Orchid biotechnology and breeding

Hong-Hwa CHEN[1,2,3]

[1] Orchid Research and Development Center, National Cheng Kung University, Tainan 701, Taiwan
[2] Department of Life Sciences, National Cheng Kung University, Tainan 701, Taiwan
[3] Institute of Tropical Plant Sciences, National Cheng Kung University, Tainan 701, Taiwan
hhchen@mail.ncku.edu.tw

Abstract – Containing more than 25,000 species, the Orchidaceae family, is one of the largest angiosperm families. The genus *Phalaenopsis*, a beautiful and popular orchid, comprises approximately 66 species. The nuclear DNA contents and karyotypes analysis from 18 *Phalaenospsis* species have been estimated by using flow cytometry and cytogenetic analysis, respectively. OrchidBase 3.0 collects the transcriptomics data of ten species distributed in the five subfamilies as well as the whole genome sequence of the tropical epiphytic crassulacean acid metabolism (CAM) orchid, *P. equestris*, a frequently used parent species for orchid breeding. *P. bellina* is a scented orchid emitting large amount of monoterpenes. GERANYL DIPHOSPHATE SYNTHASE (PbGDPS) is the key enzyme for monoterpene biosynthesis. A dual repeat in the upstream promoter fragments of *GDPS* is essential for its transcriptional activation in *Phalaenopsis* orchids. The full dual repeat was present only in the scented *Phalaenopsis* orchids, and its integrity showed strong association with the transactivation by a basic leucine zipper (bZIP) TF. As this dual repeat was close related to the monoterpene biosynthesis in *Phalaenopsis* orchids, it could be developed as a promising molecular marker for early detection of monoterpene phenotype in the offspring and thus facilitate scented orchid breeding. In addition, we have performed genotyping-by-sequence for *Phalaenopsis* genotyping from the cross between *P. aphrodite* ssp. *formosana* and *P. equestris* and their 118 F1 progenies, and set up the bioinformatics system for the marker-assisted selection on molecular breeding and gene identification.

Keywords: breeding, bZIP, dual repeat, orchid biotechnology, PbGDPS, *Phalaenopsis*

With an estimated more than 25 000 species, orchids are the most species-rich of all angiosperm families. They show a wide diversity of epiphytic and terrestrial growth forms and have successfully colonized almost every habitat on earth. The most recent common ancestor of extant orchids lived in the late Cretaceous (76-84 Mya) as dated by a fossil orchid and its pollinator (Ramirez *et al*., 2007). The radiation of the orchid family has probably took place in a comparatively short period as compared with that of most flowering plant families, which suggests that their speciation rates are presumed to be exceptionally high (Gill, 1989).

Associated with the enormous number of Orchidaceae species is extraordinary floral diversification. Orchids are renowned for an abundance of kinds, with a seemingly unending array of strange and often fantastic variations, and represent a highly advanced and terminal line of floral evolution in the monocotyledons. This spectacular diversification has been linked to the specific interaction between the orchid flower and pollinator (Cozzolino and Widmer, 2005), sequential and rapid interplay between drift and natural selection (Tremblay *et al*., 2005), the role of obligate orchid-mycorrhizal interactions (Otero and Flanagan, 2006), and Crassulacean acid metabolism and epiphytism (Silvera *et al*., 2009). In addition to their prosperity of ecological manipulations, orchids have several unique reproductive strategies that contribute to their success. These include mature pollen grains packaged as pollinia, pollination-regulated ovary/ovule development, synchronized timing of micro- and mega-gametogenesis for effective fertilization, and the release of thousands or millions of

immature embryos (seeds without endosperm) in mature pods (Yu and Goh, 2001).

Because of the thriving and prosperous orchid breeding and industry, plant scientists in Taiwan are well placed to study orchid biology and develop orchid biotechnology to apply to the orchid industry.

A better understanding of the karyotypes and DNA contents of orchid will aid in the development of new cultivars of orchids. All *Phalaenopsis* species have the same chromosome number ($2n = 2x = 38$), but their genomes vary considerably in size. Analysis of karyotypes of 9 *Phalaenopsis* species and *Doritis pulcherrima* by Feulgen- and DAPI-stained somatic metaphase chromosomes from root tips revealed that *P. aphrodite*, *P. stuartiana*, *P. equestris*, *P. cornu-cervi*, and *P. lueddemanniana* are with small and uniform chromosomes (1-2.5 μm), and all are metacentric or submetacentric. *P. venosa*, *P. amboinensis*, and *P. violacea* have bimodal karyotypes, with large and small chromosomes, and most are subtelocentric or acrocentric (Kao *et al*., 2001). Flow cytometry has proven to be an efficient and reliable method for analyzing plant genomes. The nuclear DNA contents from 18 *Phalaenospsis* species and *P. pulcherrima* are estimated by flow cytometry; 2C values ranged from 2.74 pg for *P. sanderiana* to 16.61 pg for *P. parishii* (Lin *et al*., 2001).

The OrchidBase collects the transcriptome sequences from *Phalaenopsis* cDNA libraries and assembled into 84 617 non-redundant transcribed sequences (including 8 501 contigs and 76 116 singletons) (Fu *et al*., 2011). The OrchidBase contains the transcriptome sequences derived from 11 *Phalaenopsis* orchid cDNA libraries, which are constructed from different species, including *P. aphrodite* subsp. *formosana*, *P. equestris* and *P. bellina*, and from different tissues, including developing seed, protocorm, vegetative tissue, leaf, cold-treated plantlet, pathogen-treated plantlet, inflorescence, and flower buds (Fu *et al*., 2011). The transcriptomics data collected in OrchidBase 2.0 are obtained from 10 orchid species within 5 subfamilies through both deep sequencing with ABI 3730 and NGS Roche 454 and Illumina/Solexa. Recently, the whole genome sequences are available for *P. equestris*, *Dendrobium catenatum*, and *Apostacea shengenica*, and included in the OrchidBase 3.0. The OrchidBase is freely available at http://orchidbase.itps.ncku.edu.tw and provides researchers with a high-quality genetic resource for data mining and efficient experimental studies of orchid biology and biotechnology.

The global flower industry thrives on novelty. Domestication of wild species in conjunction with traditional breeding has long been the principle path for generation of novel flowers in the industry. For orchid, traits such as flower color, shape and fragrance are primary novel markers because they are key determinants of consumer choice. However, many modern floricultural varieties have lost their scent with traditional breeding programs. Breeders of orchids in cut-flower and ornamental markets have focused on producing plants with improved vase life, shipping characteristics and visual aesthetic values (*i.e*., color and shape).

The growing cycles of *Phalaenopsis* orchids are 2-3 years. Using traditional hybridization to transmit useful traits into commercial varieties is a long process that will take years to achieve (Arditti, 1992). In addition, intraspecific and/or interspecific incompatibility limits the work of variety improvement. All 5 subgenuses of *Phalaenopsis* have the same chromosome number (2n=2x=38) that can be divided into small, medium and large chromosome groups, according to chromosome sizes and nuclear DNA contents (Kao *et al*., 2001, Lin *et al*., 2001). Most commercial cultivars are derived from species with small chromosomes, such as *P. amabilis*, *P. aphrodite*, *P. stuartiana*, *P. schilleriana*, and *P. equestris*. The species with strong scents have large chromosomes including *P. amboinnensis*, *P. bellina*, *P. venosa* and *P. violacea*. Successful crosses between species with small and with large chromosomes are difficult because of inter-specific incompatibility.

P. bellina, classified in the subgenus *Polychilos*, is native to Malaysia, and numerous commercial varieties have been bred because of the orchid's pleasant fragrance. In addition, the species has some native tetraploid species to breed scented commercial *Phalaenopsis* orchids and therefore is an important parent for breeding scented cultivars. Floral scent is a composite characteristic determined by a complex mixture of low molecular mass volatiles molecules and dominated by monoterpenoid,

sesquiterpenoid, phenylpropanoid, benzenoid compounds and fatty acid derivatives. The floral scents in *P. bellina* are rich in monoterpenes, geraniol and linalool and their derivatives (Hsiao *et al*., 2006). They include geraniol, nerol, 2,6-dimethyl-octa-3,7-diene-2,6-diol, 2,6-dimethyl-octa-1,7-diene-3,6-diol, 3,7-dimethyl-2,6-octadienal, geranic acid and 2,6-dimethyl-octa-2,6-diene-1,8-diol. In contrast, no monoterpenoid derivatives were emitted in scentless *P. equestris* flowers; fatty acid derivatives, phenylpropanoids, and benzenoids were the major volatiles. These compounds are barely detectable by the human nose.

Research into plant scents has been hampered mainly by the invisibility of this character, its dynamic nature, and complex mixtures of components that are present in very small quantities. Combining chemical analysis, genomics and bioinformatics, we have uncovered the scent biosynthesis pathway and the relevant genes in *P. bellina* flower. These include a monoterpene biosynthesis pathway of 15 steps in the *P. bellina* flower leading from glyceraldehyde-3-phosphate to monoterpenoids of geraniol, linalool and their derivatives (Figure 1, Hsiao *et al*., 2006).

Terpenoids belong to a large family of plant secondary metabolites, and their corresponding alcohols possess useful properties such as fragrance, flavour, insecticidal properties and characteristics that make them useful as pharmaceutical agents. All monoterpenes are derived from the same substrate, geranyl diphosphate (GDP, C_{10}), which is catalyzed by GDPS, a member of the short-chain trans-prenyltransferase family, via the condensation of dimethylallyl diphosphate with isopentenyl diphosphate. GDPS (Tholl *et al*., 2004) is differentially expressed in the scented species.

The full-length cDNA of *P. bellina* GDPS (PbGDPS) was isolated from a *P. bellina* floral cDNA library (Hsiao *et al*., 2006) and sequenced. *PbGDPS* was predominantly expressed in the scented species *P. bellina* and its scented offspring D. Kenneth Schubert 'Five'. In addition, its expression level was associated with the amount of scent emitted in the scented species, suggesting that *PbGDPS* play a key role in the regulation of scent production in *P. bellina* flowers (Hsiao *et al*., 2008).

Comparing the promoter sequence of from scent species *P. bellina* and scentless species *P. aphrodite*, we found that there is a dual repeat consisted of two 75 bp units present in the scent *Phalaenopsis* species. We hypothesized that the dual repeat is associated with the monoterpene production. To confirm this, another 10 *Phalaenopsis* orchids frequently used as breeding parents (Figure 2) were assessed for the correlation analysis between the dual repeat and the monoterpene production. The presence of the GDPS gene and its promoter sequence in the 12 *Phalaenopsis* orchids were then analyzed. Intriguingly, the GDPS gene was all present in these orchids regardless of their scent or scentless phenotype. It is plausible that the defects are resided in the promoter region. We then amplified the dual repeat and found its fragment length polymorphism among the 12 *Phalaenopsis* orchids. The four scented orchids with monoterpene production contain the complete dual repeat (Figure 3, the black arrowheads). In contrast, the length of the amplified fragments of the other orchids were reduced to various extents due to deletion in the dual repeat region (Chuang *et al*., 2018).

Furthermore, the dual repeat was used to screen the transcription factor library and identified the PbbZIP4. We found that PbbZIP4 was able to distinguish whether the promoter containing the dual repeat. There are two promoter fragments of *PaGDPS* in *P. aphrodite*, namely PaGDPSpA and PaGDPSpB. In the presence of PbbZIP4, it enhanced the promoter activities of both PbGDPSp and PaGDPSpA which have complete dual repats, but showed no effects on PaGDPSpB which does not have the complete dual repeat. Collectively, these results indicated that the upstream activator, bZIP4, as well as the dual repeat of *GDPS* promoter are crucial for monoterpene production in *Phalaenopsis* orchids (Chuang *et al*., 2018).

We performed genotyping-by-sequence (GBS) approach by using restriction-site associated DNA sequencing (RAD-Seq) to construct a high-density genetic map for for *P. aphrodite* ssp. *formosana* x *P. equestris* and their 118 F1 progenies. The traditional SSR marker-assisted genetic map and the reference whole genome sequence of *P. equestris* have been improved to assemble into 19 linkage groups to assist SNP mapping. Totally, 2 819 SNPs were mapped to the 19 linkage groups derived from the *Phalaenopsis* genome with the averaged 311 kp sequence containing a

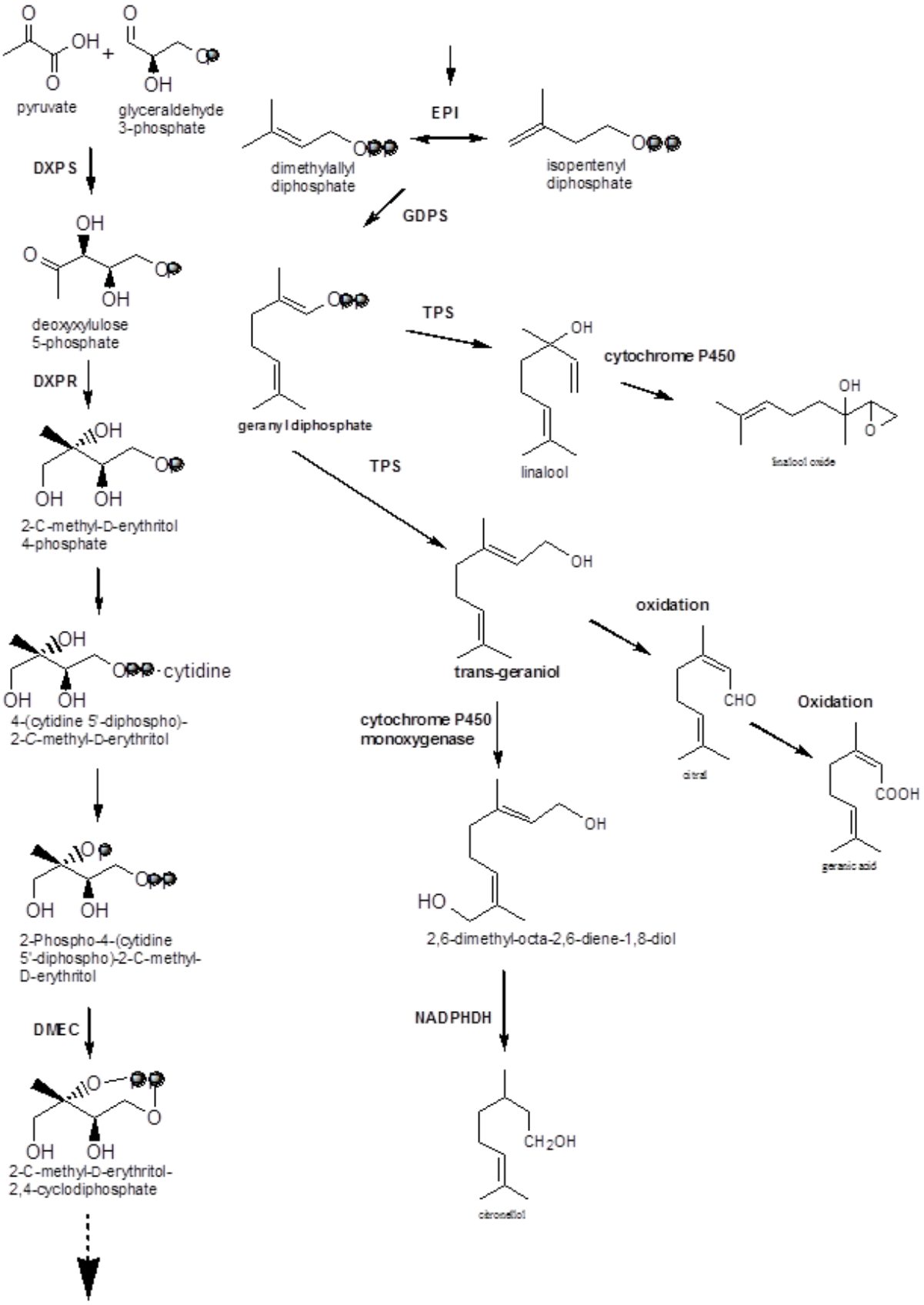

Figure 1. Putative metabolic pathway from pyruvate and glyceraldehyde-3-phosphate to scent synthesis and related enzymes in *P. bellina*. DXPS: deoxyxylulose-5-phosphate synthase; DXPR: deoxyxylulose-5-phosphate reductoisomerase; DMEC: 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate cyclase; EPI: epimerase; GDPS: geranyl diphosphate synthase; NADPHDH: NADPH dehydrogenase (Adopted from Hsiao *et al*., 2006).

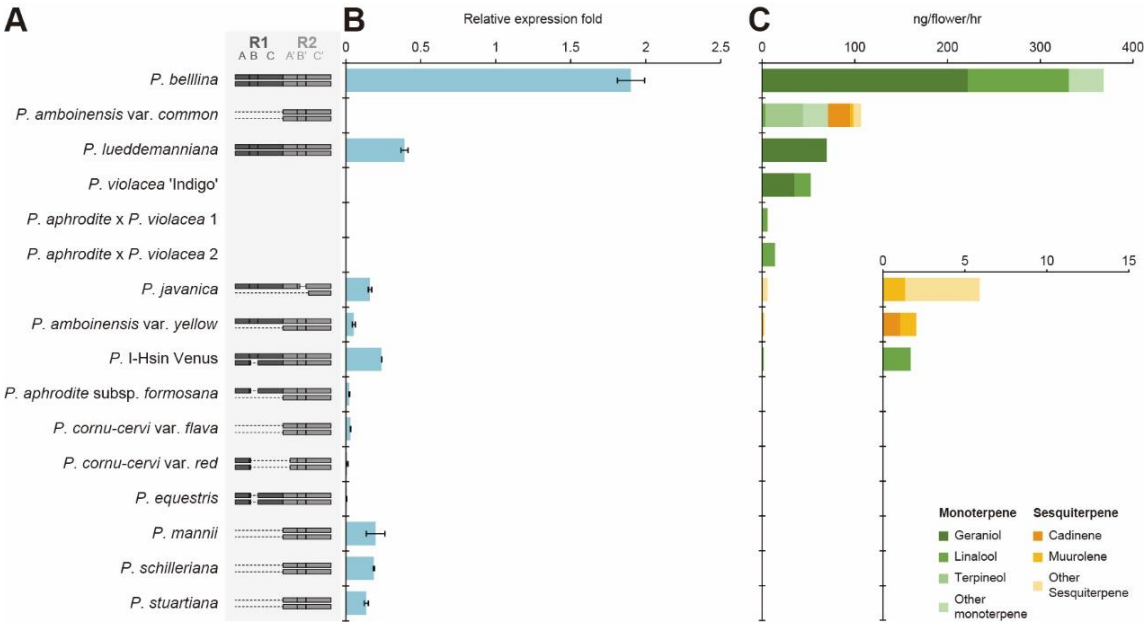

Figure 2. The 12 *Phalaenopsis* orchids used in this study. The order of the figures was followed the presentation in Figure 2. **(A)** *P.* Meidarland Bellina Age 'LM128', **(B)** *P. bellina*, **(C)** *P. lueddemaniana*, **(D)** *P.* I-Hsin Venus, **(E)** *P. javanica*, **(F)** *P. amboinensis* var. *yellow*, **(G)** *P. mannii*, **(H)** *P. schilleriana*, **(I)** *P. aphrodite* subsp. *formosana*, **(J)** *P. cornu-cervi* var. *red*, **(K)** *P. equestris* 'RO-5', and **(L)** *P. equestris* 'WY-7'. Scale bar = 1 cm.

SNP marker. GWAS analysis was performed for these 2 819 SNPs with various agricultural traits in the F1 population of *P. aphrodite* ssp. *formosana* x *P. equestris*, including flower size, sepal size, and petal color, and the functions of these genes near these SNPs were further confirmed. Therefore, we have developed a draft of genetic map that can be helpful for the genetic study in *Phalaenopsis* and benefit for orchid breeding (Chiou *et al*., unpublished data).

Funding

This work was supported by Ministry of Science and Technology, Taiwan. [Grant no. 105-2313-B-006 -002 -MY3 and 106-2321-B-006 -009 to H.H.C].

References

Arditti J. 1992. Physiology. *In*: *Fundamentals of orchid biology*. J. Arditti. (Ed.), Wiley, New York. pp. 159-181.

Chuang Y.C., Hung Y.C., Hsu C.Y., Yeh C.M., Mitsuda N., Ohme-Takagi M., Tsai W.C., Chen W.H., Chen H.H. 2018. A dual repeat *cis*-element determines expression of *GERANYL DIPHOSPHATE* SYNTHASE for monoterpene production in *Phalaenopsis* orchids. *Front. Plant Sci.*, 9: 765.

Cozzolino S., Widmer A. 2005. Orchid diversity: an evolutionary consequence of deception? *Trends Ecol. Evol.*, 20: 487-494.

Fu C.H., Chen Y.W., Hsiao Y.Y., Pan Z.J., Liu Z.J., Huang Y.M., Tsai W.C., Chen H.H. 2011. OrchidBase: a collection of sequences of the transcriptome derived from orchids. *Plant Cell Physiol.*, 52: 238-243.

Gill D.E. 1989. Fruiting failure, pollinator inefficiency, and speciation in orchids. *Sinauer Associates, Sunderland, MA*.

Hsiao Y.Y., Jeng M.F., Tsai W.C., Chuang Y.C., Li C.Y., Wu T.S., Kuoh C.S., Chen W.H., Chen H.H. 2008. A novel homodimeric geranyl diphosphate synthase from the orchids *Phalaenopsis bellina* lacking a DD(X)2-4D motif. *Plant J.*, 55: 719-733.

Hsiao Y.Y., Tsai W.C., Kuoh C.S., Huang T.H., Wang H.C., Wu T.S., Leu Y.L., Chen W.H., Chen H.H.. 2006. Comparison of transcripts in *Phalaenopsis bellina* and *Phalaenopsis equestris* (Orchidaceae) flowers to deduce monoterpene biosynthesis pathway. *BMC Plant Biol.*, 6: 14.

Kao Y.Y., Chang S.B., Lin T.Y., Hsieh C.H., Chen Y.H., Chen W.H., Chen C.C. 2001. Differential accumulation of heterochromatin as a cause of karyotype variation in *Phalaenopsis* orchids. *Ann. Bot.*, 87: 387-395.

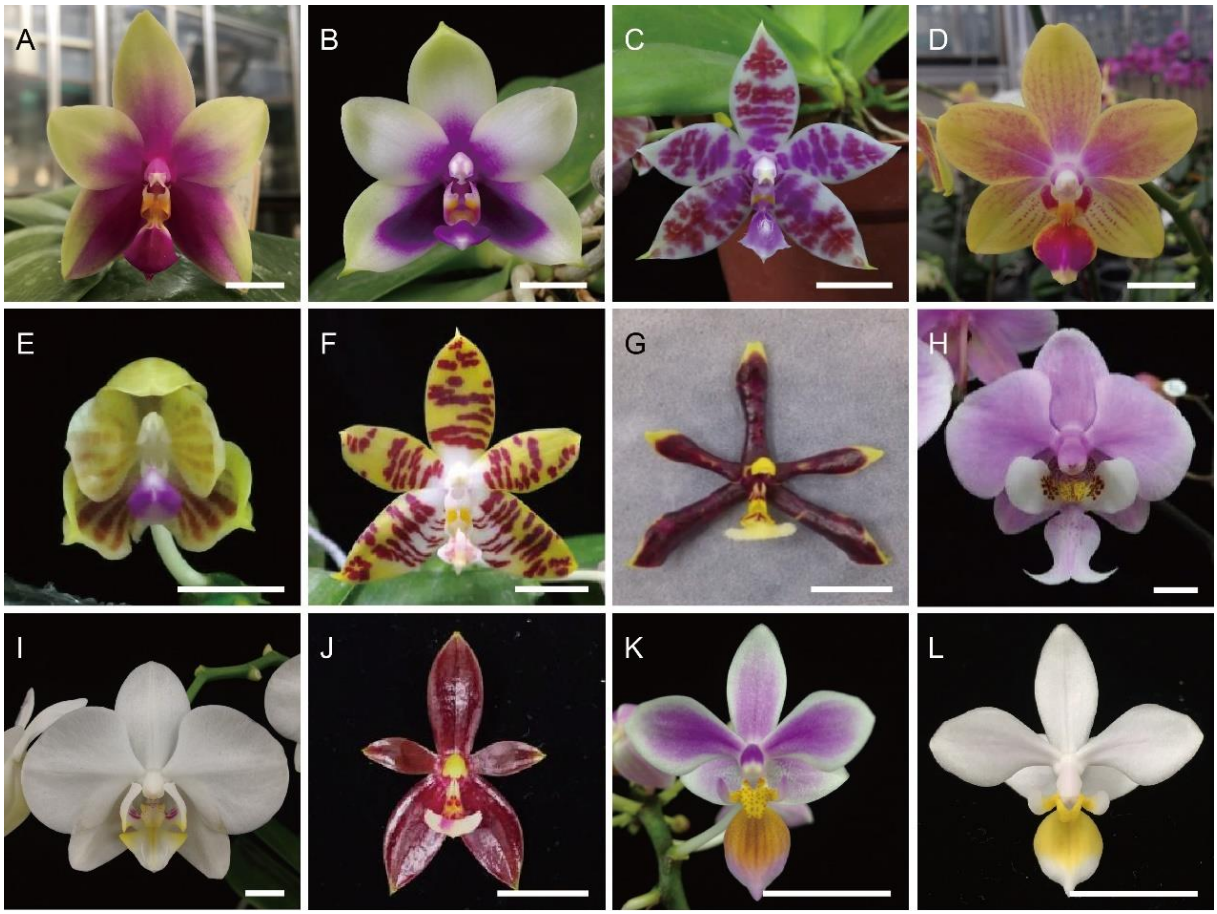

Figure 3. The analysis of *GDPS-SSU1* gene and floral volatiles in native *Phalaenopsis* orchids and hybrids. The R1R2 region structure of *GDPS-SSU1* promoter **(A)**, the expression levels of GDPS-SSU1 **(B)**, and floral volatile analysis **(C)** of native *Phalaenopsis* orchids and hybrids. Data is from single plant measurements.

Lin S., Lee H.C., Chen W.H., Chen C.C., Kao Y.Y., Fu Y.M., Chen Y.H., Lin T.Y. 2001. Nuclear DNA contents of *Phalaenopsis* species and *Doritis pulcherrima*. *J. Am. Soc. Horti. Sci.*, 126: 195-199.

Otero J.T., Flanagan N.S. 2006. Orchid diversity--beyond deception. *Trends Ecol. Evol.*, 21: 64-65.

Ramirez S.R., Gravendeel B., Singer R.B., Marshall C.R., Pierce N.E. 2007. Dating the origin of the Orchidaceae from a fossil orchid with its pollinator. *Nature*, 448: 1042-1045.

Silvera K., Santiago L.S., Cushman J.C., Winter K. 2009. Crassulacean acid metabolism and epiphytism linked to adaptive radiations in the Orchidaceae. *Plant Physiol.*, 149: 1838-1847.

Tholl D., Kish C.M., Orlova I., Sherman D., Gershenzon J., Pichersky E., Dudareva N.. 2004. Formation of monoterpenes in *Antirrhinum majus* and *Clarkia breweri* flowers involves heterodimeric geranyl diphosphate synthases. *Plant Cell*, 16: 977-992.

Tremblay R.L., Ackerman J.D., Zimmerman J.K., Calvo R.N. 2005. Variation in sexual reproduction in orchids and its evolutionary consequences: a spasmodic journey to diversification. *Biol. J. Linn. Soc.*, 84:1-54.

Yu H., Goh C.J. 2001. Molecular genetics of reproductive biology in orchids. *Plant Physiol.*, 127: 1390-1393.