The genome and transcriptome of the *Phalaenopsis* yield insights into floral organ development and flowering regulation

Jian-Zhi Huang, Chih-Peng Lin, Ting-Chi Cheng, Ya-Wen Huang, Yi-Jung Tsai, Shu-Yun Cheng, Yi-Wen Chen, Chueh-Pai Lee, Wan-Chia Chung, Bill Chia-Han Chang, Shih-Wen Chin, Chen-Yu Lee, Fure-Chyi Chen

Phalaenopsis orchid is an important potted flower with high economic value around the world. We report the 3.1 Gb draft genome assembly of an important winter flowering *Phalaenopsis* ‘KHM190’ cultivar. We generated 89.5 Gb RNA-seq and 113 million sRNA-seq reads to use these data to identify 41,153 protein-coding genes and 188 miRNA families. We also generated a draft genome for *Phalaenopsis pulcherrima* ‘B8802’, a summer flowering species, via resequencing. Comparison of genome data between the two *Phalaenopsis* cultivars allowed the identification of 691,532 single-nucleotide polymorphisms. In this study, we reveal the key role of *PhAGL6b* in the regulation of flower organ development involves alternative splicing. We also show gibberellin pathways that regulate the expression of genes control flowering time during the stage in reproductive phase change induced by cool temperature. Our work should contribute a valuable resource for the flowering control, flower architecture development, and breeding of the *Phalaenopsis* orchids.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1707v1 | CC-BY 4.0 Open Access | rec: 3 Feb 2016, publ: 3 Feb 2016

The genome and transcriptome of the *Phalaenopsis* yield insights into floral organ development and flowering regulation

Jian-Zhi Huang[1*], Chih-Peng Lin[2,4*], Ting-Chi Cheng[1], Ya-Wen Huang[1],Yi-Jung Tsai[1], Shu-Yun Cheng[1], Yi-Wen Chen[1], Chueh-Pai Lee[2], Wan-Chia Chung[2], Bill Chia-Han Chang[2,3#], Shih-Wen Chin[1#], Chen-Yu Lee[1#] & Fure-Chyi Chen[1#]

[1]Department of Plant Industry, National Pingtung University of Science and Technology, Pingtung 91201, Taiwan

[2]Yourgene Bioscience, Shu-Lin District, New Taipei City 23863, Taiwan

[3]Faculty of Veterinary Science, The University of Melbourne, Parkville Victoria 3010 Australia

[4]Department of Biotechnology, School of Health Technology, Ming Chuan University, Gui Shan District, Taoyuan 333, Taiwan

[*]These authors contributed equally to this work.

[#]Correspondence should be addressed to B-C.H.C. (bchang@yourgene.com.tw), S.-W.C. (swchin@mail.npust.edu.tw), C.-Y.L. (culee@mail.npust.edu.tw) & F.-C.C. (fchen@mail.npust.edu.tw)

Abstract

***Phalaenopsis* orchid is an important potted flower with high economic value around the world. We report the 3.1 Gb draft genome assembly of an important winter flowering *Phalaenopsis* 'KHM190' cultivar. We generated 89.5 Gb RNA-seq and 113 million sRNA-seq reads to use these data to identify 41,153 protein-coding genes and 188 miRNA families. We also generated a draft genome for *Phalaenopsis pulcherrima* 'B8802', a summer flowering species, via resequencing. Comparison of genome data between the two *Phalaenopsis* cultivars allowed the identification of 691,532 single-nucleotide polymorphisms. In this study, we reveal the key role of *PhAGL6b* in the regulation of flower organ development involves alternative splicing. We also show gibberellin pathways that regulate the expression of genes control flowering time during the stage in reproductive phase change induced by cool temperature. Our work should contribute a valuable resource for the flowering control, flower architecture development, and breeding of the *Phalaenopsis* orchids.**

Keywords: ***Phalaenopsis*, draft genome, *PhAGL6b*, flower organ development, flowering time**

INTRODUCTION

Phalaenopsis is a genus within the family Orchidaceae and comprises approximately 66 species distributed throughout tropical Asia (Christenson 2002). The predicted *Phalaenopsis* genome size is approximately 1.5 gigabases (Gb), which is distributed across 19 chromosomes (Lin et al. 2001). *Phalaenopsis* flowers have a zygomorphic floral structure, including three sepals (in the first floral whorl), two petals and one of the petals develop into a labellum in early stage of development, which is a distinctive feature of a highly modified floral part in second floral whorl unique to orchids. The gynostemium contains the male and female reproductive organs in the center (Rudall & Bateman 2002). In the ABCDE model, B-class genes play important role to perianth development in orchid species (Chang et al. 2010; Mondragon-Palomino & Theissen 2011; Tsai et al. 2004). In addition, *PhAGL6a* and *PhAGL6b*, which were expressed specifically in the *Phalaenopsis* labellum, were implied to play as a positive regulator of labellum formation (Huang et al. 2015; Su et al. 2013). However, the relationship between the function of genes involved in floral-organ development and morphological features remains poorly understood.

Phalaenopsis orchids are produced in large quantity annually and are traded as the most important potted plants worldwide. During greenhouse production of young plants, the high temperature >28°C was routinely used to promote vegetative growth and inhibit spike initiation (Blanchard & Runkle 2006). Conversely, a lower ambient temperature (24/18°C day/night) is used to induce spiking (Chen et al. 2008) to produce flowering plants. Spike induction in *Phalaenopsis* orchid by this low temperature is the key to precisely control its flowering date. Several studies have indicated that low temperatures during the night are necessary for *Phalaenopsis* orchids to flower (Blanchard & Runkle 2006; Chen et al. 1994; Chen et al. 2008; Wang 1995). Despite a number of expressed sequence tags (ESTs), RNA-seqs and sRNA-seqs from *Phalaenopsis* inflorescence, flowering buds and leaves with or without low temperature treatment have been reported and deposited in GenBank or OrchidBase (An & Chan 2012; An et al. 2011; Hsiao et al. 2011; Su et al. 2011), only a few flowering genes or miRNAs have been identified and characterized. Besides, the clues to the spike initiation during reproductive phase change in the shorten stem, which may produce signals related to flowering during cool temperature induction, have not been dealt with. So far, the molecular mechanisms leading to spiking of *Phalaenopsis* has yet to be elucidated.

Here we report a high-quality genome and transcriptomes (mRNAs and small RNAs) of *Phalaenopsis* ‘KHM190’, a winter flowering hybrid with spike formation in response to low temperature. We also provide resequencing data for summer flowering species *P. pulcherrima* ‘P8802’. Our comprehensive genomic and transcriptome analyses provide valuable insights into the molecular mechanisms of important biological processes such as floral organ development and flowering time regulation.

METHODS SUMMARY

The genome of the *Phalaenopsis* Brother Spring Dancer 'KHM190' cultivar was sequenced on the Illumina HiSeq 2000 platform. The obtained data were used to assemble a draft genome sequence using the Velvet software (Zerbino & Birney 2008). RNA-Seq and sRNA-Seq data were generated on the same platform for genome annotation and transcriptome and small RNA analyses. Repetitive elements were identified by combining information on sequence similarity at the nucleotide and protein levels and by using de novo approaches. Gene models were predicted by combining publically available *Phalaenopsis* RNA-Seq data and RNA-Seq data generated in this project. RNA-Seq data were mapped to the repeat masked genome with Tophat (Trapnell et al. 2009)and CuffLinks (Trapnell et al. 2012). The detailed methodology and associated references are available in the SI Appendix.

RESULTS AND DISCUSSION

Genome sequencing and assembly. We sequenced the genome of the *Phalaenopsis* orchid cultivar 'KHM190' (SI Appendix, Fig. S1a) using the Illumina HiSeq 2000 platform and assembled the genome with the Velvet assembler, using 300.5 Gb (90-fold coverage) of filtered high-quality sequence data (SI Appendix, Table S1). This cultivar has an estimated genome size of 3.45 Gb on the basis of a 17-mer depth distribution analysis of the sequenced reads (SI Appendix, Fig. S2 and S3 and Table S2 and S3). *De novo* assembly of the Illumina reads resulted in a sequence of 3.1 Gb, representing 89.9% of the *Phalaenopsis* orchid genome. Following gap closure, the assembly consisted of 149,151 scaffolds (≥1000 bp), with N50 lengths of 100 kb and 1.5 kb for the contigs. Approximately 90% of the total sequence was covered by 6,804 scaffolds of >100 kb, with the largest scaffold spanning 1.4 Mb (SI Appendix, Table S3-S5). The sequencing depth of 92.5% of the assembly was more than 20 reads (SI Appendix, Fig. S3), ensuring high accuracy at the nucleotide level. The GC content distribution in the *Phalaenopsis* genome was comparable with that in the genomes of *Arabidopsis* (2000), *Oryza* (2005) and *Vitis* (Jaillon et al. 2007) (SI Appendix, Fig. S4).

Gene prediction and annotation. Approximately 59.74% of the *Phalaenopsis* genome assembly was identified as repetitive elements, including long terminal repeat retrotransposons (33.44%), DNA transposons (2.91%) and unclassified repeats (21.99%) (SI Appendix, Fig. S5 and Table S6). To facilitate gene annotation, we identified 41,153 high-confidence and medium-confidence protein-coding regions with complete gene structures in the *Phalaenopsis* genome using RNA-Seq (114.1 Gb for a 157.6 Mb transcriptome assembly), based on 20 libraries representing four tissues (young floral organs, leaves, shortened stems and protocorm-like bodies (PLBs)) (SI Appendix, Table S7), and we used transcript assemblies of these regions in combination with publically available expressed sequence tags (Su et al. 2011; Tsai et al. 2013) for gene model prediction and validation (Dataset S1-S2). We predicted 41,153 genes with an average mRNA length of 1,014 bp and a mean number of 3.83 exons per gene (Table 1 and Dataset S3). In addition to protein coding genes, we identified a total of 562 ribosomal RNAs, 655 transfer RNAs, 290 small nucleolar RNAs and 263 small nuclear RNAs in the *Phalaenopsis* genome (SI Appendix, Table S8). We also obtained 92,811,417 small RNA (sRNA) reads (18-27 bp), representing 6,976,375 unique sRNA tags (SI Appendix, Fig. S6 and Dataset S6-S7). A total of 650 miRNAs distributed in 188 families were identified (Dataset S8), and a total of 1,644 miRNA-targeted genes were predicted through the alignment of conserved miRNAs to our gene models (SI Appendix, Fig. S7 and Dataset S9-S10).

The *Phalaenopsis* gene families were compared with those of *Arabidopsis* (2000), *Oryza* (2005), and *Vitis* (Jaillon et al. 2007) using OrthoMCL (Li et al. 2003). We identified 41,153 *Phalaenopsis* genes in 15,855 families, with 8,532 gene families being shared with *Arabidopsis*, *Oryza* and *Vitis*. Another 5,143 families, containing 12,520 genes, were specific to *Phalaenopsis* (figure. 1). In comparison with the 29,431 protein-coding genes estimated for the *Phalaenopsis equestris* genome (Cai et al. 2015), our gene set for *Phalaenopsis* 'KHM190' contained 11,722 more members, suggesting a more wider representation of genes in this work. This difference in gene number may be due to different approaches between *Phalaenopsis* 'KHM190' and *Phalaenopsis equestris*. To better annotate the *Phalaenopsis* genome for protein-coding genes, we generated RNA-seq reads obtained from four tissues as well as publically available expressed sequence tags for cross reference. Besides, *Phalaenopsis* 'KHM190' is a hybrid and *Phalaenopsis equestris* a species, which may also show gene number difference due to different genetic background.

We defined the function of members of these families using Gene ontology (2008), the Kyoto Encyclopedia of Genes and Genomes (Kanehisa et al. 2012) and Pfam protein motifs (Finn et al. 2014) (SI Appendix, Fig. S8 and Dataset S3-S5). Furthermore, conserved domains could be identified in 50.17% of the predicted protein sequences based on comparison against Pfam databases. In addition, we identified 2,610 transcription factors (6.34% of the total genes) and transcriptional regulators in 55 gene families (SI Appendix, Fig. S9-S11 and Dataset S11-S12).

Regulation of *Phalaenopsis* floral organ development. The relative expression of all *Phalaenopsis* genes was compared through RNA-Seq analysis of shoot tip tissues from shortened stems, leaf, floral organs and PLB samples, in addition to vegetative tissues, reproductive tissues, and germinating seeds from *P. aphrodite* (Su et al. 2011; Tsai et al. 2013) (SI Appendix, Fig. S12 and Dataset S1). *Phalaenopsis* orchids exhibit a unique flower morphology involving outer tepals, lateral inner tepals and a particularly conspicuous labellum (lip) (Rudall & Bateman 2002). However, our understanding of the regulation of the floral organ development of these species is still in its infancy. To comprehensively characterize the genes involved in the development of *Phalaenopsis* floral organs, we obtained RNA-Seq data for the sepals, petals and labella of both the wild-type and peloric mutant of *Phalaenopsis* 'KHM190' at the 0.2-cm floral bud stage, which shows early sign of differentiation. This cultivar presented an early peloric fate in its lateral inner tepals. In a peloric flower, the lateral inner tepals are converted into a lip-like morphology at this bud stage (SI Appendix, Fig. S12a and 12b). We identified 3,743 genes that were differentially expressed in the floral organs of the wild-type and peloric mutant plants. Gene Ontology analysis of the differentially expressed genes in *Phalaenopsis* floral organs revealed functions related to biological regulation, developmental processes and nucleotide binding, which were significantly altered in both genotypes (Huang et al. 2015). Transcription factors (TFs) play a role in floral organ development. Of the 3,309 putative TF genes identified in the *Phalaenopsis* genome showed differences in expression between the wild-type and peloric mutant plants (Dataset S11). Notably, the *PhAGL6b* gene was upregulated in the peloric lateral inner tepals (lip-like petals) and lip organs (Huang et al. 2015). We therefore cloned the full-length sequence of *PhAGL6b* from lip organ cDNA libraries for the wild-type, peloric mutant and big lip mutant. The big lip mutant developed a petaloid labellum instead of the regular lip observed in the wild-type flower (figure 2b). Interestingly, we identified four alternatively spliced forms of *PhAGL6b* that were specifically expressed only in the petaloid labellum of the big lip mutant (figure 2c and 2d and SI Appendix, Fig. S13-S15). The four isoforms of the encoded PhAGL6b products differ in the length of their C-terminus region (figure 2d). C-domain is important for the activation of transcription of target genes (Honma & Goto 2001) and may affect the nature of the interactions with other MADS-box proteins in multimeric complexes (Geuten et al. 2006; Gramzow & Theissen 2010). In *Oncidium*, L (lip) complex (OAP3-2/OAGL6-2/OAGL6-2/OPI) is required for lip formation (Hsu et al. 2015). The *Phalaenopsis PhAGL6b* is an orthologue of *OAGL6-2*. In our study, the PhAGL6b and its different spliced forms may each other compete the *Phalaenopsis* L-like complex to affect labellum development as reported in *Oncidium* (Hsu et al. 2015). This provides a novel clue further supporting the notion that *PhAGL6b* may function as a key floral organ regulator in *Phalaenopsis* orchids, with broad impacts on petal, sepal and labellum development (figure 2e).

Control of flowering time in *Phalaenopsis*. The flowering of *Phalaenopsis* orchids is a response to cues related to seasonal changes in light (Wang 1995), temperature (Blanchard & Runkle 2006) and other external influences (Chen et al. 1994). A cool night time temperature of 18-20°C for approximately 4 weeks will generally induce spiking in most *Phalaenopsis* hybrids, while high temperature inhibits it. To compare gene expression between a constant high-temperature (30/27°C; day/night) and inducing cool temperatures (22/18°C), we collected shoot tip tissues from shortened stems of mature *P. aphrodite* plants after treatment at a constant high temperature (BH) and a cool temperature (BL) (1 to 4 weeks) for RNA-Seq data analysis (SI Appendix, Fig.S12g-i). More than 7,500 *Phalaenopsis* genes were found to be highly expressed in the floral meristems during 4 cool temperature periods (showing at least a 2-fold difference in the expression level in the BL condition relative to BH) (Dataset S13). The identified flowering-related genes correspond to transcription factors and genes involved in signal transduction, development and metabolism (figure 3 and Dataset S14). The classification of these genes include the following categories: photoperiod, gibberellins (GAs), ambient temperature, light-quality pathways, autonomous pathways and floral pathway integrators (Fornara et al. 2010; Mouradov et al. 2002). However, the genes involved in the photoperiod, ambient temperature, light quality and autonomous pathways did not show significant changes in the floral meristems during the cool temperature treatments (SI Appendix, Fig. S16 and Dataset S14). By contrast, the expression patterns of genes involved in pathways that regulate flowering, comprising a total of 22 GA pathway-related genes, were related to biosynthesis, signal transduction and responsiveness. The GA pathway-related genes and the floral pathway integrator genes have been revealed as representative key players in the link between flowering promotion pathways and the floral transition regulation network in several plant species (Mutasa-Gottgens & Hedden 2009). In contrast to the expression patterns observed in BL and BH, the GA biosynthetic pathway and positively acting regulator genes showed high expression levels in BL. Furthermore, the expression levels of negatively acting regulator genes were suppressed by the cool temperature treatment. The genes included in the flowering promotion pathways and floral pathway integrators were generally upregulated in BL (figure 3 and SI Appendix, Fig. S16 and Dataset S11). These findings suggest that the GA pathway may play a crucial role in the regulation of flowering time in *Phalaenopsis* orchid during cool temperature.

Polymorphisms for *Phalaenopsis* orchids. The *Phalaenopsis* genome assembly also provides the basis for the development of molecular marker-assisted breeding. Analysis of the *Phalaenopsis* genome revealed a total of 532,285 simple sequence repeats (SSRs) (SI Appendix, Fig. S17 and Table S9 and Dataset S15). To enable the identification of single nucleotide polymorphisms (SNPs), we re-sequenced the genome of a summer flowering species, *P. pulcherrima* ‘B8802’, with about tenfold coverage. Comparison of the genome data from the two *Phalaenopsis* accessions (KHM190 and B8802) allowed the discovery of 691,532 SNPs, which should be valuable for future development of SNP markers for *Phalaenopsis* marker-assisted selection. (SI Appendix, Fig. S18 and Table S10 and Dataset S16).

CONCLUSION

In this study, we sequenced, de novo assembled, and extensively annotated the genome of one of the most important *Phalaenopsis* hybrid. We also annotated the genome with a wealth of RNA-seq and sRNA-seq from different tissues, and many genes and miRNAs related to floral organ development, flowering time and protocorm (embryo) development were identified. Importantly, this RNA-Seq and sRNA-seq data allowed us to further improve the genome annotation quality. In addition, mining of SSR and SNP molecular markers from the genome and transcriptomes is currently being adopted in advanced breeding programs and comparative genetic studies, which should contribute to efficient *Phalaenopsis* cultivar development. Despite the *P. equestris* genome has been reported recently (Cai et al. 2015), focus on floral organ development and flowering time regulation has not been dealt with. In our study, we obtained transcriptomes from shortened stems, which initiate spikes in response to low ambient temperature, and floral organs and generated valuable data of potentially regulate flowering time key genes and floral organ development. The genome and transcriptome informations of our work should provide a constructive reference resource to upgrade the efficiency of cultivation and genetic improvement of *Phalaenopsis* orchids.

Acknowledgements

This work was supported by grants from the Agriculture and Food Agency, Council of Agriculture, Taiwan (grant numbers 102AS-9.1.1-FD-Z2(1), 103AS-9.1.1-FD-Z2(1), and 104AS-9.1.1-FD-Z2(1)).

Author Contributions

J.-Z.H., S.-W.C., C.-Y.L. and F.-C.C. conceived the project and the strategy. C.-P.Lin, C.-P.Lee, W.-C.C. and B.-C.H.C. conducted sequencing, assembly and annotation. C.-P.Lin were involved in genome resequencing analysis. J.-Z.H., C.-P.Lin, T.-C.C., Y.-W. H., Y.-J. T., S.-Y. C. and W.-C.C. performed RNA-Seq analysis. J.-Z.H., and C.-P.Lee performed sRNA-Seq analysis. C.-P.Lin and C.-P.Lee performed gene GC content analyses. C.-P.Lin and W.-C.C. transposable-element analysis. C.-P.Lin, and C.-P.Lee performed transfer RNA and microRNA analyses. J.-Z.H., C.-P.Lin, C.-P.Lee, B.-C.H.C. S.-W.C., C.-Y.L. and F.-C.C. performed SSR and SNP markers development. J.-Z.H. and C.-P.Lin performed gene evolutionary analyses. J.-Z.H., C.-P.Lin and W.-C.C. performed gene family analyses. J.-Z.H. and T.-C.C. performed RT-PCR and real-time PCR analyses. J.-Z.H., T.-C.C., Y.-W. H., Y.-J. T., S.-Y., S.-W.C., C.-Y.L. and F.-C.C. performed plant material development, DNA or RNA extraction and phenotyping. J.-Z.H., C.-P.Lin, S.-

W.C., C.-Y.L. and F.-C.C. wrote the manuscript .

Data deposition:

The *Phalaenopsis* genome assembly, transcriptomic and sRNA-seq data were deposited in Genbank with BioProject ID PRJNA271641. The version described in this paper is the first version, JXCR00000000. All short-read data are available via Sequence Read Archive: SRR1747138, SRR1753943, SRR1753944, SRR1753945, SRR1753946, SRR1753947, SRR1753948, SRR1753949, SRR1753950, SRR1752971, SRR1753106, SRR1753165, SRR1753166 (*Phalaenopsis* 'KHM190' genomic DNA); SRR1762751, SRR1762752, SRR1762753 (*Phalaenopsis* 'B8802' genomic DNA); SRR1760428, SRR1760429, SRR1760430, SRR1760432, SRR1760433, SRR1760435, SRR1760436, SRR1760438, SRR1760439, SRX396172, SRX396784, SRX396785, SRX396786, SRX396787, SRX396788 (RNA-seq); SRR1760091, SRR1760211, SRR1760212, SRR1760213, SRR1760270, SRR1760271, SRR1760523, SRR1760524, SRR1760525, SRR1760526, SRR1760527, SRR1760528, SRR1760530, SRR1760531, SRR1760532 (small RNA)

Figure Legends

Figure 1. Venn diagram showing unique and shared gene families between and among *Phalaenopsis*, *Oryza*, *Arabidopsis* and *Vitis*.

Figure 2. Possible evolutionary relationship of *PhAGL6b* in the regulation of lip formation in *Phalaenopsis* orchid.
(a) Wild-type flower. (b) A big lip mutant of *Phalaenopsis* World Class 'Big Foot'. (c) Representative RT-PCR result showing the mRNA splicing pattern of *PhAGL6b* in wild-type (W) and big lip mutant (M). (d) Alignment of the amino acid sequences of alternatively spliced forms of *PhAGL6b*. (e) Model of *PhAGL6b* spatial expression for controlling *Phalaenopsis* floral symmetry. *PhAGL6b* ectopic expression in the distal domain (petal; pink), petal converts into a lip-like structure that leads to radial symmetry. Ectopic expression in proximal domain, (sepal; blue) sepal converts into a lip-like structure that leads to bilateral symmetry[15]. The alternative processing of *PhAGL6b* transcripts produced in proximal domain (labellum; pink), labellum converts into a petal-like structure that leads to radial symmetry. *PhAGL6b* expression patterns in *Phalaenopsis* floral organs are either an expansion or a reduction across labellum. This implies that *PhAGL6b* be a key regulator to the bilateral or radially symmetrical evolvements. Pink color: 2nd whorl of the flower; blue color: 1st whorl of the flower; fan-shaped symbol: petal or petal-like structure; triangle symbol: labellum or lip-like structure; Curved symbol: sepal.

Figure 3. Predicted pathway in the regulation of spike induction in *Phalaenopsis*.
Red color indicates that the involved genes are more highly expressed in the GA biosynthesis pathway; whereas pink color of gene names indicates their differential expression in the GA response pathway. Blue colors of gene names represent the activation of flower architecture genes. Red arrows show the steps of the GA signaling stage; Pink arrows direct the steps of inflorescence evocation stage; Blue arrows reveal the steps of flower stalk initiation stage. Black arrows indicate the genes downregulated 2X over. *GA20ox*, *GA3ox*, *GAMYB*, *FT*, *SOC1*, *LFY* and *AP1* are upregulated 2X over.

Supplementary files

SUPPLEMENTARY INFORMATION APPENDIX

Dataset 1-14

Dataset 13

Dataset 15

Dataset 16

Table 1 Statistics of the *Phalaenopsis* draft genome

Estimate of genome size	**3.45 Gb**
Chromosome number (2n)	**38**
Total size of assembled contigs	**3.1 Gb**
Number of contigs (≥1kp)	**630,316**
Largest contig	**50,944**
N50 length (contig)	**1,489**
Number of scaffolds (≥1kp)	**149,151**
Total size of assembled scaffolds	**3,104,268,398**
N50 length (scaffolds)	**100,943**
Longest scaffold	**1,402,447**
GC content	**30.7**
Number of gene models	**41,153**
Mean coding sequence length	**1,014 bp**
Mean exon length/ number	**264 bp / 3.83**
Mean intron length/ number	**3,099 bp / 2.83**
Exon GC (%)	**41.9**
Intron GC (%)	**16.1**
Number of predicted miRNA genes	**650**
Total size of transposable elements	**1,598,926,178**

REFERENCES

2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* 408:796-815. 10.1038/35048692

2005. The map-based sequence of the rice genome. *Nature* 436:793-800. 10.1038/nature03895

2008. The Gene Ontology project in 2008. *Nucleic Acids Res* 36:D440-444. 10.1093/nar/gkm883

An FM, and Chan MT. 2012. Transcriptome-wide characterization of miRNA-directed and non-miRNA-directed endonucleolytic cleavage using Degradome analysis under low ambient temperature in Phalaenopsis aphrodite subsp. formosana. *Plant Cell Physiol* 53:1737-1750. 10.1093/pcp/pcs118

An FM, Hsiao SR, and Chan MT. 2011. Sequencing-based approaches reveal low ambient temperature-responsive and tissue-specific microRNAs in phalaenopsis orchid. *PLoS One* 6:e18937. 10.1371/journal.pone.0018937

Brown K, Moreton J, Malla S, Aboobaker AA, Emes RD, and Tarlinton RE. 2012. Characterisation of retroviruses in the horse genome and their transcriptional activity via transcriptome sequencing. *Virology* 433:55-63. 10.1016/j.virol.2012.07.010

Cai J, Liu X, Vanneste K, Proost S, Tsai WC, Liu KW, Chen LJ, He Y, Xu Q, Bian C, Zheng Z, Sun F, Liu W, Hsiao YY, Pan ZJ, Hsu CC, Yang YP, Hsu YC, Chuang YC, Dievart A, Dufayard JF, Xu X, Wang JY, Wang J, Xiao XJ, Zhao XM, Du R, Zhang GQ, Wang M, Su YY, Xie GC, Liu GH, Li LQ, Huang LQ, Luo YB, Chen HH, Van de Peer Y, and Liu ZJ. 2015. The genome sequence of the orchid Phalaenopsis equestris. *Nat Genet* 47:65-72. 10.1038/ng.3149

Chang YY, Kao NH, Li JY, Hsu WH, Liang YL, Wu JW, and Yang CH. 2010. Characterization of the possible roles for B class MADS box genes in regulation of perianth formation in orchid. *Plant Physiol* 152:837-853. 10.1104/pp.109.147116

Chen WS, Liu HY, Liu ZH, Yang L, and Chen WH. 1994. Geibberllin and temperature influence carbohydrate content and flowering in Phalaenopsis. *Physiologia Plantarum* 90:391-395. 10.1111/j.1399-3054.1994.tb00404.x

Chen WH, Tseng YC, Liu YC, Chuo CM, Chen PT, Tseng KM, Yeh YC, Ger MJ, and Wang HL. 2008. Cool-night temperature induces spike emergence and affects photosynthetic efficiency and metabolizable carbohydrate and organic acid pools in Phalaenopsis aphrodite. *Plant Cell Rep* 27:1667-1675. 10.1007/s00299-008-0591-0

Christenson EA. 2001. *Phalaenopsis*: a monograph. Portland Oregon: Timber Press.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, and Punta M. 2014. Pfam: the protein families database. *Nucleic Acids Res* 42:D222-230. 10.1093/nar/gkt1223

Fornara F, de Montaigu A, and Coupland G. 2010. SnapShot: Control of flowering in Arabidopsis. *Cell* 141:550, 550 e551-552. 10.1016/j.cell.2010.04.024

Geuten K, Becker A, Kaufmann K, Caris P, Janssens S, Viaene T, Theissen G, and Smets E. 2006. Petaloidy and petal identity MADS-box genes in the balsaminoid genera Impatiens and Marcgravia. *Plant J* 47:501-518. 10.1111/j.1365-313X.2006.02800.x

Gramzow L, and Theissen G. 2010. A hitchhiker's guide to the MADS world of plants. *Genome Biol* 11:214. 10.1186/gb-2010-11-6-214

Honma T, and Goto K. 2001. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409:525-529. 10.1038/35054083

Hsiao YY, Chen YW, Huang SC, Pan ZJ, Fu CH, Chen WH, Tsai WC, and Chen HH. 2011. Gene discovery using next-generation pyrosequencing to develop ESTs for Phalaenopsis orchids. *BMC Genomics* 12:360. 10.1186/1471-2164-12-360

Hsu H-F, Hsu W-H, Lee Y-I, Mao W-T, Yang J-Y, Li J-Y, and Yang C-H. 2015. Model for perianth formation in orchids. *Nature Plants* 1. 10.1038/nplants.2015.46
http://www.nature.com/articles/nplants201546#supplementary-information

Huang JZ, Lin CP, Cheng TC, Chang BC, Cheng SY, Chen YW, Lee CY, Chin SW, and Chen FC. 2015. A de novo floral transcriptome reveals clues into phalaenopsis orchid flower development. *PLoS One* 10:e0123474. 10.1371/journal.pone.0123474

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyere C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pe ME, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quetier F, and Wincker P. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449:463-467. 10.1038/nature06148

Kanehisa M, Goto S, Sato Y, Furumichi M, and Tanabe M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40:D109-114. 10.1093/nar/gkr988

Li L, Stoeckert CJ, Jr., and Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178-2189. 10.1101/gr.1224503

Lin S, Lee HC, Chen WH, Chen CC, Kao YY, Fu YM, Chen YH, Lin TY. 2001. Nuclear DNA contents of *Phalaenopsis* sp. and *Doritis pulcherrima*. *J Amer Soc Hort Sc.i* **126:** 195-199.

Mondragon-Palomino M, and Theissen G. 2011. Conserved differential expression of paralogous DEFICIENS- and GLOBOSA-like MADS-box genes in the flowers of Orchidaceae:

refining the 'orchid code'. *Plant J* 66:1008-1019. 10.1111/j.1365-313X.2011.04560.x

Mouradov A, Cremer F, and Coupland G. 2002. Control of flowering time: interacting pathways as a basis for diversity. *Plant Cell* 14 Suppl:S111-130.

Mutasa-Gottgens E, and Hedden P. 2009. Gibberellin as a factor in floral regulatory networks. *J Exp Bot* 60:1979-1989. 10.1093/jxb/erp040

Rudall PJ, and Bateman RM. 2002. Roles of synorganisation, zygomorphy and heterotopy in floral evolution: the gynostemium and labellum of orchids and other lilioid monocots. *Biol Rev Camb Philos Soc* 77:403-441.

Su CL, Chao YT, Alex Chang YC, Chen WC, Chen CY, Lee AY, Hwa KT, and Shih MC. 2011. De novo assembly of expressed transcripts and global analysis of the Phalaenopsis aphrodite transcriptome. *Plant Cell Physiol* 52:1501-1514. 10.1093/pcp/pcr097

Su CL, Chen WC, Lee AY, Chen CY, Chang YC, Chao YT, and Shih MC. 2013. A modified ABCDE model of flowering in orchids based on gene expression profiling studies of the moth orchid Phalaenopsis aphrodite. *PLoS One* 8:e80462. 10.1371/journal.pone.0080462

Trapnell C, Pachter L, and Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25:1105-1111. 10.1093/bioinformatics/btp120

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, and Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7:562-578. 10.1038/nprot.2012.016

Tsai WC, Fu CH, Hsiao YY, Huang YM, Chen LJ, Wang M, Liu ZJ, and Chen HH. 2013. OrchidBase 2.0: comprehensive collection of Orchidaceae floral transcriptomes. *Plant Cell Physiol* 54:e7. 10.1093/pcp/pcs187

Tsai WC, Kuoh CS, Chuang MH, Chen WH, and Chen HH. 2004. Four DEF-like MADS box genes displayed distinct floral morphogenetic roles in Phalaenopsis orchid. *Plant Cell Physiol* 45:831-844. 10.1093/pcp/pch095

Wang YT. 1995. Phalaenopsis Orchid Light Requirement during the Induction of Spiking. *HortScience* 30:59-61.

Zerbino DR, and Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821-829. 10.1101/gr.074492.107

Figure 1(on next page)

Figure 1

Figure 1. Venn diagram showing unique and shared gene families between and among *Phalaenopsis*, *Oryza*, *Arabidopsis* and *Vitis*.

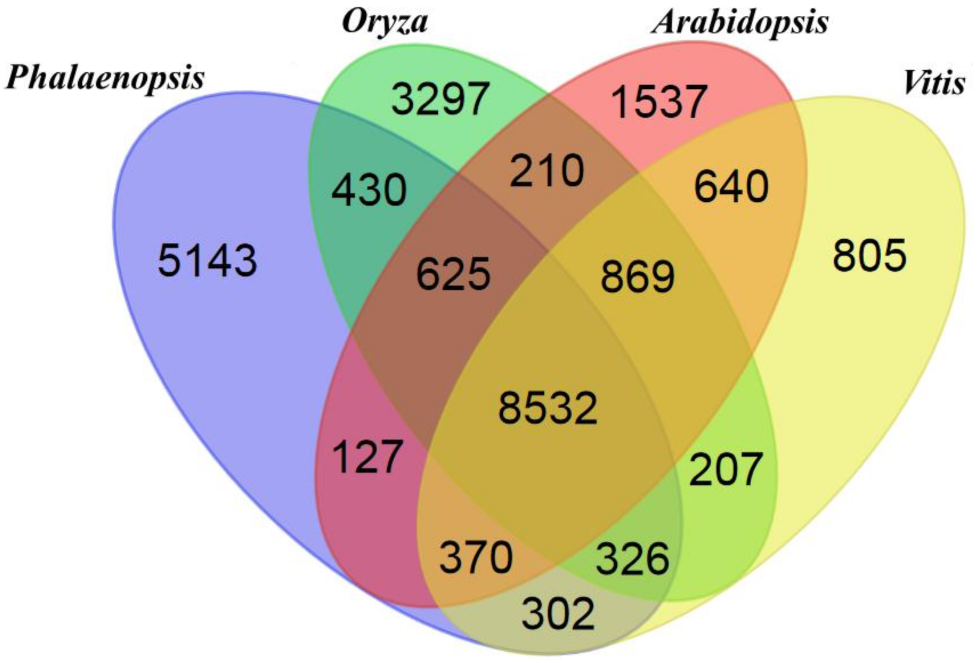

Oryza
Arabidopsis
Phalaenopsis
Vitis
3297
1537
430
210
640
5143
625
869
805
8532
127
207
370
326
302

Figure 2(on next page)

Figure 2

Figure 2. Possible evolutionary relationship of *PhAGL6b* in the regulation of lip formation in *Phalaenopsis* orchid. (a) Wild-type flower. (b) A big lip mutant of *Phalaenopsis* World Class ‘Big Foot’. (c) Representative RT-PCR result showing the mRNA splicing pattern of *PhAGL6b* in wild-type (W) and big lip mutant (M). (d) Alignment of the amino acid sequences of alternatively spliced forms of *PhAGL6b*. (e)Model of*PhAGL6b* spatial expression for controlling *Phalaenopsis* floral symmetry. *PhAGL6b* ectopic expression in the distal domain (petal; pink), petal converts into a lip-like structure that leads to radial symmetry. Ectopic expression in proximal domain, (sepal; blue) sepal converts into a lip-like structure that leads to bilateral symmetry[15]. The alternative processing of *PhAGL6b* transcripts produced in proximal domain (labellum; pink), labellum converts into a petal-like structure that leads to radial symmetry. *PhAGL6b* expression patterns in *Phalaenopsis* floral organs are either an expansion or a reduction across labellum. This implies that *PhAGL6b* be a key regulator to the bilateral or radially symmetrical evolvements. Pink color: 2nd whorl of the flower; blue color: 1st whorl of the flower; fan-shaped symbol: petal or petal-like structure; triangle symbol: labellum or lip-like structure; Curved symbol: sepal.

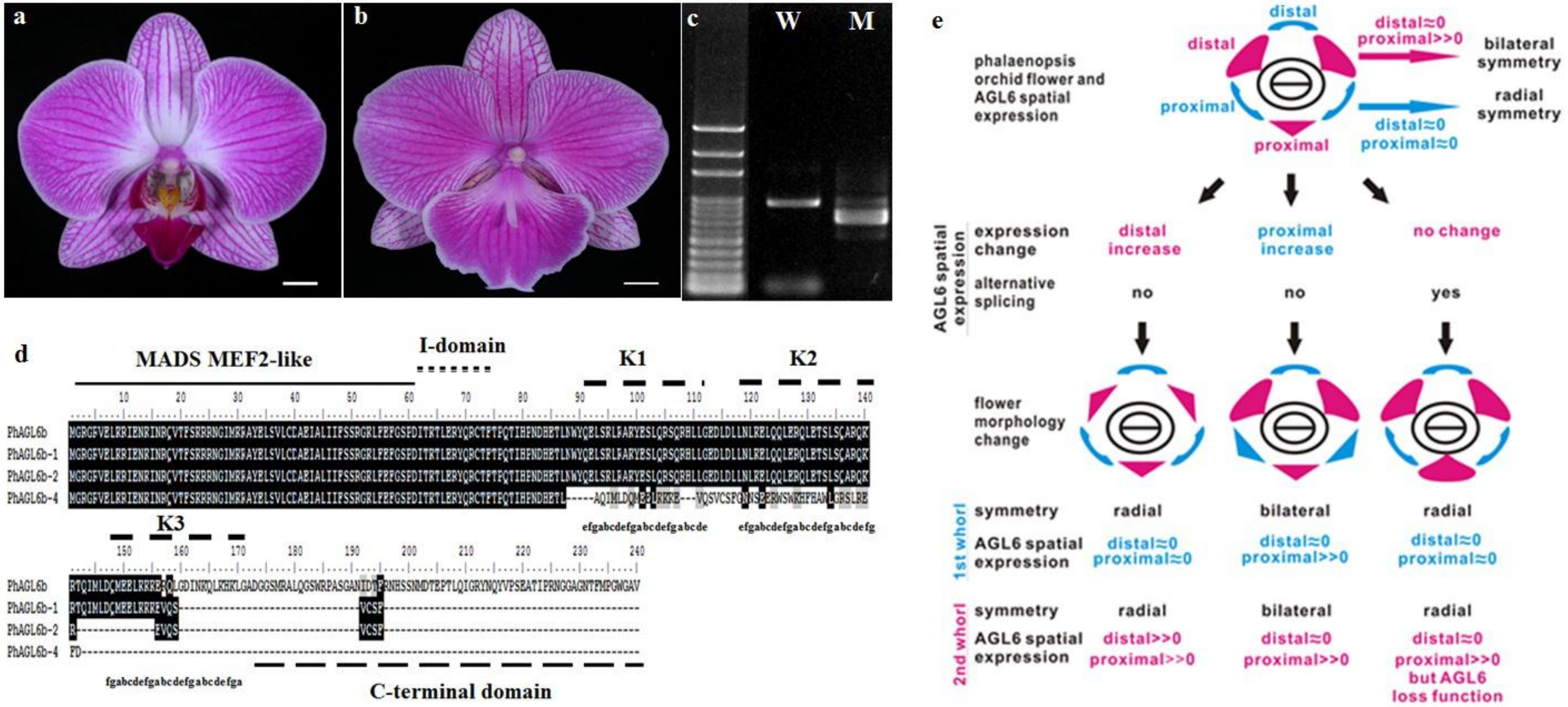
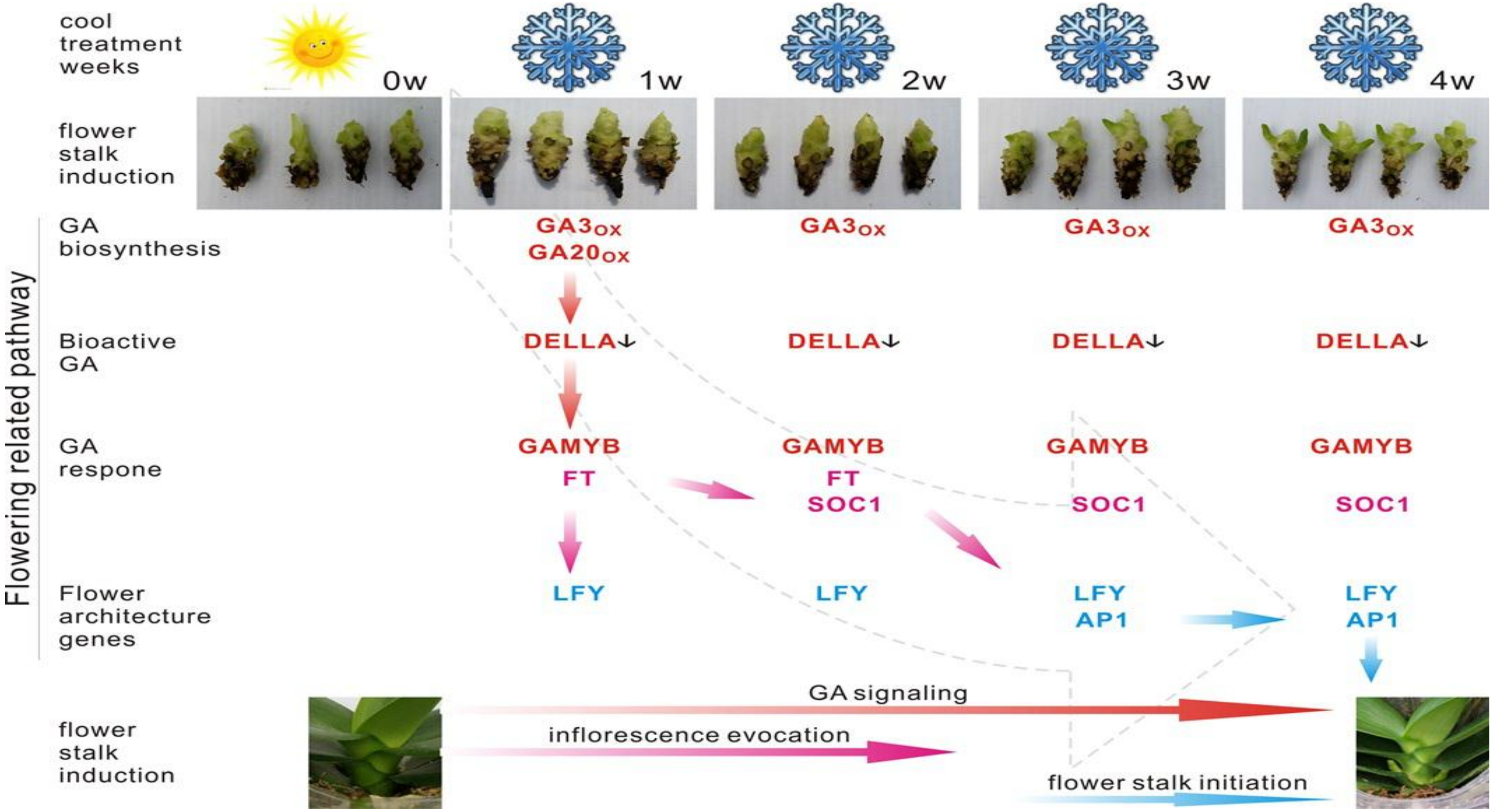
a
b
c
W
M
d
MADS MEF2-like
I-domain
K1
K2
10 20 30 40 50 60 70 80 90 100 110 120 130 140
PhAGL6b MGRGFVELRRIENKINRQVTFSKRRNGIMKKAYELSVLCDAEIALIIFSSRGRLFEFGSPDITKTLERYQRCTFTPQTIHPNDHETLNWYQELSRLKARYESLQRSQRHLLGEDLDLLNLKELQQLERQLETSLSQARQK
PhAGL6b-1 MGRGFVELRRIENKINRQVTFSKRRNGIMKKAYELSVLCDAEIALIIFSSRGRLFEFGSPDITKTLERYQRCTFTPQTIHPNDHETLNWYQELSRLKARYESLQRSQRHLLGEDLDLLNLKELQQLERQLETSLSQARQK
PhAGL6b-2 MGRGFVELRRIENKINRQVTFSKRRNGIMKKAYELSVLCDAEIALIIFSSRGRLFEFGSPDITKTLERYQRCTFTPQTIHPNDHETLNWYQELSRLKARYESLQRSQRHLLGEDLDLLNLKELQQLERQLETSLSQARQK
PhAGL6b-4 MGRGFVELRRIENKINRQVTFSKRRNGIMKKAYELSVLCDAEIALIIFSSRGRLFEFGSPDITKTLERYQRCTFTPQTIHPNDHETL-----AQIMLDQMEPLRKKE---VQSVCSPGNNSEERWSWKHPHAWLGRSLRE
efgabcdefgabcdefgabcde
efgabcdefgabcdefgabcdefg
K3
150 160 170 180 190 200 210 220 230 240
PhAGL6b RTQIMLDQMEELRKKEFQLGDINKQLKHKLGADGGSMRALQGSWRPASGANIDTERNHSSNMDTEPTLQIGRYNQYVPSEATIPRNGGAGNTFMPGWGAV
PhAGL6b-1 RTQIMLDQMEELRKKEVQS-----------------------------------VCSP--
PhAGL6b-2 R--------------EVQS-----------------------------------VCSP--
PhAGL6b-4 FD--
fgabcdefgabcdefgabcdefga
C-terminal domain
e
phalaenopsis orchid flower and AGL6 spatial expression
distal
distal
proximal
proximal
distal≈0
proximal>>0
bilateral symmetry
radial symmetry
distal≈0
proximal≈0
AGL6 spatial expression
expression change
distal increase
proximal increase
no change
alternative splicing
no
no
yes
flower morphology change
1st whorl
symmetry
radial
bilateral
radial
AGL6 spatial expression
distal≈0
proximal≈0
distal≈0
proximal>>0
distal≈0
proximal≈0
2nd whorl
symmetry
radial
bilateral
radial
AGL6 spatial expression
distal>>0
proximal>>0
distal≈0
proximal>>0
distal≈0
proximal>>0
but AGL6
loss function

Figure 3(on next page)

Figure 3

Figure 3. Predicted pathway in the regulation of spikeinduction in *Phalaenopsis*. Red color indicates that the involved genes are more highly expressed in the GA biosynthesis pathway; whereas pink color of gene names indicates their differential expression in the GA response pathway. Blue colors of gene names represent the activation of flower architecturegenes. Red arrows show the steps ofthe GA signaling stage; Pink arrows direct the steps of inflorescence evocation stage; Blue arrowsrevealthe steps offlower stalk initiation stage. Black arrows indicate the genes downregulated 2X over.*GA20ox*,*GA3ox*,*GAMYB*,*FT*,*SOC1*, *LFY*and*AP1*are upregulated 2X over.


cool treatment weeks
0w
1w
2w
3w
4w
flower stalk induction
Flowering related pathway
GA biosynthesis
GA3ox
GA20ox
GA3ox
GA3ox
GA3ox
Bioactive GA
DELLA↓
DELLA↓
DELLA↓
DELLA↓
GA respone
GAMYB
FT
GAMYB
FT
SOC1
GAMYB
SOC1
GAMYB
SOC1
Flower architecture genes
LFY
LFY
LFY
AP1
LFY
AP1
flower stalk induction
GA signaling
inflorescence evocation
flower stalk initiation