prot and Pfam databases, respectively. 15,061 (11.19%), 31,664 (23.52%) and 49,772 (36.97%) unigenes were classified into 25 function classifications in the COG, KOG, and eggNOG, respectively (Figure S4). The largest group was ‘general function prediction only’ in KOG and eggNOG classification. While in COG classification, the cluster ‘replication, recombination, and repair’ represented the largest group, followed by ‘general function prediction only’.

In GO classification, 28,872 (21.44%) unigenes were categorized into three main groups and 53 sub-functional groups (Figure S5). Dominant cellular component categories included ‘nucleolus’, ‘cell part’ and ‘organelle’. The primarily molecular function categories included ‘catalytic activity’, ‘binding’ and ‘transporter activity’. 15,697 (11.66%) unigenes were mapped to the KEGG database, and 126 biochemical pathways were obtained (Table - S3). ‘Ribosome pathway’ presented the largest number of unigenes, followed by ‘carbon metabolism’, and ‘biosynthesis of amino acids’.

Differentially expressed genes analysis in flower
Comparisons were performed between DEG libraries at adjacent stages. A total of 2,251 DEGs were detected in F1-vs-F2, with 1,599 genes up-regulated and 652 genes down-regulated. In F2-vs-F3, the largest number of DEGs (5,159) was identified, with 1,889 genes up-regulated and 3,270 genes down-regulated. In the F3-vs-F4, the fewest number of DEGs (1,317) was identified, with 574 genes up-regulated and 743 genes down-regulated. There were 1,846 DEGs detected in F4-vs-F5, with 725 genes up-regulated and 1,211 genes down-regulated (Figure 1A). The results showed that DEGs were the most active during the flower development from F2 to F3 stage.

The specific DEGs uniquely detected at different stages were analyzed. The Venn diagram of the DEGs showed the number of DEGs only identified in each comparison groups and the overlap relationship between the comparison groups. The largest number of the specific DEGs was detected only in the F2-vs-F3, with 1,491 genes up-regulated and 1,211 genes down-regulated. There were three common up-regulated DEGs shared by all four comparisons, and these genes might be closely related to the regulation of the flower development (Figure 1B). They were alternative oxidase 1 (BMK_Unigene_023270), ubiquinol oxidase 2 (BMK_Unigene_023271) and CER4 (ECERIFERUM 4) (BMK_Unigene_112129).

GO and KEGG enrichment analysis of the specific DEGs
In order to analyze the developmental characteristics of flowers at different stages, the GO-term enrichment analysis was performed on the specific DEGs that were uniquely detected in different comparisons. The results revealed that the GO terms
enriched at different stages were diverse. The GO terms for the specific up-regulated and down-regulated DEGs were mainly enriched in biological process and molecular function (Figure S6). For the specific up-regulated DEGs, several categories possibly associated with the flower development were found, including ‘cutin biosynthetic process’, ‘trehalose biosynthetic process’ and ‘sporopollenin biosynthetic process’ in F1-vs-F2, ‘chitinase activity’ and ‘chitin catabolic process’ in F2-vs-F3, ‘terpene synthase activity’ and ‘polar nucleus fusion’ in F3-vs-F4, and ‘carotene catabolic process’ in F4-vs-F5. Different categories were enriched for the specific down-regulated DEGs, and categories possibly associated with the flower development also were found, such as ‘flower morphogenesis’ and ‘floral whorl development’ in F1-vs-F2, ‘floral organ development’ and ‘reproductive process’ in F2-vs-F3, ‘cutin biosynthetic process’ and ‘fatty acid biosynthetic process’ in F3-vs-F4, and ‘carbohydrate metabolic process’ in F4-vs-F5.

KEGG analysis were performed on the specific up-regulated DEGs in different comparisons. The results ($P <= 0.01$) showed that 32 biochemical pathways were obtained (Figure 1C). Most of the specific up-regulated genes in each combination were significantly enriched in different pathways. ‘Flavone and flavonol biosynthesis’, ‘glycerolipid metabolism’, ‘brassinosteroid biosynthesis’ and ‘glycerophospholipid metabolism’ pathways were only significantly enriched in the F1-vs-F2. Seven pathways were only significantly enriched in the F2-vs-F3, such as ‘starch and sucrose metabolism’, ‘alpha-linolenic acid metabolism’, ‘citrate cycle (TCA cycle), etc. The fewest number of the specific pathways were significantly enriched in the F3-vs-F4, including ‘protein processing in endoplasmic reticulum’, ‘endocytosis’ and ‘spliceosome’. While the largest number of the specific pathways were significantly enriched in the F4-vs-F5, including eight pathways such as ‘carbon metabolism’, ‘circadian rhythm - plant’, ‘glycolysis/gluconeogenesis’, etc. The common pathways enriched in multiple comparisons were obtained, including ‘fatty acid elongation’, ‘cutin, suberine and wax biosynthesis’, ‘carotenoid biosynthesis’, ‘biosynthesis of unsaturated fatty acids’, ‘terpenoid backbone biosynthesis’, ‘biosynthesis of unsaturated fatty acids’, ‘fatty acid metabolism’, ‘galactose metabolism’, and ‘sphingolipid metabolism’.

Analysis of genes related to flower development
Based on gene expression, GO description, KEGG pathway description, and combined with related references, 32 candidate genes related to flower development were screened (Table - 2, Table - S4). The largest number of them is MADS-box gene. The phylogenetic analysis was performed using the MADS conserved domain of MADS-box genes, including the related MADS-box genes in Oncidium, rice and Arabidopsis (Arora et al., 2007). The result showed that MADS-box candidate genes in Oncidium included STK, AG, two AP1 homologs, two AP3 homologs, AGL6, two AGL12 (AGAMOUS-like 12) homologs, AGL65 (AGAMOUS-like 65), SEPI, SEP2 and SVP (SHORT VEGATATIVE PHASE) (Figure 2A). Among them, some genes were found the corresponding genes that had characterized in Oncidium, such as AP1 (BMK_Unigene_026197), two AP3 homologs and SEPI. While BMK_Unigene_020766 (AP1) and OMADS10,
BMK_Unigene_034758 (AG) and OMADS4, and BMK_Unigene_004701 (AGL6) and OMADS1 were clustered into the different branches respectively, suggesting that they probably were different genes. Some genes had not been characterized in Oncidium yet, such as SEP2, SVP, AGL12 and AGL65.

It seemed that most of MADS-box genes described by the ABCDE model were preferentially expressed in flowers. BMK_Unigene_020766 (AP1) was mainly expressed in vegetative tissues and less in flowers. In addition to being highly expressed in vegetative tissues, BMK_Unigene_026197 (AP1) is also relatively expressed at F1, F4 and F5 stages. Two AP3 homologs (BMK_Unigene_038423 and BMK_Unigene_118517) were expressed in flowers at different stages. While the expression of BMK_Unigene_038423 (AP3) was also both detected in pseudobulbs and leaves. The expressions of BMK_Unigene_034758 (AG) and BMK_Unigene_042575 (STK) were only detected in flowers. While the expression of BMK_Unigene_042575 (STK) was only detected in flowers buds at F1 and F3 stages. BMK_Unigene_040222 (SEP1) and BMK_Unigene_029684 (SEP2) were expressed in flowers at different stages. The expression of BMK_Unigene_040222 (SEP1) was also detected in young pseudobulbs. BMK_Unigene_004701 (AGL6) was expressed in flowers at different stages and in leaves at L1 and L3 stages. BMK_Unigene_113329 (AGL12) and BMK_Unigene_117532 (AGL65) were mainly expressed in flowers. While BMK_Unigene_042320 (AGL12) and BMK_Unigene_041872 (SVP) were mainly expressed in vegetative tissues (Figure 2B).

Six genes regulating flowering time, FKF1 (FLAVIN-BINDING, KELCH REPEAT, and F-BOX 1), FLK (FLOWERING LOCUS WITH KH DOMAINS), GI (GIGANTEA), HY5 (Elongated hypocotyl 5), ICU2 (INCURVATA2) and MET1 (methyltransferase 1), were screened. GI and HY5 were all expressed in flowers, pseudobulbs and leaves. FLK was not expressed in young leaves. FKF1 was not expressed in flower buds at F4 stage. ICU2 was mainly expressed in flower buds and young pseudobulbs. MET1 was only expressed in flowers at F1 and F2 stages. Four genes related to inflorescence development, AGO1 (ARGONAUTE 1), AIM1 and two CORYNE homologs, were screened. AGO1 was only expressed in flowers. AIM1 was expressed in flower buds and young pseudobulbs. The expressions of two CORYNE homologs were detected in flowers, pseudobulbs and young leaves. ACLA-2 (ATP-citrate lyase A-2), AMP1 (ALTERED MERISTEM PROGRAM 1), ARF5 (AUXIN RESPONSE FACTOR 2), GPAT6 (glycerol-3-phosphate acyltransferase 6), HK2 (histidine kinase 2), TAR2 (TRYPTOPHAN AMINOTRANSFERASE RELATED 2) and TGA8 (TGACG sequence-specific binding protein 8) were thought to be involved in floral organ development. ARF5 and TGA8 were only expressed in flowers at F1, F2 and F3 stages, and the expressions were the highest in flowers at F1 stage. ACLA-2 (BMK_Unigene_017726), HK2 and two GPAT6 homologs were expressed both in flowers and vegetative tissues at different stages. While ACLA-2 (BMK_Unigene_089599) was only expressed in flowers and young pseudobulbs. The expression of FLK could not be detected in young leaves. The expression of TAR2 could not be detected in pseudobulbs of young, mature and post-flowering plants (Figure 2B).
**Table 2. Flower development-related genes in Oncidium**

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene Symbol</th>
<th>GeneBank Description</th>
<th>Accession</th>
<th>E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMK_Unigene_017726</td>
<td>ACLA-2</td>
<td>ATP-citrate lyase A-2 [Arabidopsis thaliana]</td>
<td>NP_001320572.1</td>
<td>0</td>
</tr>
<tr>
<td>BMK_Unigene_089599</td>
<td>ACLA-2</td>
<td>ATP-citrate lyase A-2 [Arabidopsis thaliana]</td>
<td>NP_001320572.1</td>
<td>0</td>
</tr>
<tr>
<td>BMK_Unigene_033489</td>
<td>AGO1</td>
<td>Stabilizer of iron transporter SufD / Poly nucleotidyl transferase [Arabidopsis thaliana]</td>
<td>NP_001185169.1</td>
<td>0</td>
</tr>
<tr>
<td>BMK_Unigene_063492</td>
<td>AIM1</td>
<td>Enoyl-CoA hydratase/isomerase family [Arabidopsis thaliana]</td>
<td>NP_194630.1</td>
<td>0</td>
</tr>
<tr>
<td>BMK_Unigene_018210</td>
<td>AMP1</td>
<td>Peptidase M28 family protein [Arabidopsis thaliana]</td>
<td>NP_567007.1</td>
<td>0</td>
</tr>
<tr>
<td>BMK_Unigene_049059</td>
<td>ARF5</td>
<td>Transcriptional factor B3 family protein / auxin-responsive factor AUX/IAA-like protein [Arabidopsis thaliana]</td>
<td>NP_173414.1</td>
<td>0</td>
</tr>
<tr>
<td>BMK_Unigene_087654</td>
<td>CORYNE</td>
<td>Protein kinase superfamily protein [Arabidopsis thaliana]</td>
<td>NP_850812.2</td>
<td>1.00E-130</td>
</tr>
<tr>
<td>BMK_Unigene_119565</td>
<td>CORYNE</td>
<td>Protein kinase superfamily protein [Arabidopsis thaliana]</td>
<td>NP_850812.2</td>
<td>7.00E-135</td>
</tr>
<tr>
<td>BMK_Unigene_089905</td>
<td>FKF1</td>
<td>Flavin-binding, kelch repeat, I box 1 [Arabidopsis thaliana]</td>
<td>NP_564919.1</td>
<td>0</td>
</tr>
<tr>
<td>BMK_Unigene_122609</td>
<td>FLK</td>
<td>RNA-binding KH domain-containing protein [Arabidopsis thaliana]</td>
<td>NP_001325772.1</td>
<td>3.00E-10</td>
</tr>
<tr>
<td>BMK_Unigene_050492</td>
<td>GI</td>
<td>GIGANTEA [Arabidopsis thaliana]</td>
<td>NP_564180.1</td>
<td>0</td>
</tr>
<tr>
<td>BMK_Unigene_042209</td>
<td>GPAT6</td>
<td>Glycerol-3-phosphate acyltransferase 6 [Arabidopsis thaliana]</td>
<td>NP_181346.1</td>
<td>1.00E-104</td>
</tr>
<tr>
<td>BMK_Unigene_048418</td>
<td>GPAT6</td>
<td>Glycerol-3-phosphate acyltransferase 6 [Arabidopsis thaliana]</td>
<td>NP_181346.1</td>
<td>0</td>
</tr>
<tr>
<td>BMK_Unigene_008170</td>
<td>HK2</td>
<td>Histidine kinase 2 [Arabidopsis thaliana]</td>
<td>NP_568532.1</td>
<td>0</td>
</tr>
<tr>
<td>BMK_Unigene_018187</td>
<td>HY5</td>
<td>Basic-leucine zipper (bZIP) transcription factor family protein [Arabidopsis thaliana]</td>
<td>NP_568246.1</td>
<td>4.00E-43</td>
</tr>
<tr>
<td>BMK_Unigene_014851</td>
<td>ICU2</td>
<td>DNA-directed DNA polymerase; INCURVATA2 [Arabidopsis thaliana]</td>
<td>NP_001331725.1</td>
<td>0</td>
</tr>
<tr>
<td>BMK_Unigene_013169</td>
<td>MET1</td>
<td>Methyltransferase 1 [Arabidopsis thaliana]</td>
<td>NP_199727.1</td>
<td>0</td>
</tr>
<tr>
<td>BMK_Unigene_032233</td>
<td>TAR2</td>
<td>Tryptophan amino transferase related 2 [Arabidopsis thaliana]</td>
<td>NP_567706.1</td>
<td>1.00E-147</td>
</tr>
<tr>
<td>BMK_Unigene_004530</td>
<td>TGA8</td>
<td>PERIANTHIA (PAN); TGACG sequence-specific binding protein 8 (TGA8) [Arabidopsis thaliana]</td>
<td>NP_177031.1</td>
<td>6.00E-136</td>
</tr>
<tr>
<td>BMK_Unigene_034758</td>
<td>AG</td>
<td>AGAMOUS [Arabidopsis thaliana]</td>
<td>NP_567569.3</td>
<td>4E-87</td>
</tr>
<tr>
<td>BMK_Unigene_042575</td>
<td>STK</td>
<td>AGAMOUS [Arabidopsis thaliana]</td>
<td>NP_567569.3</td>
<td>7E-96</td>
</tr>
<tr>
<td>BMK_Unigene_004701</td>
<td>AGL6</td>
<td>AGAMOUS-like 6 [Arabidopsis thaliana]</td>
<td>NP_182089.1</td>
<td>4E-32</td>
</tr>
<tr>
<td>BMK_Unigene_042320</td>
<td>AGL12</td>
<td>AGAMOUS-like 12 [Arabidopsis thaliana]</td>
<td>NP_565022.1</td>
<td>9E-42</td>
</tr>
<tr>
<td>BMK_Unigene_113329</td>
<td>AGL12</td>
<td>AGAMOUS-like 12 [Arabidopsis thaliana]</td>
<td>NP_565022.1</td>
<td>6E-43</td>
</tr>
<tr>
<td>BMK_Unigene_117532</td>
<td>AGL65</td>
<td>AGAMOUS-like 65 [Arabidopsis thaliana]</td>
<td>NP_001321506.1</td>
<td>4E-77</td>
</tr>
<tr>
<td>BMK_Unigene_020766</td>
<td>AP1</td>
<td>AGAMOUS-like 8; APETALA1; FUL [Arabidopsis thaliana]</td>
<td>NP_568929.1</td>
<td>3E-81</td>
</tr>
<tr>
<td>BMK_Unigene_026197</td>
<td>AP1</td>
<td>AGAMOUS-like 7; APETALA1 [Arabidopsis thaliana]</td>
<td>NP_001320362.1</td>
<td>3E-79</td>
</tr>
<tr>
<td>BMK_Unigene_038423</td>
<td>AP3</td>
<td>APETALA3 [Arabidopsis thaliana]</td>
<td>NP_191002.1</td>
<td>6E-64</td>
</tr>
<tr>
<td>BMK_Unigene_118517</td>
<td>AP3</td>
<td>APETALA3 [Arabidopsis thaliana]</td>
<td>NP_191002.1</td>
<td>3E-66</td>
</tr>
<tr>
<td>BMK_Unigene_040222</td>
<td>SEP1</td>
<td>AGAMOUS-like 2; SEPALLATA1 [Arabidopsis thaliana]</td>
<td>NP_568322.1</td>
<td>3E-100</td>
</tr>
<tr>
<td>BMK_Unigene_029684</td>
<td>SEP2</td>
<td>AGAMOUS-like 4; SEPALLATA 2 [Arabidopsis thaliana]</td>
<td>NP_186880.1</td>
<td>3E-91</td>
</tr>
<tr>
<td>BMK_Unigene_041872</td>
<td>SVP</td>
<td>AGAMOUS-like 22; SHORT VEGATATIVE PHASE [Arabidopsis thaliana]</td>
<td>NP_001324584.1</td>
<td>1E-77</td>
</tr>
</tbody>
</table>
Figure-2. Genes related to the flower development. A. Phylogenetic tree of MADS-box genes in Arabidopsis, rice and Oncidium. OMADS1-10 have been characterized in previous researches. All the MADS-box protein domain sequences were aligned using ClustalW program. The neighbor-joining tree was constructed using MEGAX64 program with bootstrap analysis of 1000 replicates. The branch length represented the distance length. The number represented the percent of data coverage for inter notes. B. Heatmap and clustering diagram showing the expression patterns of 32 candidate genes related to the flower development in different tissues. Log2-transformed FPKM values were used to plot the heatmap. The color scale indicated gene expression. Red indicated high expression, and green indicated low expression.

Analysis of genes related to cutin biosynthesis
The morphology of floral surfaces depends on the synthesis of cutin (Li-Beisson et al., 2009). In this study, the cutin, suberine and wax biosynthesis as a common pathway was enriched in multiple comparisons. 11 genes were screened in this pathway, including ACEITH (ADHESION OF CALYX EDGES/HOTHEAD), CYP86A1, CYP86A4, four CER1 (ECERIFERUM 1) homologs, two FAR1 (alcohol-forming fatty acyl-CoA reductase 1) homologs and two FAR4 (alcohol-forming fatty acyl-CoA reductase 4) homologs (Table - 3, Figure 3). The expression of ACEITH was detected in flowers, pseudobulbs and leaves, and the expression in flowers was higher than other tissues. CYP86A1 was expressed in flowers at different stages, pseudobulbs of young and mature plants and young leaves. CYP86A4 had a similar expression pattern to that of CYP86A1. Four CER1 homologs had different expression patterns. The expression of BMK_Unigene_120875 (CER1) was only detected in flowers, while the expression of BMK_Unigene_045810 (CER1) was detected in all tissues. BMK_Unigene_001977 (CER1) had a similar expression pattern to BMK_Unigene_045810 (CER1), except for that in pseudobulbs of the post-flowering plants. BMK_Unigene_054532 (CER1) was expressed mainly in flowers and certain vegetative tissues. FAR1 (BMK_Unigene_029767 and BMK_Unigene_082189) had same expression
Nengyan Fang et al.

patterns. Their expressions were detected in flowers and young pseudobulbs, and showed up-regulated in flowers with the flower development. **BMK_Unigene_029767** (FAR4) was only expressed in flowers at F3, F4 and F5 stages. **BMK_Unigene_082189** (FAR4) was only expressed in flowers at F4 and F5 stages. The expressions of both two FAR4 homologs showed up-regulated with the flower development.

**Table 3. Genes involving cutin suberine and wax biosynthesis**

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene Symbol</th>
<th>GeneBank Description</th>
<th>Accession</th>
<th>E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMK_Unigene_066611</td>
<td>CYP86A1</td>
<td>CYP86A1 [Arabidopsis thaliana]</td>
<td>NP_200694</td>
<td>4.00E-47</td>
</tr>
<tr>
<td>BMK_Unigene_129345</td>
<td>CYP86A4</td>
<td>CYP86A4 [Arabidopsis thaliana]</td>
<td>NP_171666</td>
<td>2.00E-60</td>
</tr>
<tr>
<td>BMK_Unigene_049557</td>
<td>ACE</td>
<td>ACE [Arabidopsis thaliana]</td>
<td>NP_565050</td>
<td>0</td>
</tr>
<tr>
<td>BMK_Unigene_001977</td>
<td>CER1</td>
<td>CER1 protein [Arabidopsis thaliana]</td>
<td>NP_001320</td>
<td>9.00E-101</td>
</tr>
<tr>
<td>BMK_Unigene_054532</td>
<td>CER1</td>
<td>CER1 protein [Arabidopsis thaliana]</td>
<td>NP_001320</td>
<td>3.00E-105</td>
</tr>
<tr>
<td>BMK_Unigene_120875</td>
<td>CER1</td>
<td>CER1 protein [Arabidopsis thaliana]</td>
<td>NP_001320</td>
<td>1.00E-105</td>
</tr>
<tr>
<td>BMK_Unigene_029767</td>
<td>FAR1</td>
<td>FAR1 [Arabidopsis thaliana]</td>
<td>NP_001331</td>
<td>3.00E-145</td>
</tr>
<tr>
<td>BMK_Unigene_082189</td>
<td>FAR1</td>
<td>FAR1 [Arabidopsis thaliana]</td>
<td>NP_197642</td>
<td>1.00E-39</td>
</tr>
<tr>
<td>BMK_Unigene_131954</td>
<td>FAR4</td>
<td>FAR4 [Arabidopsis thaliana]</td>
<td>NP_190040</td>
<td>5.00E-35</td>
</tr>
<tr>
<td>BMK_Unigene_133398</td>
<td>FAR4</td>
<td>FAR4 [Arabidopsis thaliana]</td>
<td>NP_190040</td>
<td>2.00E-36</td>
</tr>
</tbody>
</table>

**qRT-PCR validation**

To validate the gene expression levels obtained from Illumina RNA-Seq results, nine unigenes including two AP1, two AP3, AG, STK, SEP1, SEP2 and AGL6, were selected for real-time PCR analysis. The results showed that AP3, AG, STK, SEP1, SEP2 and AGL6 were mainly expressed in flowers, and displayed lower expression level in vegetative tissues. While **BMK_Unigene_020766** (AP1) and **BMK_Unigene_026197** (AP1) was mainly expressed in leaves and pseudobulbs, and **BMK_Unigene_020766** (AP1) displayed higher expression level in pseudobulbs than that in leaves. Overall, the expression patterns of these nine genes were consistent with the RNA-Seq assay, verifying the reliability of the Illumina sequencing data (Figure S7).

**Discussion**

Currently, there are few available genomic information of Oncidium that can provide insights into the mechanisms behind the flower development. In this study, a large number of cDNA sequencing data were generated, facilitating the identification of genes regulating the flower development in

Figure 3. Regulatory changes in the pathway of cutin, suberin and wax biosynthesis during the flower development. Colored boxes, genes identified in our data; blue boxes, mixed regulated genes in different stages; red boxes, up-regulated genes; dark blue boxes, down-regulated genes. Log2-transformed FPKM values were used to plot the heatmap. The color scale presents gene expression, where red indicates high expression and green indicates low expression.
Oncidium. A set of 621 million clean reads were generated, and assembled into 134,640 unigenes. After annotation against protein databases, we concluded that 40.27% of unigenes were annotated as putative functions and > 50% of unigenes had no homologs in the databases used. These transcriptome data provided us more and novel unigenes to predict the flower development in Oncidium.

A limited number of MADS-box genes have been characterized in Oncidium. In A class, AP1/FUL is required for establishment of the floral meristem (Mandel et al., 1992). In Oncidium, OMADS10 is characterized as a putative API ortholog, and is expressed in lip, carpel/ovule of flowers and leaves (Chang et al., 2009). In this study, Two API homologs (BMK_Unigene_020766 and BMK_Unigene_026197) were obtained. BMK_Unigene_026197 was expressed in floral meristems, mature flowers and vegetative tissues, similar to the expression pattern of OMADS10 (Chang et al., 2009). Based on the phylogenetic analysis and expression results, it was inferred that BMK_Unigene_026197 probably was OMADS10. While BMK_Unigene_020766 had a different expression pattern, with high expression in vegetative tissues rather than in flowers. The phylogenetic analysis result showed that BMK_Unigene_020766 placed different branch with OMADS10. The results suggested that BMK_Unigene_020766 was different form OMADS10, and there were at least two genes in A class in Oncidium. BMK_Unigene_020766 probably had a different function in regulating the flower development. In class B, OMADS3, OMADS5 and OMADS9 have been characterized as AP3-like genes in previous studies (Chang et al., 2010). OMADS8 has been characterized as PI homolog (Chang et al., 2010). OMADS3 and OMADS8 are expressed in floral organs as well as in vegetative leaves. OMADS5 and OMADS9 are strongly expressed in sepal and petal of flowers (Chang et al., 2010). In this study, only two AP3 homologs (BMK_Unigene_038423 and BMK_Unigene_118517) were obtained. Their expression patterns are similar to those of OMADS5 and OMADS9, and they are mainly expressed in flowers. The phylogenetic analysis result showed that BMK_Unigene_038423 and OMADS5, BMK_Unigene_118517 and OMADS9 were clustered in the same branch, respectively. The results suggested that BMK_Unigene_038423 probably was OMADS5 and BMK_Unigene_118517 probably was OMADS9. No PI homologs were obtained in this study. In C class, AG is required for determinacy of stamen and carpel development (Yanofsky et al., 1990). OMADS4 is characterized as AG homolog, and is expressed in the reproductive organ column (Hsu et al., 2010). In this study, BMK_Unigene_034758 was characterized as AG homolog, and its expression was mainly detected in flowers at F2, F4 and F5 stages. The phylogenetic analysis result showed that BMK_Unigene_034758 and OMADS4 were clustered into different branches, and their genetic relationship was relatively distant. The results suggested that BMK_Unigene_034758 and OMADS4 were different genes, and there were at least two C class genes in Oncidium. In D class, OMADS2 is putative STK-like gene, and is expressed in the stigmatic cavity of column and ovary (Hsu et al., 2010). In this study, BMK_Unigene_042575 was characterized as STK homolog. BMK_Unigene_042575 was only expressed in flowers. There was a certain distance between BMK_Unigene_042575 and OMADS2 in the phylogenetic tree, suggesting that they were different genes. The results showed that there were at least two genes in D class in Oncidium. In E class, OMADS6 as SEP3 homolog, and OMADS11 as SEP1/2 homolog have been characterized, and they are expressed in flowers and but not in leaves (Hsu et al., 2003; Chang et al., 2009). BMK_Unigene_040222 was characterized as SEP1. The phylogenetic analysis result showed that BMK_Unigene_040222 and OMADS11 were clustered in the same branch, and their expression patterns were the same, indicating that BMK_Unigene_040222 probably was OMADS11. BMK_Unigene_029684 was characterized as SEP2 homolog, and its expression pattern in flowers was different from that of BMK_Unigene_040222, suggesting that it exhibited different roles in regulating the flower development. SEP3 homolog and SEP4 homolog were not found in this study. OMADS1 and OMADS7 have been characterized as AGL6 homologs, and their expressions could be detected in floral organs, but not in leaves (Hsu et al., 2003; Chang et al., 2009). In this study, BMK_Unigene_004701 was characterized as AGL6 homolog. BMK_Unigene_004701 was expressed in both flowers and leaves, which was different from the expression patterns of OMADS1 and OMADS7. In addition, the phylogenetic analysis result showed that BMK_Unigene_004701, OMADS1 and OMADS7 were not clustered into the same
branch. The results suggested that BMK_Unigene_004701 was not OMADS1 or OMADS7. In all, the pattern of these MADS-box genes we inferred regulating the flower development of Oncidium was shown in the Figure S8.

In addition to MADS-box genes, genes regulating flowering time were obtained, such as GI, FKF1 and HY5 in photoperiod pathway, SVP in ambient temperature pathway, ICU2 in vernalization pathway, and FLK in the autonomous pathway. GI and FKF1 are related to the circadian rhythm. They can form a complex and promote the expression of CO (CONSTANS), thus regulating the flowering time of plants (Imaizumi et al., 2005; Sawa et al., 2007). HY5 can bind directly to the promoter regions of PIF4 (PHYTOCHROME-INTERACTING FACTOR 4) and COL5 (CONSTANS-LIKE 5), and then represses their expressions, varying flowering time in Arabidopsis (Chu et al., 2022). GI, FKF1 and HY5 were expressed in flowers, pseudobulbs and leaves of Oncidium at different stages, but their expression patterns were different. The expression level of GI in Oncidium leaves and pseudobulbs increased with the growth of Oncidium, while the expression level in flowers decreased gradually from F1 to F4 stage, and then increased. The expression level of FKF1 was lower than that of GI. The expression pattern was similar to that of GI in leaves, but it fluctuated in pseudobulbs and flowers. With the growth of Oncidium, the expression level of HY5 decreased first and then increased in flowers, pseudobulbs and leaves, and the highest expression levels were found in mature flowers, post-flowering pseudobulbs and young leaves, respectively. SVP, an important role in the response of plants to ambient temperature changes, controls flowering time by negatively regulating the expression of a floral integrator, FT (FLOWERING LOCUS T) (Lee et al., 2007). In Oncidium, SVP was highly expressed in pseudobulbs and leaves, and lowly expressed in flowers at F1 stage. From vegetative growth to flowering, the expression level of SVP increased in pseudobulbs, while the expression level in leaves decreased first and then increased, suggesting that SVP functioned mainly in leaves in regulating flowering time. ICU2 can regulate flowering in vernalization pathway, and its mutants show an early flowering phenotype under both long-day and short-day photoperiod conditions (Hyun et al., 2013). The expression level of ICU2 in flowers decreased with the flower development, and in vegetative tissues was only detected in young pseudobulbs, which was different from the expression pattern of ICU2 in Arabidopsis. This indicated that ICU2 had different functions in Oncidium that was, it mainly regulated the flower development. FLK promotes flowering in the autonomous pathway by negatively regulating the MADS-box floral repressor FLC (FLOWERING LOCUS C) (Michaels and Amasino, 1999; Lim et al., 2004). In Oncidium, FLK was expressed in flowers, pseudobulbs and leaves, but the expression patterns were different. The expression level of FLK in leaves increased with the growth of Oncidium. Its expression level in pseudobulbs increased before flowering, decreased during flowering, and increased after flowering. While in flowers, the expression was increased first and then decreased.

The morphology of floral surfaces depends on the synthesis of cutin (Li-Beisson et al., 2009). Cutin that function is controlled by wax composition, can also block excessive water loss (Kerstiens, 1996; Riederer and Schreiber, 2001). The rich cutin and wax on flower surface of Oncidium may be the reason for the flower shape and long vast life. Cutin, suberin and wax biosynthesis pathway was common enriched in multiple comparisons in this study, indicating that the pathway played a role in the whole process of the flower development in Oncidium. In the pathway, CYP86AI, CYP86A4, ACE/HTH, CER1, FAR1 and FAR4 were screened. Most of them were mainly expressed in flowers, indicating that they played a role in flowers.

Conclusion

In this study, the transcriptome of different tissues from Oncidium using RNA-Seq was characterized, and the specific and common DEGs and pathways were identified in flowers. 32 candidate genes related flower development were screened and their expression patterns were analyzed. Among them, 13 MADS-box genes in Oncidium were identified, and most of them described by the ABCDE model showed tissue specificity. Phylogenetic analysis result showed that several MADS-box genes were not characterized in previous research. The cutin, suberin and wax biosynthesis pathway was enriched in multiple comparisons, and 11 key genes were obtained in this pathway. In all, these results greatly improve our understanding of the mechanisms underlying the flower development of Oncidium. The genes identified in this study could provide valuable
Nengyan Fang et al.

insights into their functions and contribute to understanding the flower development, and probably be applied to in breeding in the future.

Disclaimer: None.
Conflict of Interest: None.
Source of Funding: This research was supported by grants from the Natural Science Foundation of Fujian Province (2020J011371), The Scientific Research Projects of Fujian Academy of Agricultural Sciences (DEC201915), The ‘5511’ Collaborative Innovation Project (XTCXGC2021016) and The Scientific and Technological Innovation Team Construction Project of Fujian Academy of Agricultural Sciences (CXTD2021010-2).

Figure-S1. Samples from Oncidium Jinhui were used to transcriptome sequencing. A. The young plant with shoot-pseudobulb in the vegetative stage. B. The mature plant with inflated pseudobulbs. C. The flowering plant. D. The plant after flowering. E. Five development stages of flower. Bar = 1 cm.
Figure S2. The length distribution of assembled unigenes from Illumina reads.

Figure S3. BLAST outputs of unigene sequences against the Nr database. A. The distribution of E-values. B. The distribution of similarity scores. C. The mix of species contributing sequences to the analysis.

Figure S4. Unigenes annotation. A. KOG function classification of unigenes; B. COG function classification.
Figure-S5. GO function classification of unigenes. The left y-axis indicates the percentage of a specific category of genes in each main category. The right y-axis indicates the number of genes in the same category.
Figure S6. The most enriched GO terms for specific DEGs in different comparisons. The left y-axis indicates the percentage of a specific category of genes in each main category. The right y-axis indicates the number of genes in the same category.
Figure-S7. qRT-PCR validation of nine DEGs selected from the RNA-Seq. The blue bar charts showed the relative expression levels of AP1, AP3, AG, STK, SEP1, SEP2, and AGL6 genes, respectively, and the line charts showed the corresponding FPKM values of these genes.

Figure-S8. The pattern of MADS-box genes regulating the flower development of Oncidium. E class genes SEP and AGL6 (except for OMADS1) regulated sepal, petal, lip and gynostemium development, while OMADS1 only regulated lip development. C/D class genes AG and STK only regulated gynostemium development. B class gene AP3 (OMADS3) and PI regulated sepal, petal and lip development. OMADS5 only regulated sepal development, and OAMDS9 regulated petal and lip development.
References


Hsu HF, Huang CH, Chou LT and Yang CH, 2003. Ectopic expression of an orchid (Oncidium Gower Ramsey) AGL6-like gene promotes flowering by activating flowering time genes in

Asian J Agric & Biol. xxxx(x).
Nengyan Fang et al.

https://doi.org/10.1093/pcp/pcg099.

https://doi.org/10.1093/pcp/pcq052.

https://doi.org/10.1242/dev.084624.

https://doi.org/10.1093/nar/gkh063.

https://doi.org/10.1126/science.1110586.

https://doi.org/10.1093/jxb/47.12.1813.


https://doi.org/10.1105/tpc.11.5.949.

https://doi.org/10.1101/gad.1518407.

https://doi.org/10.1093/aob/mcp003.

https://doi.org/10.1186/1471-2105-12-323.

https://doi.org/10.1073/pnas.090990106.

https://doi.org/10.1105/tpc.019331.


https://doi.org/10.1038/360273a0.

https://doi.org/10.1101/gr.079558.108.


https://doi.org/10.1126/science.1146994.


**Contribution of Authors**

Fang N & Zhong H: Conceived idea, designed research methodology, conducted experiments, collected & analyzed data and wrote the manuscript.

Luo Y, Fan R & Ye X: Prepared plants and performed q-PCR analysis.

Huang M: Conceived idea, designed research methodology and analyzed data.