

REVIEW

Open Access



Post genomics era for orchid research

Wen-Chieh Tsai^{1,2,3}, Anne Dievart^{4,5}, Chia-Chi Hsu³, Yu-Yun Hsiao^{2,3}, Shang-Yi Chiou³, Hsin Huang³ and Hong-Hwa Chen^{1,2,3*}

Abstract

Among 300,000 species in angiosperms, Orchidaceae containing 30,000 species is one of the largest families. Almost every habitats on earth have orchid plants successfully colonized, and it indicates that orchids are among the plants with significant ecological and evolutionary importance. So far, four orchid genomes have been sequenced, including *Phalaenopsis equestris*, *Dendrobium catenatum*, *Dendrobium officinale*, and *Apostasia shenggen*. Here, we review the current progress and the direction of orchid research in the post genomics era. These include the orchid genome evolution, genome mapping (genome-wide association analysis, genetic map, physical map), comparative genomics (especially receptor-like kinase and terpene synthase), secondary metabolomics, and genome editing.

Keywords: Comparative genomics, Genome editing, Genome evolution, GWAS, Orchidaceae, *Phalaenopsis*, Post genomics era, Receptor-like kinase, Secondary metabolomics, Terpene synthase

Background

Containing about 30,000 species, Orchidaceae plants account for one tenth of all angiosperms containing 300,000 species (Hsiao et al. 2011b). More than 70% of orchids are epiphytic growth with distinct physiological characteristics (Hsiao et al. 2011b). Orchids are highly adapted to unfavorable environment. They have colonized successfully every habitat on earth wherever the sun shines. According to the molecular clock revealed by whole genome sequencing of *Phalaenopsis equestris* (Cai et al. 2015), the emergence of orchids have occurred in the late Cretaceous (76 Mya), and allowed them to cross the border of Jurassic mass extinction (66 Mya). This is consistent to the observation of the date for the most ancient fossil record (76–84 Mya) (Ramirez et al. 2007). It has been speculated that the speciation rates of orchids are exceptionally high (Gill 1989), with the fact that, even now, new species of orchids are still recorded worldwide suggesting that the evolution of orchids has never ceased.

The enormous number of Orchidaceae species allows it divided into five subfamilies: Apostasioideae,

Vanilloideae, Cyripedioideae, Orchidoideae, and Epidendroideae (Fig. 1). They are all extraordinary floral diversified, and this heterogeneity has been related to the specialized interaction between the pollinators and orchid flowers (Cozzolino and Widmer 2005). The unique features for orchids include the obligate interactions between orchids and mycorrhizal fungi (Otero and Flanagan 2006), with C3 and/or crassulacean acid metabolism photosynthesis (Mascher et al. 2017), and epiphytic growth forms (Silvera et al. 2009). Orchids have exclusive reproductive strategies contributing to their successful adaptation to their ecological exploitations (Yu and Goh 2000).

The orchid cultivation and hybridization in Taiwan is very popular to the worldwide orchid market. The elegant appearance, some even with charming fragrance, and prolonged long life for orchid flowers have promoted attractiveness of *Phalaenopsis* orchids among breeders, nurseries and customers. For the past 20 years, orchid researchers have devoted to establish the foundation of orchid genomics research. These include karyotype analysis (Kao et al. 2001; Lin et al. 2001), genome size analysis (Chen et al. 2013, 2014), establishment of expressed sequence tags (ESTs) (Hsiao et al. 2006), genomics and transcriptomics databases (Fu et al. 2011; Su et al. 2011, 2013; Tsai et al. 2013), bacterial artificial chromosome

*Correspondence: hhchen@mail.ncku.edu.tw

³ Department of Life Sciences, National Cheng Kung University, Tainan 701, Taiwan

Full list of author information is available at the end of the article

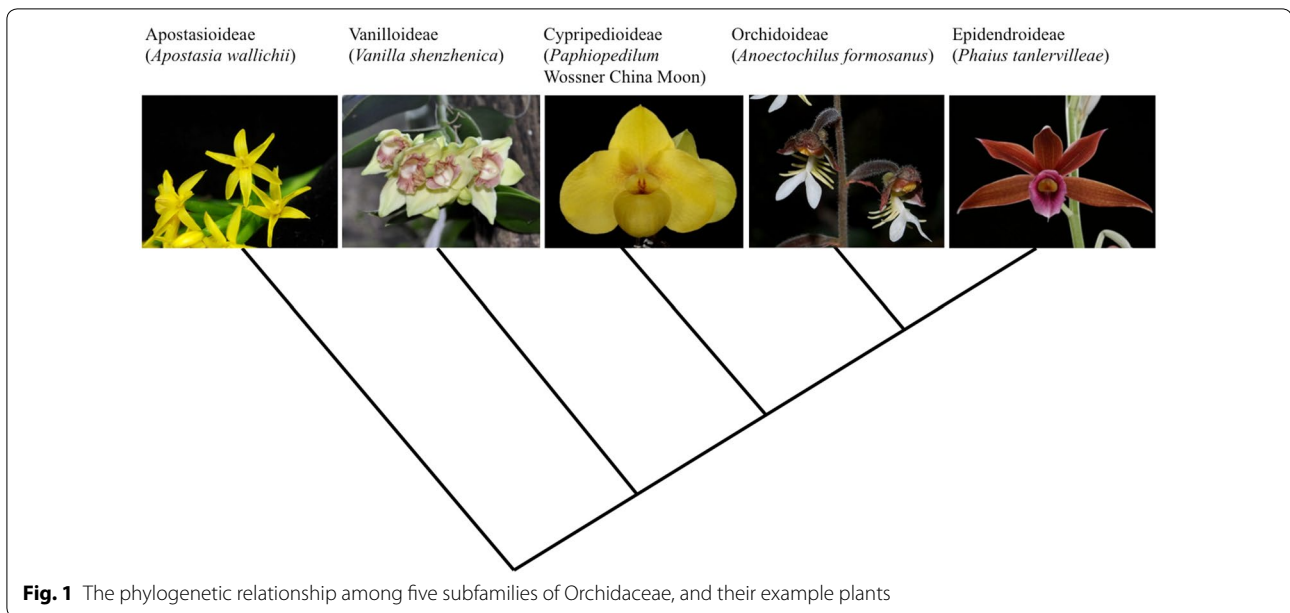


Fig. 1 The phylogenetic relationship among five subfamilies of Orchidaceae, and their example plants

(BAC) end sequencing (BES) (Hsu et al. 2011), chloroplast genome sequencing (Chang et al. 2006; Pan et al. 2012), miRNA (Lin et al. 2013a), and whole genome sequencing (Cai et al. 2015; Yan et al. 2015; Zhang et al. 2016a, 2017). In addition, orchid functional genomics studies has been available with virus-induced gene silencing using *Cymbidium* mosaic virus infectious clones for the assessing gene functions involved in flower color, floral morphogenesis and floral scent studies (Lu et al. 2007; Hsieh et al. 2013a, b). In micro-propagation, the induction of polyploidy has been developed to circumvent the hybrid incompatibility (Sattler et al. 2016).

In this article, we review the relevant orchid research progress and the future directions of the orchid investigation at the post genomics era. These include the orchid genome evolution, genome mapping, comparative genomics, secondary metabolomics, and genome editing.

Orchid genome evolution

Genome size variation

Angiosperms have variable genome sizes ranging nearly 2400-fold from *Genlisea margaretae* (Lentibulariaceae) with just 0.065 pg to the huge genome in *Paris japonica* (Melanthiaceae) with 152.23 pg. Many species with large genomes are observed in monocots, such as species in Alliaceae, Asparagaceae, Liliaceae, Melanthiaceae and Orchidaceae (Leitch et al. 2009). Among these, Orchidaceae with the genome sizes ranging 168-fold (1C = 0.33–55.4 pg) are perhaps the most diverse angiosperm families (Leitch et al. 2009).

As the species-richest subfamily, Epidendroideae with genome contents ranging over 60-fold (1C = 0.3–19.8 pg)

harbor the most variable genome size in Orchidaceae. Orchidoideae, where the largest descending/offspringing from species in subtribe Orchidinae, are pictured by a more restricted range of genomes (1C = 2.9–16.4 pg) varying not more than sixfold. Cypridioideae show genome sizes ranging only tenfold (1C = 4.1–43.1 pg). Cypridioideae contain the largest mean genome size (1C = 25.8 pg) among all the subfamilies. Only genome size of few species in Vanilloideae was estimated, ranging from 1C = 7.3 to 55.4 pg. In this subfamily, *Pogonia ophioglossoides* presents the largest genome size (1C = 55.4 pg) (Leitch et al. 2009). Apostasioideae, the primitive subfamilies, contain calculated 1C-values ranging from 0.38 pg in *Apostasia nuda* to 5.96 pg in *Neuwiedia zollingeri* var. *javanica*, a close to 16-fold range (Jersáková et al. 2013). *P. equestris* and *P. aphrodite* subsp. *formosana*, the two native *Phalaenopsis* species usually be used as parents for breeding in Taiwan, respectively have a relative small genome size of 1.6 and 1.4 pg/1C (Chen et al. 2013, 2014).

Recent progress of orchid transcriptomic sequencing

High-throughput EST sequencing provides a gateway into the genome by reasons of the many data covered in the genomewide expression information. Before the next generation sequencing technologies (Zhang et al. 2013) were developed, the most popular sequencing method was the Sanger method, which was applied to the EST sequencing project. 1080 subtractive ESTs were obtained from an *Oncidium* Gower Ramsey pseudobulb subtractive EST library (Tan et al. 2005). Most ESTs were revealed as being related to carbohydrate metabolism

and regulatory function, biosynthesis of mannose, pectin and starch, stress-related, and transportation (Tan et al. 2005). To illustrate expressed genes in reproductive organs of *Phalaenopsis*, *P. equestris* mature flower buds were collected and 5593 ESTs were sequenced and annotated (Tsai et al. 2006). In addition, 2359 ESTs were sequenced from scented *P. bellina* flower buds cDNA library to deduced ESTs involved in scent biosynthesis pathway (Hsiao et al. 2006).

The sudden rise of rapid and low-cost next-generation sequencing technologies is substantially promoting our competence to examine the sequences information comprehensively at unparalleled resolution and depth in a cell (Delseny et al. 2010). The technologies were adopted rapidly for orchid transcriptome analysis (Table 1). 454 technology was independently applied to generate 8233 contigs and 34,630 singletons sequenced from the mixed tissues of three *Phalaenopsis* species (Tsai et al. 2013), and 50,908 contig sequences released from six different tissues of *O. Gower Ramsey* (Chang et al. 2011). These data set expansively increased information of expressed genes in *Phalaenopsis* and *Oncidium* and speed identifying sets of ESTs associated with a broad range of biological processes (Chang et al. 2011; Hsiao et al. 2011a; Huang et al. 2015). A total of 121,917 unique ESTs were obtained from the *Ophrys* species by using 454 pyrosequencing and Illumina (Solexa) technologies to identify genes responding to pollinator attraction (Sedeek et al. 2013). A traditional Chinese herb together with 454 pyrosequencing and Illumina technology were able to generate plentiful ESTs for mining the genes participated in alkaloid biosynthetic pathway and polysaccharide biosynthesis in *Dendrobium officinale* (Guo et al. 2013; Zhang et al. 2016b). To provide a general resource for studying the pod development of *Vanilla planifolia*, one of the most valued flavour species for its flavour qualities, the combined 454 and Illumina RNA-seq technologies produced de novo transcriptome with high quality assembly for this important orchid cash crop (Rao et al. 2014). In addition, to improve the horticultural value of *Phalaenopsis* and *Cymbidium*, transcriptome derived from browning leaf of *Phalaenopsis* explant (sequencing by Illumina HiSeq 2000), and variable colour of *Cymbidium* leaf (sequencing by 454 pyrosequencing) were investigated (Xu et al. 2015; Zhu et al. 2015). To study the symbiotic orchid–fungus relationship and the molecular mechanism of orchid seed germination, 454 and Illumina were adopted to explore transcriptomes derived from *Serapias vomeracea* (Perotto et al. 2014), *C. hybridum* (Zhao et al. 2014), *Anoectochilus roxburghii* (Liu et al. 2015), and *Gastrodia elata* (Tsai et al. 2016).

In Orchidaceae, about 40% species adopt crassulacean acid metabolism (Mascher et al.) to fix carbon dioxide

suggesting Orchidaceae to be the largest CAM clade in plants (Silvera et al. 2009). To illuminate the origin and evolution of CAM pathway, transcriptomes derived from leaves of CAM orchids *P. equestris*, *D. terminale* and *C. manni* were sampled at different time interval and sequenced by Illumina HiSeq 2000 (Deng et al. 2016; Zhang et al. 2016c). Their results showed that key carbon fixation pathway genes might primarily evolve by changes at the transcription level in CAM plants. Several techniques were applied to study the development of spectacular orchid flower morphology. This includes the developing floral transcriptomes originating from *Phalaenopsis* (Hsiao et al. 2011a), *Cymbidium* (Zhang et al. 2013; Li et al. 2014; Yang and Zhu 2015), and *Orchis* (De Paolo et al. 2014). The molecular model of MADS-box genes associated with floral development was proposed and discussed.

Recently, root transcriptome from *Paphiopedilum concolor* was also produced to explore genes involved in orchid root development (Li et al. 2015b). Over 1195 unique genes participating in secondary metabolic pathways, and 609 ESTs involved in plant hormone biosynthesis, and plant signal transduction were revealed. The accumulated transcribed sequences could be directly used to develop microarray platform. It is also a resource for phylogenetic analysis. For example, an developed oligomicroarray harboring 14,732 unique expressed sequences based on the information of ESTs collected from *Phalaenopsis* orchids was applied to compare transcriptomes among sepal, petal, and labellum (Hsiao et al. 2013). Three hundred fifteen single-copy orthologs characterized in the transcriptomes of 10 species distributed in all five subfamilies of Orchidaceae were used to investigate the phylogenetic association of orchids (Deng et al. 2015). The results indicated this strategy appeared to be more reliable and efficient than using a few markers of genes for phylogenetic analyses, particularly for those orchids whose DNA sequences are difficult to be amplified or the holomycotrophic species (Deng et al. 2015).

NGS technologies are not only applied to characterize orchid transcriptomes but also used to systematically analyze small RNAs in orchids (Table 1). The roles of small RNAs were studied on the regulation of flowering in *P. aphrodite* and *Erycina pusilla* (An et al. 2011; An and Chan 2012; Chou et al. 2013; Lin et al. 2013a), flower development in *Orchis italica* and *Cymbidium ensifolium* (Aceto et al. 2014; De Paolo et al. 2014; Li et al. 2015c), and interaction between the fungus *Piriformospora indica* with an *Oncidium* hybrid orchid (Ye et al. 2014). Later, comprehensive collection of small RNAs derived from *P. aphrodite* (Chao et al. 2014), and *D. officinale* (Meng et al. 2016) were performed. These efforts provided valuable messages about the expression,

Table 1 Characteristics of findings in the literature for the application of next generation sequencing (NGS) to orchid transcriptomes

Subfamily	Species	Sequencing platform	Tissue	Study aim	References
Apostasioideae	<i>Apostasia shenzhenica</i>	Illumina/Solexa	Mature flower buds	Study of floral development and evolutionary trends of orchid flowers	Tsai et al. (2013)
Apostasioideae	<i>Neuwiedia malipoensis</i>	Illumina/Solexa	Mature flower buds		
Vanilloideae	<i>Vanilla shenzhenica</i>	Illumina/Solexa	Mature flower buds		
Vanilloideae	<i>Galeola faberi</i>	Illumina/Solexa	Mature flower buds		
Cypripedioideae	<i>Paphiopedilum armeniacum</i>	Illumina/Solexa	Mature flower buds		
Cypripedioideae	<i>Cypripedium singchii</i>	Illumina/Solexa	Mature flower buds		
Orchidoideae	<i>Habenaria delavayi</i>	Illumina/Solexa	Mature flower buds		
Orchidoideae	<i>Hemipilia forrestii</i>	Illumina/Solexa	Mature flower buds		
Epidendroideae	<i>Phalaenopsis equestris</i>	Illumina/Solexa	Mature flower buds		
Epidendroideae	<i>Cymbidium sinense</i>	Illumina/Solexa	Mature flower buds		
Vanilloideae	<i>Vanilla planifolia</i>	Illumina/Solexa Roche/454	Pod tissues, seeds	Study of biosynthetic routes to flavor components	Rao et al. (2014)
Cypripedioideae	<i>Paphiopedilum concolor</i>	Illumina HiSeq 2000	Roots	Identify the genes that control root growth and development	Li et al. (2015b)
Orchidoideae	<i>Ophrys</i> species	Roche/454	Flowers, labellums, leaves, flower organ from open flowers and buds	Identify genes responding for pollinator attraction	Sedeek et al. (2013)
Orchidoideae	<i>Orchis italica</i>	Illumina/Solexa (MiSeq)	Inflorescences	The roles of small RNAs on the flower development	Aceto et al. (2014)
Orchidoideae	<i>Orchis italica</i>	Illumina HiSeq 2500	Florets of inflorescence before anthesis	Analyzing transcripts potentially involved in flower development	De Paolo et al. (2014)
Orchidoideae	<i>Serapias vomeracea</i>	Roche/454	Protocorms	Investigate the molecular bases of the orchid response to mycorrhizal invasion	Perotto et al. (2014)
Orchidoideae	<i>Anoectochilus roxburghii</i> (Wall.) Lindl.	Illumina HiSeq 4000	Dry seeds, seeds from symbiotic or germination	Study of seed germination process	Liu et al. (2015)
Orchidoideae	<i>Gastrodia elata</i> Blume	Illumina HiSeq 2000	Vegetative tissues, corms, juvenile tubers	Address the gene regulation mechanism in gastrodin biosynthesis	Tsai et al. (2016)
Epidendroideae	<i>Phalaenopsis aphrodite</i>	Sanger: EST	Protocorms	Gene discovery and genomic annotation	Fu et al. (2011), Hsiao et al. (2011a)
Epidendroideae	<i>Phalaenopsis equestris</i>	Sanger: EST	Mature flower buds		
Epidendroideae	<i>Phalaenopsis bellina</i>	Sanger: EST	Mature flower buds without column		
Epidendroideae	<i>Phalaenopsis aphrodite</i> <i>Phalaenopsis equestris</i> <i>Phalaenopsis bellina</i>	Roche/454	Mixed tissues		
Epidendroideae	<i>Phalaenopsis equestris</i>	Illumina/Solexa	Leaves		
Epidendroideae	<i>Phalaenopsis aphrodite</i>	Roche/454 Illumina/Solexa	Leaves, stems, roots, young inflorescences, stalks, flower buds, flowers, germinating seeds	Investigate expressed genes involved in many biological processes of orchids	Su et al. (2011)

Table 1 continued

Subfamily	Species	Sequencing platform	Tissue	Study aim	References
Epidendroideae	<i>Phalaenopsis aphrodite</i>	Illumina/Solexa	Leaves, stalks, flower buds	Study the roles of small RNAs on the regulation of flowering	An et al. (2011), An and Chan (2012)
Epidendroideae	<i>Phalaenopsis aphrodite</i>	Illumina/Solexa	Leaves, roots, flowers, germinating seeds, young inflorescences	Identify species- and tissue-specific miRNAs	Chao et al. (2014)
Epidendroideae	<i>Phalaenopsis</i> Brother Spring Dancer 'KHM190'	Illumina HiSeq 2000	Petals, sepals or labellums from flower buds of wild-type and peloric petal mutant plants	Study regulation of floral-organ development	Huang et al. (2015)
Epidendroideae	<i>Phalaenopsis</i> sp.	Illumina HiSeq 2000	Explants	Examine <i>Phalaenopsis</i> leaf explant browning	Xu et al. (2015)
Epidendroideae	<i>Oncidium</i> 'Gower Ramsey'	Roche/454	Leaves, pseudobulbs, young inflorescences, inflorescences, flower buds, mature flowers	Identify genes associated with flowering time	Chang et al. (2011)
Epidendroideae	<i>Oncidium</i> 'Gower Ramsey'	Illumina/Solexa	Roots with or without fungus	Study the roles of small RNAs on the interaction between root and the fungus	Ye et al. (2014)
Epidendroideae	<i>Erycina pusilla</i>	Illumina/Solexa	Roots, leaves, peduncles, flowers, capsules	Investigate photoperiod-dependent flowering genes	Chou et al. (2013)
Epidendroideae	<i>Erycina pusilla</i>	Illumina/Solexa	Roots, leaves, peduncles, flowers, capsules	Study the roles of small RNAs on the regulation of flowering	Lin et al. (2013a)
Epidendroideae	<i>Cymbidium ensifolium</i> 'Tiegusu'	Illumina HiSeq 2000	Flower buds, mature flower	Identify genes associated with floral development	Li et al. (2014)
Epidendroideae	<i>Cymbidium sinense</i> 'Qi Jian Bai Mo'	Illumina HiSeq 2000	Plants in vegetative phase/floral differentiation phase/reproductive phase	Identify genes associated with floral development	Zhang et al. (2013)
Epidendroideae	<i>Cymbidium hybridum</i> 'Golden Boy'	Illumina HiSeq 2000	Roots with or without fungus	Study of orchid-mycorrhizal fungi interactions	Zhao et al. (2014)
Epidendroideae	<i>Cymbidium ensifolium</i> 'Tiegusu'	Illumina/Solexa	Flower bud	Identify miRNAs related to floral development	Li et al. (2015c)
Epidendroideae	<i>Cymbidium ensifolium</i> 'tiansu'	Roche/454	Sepals, petals, labellums, gynostemium from flower buds and mature flowers	Reveal genes associated with floral organ differentiation	Yang and Zhu (2015)
Epidendroideae	<i>Cymbidium sinense</i> 'Dharma'	Roche/454	Roots, leaves, pseudobulbs, flowers	Analyze molecular mechanism underlying leaf-color variations	Zhu et al. (2015)
Epidendroideae	<i>Cymbidium sinense</i> <i>Cymbidium atropurpureum</i> <i>Cymbidium manii</i>	Illumina HiSeq 2000	Leaves	Explore the evolution and molecular regulation of CAM plants	Zhang et al. (2016c)
Epidendroideae	<i>Dendrobium officinale</i>	Roche/454	Stems	Study of alkaloid biosynthesis	Guo et al. (2013)
Epidendroideae	<i>Dendrobium officinale</i>	Illumina HiSeq 2000	Juvenile and adult plants	Identify genes associated with polysaccharide synthesis	Zhang et al. (2016b)
Epidendroideae	<i>Dendrobium officinale</i>	Illumina HiSeq 2500	Flower, roots, leaves, stems	Study of the regulatory networks of the production and accumulation of the medicinal constituents	Meng et al. (2016)

composition and function of small RNAs which help us to better understand functional genomics of orchids.

For the storage and manage the massive expressed gene sequences from orchid, OrchidBase was developed to collect the transcriptomic sequences from 11 various tissues/organs of *Phalaenopsis* spp. and flower tissues of 10 species distributed in five subfamilies of Orchidaceae (Fu et al. 2011; Tsai et al. 2013; Niu et al. 2016). Both deep sequencing with ABI 3730, Roche 454 and Illumina/Solexa were applied to generate EST sequences collected in OrchidBase. OrchidBase is generously accessible at <http://orchidbase.itps.ncku.edu.tw/>. The database delivers a prominent feature of genetic resource for both data mining and experimental researches of orchid biology and biotechnology. Orchidstra (<http://orchidstra.abrc.sinica.edu.tw>), another orchid transcriptomic database, was developed to collect 233,924 unique contigs of *P. aphrodite* transcriptomic sequences by use of a Illumina/Solexa and Roche 454 platform. Profiling analysis with RNA-Seq was applied to categorize the genes with tissue-tropism expression patterns (Su et al. 2011, 2013). In addition, 50,908 contigs of sequences generated by using Roche 454 platform from various organs of *Oncidium* were assembled into the *OncidiumOrchidGenomeBase* (<http://predictor.nchu.edu.tw/oogb/>) (Chang et al. 2011). These EST dataset are valuable for the identification of specificity of orchids, annotation of genes for genomic sequencing, and assistance in the organization of the orchid genome.

Current status of orchid genome sequencing

With the quick development and lower cost of NGS, whole genome sequencing of non-model species can be implemented. The first milestone was the sequencing of the *Phalaenopsis equestris*, a tropical epiphytic orchid and recurrently be used as parent species for orchid breeding (Cai et al. 2015). The *P. equestris* genome was whole-genome shotgun and sequenced by Illumina technology. The genome size was estimated to 1.16 Gb containing 29,431 predicted protein-coding genes (Cai et al. 2015). Analysis of the *P. equestris* genome showed that the majority of the genome (about 62%) was occupied by repetitive DNAs, mostly transposable elements (TEs). In addition, before the radiation of most of the orchid clades, an orchid-specific paleopolyploidy event was discovered. This species is also the first CAM plant that has been whole-genome sequenced, and the gene family (α carbonic anhydrase) involved in CAM pathway has an obvious expansion. The result suggested that the evolution of CAM photosynthesis in *P. equestris* might associate with gene duplication. In addition, genes located at the heterozygous regions might relate to self-incompatibility. Genes in type II MADS-box clades, including the

E-class, C/D-class, B-class AP3 and AGL6 clades, contain extra orthologs than other plant species. These expanded clades are involved in orchid floral organ development that can provide the unique evolutionary paths of these floral organ identity genes accorded with the innovative development of lip and column in orchids. Furthermore, the *Phalaenopsis* genome sequence was practical to identify MYB genes controlling floral pigmentation patterning (Hsu et al. 2015) and TCP genes participated in the ovule development (Lin et al. 2016).

Dendrobium officinale, with both ornamental value and therapeutic effects, is the second sequenced orchid plant by joining both the NGS Illumina Hiseq 2000 and the third-generation PacBio machineries (Yan et al. 2015). The assembled genome of *D. officinale* had a predicted gene number of 35,567, and that was higher than that in *Phalaenopsis*. For example, the number of B-class MADS-box genes presented in *D. officinale* was much higher than that in *Phalaenopsis* with its four members in B-class AP3-like subfamily, and one member in B-class PI-like subfamily. In contrast, 19 AP3-like genes and five PI-like genes were present in this *Dendrobium* genome. It is possible that the *D. officinale* plants used for the whole genome sequencing were hybrids rather than a native species. Later, another *Dendrobium* species (*Dendrobium catenatum*) was whole genome sequenced by Illumina HiSeq 2000 platform (Zhang et al. 2016a). The predicted 28,910 protein-coding genes were comparable with those of *Phalaenopsis*, and a whole genome duplication event could be shared with *Phalaenopsis*. Adaptation to a wide range of ecological niches of *Dendrobium* might relate to the expansion of many resistance-related genes. In addition, it seems that extensive duplication of genes encoding glucomannan synthase associated with the synthesis of medicinal polysaccharides. It also found that MADS-box gene clades ANR1, StMADS11, and MIKC* were expanded in the *Dendrobium* suggesting the function of these clades might relate to the astonishing diversity of plant architecture in *Dendrobium* (Zhang et al. 2016a).

Most recently, the primitive orchid *A. shenzhenica*, a representative of one of two genera that form a sister lineage to the rest of the Orchidaceae, was whole genome sequenced to provide a reference for improving our understanding of orchid origins and evolution (Zhang et al. 2017). The *A. shenzhenica* genome was sequenced by use of a combination of different approaches including Illumina, PacBio and 10 \times genomics technologies. The total length of the genome assembly was 349 Mb containing 21,841 protein-coding genes. *A. shenzhenica* shows changes within MADS-box gene classes, which control a diverse suite of developmental processes, during orchid evolution. This study provides new insights on the genetic evidence underlying key orchid innovations,

including the development of the labellum and gynostemium, pollinia, and seeds without endosperm, as well as the evolution of epiphytism (Zhang et al. 2017).

Whole-genome duplication

A striking feature of plant genomes is the whole genome duplication (WGD) has occurred for many times (Cui et al. 2006; Van de Peer et al. 2009). WGD is a major evolutionary force in angiosperm genomes. The complete genome sequences of angiosperm trees have provided information on polyploidy and genome evolution. It is now generally recognized that in the predecessor of all seed plants, one WGD occurred, and in the ancestor of all angiosperms, an extra one happened (Jiao et al. 2011). Furthermore, a hexaploidy event preceded the radiation of core eudicots (Jiao et al. 2012). Whereas a WGD shared by most of the monocots was also suggested (Paterson et al. 2004). Based on the *Phalaenopsis* and *Dendrobium* genome sequences, two turns of WGDs in the *D. catenatum* lineage was inferred from the examination of both the allocation of synonymous substitutions per synonymous site (Ks) throughout all paralogous genes and for duplicated genes sitting in synteny blocks (Zhang et al. 2016a). This most recent WGD event is shared with *Dendrobium*, *Phalaenopsis*, and *Apostasia* and might occur near the Cretaceous–Paleogene (K/Pg) boundary. Putative peaks at older Ks age distribution might point to further ancient WGD events in the monocot lineage that could be shared by monocot ancestors (Cai et al. 2015; Zhang et al. 2016a, 2017).

Since it is suggested that species diversification might be facilitated by WGDs (Van de Peer et al. 2009), it is intriguing to perceive whether the WGD is common with the subfamily Orchidoideae containing 3630 species, and the subfamilies Cyripedioideae, and Vanilloideae, which include 180, and 185 species, respectively. Orchidoideae separated from the Epidendroideae (about 20,000 species including *P. equestris* and *D. catenatum*) about 59 million years ago (Gustafsson et al. 2010). Cyripedioideae and the ancestor of Orchidoideae and Epidendroideae subfamilies are predicted to have disconnected from each other around 68 million years ago (Gustafsson et al. 2010). Fortunately, extensive transcriptomic data sets and/or whole genome sequences from members of these other subfamilies are ongoing. It will be optimistic for uncovering the mystery of orchid genome evolution after more orchid genomic and transcriptomic data obtained.

Transposable elements

Complements of genomic elements and extraordinary variations in sizes were observed in the genomes of flowering plant containing approximately 300,000 species (Wendel et al. 2016). An obvious route to genome

expansion is WGD events, but many species with great genomes are diploid. For example, *Oryza australiensis* and *O. sativa* are from the same genus, however, the former is more than twice the size of the latter due to the addition of ~ 400 Mbp of DNA majorly by three individual retrotransposable element families in the past few million years (Piegu et al. 2006). And thus, the dynamics of transposable element proliferation and clearance might play the important role in the majority of variation in plant genome size, which is in turn superimposed by the history of WGD (Bennetzen and Wang 2014).

The assembled *P. equestris* genome is composed of about 62% repetitive DNA. This is relatively higher than those found in rice (29%) and grape (41%), but similar in sorghum (61%). About 59% of the genome is occupied by interspersed repeats and TEs and 3% is accounted for tandem repeats. Among the TEs, ~ 46% of the genome is occupied by long terminal repeats (LTRs). Long interspersed nuclear elements (LINEs) account for ~ 8% of the genome (Cai et al. 2015). Most LTRs (71%) appeared in the *P. equestris* genome during 11.7–43 million years ago, far behind the origin of the last common ancestor of orchids (74–86 million years ago). *Copia* and *Gypsy* TEs might experience a recent burst (Cai et al. 2015).

About 78.1% of the *D. catenatum* genome is occupied by 789 Mb repetitive elements. Retrotransposable elements constitute a huge portion of the *D. catenatum* genome including LINE/RTE (5.68%), LINE/L1 (8.44%), LTR/Gypsy (18.49%), and LTR/Copia (27.36%). In addition, the quantity of de novo predicted repeats was predominantly more than that composed based on Repbase, demonstrating that *D. catenatum* has arose numerous unique repeats contrasted to other sequenced plant genomes (Zhang et al. 2016a). During the last 5 million years, a burst of LTR activity was observed, suggesting that these LTRs were integrated into the genome of *D. catenatum* after it was separated from *P. equestris* (22.6–59.6 million years ago) (Zhang et al. 2016a).

Angiosperms have a high frequency of hybridization and polyploidy, with resultant increased TE activity and genome shuffling, and profitable gene duplication (Oliver et al. 2013). TE activity enhanced genomic plasticity provides a more satisfactory and complete explanation to “abominable mystery” described by Darwin. Although explanations for orchid diversity are persuasive and valid, the extreme diversity of orchids still cannot fully understand. However, due to a few of genomic data of orchids, the role of TE involved in orchid diversity and evolution has not been accessed yet. Probably, artificial arena of *Phalaenopsis* orchid breeding might supply direct and relatively recent clues for the pivotal of TEs in the generation of selectable cultivars in *Phalaenopsis* orchids.

Genome mapping

Several fundamental studies are necessary to well-assemble the whole-genome sequences. These studies include high-density genetic map and bacteria artificial chromosomes (BACs) or yeast artificial chromosomes (YACs)-based physical map.

Genetic map

Even though the whole-genome sequence of *Phalaenopsis equestris* is published (Cai et al. 2015), the current scaffold number of *P. equestris* genome is still high (236,185 scaffolds), and the assembly of *P. equestris* scaffolds remains to be improved. A saturated genetic linkage map is regarded as an effective tool to facilitate scaffolds assembly. Besides, genetic mapping is referred to a step-in genome mapping, which can integrate with a physical map and a cytogenetic map to access genes in genome. For many important crops, genome mapping using genetic linkage maps, physical maps and cytogenetic maps are well developed, such as rice, wheat, barley, etc. (Kurata et al. 1994; Ramsay et al. 2000; Hearnden et al. 2007; Li et al. 2015a). In Orchidaceae, the genetic linkage maps have been constructed for the medicinal orchids, *Dendrobium* (Xue et al. 2010; Lu et al. 2012a, b; Feng et al. 2013), but not for the ornamental orchids, *Phalaenopsis*. Hence, it is vital needed to construct a saturated genetic map for *Phalaenopsis*.

Usually, a genetic linkage map is constructed based on the crossing over events during meiosis. By estimating the recombination frequency between molecular markers, linkage groups can be built while several markers are linked together. A genetic linkage map is comprised of the numbers of linkage groups corresponding to the chromosome numbers.

First, to identify the crossing over events, one of the significant factors is a segregated mapping population. For constructing genetic maps in rice and wheat, F_2 population or recombinant inbred lines (RILs) are often developed as suitable mapping populations (Kurata et al. 1994; Li et al. 2015a). However, unlike other crops, *Phalaenopsis* orchids have long juvenile periods. It usually takes 2–3 years to generate a new generation. Consequently, it is inefficient and time-consuming to generate *Phalaenopsis* F_2 population or RILs. As an alternative approach, an outcrossing F_1 population is used as a mapping population. Yet, the segregation pattern may be more complex due to the segregated alleles that may vary in number. Currently, a F_1 population with 116 progenies from the cross between *P. aphrodite* subsp. *formosana* and *P. equestris*, the two native species in Taiwan and important parents in breeding, has been adopted as the mapping population for constructing a genetic map for *Phalaenopsis*.

After determination of the mapping population, various types of molecular markers are applied for linkage analysis. For instance, restriction fragment length polymorphisms (RFLPs), amplified fragment length polymorphisms (AFLPs), expressed sequence tags (ESTs), and simple sequence repeats (SSRs) are widely utilized in the constructing genetic maps for several crops. Among them, SSR markers particularly have benefits for classification and mapping of genes due to their extraordinary reproducibility, co-dominant inheritance, relative large quantity, extreme polymorphism, simplicity of genotyping and easily amplified through PCR (Varshney et al. 2005). Previously, 950 SSRs from *P. equestris* have been identified, and among them 206 SSRs primer sequences were developed for *Phalaenopsis* genetic mapping (Hsu et al. 2011). *P. equestris* contains 38 chromosomes ($2N = 2X = 38$) and its genome size is 1600 Mb per haploid genome. It is a relatively large genome compared to other crops, *ex.*: rice 400–430 Mb, and sorghum 750–770 Mb (Eckardt 2000). Hence, development of abundant molecular markers for *Phalaenopsis* genome mapping is necessary. Nowadays, as NGS technologies are rapidly developed and spread, huge amount of binary single nucleotide polymorphism (SNP) markers is easily generated. Due to this characteristic of SNP markers generation, constructing a saturated genetic map in a relative short time is feasible.

After a saturated, high-density *Phalaenopsis* genetic map is plotted, it is capable of sequence alignment and boosting *P. equestris* scaffolds assembly. In addition, a genetic linkage map can be adopted for useful trait analysis and applied for quantitative trait loci (QTL) mapping. Combining the phenotypic data of the mapping population, QTLs may be able to be mapped on the saturated genetic linkage map. Moreover, after QTL mapping, the molecular markers corresponding to QTLs will be adopted for marker-assisted selection (MAS) for interested agronomic traits. For orchid marketing, *Phalaenopsis* breeders usually constant cross or backcross for about 5–10 generations to select potentially commercial cultivars. Considering 2–3 years for each generation, a period up to 20–30 years is required to breed valuable and popular cultivars, such as *P. Sogo Yukidian* “V3” and *P. I-Hsin Sesame*, for markets. By applying the saturated genetic map to MAS, the time-consuming breeding process will be shortened. Thus, a saturated genetic map is considered as the first step for breeding by design and to reveal interested traits controlled by QTLs and genes.

Physical map

The construction of physical map is using restriction enzymes (RE) digestion of large insert clone libraries, such as BACs and YACs (Chaney et al. 2016). For

example, the BAC libraries are treated with one RE, and the digested fragment patterns are analyzed by using electrophoresis. Different BAC clones with similar fragment patterns mean that they contain overlapping regions and can be assembled together to form fingerprinting contigs (FPCs). Therefore, physical map provides a framework for accelerated position cloning of genes for important agronomic traits, and assembly of genome sequence. With the integration of genetic map, the molecular markers identified on the genetic map can be aligned with physical map, and show the physical location of these markers.

Alternatively, optical mapping is another approach to construct a physical map. Similar to FPCs, optical mapping uses RE to digest DNA and stained with an intercalating dye. Then DNA fragments are imaged, separated according to their sizes, and then ordered with restriction map. Although optical mapping is effective, it is too complex and slow in lengthening the DNA, data imaging and processing. Hence, the optical mapping has mainly applied for small genome organisms, ex.: bacteria (Zhou et al. 2004; Haas et al. 2009). Further improvement of automatic technology for imaging and data processing is now available, so that optical mapping is granted for mapping of large genomes, such as human genome (Jo et al. 2009; Teague et al. 2010).

As an improved high throughput layout, the BioNano Genomics Irys system is much efficient than the traditional optical mapping. This system uses a modified RE to introduce DNA breakage on a single strand or incisions in DNA at sequences-specific site. The breaks are then labeled with fluorescence, and the DNA molecular is passed through automated electrophoresis system to form linearized DNA in nanochannels. For each molecular, the length and distances between labeled sites are measured and analyzed to create a molecular map. Two DNA molecules containing similar distance patterns are assembled into a large contig. Therefore, the BioNano Genomics Irys system is automated and more efficient for mapping complex plant genomes, including wheat and barley (Xiao et al. 2015; Stankova et al. 2016; Mascher et al. 2017). The physical mapping of *Phalaenopsis* using the upmost technology is currently underway.

Genome-wide association studies (GWAS)

Plant breeding is the practice to produce novel varieties with better phenotypes which are then applied in farming and husbandry, and for human utility (Pawelkiewicz et al. 2016). The basic strategy for plant breeding is to select the best individuals and their progenies containing the desired agricultural traits. The association mapping strategy commonly evaluates the statistical analyses to calculate the significance of the association between

various phenotypes and the genetic polymorphism in a set of defined persons with genetic variations (Ogura and Busch 2015).

In the past, a number of molecular markers have already been applied for genome mapping or association analysis, such as SSRs, AFLPs, and various kinds of SNPs. With the emergence of NGS technologies, genome-wide association studies (GWAS) is developed for the application of genetic variations that are largely and compactly distributed all over the genome. Therefore, GWAS has replaced the traditional mapping strategy with more opportunities to identify large collections of molecular markers and find new genes and regulatory sequences responsible for specific traits (Pawelkiewicz et al. 2016).

With the great progresses in the methods of detecting trait differences and genetic variations, GWAS has emerged as a promising approach to inspect the relationships between complex genetic polymorphisms and distinctive phenotypes for most species (Slate et al. 2009). Over last decades, the employment of the automated image system encourages high-throughput phenotyping of a lots of plants by using large-scale image snapping and phenotypic trait quantification from the images (Ogura and Busch 2015). GWAS can combine the quantitative traits to their genetic variations in distinct individuals, and inspire the quantitative regulation of the growth and development.

As NGS technologies emerge, the genotyping-by-sequencing (GBS) approach rapidly gets hundreds of thousands to millions of markers to construct the high-resolution genetic map, and generates sufficient information and coverage of plant genomes (Edwards and Batley 2010). GWAS through the GBS approach provides major assistant for traditional crop improvement with the identification of great amount of SNP molecular markers and the production of genetic linkage maps with high-density markers (Poland and Rife 2012).

GBS method is a useful and powerful approach to rapidly get lots of markers to construct genetic maps with high resolution. This tactics generates sufficient information and coverage to the plant genomes (Edwards and Batley 2010). The first step of GBS is to digest genome DNAs from different plants in a population of genetically distinct individuals with an appropriate RE (Elshire et al. 2011). Each digested genome DNA is ligated with unique, short DNA sequences (barcodes adaptors) to make different plant DNAs be assembled and sequence determined together in a sole sequencing lane. Then the genome DNAs were recovered and prepared for NGS to characterize high amount of SNPs between each individual in this population, and these SNPs are applied as molecular markers for the construction of a genetic map with high density (Elshire et al. 2011).

There are three advantages by using GBS for genome-wide genetic analysis. First, the usage of unique short barcode for each plant sample makes all the DNAs be pooled together for sequencing. Second, an appropriate RE is necessary for getting highly reproducible results and important genome regions, as well as reducing the repetitive regions and genome complexity. Third, with the NGS technology, a high amount of SNP molecular markers will be identified from the sequencing data, and be applied to construct a genetic map with high density.

For example, by using the GBS approach, 41,371 SNP markers were identified in 254 advanced wheat breeding lines from the wheat breeding program of International Wheat and Maize Improvement Center (Poland et al. 2012b). In addition, 12.4 Gb of high-quality genome sequence and 129,156 SNP markers have been identified for potato. These SNPs were mapped to 2.1 Mb of potato reference genome, and the average read depth was 63× coverage per cultivar (Uitdewilligen et al. 2013). In durum wheat, 9983 putative SNP markers were characterized between their two parents and used for genotyping 91 RILs (van Poecke et al. 2013).

Moreover, GBS approach is a powerful method to develop high-density markers even lack of the sequenced reference genome or previous identification of DNA polymorphism. The high-density markers identified from GBS can be used to anchor the orders of physical maps as well as the whole genome shotgun sequence to improve the genome assembly (Poland et al. 2012a). The application of genomics-assisted breeding to valuable cash crops especially with complex genomes is a very important process. Similar transitions are found in rapeseed, lupin (Fabacea), lettuce, switchgrass (perennial warm season bunchgrass in north America), soybean, and maize (Bus et al. 2012; Poland and Rife 2012; Truong et al. 2012; Yang et al. 2012; Lu et al. 2013; Sonah et al. 2015). For example, in *Manihot esculenta* Crantz (ICGMC 2015), the complex 2412-cM map is incorporated into 10 biparental maps (containing 3480 meioses), and 22,403 genetic markers are anchored on 18 chromosomes (ICGMC 2015). This map was arranged 71.9% of the draft genome assembly and 90.7% of the predicted protein-coding genes (ICGMC 2015). It is beneficial with the chromosome-anchored genome sequence for breeding amendment by providing the prominent characterization of markers linked to chief traits.

In *Phalaenopsis* orchids, the GBS analysis has been carried out in the first generation offsprings (F₁) from the cross between *P. aphrodite* subsp. *formosana* (white-flower) and *P. equestris* (red-flower). This is beneficial to the assembly of whole-genome sequences of *P. equestris* into 19 linkage groups (unpublished data). The regulatory regions for various phenotypes in the F₁ plants, including

plant size, flower color and size, have been investigated (unpublished data).

Most economically important traits are usually inherited in a quantitative manner, such as the height, weight, and stress resistance. The genetic basis of the quantitative traits are regulated by many genes, the synergistic effects of genes, and the mutual interactions between genes and environment. The major challenge to identify the QTL is to precisely locate the gene via high-density linkage map. This is an extremely costly and time-consuming process. GWAS with NGS technologies offer an opportunity for a powerful strategy to identify large numbers of SNP markers, and construct high-density genetic association map that is suitable for underling these important and complex QTLs. For example, complete genotyping of 2815 maize inbred accessions have characterized 681,257 SNP markers distributed across all over the maize genome. Among them, SNPs associated to the plausible candidate genes are identified for kernel color, sweetness, and flowering time (Romay et al. 2013).

The orchid floral scent trait could be regulated by QTL. Recently, GBS has been applied to study the deceptive orchid *Orphis*, and found several SNPs linked to odour related genes (Sedeek et al. 2014). The study of the ecological speciation was surveyed to understand the reproductive hurdles and variations in floral phenotypes in four closely related species, sexually deceptive *Ophrys* species with various flower morphology and distinct labellum coloration. Regardless of the flower odour chemistry may fundamentally cause the reproductive hurdles, GBS showed common polymorphism all over the *Ophrys* genome but highly distinguished polymorphisms in genes involved in floral odour biosynthesis (Sedeek et al. 2014). This finding suggests that these species are marked mostly by genic divergence rather than genome-wide variations. This result may be applied to other orchids with sexual deception, where floral odour genes may be amongst the first to differentiate (Sedeek et al. 2014).

Recently, the molecular omics, such as epigenome, transcriptome, proteome, and metabolome, have been used as complementary manners to improve the SNP-trait association studies. Epigenomic changes, that are responded to environment stimulate and reflect in phenotypic changes, provide another meaning of the “molecular phenotype” definition. Recent advances in population epigenomics studied the association between genetic SNPs and epigenomic variations in a global accession collection of *Arabidopsis thaliana* (Kawakatsu et al. 2016). Moreover, the transcriptomic variations, which reflect the variation in both genetic and epigenetic regulatory and other omics studies of the proteome and metabolome, have proved to be great resources as molecular phenotype (He et al. 2013).

Alternatively, it would also be valuable if the epigenome, transcriptome, proteome, and metabolome are applied as molecular markers, the “genotypes”, to calculate their associations to the downstream phenotypic traits. For example, the differential transcriptomes from 368 maize diverse inbred lines was used to identify the expression presence/absence variation (ePAV; genes were expressed, or not) and served as as “phenotype” to perform association analysis with 15 agronomic phenotypes and 526 metabolic traits (Jin et al. 2016). As for the quantitative traits of QTLs, the genomic variations from these molecular omics are also quantitative and range mutable. However, they are not only binary or with limited numbers of alleles. Therefore, the quantitative GWAS (qGWAS) has recently been proposed to solve the continuous genotype issue. It has also been used to explore the regulatory network by treating expression level of genes as both “genotypes” and “phenotypes” (Wen et al. 2016).

The major goal for plant breeding is to select the best progenies with the desired traits. GWAS can guild the genetic variations with the high-density SNP markers for mapping the regulatory locus of these traits. In *Phalaenopsis* orchids, several agricultural traits have been emphasized, including flower sizes, length of floral inflorescence, flower colors and pigmentation patterning, and floral fragrances. With the GWAS analysis, it will be plausible to construct the high-density genetic map, and the regulatory regions responsive for the important traits can be identified. Moreover, this will be beneficial for the *Phalaenopsis* breeding for new varieties with the traits of interests.

Comparative genomics

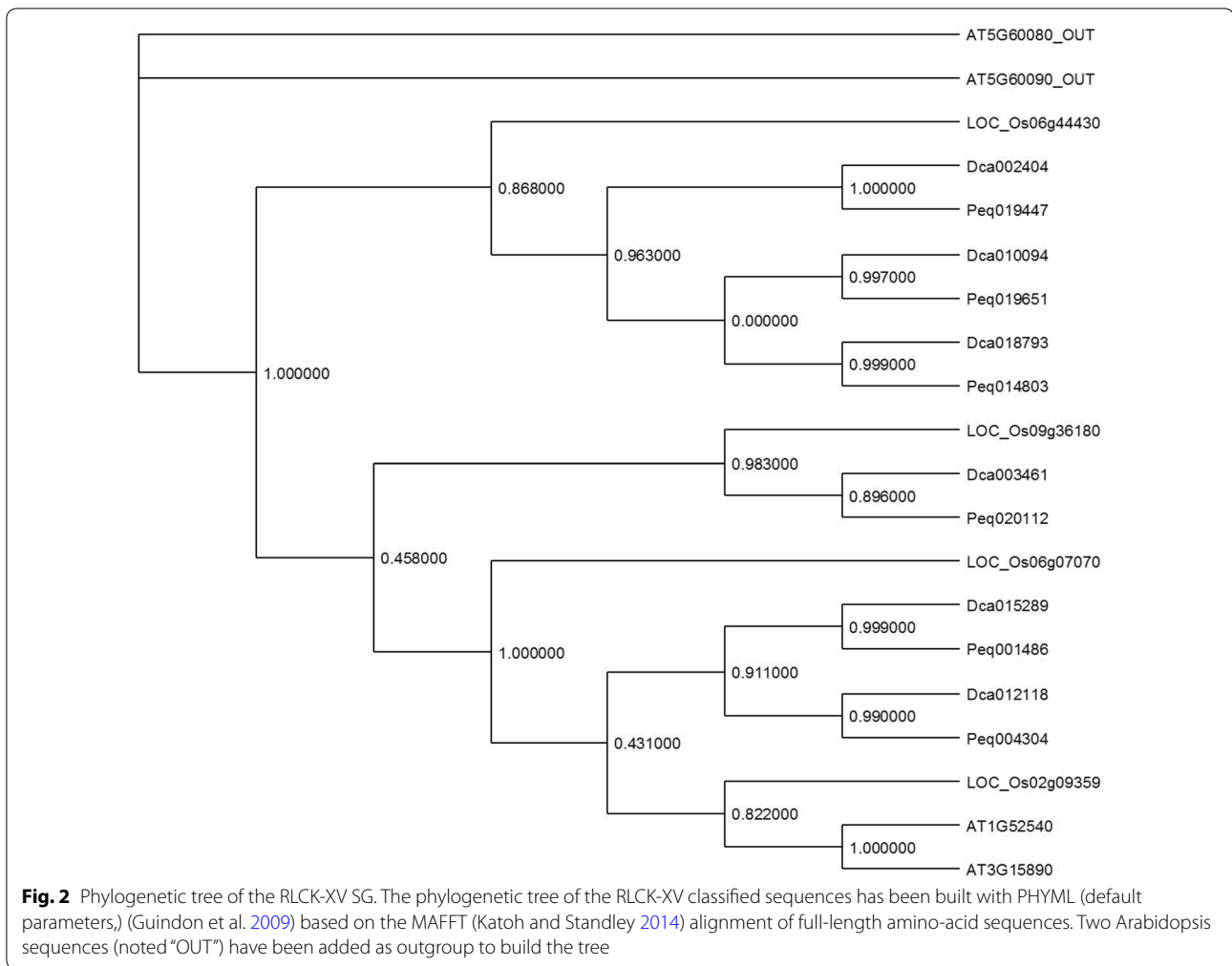
Receptor-like kinases in orchid genomes

Phosphorylation is a reversible addition of phosphate to proteins. This post-translational modification is involved in all signaling pathways and every cellular activity in living organisms. In eukaryotes, this process is conducted by a superfamily of protein kinases (ePKs for eukaryotic protein kinases) which represent ~ 1.5 to 2.5% of all genes in average, making ePK one of the largest protein families in eukaryotes (Manning et al. 2002). In *Arabidopsis thaliana* and rice (*Oryza sativa*), the two model plant species, ~ 1000 and ~ 1500 Ser/Thr protein kinases have been identified, corresponding to 2.9 and 2.3% of the proteomes in respective organisms (<http://plantsp.genomics.purdue.edu/index.html>) (Shiu et al. 2004; Dardick et al. 2007).

According to the taxonomy of Hanks and Hunter and the studies based on kinase domain similarity and phylogeny of eukaryotes, the kinomes of many eukaryotic genomes have been divided into subfamilies (Hanks and

Hunter 1995; Martin et al. 2009). In the plant kingdom, the largest subfamily is composed of plant-specific receptors named receptor-like kinase (RLK) (Shiu et al. 2004; Dardick et al. 2007). These receptors play important roles in signal transduction to relay external signals across cell membranes. They are involved in developmental processes and/or represent guard molecules which are able to recognize pathogen attacks in a process called pathogen-associated molecular patterns (PAMP)-triggered immunity (PTI) (Boller and Felix 2009; Gish and Clark 2011; Wu and Zhou 2013; Antolín-Llovera et al. 2014; Haruta and Sussman 2017). Many of the RLKs are also activated under a whole range of abiotic stress responses (Ye et al. 2017). These receptors typically contain an amino-terminal extracellular domain (ECD), a transmembrane (TM) domain, and an intracellular domain composed of the kinase domain (KD). Several phylogenetic studies of the RLK subfamily have been conducted. They focus mainly on *Arabidopsis* but also on other plant species (Shiu and Bleecker 2001, 2003; Shiu et al. 2004; Lehti-Shiu et al. 2009; Liu et al. 2009; Sakamoto et al. 2012; Zan et al. 2013). *Arabidopsis* RLK phylogenetic analyses inferred from KD alignments have led to the classification into 12 subgroups. These subgroups contain similar motifs in their ECD, including CRINKLY 4 like (CR4L), CrRLK1L (named after the first member identified in *Catharanthus roseus*), cystein-rich (CRK), extensin, lectin (Pawelkiewicz et al.), leucine-rich repeats (LRR), lysin motif (LysM), proline-rich extensin-like (PERK), RKF3, wall-associated (Kawakatsu et al.), LRK10L-2 and receptor-like cytoplasmic kinase (RLCK) (Shiu and Bleecker 2001, 2003; Shiu et al. 2004).

To analyze the RLK subfamily in *Dendrobium catenatum* and *Phalaenopsis equestris*, first the hmmsearch program was used to seek the kinase hidden Markov model (HMM) profile (PF00069.16) within the *Arabidopsis*, rice, and the two orchid proteomes (Sonnhammer et al. 1998; Eddy 2009). The KD sequences of these proteins were extracted and aligned using the MAFFT program (Katoh and Standley 2013). A phylogenetic tree was then built by the maximum likelihood method FastTree (Price et al. 2009). The tree leaves were annotated according to previous *Arabidopsis* and rice annotations to classify orchid sequences within one of the RLK subgroups (Fig. 2) (Shiu and Bleecker 2003; Shiu et al. 2004). This analysis confirms previous observations that with ~ 650 and ~ 1100 RLK genes in *Arabidopsis* and rice respectively. The expansion rate in rice is approximately twice that of *Arabidopsis*. The subgroup expansions observed in rice have been shown to mainly involve receptors with roles in PTI, like the WAK or LEC subfamilies (Lehti-Shiu et al. 2009; Vaid et al. 2013; Delteil et al. 2016). In the orchid genomes, ~ 400 and ~ 300 RLKs have been detected in



Dendrobium and *Phalaenopsis* respectively, accounting for 1.4 and 1% of the respective proteomes (containing 29,400 and 29,679 annotated proteins) (see Table 2 for the complete list of accessions and their classification). Compared to the Arabidopsis and rice proportion of RLKs in genomes (~ 1.8 and ~ 1.7% respectively), the number of RLK in orchid genomes is lower. Considering that the *Phalaenopsis* proteome has been established on ~ 85% of the estimated total genome, ~ 15% more RLK sequences could be present in the *Phalaenopsis* genome, increasing the proportion of RLK in this genome to ~ 1.1%. Nevertheless, these proportions are below the range of what has been observed previously in other plant genomes. These results suggest that the *Dendrobium* and *Phalaenopsis* RLK subgroups did not experience large-scale expansions such as the ones observed in the rice genome. Looking at small-scale expansions, it has to be noted that one RLCK clade seems to have expanded specifically in the orchid genomes. Indeed, in the Arabidopsis and rice genomes, 2 and 4 genes are classified into the RLCK-XV subset,

while 6 genes belonging to this SG have been retrieved in each orchid genome. The RLCK SG has a particularity among the RLKs since these receptors are lacking the ECD and TM domains. Some of them have been shown to be membrane-anchored and physically associated with membrane-spanning RLK. These protein kinases are then involved in signal relays (not signal perception) via transphosphorylations with other receptors to modulate signaling related to development and stress responses (Lin et al. 2013b). In Arabidopsis, the function of the two genes belonging to the RLCK-XV SG has not been characterized yet. In silico expression data show that these genes are ubiquitously expressed along all developmental stages (Hruz et al. 2008).

Terpene synthases (TPSs) in orchid genomes

Terpenoids are the largest class of metabolites found to date, including over 40,000 structures (Croteau et al. 2000; Chappell 2002; Gershenzon and Dudareva 2007), responsible to the interaction between plants and

Table 2 Proportion of *Arabidopsis*, *Oryza* and orchid sequences belonging to RLK subgroups

RLK subgroups	<i>Arabidopsis thaliana</i>	<i>Oryza sativa</i>	<i>Dendrobium catenatum</i>	<i>Phalaenopsis equestris</i>
LEC	8.8	19.1	8.9	6.3
CR4L	1.3	1.9	1.7	1.7
CRK	11.6	7.8	8.6	7.3
CrRLK1L	2.4	1.6	3.2	3.3
Extensin	0.5	0.4	1.5	0.3
LRK10L-2	2.1	4.6	2.7	0.7
LysM	1.0	1.0	1.5	2.0
PERK	2.9	1.8	2.2	4.0
RKF3	0.3	0.3	0.5	0.7
LRR	39.3	33.7	38.7	37.2
WAK	4.5	11.9	2.5	2.7
RLCK	24.4	15.4	26.8	32.2

environment, including defense to insects, pathogens, diseases and other stresses, also work to attract pollinators. These compounds are now utilized as pharmaceuticals (e.g., taxol), fragrance (e.g., limonene from orange or lemon oil), industrial materials (e.g., diterpene resin acid), and even as biofuel ingredient (Chen et al. 2011). With the great use of the terpene, understanding the genes that related to the biosynthesis of terpenes is now under study. Terpene synthases (TPSs), the key enzymes that generate structure diversity of terpene are usually followed by cytochromes P450 (CYPs), which further modify the products from TPSs and cause much more diverse molecules of terpenes.

The *TPS* gene family number are different among species, *Physcomitrella patens* (moss) has only one functional *TPS* gene, while in *Eucalyptus grandis* (flooded gum) 113 putative functional genes are found (Külheim et al. 2015). However, through the analysis of plant genome, researchers have shown most of the plant *TPS* gene family have their gene numbers ranging from 20 to 150, belonging to mid-size family (Chen et al. 2011). The functions of TPSs can be mono- or multi-functional, and the enzymes can be highly identical to each other. For instance, the diterpene synthases levopimaradiene/abietadiene synthase and isopimaradiene synthase showed 91% identity in Norway spruce, moreover, the functional bifurcation of these two enzymes was proved to cause by four amino acid residues only (Keeling et al. 2008). This suggests that *TPS* genes have gone through gene duplication, neofunctionalization, and/or subfunctionalization, so that lead to specialized metabolites of large family of *TPS* genes.

The large *TPS* family is divided into seven clades, from *TPS*-a to *TPS*-g, according to their protein sequence (Bohlmann et al. 1998; Dudareva et al. 2003; Martin

et al. 2004). *Phalaenopsis equestris* genome has 23 *TPS*s belonging to *TPS*-a, -b, -c, e/f, -g (Figs. 3, 4). We investigated the *TPS* evolutionary relationship among orchids, to see whether duplication and then sub- or neo-functionalization of the *TPS*s have occurred during evolution. Cao et al. (2010) has proposed that diterpene synthases are the origin of mono- and sesqui-terpene synthases. Interestingly, the evolution of *TPS*s with other related genes may also create unexpected mechanism. For instance, the evolution pathways of *TPS/CYP* pairs are different in monocot and dicot: *TPS/CYP* pairs duplicated with ancestral *TPS/CYP* pairs as template to evolve in dicot, but evolutionary mechanism of monocot showed genome rearrangement of *TPS* and *CYP* individually (Boutanaev et al. 2015). Moreover, *TPS* cluster density showed that *TPS*s occupied 0.008 gene/Mb in *P. patens* genome, 0.07 gene/Mb in Sorghum, and 0.3 gene/Mb in *Arabidopsis* and *Vitis vinifera* (Chen et al. 2011). *Phalaenopsis* tends to have dense *TPS* cluster compared to others, with 1.4 genes/Mb found in *P. equestris* and 2.09 genes/Mb in *Dendrobium*. It is possible that the higher density of *TPS*s genes in both *Phalaenopsis* and *Dendrobium* are related to that they both are CAM plant, and requires the *TPS* cluster to synthesize terpenoid compounds quickly for them to circumvent the adverse environment.

The *TPS* number, expression pattern, gene structure and phylogeny relationship among orchids may reveal the initial architecture of the *TPS* genes role in the orchids. Based on functions of these *TPS*s, mono-, di- and sesqui-terpene synthases are found in orchids, *Arabidopsis*, and *Oryza sativa*. Total numbers of *TPS*s found in *P. equestris* are lesser than in *Arabidopsis*, rice, and *Dendrobium*, but the annotated monoterpene synthases are more than the others. *Dendrobium* seems to have significant larger

TPS subfamily

- TPS-a
- TPS-b
- TPS-c
- TPS-d
- TPS-e/f
- TPS-g
- TPS-h

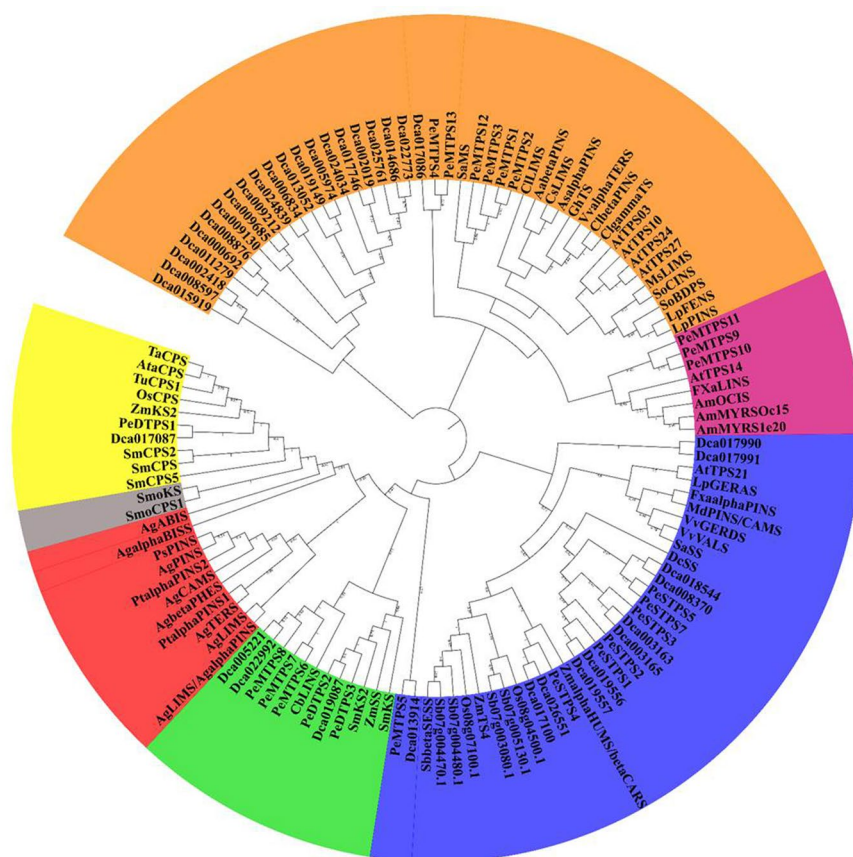


Fig. 3 Phylogeny of *P. equestris* and *Dendrobium catenatum*'s TPS subfamily. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei 1987). The bootstrap consensus tree inferred from 100 replicates (Felsenstein 1985) is taken to represent the evolutionary history of the taxa analyzed (Felsenstein 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The evolutionary distances were computed using the JTT matrix-based method (Jones et al. 1992) and are in the units of the number of amino acid substitutions per site. The analysis involved 123 amino acid sequences. All ambiguous positions were removed for each sequence pair. There were a total of 1411 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al. 2016)

group of annotated diterpene synthases than *Arabidopsis*, rice, and even the closest related *P. equestris* (Table 3).

The genes related in terpene synthesis are usually found to be lined together, forming functional clusters in plants (Matsuba et al. 2013). The functional clusters of TPS genes are already found in several species, such as *Eucalyptus*, *solanum*, *Vitis vinifera*, *Arabidopsis*, rice, etc. (Shimura et al. 2007). Based on functions of these TPSs, mono-, di- and sesquiterpene synthases are found in orchids, *Arabidopsis*, and *Oryza sativa* (Table 3). In some plant, such as *Eucalyptus*, only the TPSs in same subfamily would form clusters (Külheim et al. 2015). Moreover, the TPS can form cluster with genes that are related within the same biosynthesis pathway. In *solanum*, TPS form functional cluster with *cis*-prenyl transferase (Matsuba et al. 2013). In our analysis of *P. equestris* genome, total 23 TPSs were found and predicted as mono-, di-,

and sesqui-terpene synthase. In addition, those TPSs form clusters on four different scaffolds, and each scaffold only has the terpene synthase with the same function (Fig. 4).

Secondary metabolomics

Plants produce assorted specified metabolites, but the biosynthesis genes responsible for the production and regulation of these metabolites stay mostly unknown, hampering attempts to monitor plant pharmacopeia. Assumed that genes encompassing particular metabolites pathways display environmentally reliant co-regulation, we supposed that genes participated in a specified metabolites biosynthesis pathway are arranged in strong relations (modules) in gene coexpression networks, accelerating their discovery (Wisecaver et al. 2017). Many geneticists have developed an effective method

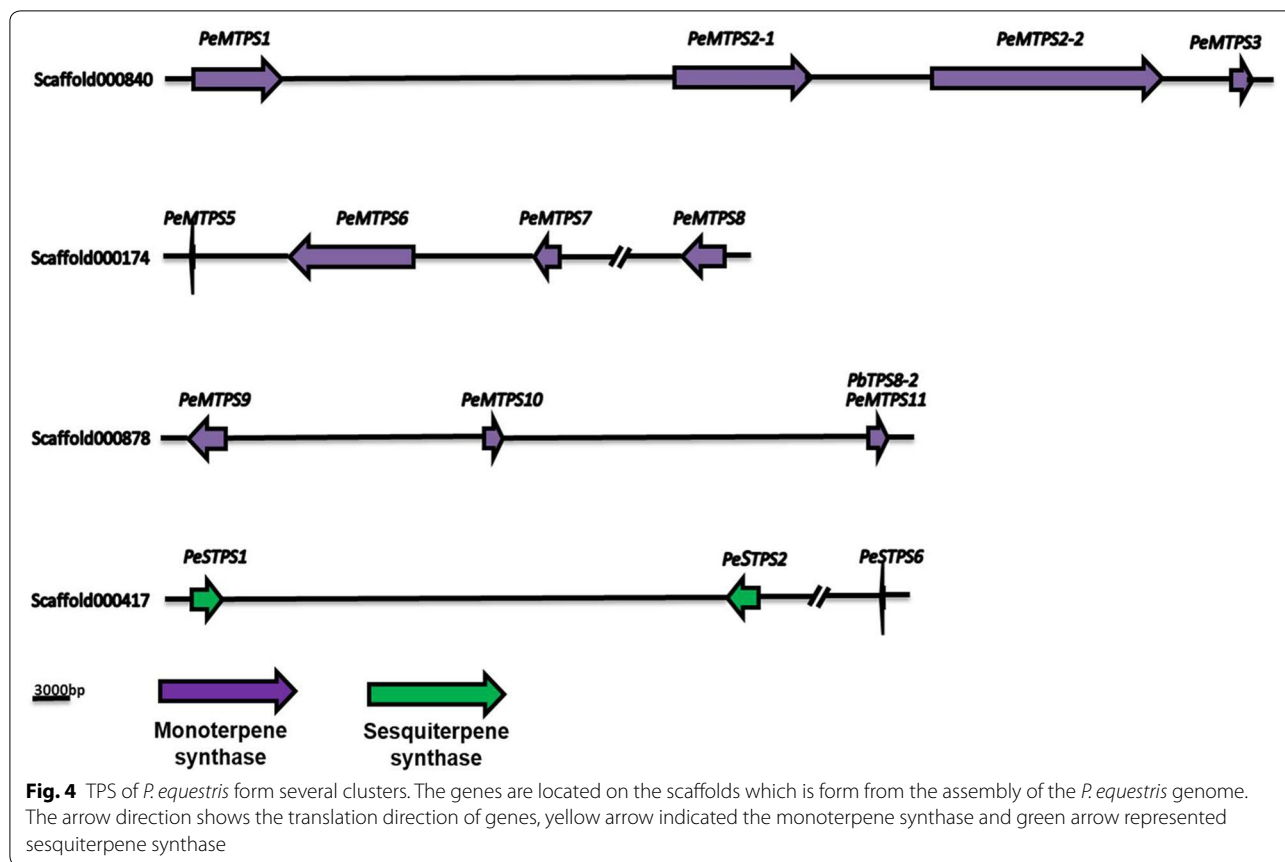


Table 3 The annotated function of TPS of *Arabidopsis*, *Oryza* and orchids

Annotated function	<i>Arabidopsis thaliana</i>	<i>Oryza sativa</i>	<i>Dendrobium catenatum</i>	<i>Phalaenopsis equestris</i>
Monorpen synthase	7	2	1	13
Sesquiterpene synthase	22	18	14	7
Diterpene synthase	3	12	21	3

for identifying the plant genes that produce the chemical compounds to protect plants from predation, and it is a natural source of many important drugs.

Floral scent

Floral volatile compounds (VOCs) have been characterized in several orchids, such as lilac aldehydes in *Platanthera bifolia*, phenylacetaldehyde in *Gymnadenia odoratissima*, the green-leaf volatiles in *Epipactis helleborine*, chiloglottones in *Chiloglottis*, 1-octen-3-ol in *Dracula lafleuri*, and monoterpenes in *P. bellina*.

Identification of the fragrant vandaceous orchids that are endemic to Malaysia is performed by using similar strategy to excavate probable fragrance-related EST-SSRs as the molecular markers in the *Vanda Mimi* Palmer (Teh

et al. 2011). The unique transcripts were obtained from the *Ophrys* species by using 454 pyrosequencing and Illumina (Solexa) technologies to identify genes responding for pollinator attraction (Sedeek et al. 2013).

Medicinal orchids

To study the genes involved in alkaloid biosynthetic pathway and polysaccharide biosynthesis in *Dendrobium officinale*, an important traditional Chinese herb, 454 pyrosequencing and Illumina technology was respectively applied to generate plentiful ESTs (Guo et al. 2013; Zhang et al. 2016b). Among them, 69 sequences denoting 25 genes in the biosynthesis of alkaloid backbone are detected, and 170 and 37 genes encoding to glycosyl-transferase and cellulose synthase, respectively, showed

differential expression patterns (Guo et al. 2013; Zhang et al. 2016b). *Dendrobium catenatum* whole genome sequence predicted 28,910 protein-coding genes are comparable with those of *Phalaenopsis*. In *Dendrobium*, the expansion of several resistance-related genes indicate a potent immune system accountable for adaptation to a broad extent of ecological niches. In addition, the genes participating glucomannan synthase actions are generally linked to the biosynthesis of medicinal polysaccharides (Zhang et al. 2016b).

Mycoheterotrophic metabolomics

Orchidaceae contain typical mycorrhizal plants. Their seeds possess no endosperm therefore they are devoided of nutrient supply. For seed germination and growth during seedling stage, the majority of orchid species are reliant on mycorrhizal fungi. Orchids are absolutely mycoheterotrophic (reliant on symbiotic fungi for the supply of carbon and nitrogen) during the achlorophyllous protocorm stage that follows seed germination in nature (Rasmussen and Rasmussen 2009; Fochi et al. 2017; Suetsugu et al. 2017).

Orchids are unique among plants in that during their life history stages, they require mycorrhizal symbioses with soil fungi from seed germination to adulthood. To identify the molecular process of orchid seed germination and the symbiotic orchid–fungus correlation, 454 and Illumina have been applied to explore transcriptomes derived from *Serapias vomeracea* (Perotto et al. 2014), *Cymbidium hybridum* (Zhao et al. 2014), *Anoectochilus roxburghii* (Liu et al. 2015), and *Gastrodia elata* (Tsai et al. 2016). The genes related to mycorrhizal symbiosis in autotrophic orchids and arbuscular mycorrhizal plants exist with common molecular mechanisms among various mycorrhizal types.

Proteome analysis of 2D-LC–MS/MS different nutrient changed stages of symbiotic germination has been performed in *Oncidium sphacelatum* (Valadares et al. 2014). A bidirectional carbon flow even in the mycoheterotrophic symbiosis in *Epipactis helleborine* has been indicated (Suetsugu et al. 2017). This is in contrast to most fully mycoheterotrophic and partially mycoheterotrophic species, which interact mainly with ectomycorrhiza-forming fungi, such as the *Sebacinales*, *Russulaceae* and *Thelephoraceae*, suggesting that their ultimate source of carbon is the photosynthate produced by nearby trees (Lee et al. 2015; Gebauer et al. 2016).

Genome editing in orchids

In the past decade, new technologies for genetic modification have emerged and known as genome-editing technologies. These technologies depend on engineered endonucleases that show sequence-specific manner

of DNA cleavage due to the presence of a sequence-specific DNA-binding domain or RNA sequence (Gaj et al. 2013; Carroll 2014). Via the binding of the specific DNA sequence, the engineered nucleases can proficiently cleave the targeted genes. The cleaved DNA with double-strand breaks (DSB) subsequently result in DNA repair and lead to gene modification at the target sites. The DNA repair mechanisms include homology-directed repair and non-homologous end joining breaks (Wyman and Kanaar 2006). Recently, the CRISPR (clustered regularly interspaced short palindromic repeats)/Cas (CRISPR-associated) system was derived from the adaptive immune system (type II) of prokaryotic organism (Jinek et al. 2012). CRISPRs were identified as an unusual sequence element containing a series of 29-nucleotide repeats separated with 32-nucleotide “spacer” sequences in *Escherichia coli* genome (Ishino et al. 1987; Wiedenheft et al. 2012), and function as RNA interference mechanism to bind and cleave target DNA. A short CRISPR RNA (crRNA), the type II CRISPR/Cas from *Streptococcus pyogenes*, can recognize a complementary strand in foreign DNA with sequence specificity. Furthermore, formation of a ribonucleoprotein complex with Cas9 nuclease to generate site-specific DSBs requires a transactivating crRNA (tracrRNA) (Walsh and Hochedlinger 2013). The modules of crRNA and tracrRNA are joined into a single RNA molecule, and termed as guide RNA (Mali et al. 2013). Effective cleavage includes the presence of the protospacer adjacent motif (PAM) in the complementary strand of foreign DNA succeeding the recognition sequence (Jinek et al. 2012).

Genome editing in other crops

The crucial requirement of genome editing is the accessibility of accurate genomic data as well as gene functions which are mostly available in model plants. The deficiency for the genomic information on horticultural crops has significantly limited breeding efficiency. So far, several cashcrops for horticulture have been whole-genome sequenced (Bolger et al. 2014). These comprise grapevine, papaya, strawberry, sweet orange, etc. In addition, a substantial increased number of transcriptomes of many horticultural crops are also feasible. The genomics information include reference genome data, transcriptomic sequences and genomic resequencing data for several horticultural crops, may provide unrestricted targets for genome editing for depicting the gene functions, which sequentially can assist engaging the CRISPR/Cas technology to design better crops.

Plant genome editing alleviates governing worries associated to genetically modified (GM) plants. Up to now, plant protoplasts of *Arabidopsis thaliana*, tobacco, lettuce and rice have been transfected using CRISPR/Cas9

system and achieved targeted mutagenesis in regenerated plants at frequencies up to 46% (Woo et al. 2015). The genome targeted sites enclosed small insertions or deletions cannot be distinguished from the naturally occurring genetic variations, and that they are germline-transmissible (Woo et al. 2015).

Genome editing in orchids

The two criteria for a crop to perform genome editing are the accessibility of genomic data and confirmation of gene functions as well as the availability for the platform of genetic transformation and regeneration process. So far, the two sequenced orchid genomes are *Phalaenopsis equestris* and *Dendrobium officinale*. The transformation of *D. officinale* is well established and has a shorter regeneration time. For *P. equestris*, the transformation system has been established, but the regeneration time is long, about 2–3 years. With these advantages, also being a popular medicinal plant for multiple pharmaceutical effectives such as immunomodulation, anti-oxidation, anti-fatigue, genome editing has been recently reported in *D. officinale* to knockout the expression of several lignocellulose biosynthesis genes, including C3H, C4H, 4CL, CCR and IRX (Kui et al. 2016). In addition, Kui et al., has adopted *Agrobacterium* to deliver the CRISPR/Cas9 construct and compare five promoters with Cauliflower mosaic virus 35S promoter, and identified their compatible promoter activities. Even though Kui et al. showed 100% success rate for the genome editing, they did not show the reduction of the content of lignocellulose in the knockout plants.

In *Phalaenopsis* orchids, the whole genome sequence is available, and the gene functions for floral morphogenesis, flower color, and floral scent are well studied. These have laid the groundwork for genome editing in *Phalaenopsis* orchids with the genetic transformation system. It is well expected that breeding of many more cultivars can be feasible via genome editing of these trait genes.

Conclusions and perspectives

With the available whole genome sequences of *P. equestris*, *D. catenatum*, *D. officinale*, and *A. shenzhenica*, the genetic blueprint of orchids provides a fundamental knowledge of the genetic basis of orchids. Furthermore, the whole genome sequences of one of the most popular aromatic orchids, *Vanilla*, will be available soon. The efforts by many scientists to use a plethora of genome information and genomics tools will lead to a promising understanding of the biological, physiological, molecular and genetic mechanisms of orchids in years to come. Comparative genomics analysis revealed expanded RLCK-XV clade was detected in *Phalaenopsis* genome, and the investigation of their expression patterns and

putative interactors should give new insights into their functions.

In addition, the genome sequences will also be an important resource for genetic transformation for molecular breeding, including molecular marker-assisted breeding (MAB), or genome-assisted breeding (Hruz et al. 2008), and the production of transgenic or genome-edited plants. These are necessary to aid orchid horticultural research and speed up the orchid breeding. The major goal for plant breeding is to select the best progenies with the desired traits, and GWAS can guild the genetic variations with the high-density SNP markers for mapping the regulatory locus of these traits. In *Phalaenopsis* orchids, several agricultural traits have been focused, the flower sizes, length of floral inflorescence, flower colors and pigmentation patterning, and floral fragrances. With the GWAS analysis, the high-density genetic map can be constructed, and the regulatory regions responsive for the important traits can be identified. Moreover, this will be benefit for the *Phalaenopsis* breeding for new varieties with the traits of interests.

Abbreviations

BACs: bacteria artificial chromosomes; CAM: crassulacean acid metabolism; Cas: CRISPR-associated system; CRISPR: clustered regularly interspaced short palindromic repeats; ECD: extracellular domain; ePAV: expression presence/absence variation; ESTs: expressed sequence tags; FPCs: fingerprint contigs; GBS: genotyping-by-sequencing; GDPS: geranyl diphosphate synthase; GWAS: genome-wide association studies; HDR: homology-directed repair; HMM: hidden Markov model; LTRs: long terminal repeats; MAS: marker-assisted selection; NHEJ: non-homologous end joining breaks; NGS: next-generation sequencing; PTI: pathogen-associated molecular patterns (PAMP)-triggered immunity; QTL: quantitative trait loci; RE: restriction enzyme; RLK: receptor-like kinase; SNP: single nucleotide polymorphism; SSRs: simple sequence repeats; TEs: transposable elements; TPSs: terpene synthases; VOCs: volatile compounds; WGDs: whole genome duplications; YACs: yeast artificial chromosomes.

Authors' contributions

HHC, WCT, and YYH conceived the manuscript. WCT wrote the section of "Orchid genome evolution". CCH and SYC wrote the section of "Genome mapping". HHC, AD, and HH wrote the section of "Comparative genomics". YYH wrote the section of "Secondary metabolomics". HHC wrote the section of "Genome editing in orchids". All authors read and approved the final manuscript.

Author details

¹ Institute of Tropical Plant Sciences, National Cheng Kung University, Tainan 701, Taiwan. ² Orchid Research and Development Center, National Cheng Kung University, Tainan 701, Taiwan. ³ Department of Life Sciences, National Cheng Kung University, Tainan 701, Taiwan. ⁴ CIRAD, UMR AGAP, TA A 108/03, Avenue Agropolis, 34398 Montpellier, France. ⁵ Present Address: School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Road, Life Sciences Building, Room 3-117, Shanghai 200240, People's Republic of China.

Acknowledgements

We thank Ching-Jen Su's assistance in preparing the manuscript, Morris Ng provides the picture of *Anoectochilus formosanus*, and Dr. Yung-I Lee provides the picture of *Apostasia wallichii*.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Not applicable.

Consent for publication

Yes.

Ethics approval and consent to participate

Not applicable.

Funding

This work was supported by the Grants 105-2313-B-006-002-MY3, and 105-2321-B-006-026- to H.H.C. and W.C.T., respectively from Ministry of Science and Technology, Taiwan.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 8 June 2017 Accepted: 1 December 2017

Published online: 12 December 2017

References

- Aceto S, Sica M, De Paolo S, D'Argenio V, Cantiello P, Salvatore F, Gaudio L (2014) The analysis of the inflorescence miRNome of the orchid *Orchis italica* reveals a DEF-Like MADS-box gene as a new miRNA target. *PLoS ONE* 9:e97839
- An FM, 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
- An FM, Hsiao SR, Chan MT (2011) Sequencing-based approaches reveal low ambient temperature-responsive and tissue-specific microRNAs in *Phalaenopsis* orchid. *PLoS ONE* 6:e18937
- Antolín-Llovera M, Petutsching EK, Ried MK, Lipka V, Nürnberger T, Robatzek S, Parniske M (2014) Knowing your friends and foes—plant receptor-like kinases as initiators of symbiosis or defence. *New Phytol* 204:791–802
- Bennetzen JL, Wang H (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu Rev Plant Biol* 65:505–530
- Bohlmann J, Meyer-Gauen G, Croteau R (1998) Plant terpenoid synthases: molecular biology and phylogenetic analysis. *Proc Natl Acad Sci USA* 95:4126–4133
- Bolger ME, Weisshaar B, Scholz U, Stein N, Usadel B, Mayer KF (2014) Plant genome sequencing—applications for crop improvement. *Curr Opin Biotechnol* 26:31–37
- Boller T, Felix G (2009) A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu Rev Plant Biol* 60:379–406
- Boutanaev AM, Moses T, Zi J, Nelson DR, Mugford ST, Peters RJ, Osbourn A (2015) Investigation of terpene diversification across multiple sequenced plant genomes. *Proc Natl Acad Sci USA* 112:E81–E88
- Bus A, Hecht J, Huettel B, Reinhardt R, Stich B (2012) High-throughput polymorphism detection and genotyping in *Brassica napus* using next-generation RAD sequencing. *BMC Genomics* 13:281
- 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, Liu ZJ (2015) The genome sequence of the orchid *Phalaenopsis equestris*. *Nat Genet* 47:65–72
- Cao R, Zhang Y, Mann FM, Huang C, Mukkamala D, Hudock MP, Mead ME, Prisc S, Wang K, Lin F-Y, Chang T-K, Peters RJ, Oldfield E (2010) Diterpene cyclases and the nature of the isoprene fold. *Proteins* 78:2417–2432
- Carroll D (2014) Genome engineering with targetable nucleases. *Annu Rev Biochem* 83:409–439
- Chaney L, Sharp AR, Evans CR, Udall JA (2016) Genome mapping in plant comparative genomics. *Trends Plant Sci* 21:770–780
- Chang CC, Lin HC, Lin IP, Chow TY, Chen HH, Chen WH, Cheng CH, Lin CY, Liu SM, Chang CC, Chaw SM (2006) The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol Biol Evol* 23:279–291
- Chang YY, Chu YW, Chen CW, Leu WM, Hsu HF, Yang CH (2011) Characterization of *Oncidium* 'Gower Ramsey' transcriptomes using 454 GS-FLX pyrosequencing and their application to the identification of genes associated with flowering time. *Plant Cell Physiol* 52:1532–1545
- Chao YT, Su CL, Jean WH, Chen WC, Chang YCA, Shih MC (2014) Identification and characterization of the microRNA transcriptome of a moth orchid *Phalaenopsis aphrodite*. *Plant Mol Biol* 84:529–548
- Chappell J (2002) The genetics and molecular genetics of terpene and sterol origami. *Curr Opin Plant Biol* 5:151–157
- Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J* 66:212–229
- Chen WH, Kao YK, Tang CY, Tsai CC, Lin TY (2013) Estimating nuclear DNA content within 50 species of the genus *Phalaenopsis* Blume (Orchidaceae). *Sci Hortic (Amsterdam)* 161:70–75
- Chen WH, Kao YK, Tang CY (2014) Variation of the genome size among *Phalaenopsis* species using DAPI. *J Taiwan Soc Hort Sci* 60:115–123
- Chou ML, Shih MC, Chan MT, Liao SY, Hsu CT, Haung YT, Chen JJ, Liao DC, Wu FH, Lin CS (2013) Global transcriptome analysis and identification of a *CONSTANS*-like gene family in the orchid *Erycina pusilla*. *Planta* 237:1425–1441
- Cozzolino S, Widmer A (2005) Orchid diversity: an evolutionary consequence of deception? *Trends Ecol Evol* 20:487–494
- Croteau R, Kutchan TM, Lewis NG (2000) Natural products (secondary metabolites). *Biochem Mol Biol Plants* 24:1250–1319
- Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A (2006) Widespread genome duplications throughout the history of flowering plants. *Genome Res* 16:738–749
- Dardick C, Chen J, Richter T, Ouyang S, Ronald P (2007) The rice kinase database. A phylogenomic database for the rice kinome. *Plant Physiol* 143:579–586
- De Paolo S, Salvemini M, Gaudio L, Aceto S (2014) De novo transcriptome assembly from inflorescence of *Orchis italica*: analysis of coding and non-coding transcripts. *PLoS ONE* 9:e102155
- Delseny M, Han B, Hsing YI (2010) High throughput DNA sequencing: the new sequencing revolution. *Plant Sci* 179:407–422
- Delteil A, Gobbato E, Cayrol B, Estevan J, Michel-Romiti C, Dievart A, Kroj T, Morel JB (2016) Several wall-associated kinases participate positively and negatively in basal defense against rice blast fungus. *BMC Plant Biol* 16:17
- Deng H, Zhang GQ, Lin M, Wang Y, Liu ZJ (2015) Mining from transcriptomes: 315 single-copy orthologous genes concatenated for the phylogenetic analyses of Orchidaceae. *Ecol Evol* 5:3800–3807
- Deng H, Zhang LS, Zhang GQ, Zheng BQ, Liu ZJ, Wang Y (2016) Evolutionary history of PEPC genes in green plants: implications for the evolution of CAM in orchids. *Mol Phylogenet Evol* 94:559–564
- Dudareva N, Martin D, Kish CM, Kolosova N, Gorenstein N, Faldt J, Miller B, Bohlmann J (2003) (E)-beta-ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: function and expression of three terpene synthase genes of a new terpene synthase subfamily. *Plant Cell* 15:1227–1241
- Eckardt NA (2000) Sequencing the rice genome. *Plant Cell* 12:2011–2017
- Eddy SR (2009) A new generation of homology search tools based on probabilistic inference. *Genome Inform* 23:205–211
- Edwards D, Batley J (2010) Plant genome sequencing: applications for crop improvement. *Plant Biotechnol J* 8:2–9
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791
- Feng S, Zhao H, Lu J, Liu J, Shen B, Wang H (2013) Preliminary genetic linkage maps of Chinese herb *Dendrobium nobile* and *D. moniliforme*. *J Genet* 92:205–212

- Fochi V, Chitarra W, Kohler A, Voyron S, Singan VR, Lindquist EA, Barry KW, Girlanda M, Grigoriev IV, Martin F, Balestrini R, Perotto S (2017) Fungal and plant gene expression in the *Tulasnella calospora*–*Serapias vomeracea* symbiosis provides clues about nitrogen pathways in orchid mycorrhizas. *New Phytol* 213:365–379
- Fu CH, Chen YW, Hsiao YY, Pan ZJ, Liu ZJ, Huang YM, Tsai WC, Chen HH (2011) OrchidBase: a collection of sequences of the transcriptome derived from orchids. *Plant Cell Physiol* 52:238–243
- Gaj T, Gersbach CA, Barbas III CF (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* 31:397–405
- Gebauer G, Preiss K, Gebauer AC (2016) Partial mycoheterotrophy is more widespread among orchids than previously assumed. *New Phytol* 211:11–15
- Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. *Nat Chem Biol* 3:408–414
- Gill DE (1989) Fruiting failure, pollinator inefficiency, and speciation in orchids. *Speciat Conseq* 458:481
- Gish LA, Clark SE (2011) The RLK/Pelle family of kinases. *Plant J* 66:117–127
- Guindon S, Delsuc F, Dufayard JF, Gascuel O (2009) Estimating maximum likelihood phylogenies with PhyML. In: Posada D (ed) *Bioinformatics for DNA sequence analysis*. Humana Press, Totowa, pp 113–137
- Guo X, Li Y, Li C, Luo H, Wang L, Qian J, Luo X, Xiang L, Song J, Sun C, Xu H, Yao H, Chen S (2013) Analysis of the *Dendrobium officinale* transcriptome reveals putative alkaloid biosynthetic genes and genetic markers. *Gene* 527:131–138
- Gustafsson ALS, Verola CF, Antonelli A (2010) Reassessing the temporal evolution of orchids with new fossils and a Bayesian relaxed clock, with implications for the diversification of the rare South American genus *Hoffmannseggella* (Orchidaceae: Epidendroideae). *BMC Evol Biol* 10:177
- Haas BJ, Kamoun S, Zody MC, Jiang RH, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, Bozkurt TO, Ah-Fong AM, Alvarado L, Anderson VL, Armstrong MR, Avrova A, Baxter L, Beynon J, Boevink PC, Bollmann SR, Bos JL, Bulone V, Cai G, Cakir C, Carrington JC, Chawner M, Conti L, Costanzo S, Ewan R, Fahlgren N, Fischbach MA, Fugelstad J, Gilroy EM, Gnerre S, Green PJ, Grenville-Briggs LJ, Griffith J, Grunwald NJ, Horn K, Horner NR, Hu CH, Huitema E, Jeong DH, Jones AM, Jones JD, Jones RW, Karlsson EK, Kunjeti SG, Lamour K, Liu Z, Ma L, Maclean D, Chibucos MC, McDonald H, McWalters J, Meijer HJ, Morgan W, Morris PF, Munro CA, O'Neill K, Ospina-Giraldo M, Pinzon A, Pritchard L, Ramsahoye B, Ren Q, Restrepo S, Roy S, Sadanandom A, Savidor A, Schornack S, Schwartz DC, Schumann UD, Schwessinger B, Seyer L, Sharpe T, Silvar C, Song J, Studholme DJ, Sykes S, Thines M, van de Vondervoort PJ, Phuntumart V, Wawra S, Weide R, Win J, Young C, Zhou S, Fry W, Meyers BC, van West P, Ristaino J, Govers F, Birch PR, Whisson SC, Judelson HS, Nusbaum C (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461:393–398
- Hanks SK, Hunter T (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J* 9:576–596
- Haruta M, Sussman MR (2017) Chapter ten—ligand receptor-mediated regulation of growth in plants. In: Andreas J (ed) *Current topics in developmental biology*, vol 123. Academic Press, Cambridge, pp 331–363
- He G, Chen B, Wang X, Li X, Li J, He H, Yang M, Lu L, Qi Y, Wang X, Wang Deng X (2013) Conservation and divergence of transcriptomic and epigenomic variation in maize hybrids. *Genome Biol* 14:R57
- Hearnden PR, Eckermann PJ, McMichael GL, Hayden MJ, Eglinton JK, Chalmers KJ (2007) A genetic map of 1000 SSR and DArT markers in a wide barley cross. *Theor Appl Genet* 115:383
- Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W, Zimmermann P (2008) Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinform* 2008:5
- Hsiao YY, Tsai WC, Kuoh CS, Huang TH, Wang HC, Wu TS, Leu YL, Chen WH, Chen HH (2006) Comparison of transcripts in *Phalaenopsis bellina* and *Phalaenopsis equestris* (Orchidaceae) flowers to deduce monoterpene biosynthesis pathway. *BMC Plant Biol* 6:14
- Hsiao YY, Chen YW, Huang SC, Pan ZJ, Fu CH, Chen WH, Tsai WC, Chen HH (2011a) Gene discovery using next-generation pyrosequencing to develop ESTs for *Phalaenopsis* orchids. *BMC Genomics* 12:360
- Hsiao YY, Pan ZJ, Hsu CC, Yang YP, Hsu YC, Chuang YC, Shih HH, Chen WH, Tsai WC, Chen HH (2011b) Research on orchid biology and biotechnology. *Plant Cell Physiol* 52:1467–1486
- Hsiao YY, Huang TH, Fu CH, Huang SC, Chen YJ, Huang YM, Chen WH, Tsai WC, Chen HH (2013) Transcriptomic analysis of floral organs from *Phalaenopsis* orchid by using oligonucleotide microarray. *Gene* 518:91–100
- Hsieh MH, Lu HC, Pan ZJ, Yeh HH, Wang SS, Chen WH, Chen HH (2013a) Optimizing virus-induced gene silencing efficiency with *Cymbidium* mosaic virus in *Phalaenopsis* flower. *Plant Sci* 2011–202:25–41
- Hsieh MH, Pan ZJ, Lai PH, Lu HC, Yeh HH, Hsu CC, Wu WL, Chung MC, Wang SS, Chen WH, Chen HH (2013b) Virus-induced gene silencing unravels multiple transcription factors involved in floral growth and development in *Phalaenopsis* orchids. *J Exp Bot* 64:3869–3884
- Hsu CC, Chung YL, Chen TC, Lee YL, Kuo YT, Tsai WC, Hsiao YY, Chen YW, Wu WL, Chen HH (2011) An overview of the *Phalaenopsis* orchid genome through BAC end sequence analysis. *BMC Plant Biol* 11:3
- Hsu CC, Chen YY, Tsai WC, Chen WH, Chen HH (2015) Three R2R3-MYB transcription factors regulate distinct floral pigmentation patterning in *Phalaenopsis* spp. *Plant Physiol* 168:175–191
- Huang JZ, Lin CP, Cheng TC, Chang BC, Cheng SY, Chen YW, Lee CY, Chin SW, Chen FC (2015) A *de novo* floral transcriptome reveals clues into *Phalaenopsis* orchid flower development. *PLoS ONE* 10:e0123474
- ICGMC (2015) High-resolution linkage map and chromosome-scale genome assembly for cassava (*Manihot esculenta* Crantz) from 10 populations. *G3 (Bethesda)* 5(1):133–144
- Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A (1987) Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol* 169:5429–5433
- Jersáková J, Trávníček P, Kubátová B, Krejčíková J, Urfus T, Liu ZJ, Lamb A, Ponert J, Schulte K, Čurn V, Vrána J, Leitch IJ, Suda J (2013) Genome size variation in Orchidaceae subfamily Apostasioideae: filling the phylogenetic gap. *Bot J Linn Soc* 172:95–105
- Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, de Pamphilis CW (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature* 473:97–100
- Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, McNeal J, Rolf M, Ruzicka DR, Wafula E, Wickett NJ, Wu X, Zhang Y, Wang J, Zhang Y, Carpenter EJ, Deyholos MK, Kutchan TM, Chanderbali AS, Soltis PS, Stevenson DW, McCombie R, Pires JC, Wong GKS, Soltis DE, dePamphilis CW (2012) A genome triplication associated with early diversification of the core eudicots. *Genome Biol* 13:R3
- Jin M, Liu H, He C, Fu J, Xiao Y, Wang Y, Xie W, Wang G, Yan J (2016) Maize pan-transcriptome provides novel insights into genome complexity and quantitative trait variation. *Sci Rep* 6:18936
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–821
- Jo K, Schramm TM, Schwartz DC (2009) A single-molecule barcoding system using nanoslits for DNA analysis: nanocoding. *Methods Mol Biol* 544:29–42
- Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 8:275–282
- Kao YY, Chang SB, Lin TY, Hsieh CH, Chen YH, Chen WH, Chen CC (2001) Differential accumulation of heterochromatin as a cause for karyotype variation in *Phalaenopsis* orchids. *Ann Bot* 87:387–395
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and ssability. *Mol Biol Evol* 30:772–780
- Katoh K, Standley DM (2014) MAFFT: iterative refinement and additional methods. In: Russell DJ (ed) *Multiple sequence alignment methods*. Humana Press, Totowa, pp 131–146
- Kawakatsu T, Huang SC, Jupe F, Sasaki E, Schmitz RJ, Urlich MA, Castanon R, Nery JR, Barragan C, He Y, Chen H, Dubin M, Lee C-R, Wang C, Bemm F, Becker C, O'Neil R, O'Malley RC, Quarless DX, Alonso-Blanco C, Andrade J, Becker C, Bemm F, Bergelson J, Borgwardt K, Chae E, Dezaan T, Ding W, Ecker JR, Expósito-Alonso M, Farlow A, Fitz J, Gan X, Grimm DG, Hancock A, Henz SR, Holm S, Horton M, Jarsulic M, Kerstetter RA, Korte A, Korte P, Lanz C, Lee C-R, Meng D, Michael TP, Mott R, Muliya FX, Nägele T, Nagler M, Nizhynska V, Nordborg M, Novikova P, Picó FX,

- Platzer A, Rabanal FA, Rodriguez A, Rowan BA, Salomé PA, Schmid K, Schmitz RJ, Seren Ü, Sperone FG, Sudkamp M, Svardal H, Tanzer MM, Todd D, Volchenboum SL, Wang C, Wang G, Wang X, Weckwerth W, Weigel D, Zhou X, Schork NJ, Weigel D, Nordborg M, Ecker JR (2016) Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell* 166:492–505
- Keeling CI, Weisshaar S, Lin RPC, Bohlmann J (2008) Functional plasticity of paralogous diterpene synthases involved in conifer defense. *Proc Natl Acad Sci USA* 105:1085–1090
- Kui L, Chen H, Zhang W, He S, Xiong Z, Zhang Y, Yan L, Zhong C, He F, Chen J, Zeng P, Zhang G, Yang S, Dong Y, Wang W, Cai J (2016) Building a genetic manipulation tool box for orchid biology: identification of constitutive promoters and application of CRISPR/Cas9 in the orchid, *Dendrobium officinale*. *Front Plant Sci* 7:2036
- Külheim C, Padovan A, Hefer C, Krause ST, Köllner TG, Myburg AA, Degenhardt J, Foley WJ (2015) The *Eucalyptus* terpene synthase gene family. *BMC Genomics* 16:450
- Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874
- Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA, Shomura A, Shimizu T, Lin SY, Inoue T, Fukuda A, Shimano T, Kuboki Y, Toyama T, Miyamoto Y, Kirihara T, Hayasaka K, Miyao A, Monna L, Zhong HS, Tamura Y, Wang ZX, Momma T, Umehara Y, Yano M, Sasaki T, Minobe Y (1994) A 300 kilobase interval genetic map of rice including 883 expressed sequences. *Nat Genet* 8:365–372
- Lee YI, Yang CK, Gebauer G (2015) The importance of associations with saprotrophic non-Rhizoctonia fungi among fully mycoheterotrophic orchids is currently under-estimated: novel evidence from sub-tropical Asia. *Ann Bot* 116:423–435
- Lehti-Shiu MD, Zou C, Hanada K, Shiu SH (2009) Evolutionary history and stress regulation of plant receptor-like kinase/pelle genes. *Plant Physiol* 150:12–26
- Leitch IJ, Kahandawala I, Suda J, Hanson L, Ingrouille MJ, Chase MW, Fay MF (2009) Genome size diversity in orchids: consequences and evolution. *Ann Bot* 104:469–481
- Li X, Luo J, Yan T, Xiang L, Jin F, Qin D, Sun C, Xie M (2014) Deep sequencing-based analysis of the *Cymbidium ensifolium* floral transcriptome. *PLoS ONE* 8:e85480
- Li C, Bai G, Chao S, Wang Z (2015a) A high-density SNP and SSR consensus map reveals segregation distortion regions in wheat. *Biomed Res Int* 2015:10
- Li D, Zhao C, Liu X, Liu X, Lin Y, Liu J, Chen H, Lü F (2015b) *De novo* assembly and characterization of the root transcriptome and development of simple sequence repeat markers in *Paphiopedilum concolor*. *Genet Mol Res* 14:6189–6201
- Li X, Jin F, Jin L, Jackson A, Ma X, Shu X, Wu D, Jin G (2015c) Characterization and comparative profiling of the small RNA transcriptomes in two phases of flowering in *Cymbidium ensifolium*. *BMC Genomics* 16:622
- 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 Am Soc Hort Sci* 126:195–199
- Lin CS, Chen JJW, Huang YT, Hsu CT, Lu HC, Chou ML, Chen LC, Ou CI, Liao DC, Yeh YY, Chang SB, Shen SC, Wu FH, Shih MC, Chan MT (2013a) Catalog of *Erycina pusilla* miRNA and categorization of reproductive phase-related miRNAs and their target gene families. *Plant Mol Biol* 82:193–204
- Lin W, Ma X, Shan L, He P (2013b) Big roles of small kinases: the complex functions of receptor-like cytoplasmic kinases in plant immunity and development. *J Integr Plant Biol* 55:1188–1197
- Lin YF, Chen YY, Hsiao YY, Shen CY, Hsu JL, Yeh CM, Mitsuda N, Ohme-Takagi M, Liu ZJ, Tsai WC (2016) Genome-wide identification and characterization of TCP genes involved in ovule development of *Phalaenopsis equestris*. *J Exp Bot* 67:5051–5066
- Liu P, Wei W, Ouyang S, Zhang JS, Chen SY, Zhang WK (2009) Analysis of expressed receptor-like kinases (RLKs) in soybean. *J Genet Genomics* 36:611–619
- Liu SS, Chen J, Li SC, Zeng X, Meng ZX, Guo SX (2015) Comparative transcriptome analysis of genes involved in GA-GID1-DELLA regulatory module in symbiotic and asymbiotic seed germination of *Anoectochilus roxburghii* (Wall.) Lindl. (Orchidaceae). *Int J Mol Sci* 16:26224
- Lu HC, Chen HH, Tsai WC, Chen WH, Su HJ, Chang DC, Yeh HH (2007) Strategies for functional validation of genes involved in reproductive stages of orchids. *Plant Physiol* 143:558–569
- Lu JJ, Wang S, Zhao HY, Liu JJ, Wang HZ (2012a) Genetic linkage map of EST-SSR and SRAP markers in the endangered Chinese endemic herb *Dendrobium* (Orchidaceae). *Genet Mol Res* 11:4654–4667
- Lu JJ, Zhao HY, Suo NN, Wang S, Shen B, Wang HZ, Liu JJ (2012b) Genetic linkage maps of *Dendrobium moniliforme* and *D. officinale* based on EST-SSR, SRAP, ISSR and RAPD markers. *Sci Hortic (Amsterdam)* 137:1–10
- Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, Buckler ES, Costich DE (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genet* 9:e1003215
- Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. *Science* 339:823–826
- Manning G, Plowman GD, Hunter T, Sudarsanam S (2002) Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci* 27:514–520
- Martin DM, Faldt J, Bohlmann J (2004) Functional characterization of nine Norway spruce TPS genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. *Plant Physiol* 135:1908–1927
- Martin DMA, Miranda-Saavedra D, Barton GJ (2009) Kinomer v. 1.0: a database of systematically classified eukaryotic protein kinases. *Nucleic Acids Res* 37:D244–D250
- Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J, Bayer M, Ramsay L, Liu H, Haberer G, Zhang XQ, Zhang Q, Barrero RA, Li L, Taudien S, Groth M, Felder M, Hastie A, Simkova H, Stankova H, Vrana J, Chan S, Munoz-Amatriaín M, Ounit R, Wanamaker S, Bolser D, Colmsee C, Schmutz T, Aliyeva-Schnorr L, Grasso S, Tanskanen J, Chailyan A, Sampath D, Heavens D, Clissold L, Cao S, Chapman B, Dai F, Han Y, Li H, Li X, Lin C, McCooke JK, Tan C, Wang P, Wang S, Yin S, Zhou G, Poland JA, Bellgard MI, Borisjuk L, Houben A, Dolezel J, Ayling S, Lonardi S, Kersey P, Langridge P, Muehlbauer GJ, Clark MD, Caccamo M, Schulman AH, Mayer KFX, Platzer M, Close TJ, Scholz U, Hansson M, Zhang G, Braumann I, Spannagl M, Li C, Waugh R, Stein N (2017) A chromosome conformation capture ordered sequence of the barley genome. *Nature* 544:427–433
- Matsuba Y, Nguyen TT, Wiegert K, Falara V, Gonzales-Vigil E, Leong B, Schafer P, Kudrna D, Wing RA, Bolger AM, Usadel B, Tissier A, Fernie AR, Barry CS, Pichersky E (2013) Evolution of a complex locus for terpene biosynthesis in solanum. *Plant Cell* 25:2022–2036
- Meng Y, Yu D, Xue J, Lu J, Feng S, Shen C, Wang H (2016) A transcriptome-wide, organ-specific regulatory map of *Dendrobium officinale*, an important traditional Chinese orchid herb. *Sci Rep* 6:18864
- Niu SC, Xu Q, Zhang GQ, Zhang YQ, Tsai WC, Hsu JL, Liang CK, Luo YB, Liu ZJ (2016) *De novo* transcriptome assembly databases for the butterfly orchid *Phalaenopsis equestris*. *Sci Data* 3:160083
- Ogura T, Busch W (2015) From phenotypes to causal sequences: using genome wide association studies to dissect the sequence basis for variation of plant development. *Curr Opin Plant Biol* 23:98–108
- Oliver KR, McComb JA, Greene WK (2013) Transposable elements: powerful contributors to angiosperm evolution and diversity. *Genome Biol Evol* 5:1886–1901
- Otero JT, Flanagan NS (2006) Orchid diversity—beyond deception. *Trends Ecol Evol* 21:64–65 (author reply 65–66)
- Pan IC, Liao DC, Wu FH, Daniell H, Singh ND, Chang C, Shih MC, Chan MT, Lin CS (2012) Complete chloroplast genome sequence of an orchid model plant candidate: *Erycina pusilla* apply in tropical *Oncidium* breeding. *PLoS ONE* 7:e34738
- Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc Natl Acad Sci USA* 101:9903–9908
- Pawelkiewicz M, Zieliński K, Zielińska D, Płader W, Yagi K, Wojcieszek M, Siedlecka E, Bartoszewski G, Skarzyńska A, Przybecki Z (2016) Next generation sequencing and omics in cucumber (*Cucumis sativus* L.) breeding directed research. *Plant Sci* 242:77–88
- Perotto S, Rodda M, Benetti A, Sillo F, Ercole E, Rodda M, Girlanda M, Murat C, Balestrini R (2014) Gene expression in mycorrhizal orchid protocorms suggests a friendly plant–fungus relationship. *Planta* 239:1337–1349

- Piegu B, Guyot R, Picault N, Roulin A, Saniyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA (2006) Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res* 16:1262–1269
- Poland JA, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* 5:92–102
- Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sánchez-Villeda H, Sorrells M, Jannink JL (2012a) Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* 5:103–113
- Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012b) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 7:e32253
- Price MN, Dehal PS, Arkin AP (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 26:1641–1650
- Ramirez SR, Gravendeel B, Singer RB, Marshall CR, Pierce NE (2007) Dating the origin of the Orchidaceae from a fossil orchid with its pollinator. *Nature* 448:1042–1045
- Ramsay L, Macaulay M, Ivanissevich Sd, MacLean K, Cardle L, Fuller J, Edwards KJ, Turesson S, Morgante M, Massari A, Maestri E, Marmiroli N, Sjakste T, Ganai M, Powell W, Waugh R (2000) A simple sequence repeat-based linkage map of barley. *Genetics* 156:1997–2005
- Rao X, Krom N, Tang Y, Widiez T, Havkin-Frenkel D, Belanger FC, Dixon RA, Chen F (2014) A deep transcriptomic analysis of pod development in the vanilla orchid (*Vanilla planifolia*). *BMC Genomics* 15:964
- Rasmussen HN, Rasmussen FN (2009) Orchid mycorrhiza: implications of a mycophagous life style. *Oikos* 118:334–345
- Romay MC, Millard MJ, Glaubit JC, Peiffer JA, Swarts KL, Casstevens TM, Elshire RJ, Acharya CB, Mitchell SE, Flint-Garcia SA, McMullen MD, Holland JB, Buckler ES, Gardner CA (2013) Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol* 14:R55
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Sakamoto T, Deguchi M, Brustolini OJ, Santos AA, Silva FF, Fontes EP (2012) The tomato RLK superfamily: phylogeny and functional predictions about the role of the LRRIL-RLK subfamily in antiviral defense. *BMC Plant Biol* 12:229
- Sattler MC, Carvalho CR, Clarindo WR (2016) The polyploidy and its key role in plant breeding. *Planta* 243:281–296
- Sedek KEM, Qi W, Schauer MA, Gupta AK, Poveda L, Xu S, Liu ZJ, Grossniklaus U, Schiestl FP, Schlüter PM (2013) Transcriptome and proteome data reveal candidate genes for pollinator attraction in sexually deceptive orchids. *PLoS ONE* 8:e64621
- Sedek KEM, Scopece G, Staedler Y, Schönenberger J, Cozzolino S, Schiestl FP, Schlüter PM (2014) Genic rather than genome-wide differences between sexually deceptive *Ophrys* orchids with different pollinators. *Mol Ecol* 23:6192–6205
- Shimura K, Okada A, Okada K, Jikumaru Y, Ko KW, Toyomasu T, Sassa T, Hasegawa M, Kodama O, Shibuya N, Koga J, Nojiri H, Yamane H (2007) Identification of a biosynthetic gene cluster in rice for momilactones. *J Biol Chem* 282:34013–34018
- Shiu SH, Bleecker AB (2001) Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc Natl Acad Sci USA* 98:10763–10768
- Shiu SH, Bleecker AB (2003) Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in *Arabidopsis*. *Plant Physiol* 132:530–543
- Shiu SH, Karlowski WM, Pan R, Tzeng YH, Mayer KF, Li WH (2004) Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. *Plant Cell* 16:1220–1234
- Silvera K, Santiago LS, Cushman JC, Winter K (2009) Crassulacean acid metabolism and epiphytism linked to adaptive radiations in the Orchidaceae. *Plant Physiol* 149:1838–1847
- Slate J, Gratten J, Beraldi D, Stapley J, Hale M, Pemberton JM (2009) Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica* 136:97–107
- Sonah H, O'Donoghue L, Cober E, Rajcan I, Belzile F (2015) Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnol J* 13:211–221
- Sonnhammer EL, Eddy SR, Birney E, Bateman A, Durbin R (1998) Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res* 26:320–322
- Stankova H, Hastie AR, Chan S, Vrana J, Tulpova Z, Kubalaková M, Visendi P, Hayashi S, Luo M, Batley J, Edwards D, Dolezel J, Simkova H (2016) BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes. *Plant Biotechnol J* 14:1523–1531
- Su CL, Chao YT, Alex Chang YC, Chen WC, Chen CY, Lee AY, Hwa KT, Shih MC (2011) *De novo* assembly of expressed transcripts and global analysis of the *Phalaenopsis aphrodite* transcriptome. *Plant Cell Physiol* 52:1501–1514
- Su CL, Chao YT, Yen SH, Chen CY, Chen WC, Chang YC, Shih MC (2013) Orchidstra: an integrated orchid functional genomics database. *Plant Cell Physiol* 54:e11
- Suetsugu K, Yamato M, Miura C, Yamaguchi K, Takahashi K, Ida Y, Shigenobu S, Kaminaka H (2017) Comparison of green and albino individuals of the partially mycoheterotrophic orchid *Epipactis helleborine* on molecular identities of mycorrhizal fungi, nutritional modes and gene expression in mycorrhizal roots. *Mol Ecol* 26:1652–1669
- Tan J, Wang HL, Yeh KW (2005) Analysis of organ-specific, expressed genes in *Oncidium* orchid by subtractive expressed sequence tags library. *Biotechnol Lett* 27:1517–1528
- Teague B, Waterman MS, Goldstein S, Potamouis K, Zhou S, Reslewic S, Sarkar D, Valouev A, Churas C, Kidd JM, Kohn S, Runnheim R, Lamers C, Forrest D, Newton MA, Eichler EE, Kent-First M, Surti U, Livny M, Schwartz DC (2010) High-resolution human genome structure by single-molecule analysis. *Proc Natl Acad Sci USA* 107:10848–10853
- Teh SL, Chan WS, Abdullah JO, Namasivayam P (2011) Development of expressed sequence tag resources for *Vanda Mimi* Palmer and data mining for EST-SSR. *Mol Biol Rep* 38:3903–3909
- Truong HT, Ramos AM, Yalcin F, de Ruiter M, van der Poel HJA, Huvenaars KHJ, Hogers RCJ, van Enckevort LJJ, Janssen A, van Orsouw NJ, van Eijk MJT (2012) Sequence-based genotyping for marker discovery and co-dominant ccoing in germplasm and populations. *PLoS ONE* 7:e37565
- Tsai WC, Hsiao YY, Lee SH, Tung CW, Wang DP, Wang HC, Chen WH, Chen HH (2006) Expression analysis of the ESTs derived from the flower buds of *Phalaenopsis equestris*. *Plant Sci* 170:426–432
- Tsai WC, Fu CH, Hsiao YY, Huang YM, Chen LJ, Wang M, Liu ZJ, Chen HH (2013) OrchidBase 2.0: comprehensive collection of Orchidaceae floral transcriptomes. *Plant Cell Physiol* 54:e7
- Tsai CC, Wu KM, Chiang TY, Huang CY, Chou CH, Li SJ, Chiang YC (2016) Comparative transcriptome analysis of *Gastrodia elata* (Orchidaceae) in response to fungus symbiosis to identify gastrodin biosynthesis-related genes. *BMC Genomics* 17:212
- Uitdewilligen JGAML, Wolters A-MA, D'hoop BB, Borm TJA, Visser RGF, van Eck HJ (2013) A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS ONE* 8:e62355
- Vaid N, Macovei A, Tuteja N (2013) Knights in action: lectin receptor-like kinases in plant development and stress responses. *Mol Plant* 6:1405–1418
- Valadares RBS, Perotto S, Santos EC, Lambais MR (2014) Proteome changes in *Oncidium sphacelatum* (Orchidaceae) at different trophic stages of symbiotic germination. *Mycorrhiza* 24:349–360
- Van de Peer Y, Maere S, Meyer A (2009) The evolutionary significance of ancient genome duplications. *Nat Rev Genet* 10:725–732
- van Poecke RMP, Maccaferri M, Tang J, Truong HT, Janssen A, van Orsouw NJ, Salvi S, Sanguineti MC, Tuberosa R, van der Vossen EAG (2013) Sequence-based SNP genotyping in durum wheat. *Plant Biotechnol J* 11:809–817
- Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends Biotechnol* 23:48–55
- Walsh RM, Hochedlinger K (2013) A variant CRISPR-Cas9 system adds versatility to genome engineering. *Proc Natl Acad Sci USA* 110:15514–15515
- Wen W, Liu H, Zhou Y, Jin M, Yang N, Li D, Luo J, Xiao Y, Pan Q, Tohge T, Fernie AR, Yan J (2016) Combining quantitative genetics approaches with regulatory network analysis to dissect the complex metabolism of the maize kernel. *Plant Physiol* 170:136–146
- Wendel JF, Jackson SA, Meyers BC, Wing RA (2016) Evolution of plant genome architecture. *Genome Biol* 17:37

- Wiedenheft B, Sternberg SH, Doudna JA (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482:331–338
- Wisecaver JH, Borowsky AT, Tzin V, Jander G, Kliebenstein DJ, Rokas A (2017) A global co-expression network approach for connecting genes to specialized metabolic pathways in plants. *Plant Cell* 29:944–959
- Woo JW, Kim J, Kwon SJ, Corvalan C, Cho SW, Kim H, Kim SG, Kim ST, Choe S, Kim JS (2015) DNA-free genome editing in plants with preassembled CRISPR–Cas9 ribonucleoproteins. *Nat Biotechnol* 33:1162–1164
- Wu Y, Zhou JM (2013) Receptor-like kinases in plant innate immunity. *J Integr Plant Biol* 55:1271–1286
- Wyman C, Kanaar R (2006) DNA double-strand break repair: all's well that ends well. *Annu Rev Genet* 40:363–383
- Xiao S, Li J, Ma F, Fang L, Xu S, Chen W, Wang ZY (2015) Rapid construction of genome map for large yellow croaker (*Larimichthys crocea*) by the whole-genome mapping in BioNano Genomics Irys system. *BMC Genomics* 16:670
- Xu C, Zeng B, Huang J, Huang W, Liu Y (2015) Genome-wide transcriptome and expression profile analysis of *Phalaenopsis* during explant browning. *PLoS ONE* 10:e0123356
- Xue D, Feng S, Zhao H, Jiang H, Shen B, Shi N, Lu J, Liu J, Wang H (2010) The linkage maps of *Dendrobium* species based on RAPD and SRAP markers. *J Genet Genomics* 37:197–204
- Yan L, Wang X, Liu H, Tian Y, Lian J, Yang R, Hao S, Wang X, Yang S, Li Q, Qi S, Kui L, Okpekum M, Ma X, Zhang J, Ding Z, Zhang G, Wang W, Dong Y, Sheng J (2015) The genome of *Dendrobium officinale* illuminates the biology of the important traditional Chinese orchid herb. *Mol Plant* 8:922–934
- Yang F, Zhu G (2015) Digital gene expression analysis based on *de novo* transcriptome assembly reveals new genes associated with floral organ differentiation of the orchid plant *Cymbidium ensifolium*. *PLoS ONE* 10:e0142434
- Yang H, Tao Y, Zheng Z, Li C, Sweetingham MW, Howieson JG (2012) Application of next-generation sequencing for rapid marker development in molecular plant breeding: a case study on anthracnose disease resistance in *Lupinus angustifolius* L. *BMC Genomics* 13:318
- Ye W, Shen CH, Lin Y, Chen PJ, Xu X, Oelmüller R, Yeh KW, Lai Z (2014) Growth promotion-related miRNAs in *Oncidium* orchid roots colonized by the endophytic fungus *Piriformospora indica*. *PLoS ONE* 9:e84920
- Ye Y, Ding Y, Jiang Q, Wang F, Sun J, Zhu C (2017) The role of receptor-like protein kinases (RLKs) in abiotic stress response in plants. *Plant Cell Rep* 36:235–242
- Yu H, Goh CJ (2000) Identification and characterization of three orchid *MADS*-box genes of the *AP1/AGL9* subfamily during floral transition. *Plant Physiol* 123:1325–1336
- Zan Y, Ji Y, Zhang Y, Yang S, Song Y, Wang J (2013) Genome-wide identification, characterization and expression analysis of populus leucine-rich repeat receptor-like protein kinase genes. *BMC Genomics* 14:318
- Zhang J, Wu K, Zeng S, Teixeira da Silva JA, Zhao X, Tian C-E, Xia H, Duan J (2013) Transcriptome analysis of *Cymbidium sinense* and its application to the identification of genes associated with floral development. *BMC Genomics* 14:279
- Zhang GQ, Xu Q, Bian C, Tsai WC, Yeh CM, Liu KW, Yoshida K, Zhang LS, Chang SB, Chen F, Shi Y, Su YY, Zhang YQ, Chen LJ, Yin Y, Lin M, Huang H, Deng H, Wang ZW, Zhu S, Zhao X, Deng C, Niu SC, Huang J, Wang M, Liu GH, Yang HJ, Xiao XJ, Hsiao YY, Wu WL, Chen YY, Mitsuda N, Ohme-Takagi M, Luo YB, Van de Peer Y, Liu ZJ (2016a) The *Dendrobium catenatum* Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. *Sci Rep* 6:19029
- Zhang J, He C, Wu K, Teixeira da Silva JA, Zeng S, Zhang X, Yu Z, Xia H, Duan J (2016b) Transcriptome analysis of *Dendrobium officinale* and its application to the identification of genes associated with polysaccharide synthesis. *Front Plant Sci* 7:5
- Zhang L, Chen F, Zhang GQ, Zhang YQ, Niu SC, Xiong JS, Lin Z, Cheng ZM, Liu ZJ (2016c) Origin and mechanism of crassulacean acid metabolism in orchids as implied by comparative transcriptomics and genomics of the carbon fixation pathway. *Plant J* 86:175–185
- Zhang GQ, Liu KW, Li Z, Lohaus R, Hsiao YY, Niu SC, Wang JY, Lin YC, Xu Q, Chen LJ, Yoshida K, Fujiwara S, Wang ZW, Zhang YQ, Mitsuda N, Wang M, Liu GH, Pecoraro L, Huang HX, Xiao XJ, Lin M, Wu XY, Wu WL, Chen YY, Chang SB, Sakamoto S, Ohme-Takagi M, Yagi M, Zeng SJ, Shen CY, Yeh CM, Luo YB, Tsai WC, Van de Peer Y, Liu ZJ (2017) The *apostasia* genome and the evolution of orchids. *Nature* 549:379–383
- Zhao X, Zhang J, Chen C, Yang J, Zhu H, Liu M, Lv F (2014) Deep sequencing-based comparative transcriptional profiles of *Cymbidium hybridum* roots in response to mycorrhizal and non-mycorrhizal beneficial fungi. *BMC Genomics* 15:747
- Zhou S, Kile A, Bechner M, Place M, Kvikstad E, Deng W, Wei J, Severin J, Runnheim R, Churas C, Forrest D, Dimalanta ET, Lamers C, Burland V, Blattner FR, Schwartz DC (2004) Single-molecule approach to bacterial genomic comparisons via optical mapping. *J Bacteriol* 186:7773–7782
- Zhu G, Yang F, Shi S, Li D, Wang Z, Liu H, Huang D, Wang C (2015) Transcriptome characterization of *Cymbidium sinense* 'Dharma' using 454 pyrosequencing and its application in the identification of genes associated with leaf color variation. *PLoS ONE* 10:e0128592

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com