the plant journal

SEB Society for Experimental Biology

The Plant Journal (2016) **86,** 175–185 doi: 10.1111/tpj.13159

Origin and mechanism of crassulacean acid metabolism in orchids as implied by comparative transcriptomics and genomics of the carbon fixation pathway

Liangsheng Zhang[1,*,†], Fei Chen[2,3,†], Guo-Qiang Zhang[4], Yong-Qiang Zhang[4], Shance Niu[4], Jin-Song Xiong[2,3], Zhenguo Lin[5], Zong-Ming (Max) Cheng[2,3,*] and Zhong-Jian Liu[4,*]

[1]*Center for Genomics and Biotechnology, Haixia Institute of Science and Technology, Fujian Agriculture and Forestry University, Fuzhou 350002, China,*

[2]*College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China,*

[3]*Department of Plant Sciences, The University of Tennessee, Knoxville, TN 37996, USA,*

[4]*Shenzhen Key Laboratory for Orchid Conservation and Utilization, The National Orchid Conservation Center of China and The Orchid Conservation & Research Center of Shenzhen, Shenzhen 518114, China, and*

[5]*Department of Biology, Saint Louis University, St Louis, MO 63103, USA*

Received 23 October 2015; revised 3 March 2016; accepted 4 March 2016; published online 9 March 2016.
*For correspondence (e-mails fafuzhang@163.com; zmc@njau.edu.cn; liuzj@sinicaorchid.org).
†Co-first author.

SUMMARY

Crassulacean acid metabolism (CAM) is a CO_2 fixation pathway that maximizes water-use efficiency (WUE), compared with the C3/C4 CO_2 pathway, which permits CAM plants to adapt to arid environments. The CAM pathway provides excellent opportunities to genetically design plants, especially bioenergy crops, with a high WUE and better photosynthetic performance than C3/C4 in arid environments. The information available on the origin and evolution of CAM is scant, however. Here, we analyzed transcriptomes from 13 orchid species and two existing orchid genomes, covering CAM and C3 plants, with an emphasis on comparing 13 gene families involved in the complete carbon fixation pathway. The dosage of the core photosynthesis-related genes plays no substantial role in the evolution of CAM in orchids; however, CAM may have evolved primarily by changes at the transcription level of key carbon fixation pathway genes. We proposed that in both dark and light, CO_2 is primarily fixed and then released through two metabolic pathways via known genes, such as *PPC1*, *PPDK* and *PPCK*. This study reports a comprehensive comparison of carbon fixation pathway genes across different photosynthetic plants, and reveals the importance of the level of expression of key genes in the origin and evolution of CAM.

Keywords: carbon fixation, crassulacean acid metabolism (CAM), gene family, Orchidaceae, photosynthesis, transcriptome.

INTRODUCTION

Crassulacean acid metabolism (CAM) is a photosynthetic specialization beyond C3 and C4 that permits the opening of the stomatal aperture and net CO_2 uptake at night. Therefore, CAM improves the water-use efficiency (WUE) and facilitates adaptation to arid habitats (Kluge and Ting, 1978). CAM plants fix carbon in two steps (Borland *et al.*, 2009). First, CO_2 is converted into HCO_3^- by beta carbonic anhydrase (β-CA/BCA) and primarily fixed by phosphoenolpyruvate carboxylase (PEPC/PPC). PPC is tightly regulated by PEP carboxylase phosphatase (PPCP) and phosphoenolpyruvate carboxylase kinase (PPCK). As a phosphatase, PPCP inhibits PEPC by the removal of its phosphate group to stop its activity during the day. In contrast, PPCK activates PEPC by phosphorylation at night. Second, the primary and unstable product oxaloacetate (OAA) is converted into the stable form malate by NAD-dependent malate dehydrogenase [NAD(P)-MDH] and stored in a vacuole. During the day, malate is pumped out of the vacuole and converted into OAA by NAD(P)-MDH, or into pyruvate (PYR) by the malate enzyme (ME). The intermediate C4 acids are decarboxylated into phosphoenolpyruvate (PEP) through phosphoenolpyruvate carboxykinase (PCK) and pyruvate orthophosphate dikinase (PPDK), respectively. The released CO_2 enters chloro-

plasts where it is fixed by RuBisCO via the Calvin–Benson–Bassham (CBB) cycle, which is used by C3 plants as the sole CO_2 fixation step (Borland *et al.*, 2009; Silvera *et al.*, 2010; Borland and Yang, 2013; Depaoli *et al.*, 2014). These gene families/subfamilies involved in the CAM carbon fixation pathway have been described in biochemical studies only (Figure 1). By comparison, few genetic and transcriptomic analyses have been reported in CAM plants (Ming *et al.*, 2015).

In C4 plants the two steps of carbon fixation are spatially separated into the bundle sheath and mesophyll cells, whereas they are performed independently during the day and night in CAM plants, respectively. Compared with C4 and C3 plants, CAM plants use 20–80% less water to produce the same quantities of biomass (Borland *et al.*, 2009), and as such they have been remarkably successful in occupying water-restricted ecosystems, such as deserts. A full understanding of CAM may offer a blueprint for engineering CAM into other crop plants or trees via the tools of synthetic biology, to potentially boost crop yields in arid regions and in increasingly unpredictable global environments (Borland *et al.*, 2009; Borland and Yang, 2013; Depaoli *et al.*, 2014; Yang *et al.*, 2015). One prerequisite for such cross-photosystem engineering is the complete understanding of the evolution and molecular regulation of CAM (Smith and Winter, 1996; Silvera *et al.*, 2014), which remains obscure and largely elusive (Borland and Yang, 2013), however, because of the lack of a genome-scale analysis of CAM plant genomes and transcriptome. Even today, among the CAM regulation network, only detailed functional studies of *PPC* genes are available (Silvera *et al.*, 2014). In this study, we aim to identify additional putative CAM-related genes in the carbon fixation pathway and to identify the genetic background and regulation of CAM through a comparative study among plants that use various types of photosynthesis.

The Orchidaceae is the second largest angiosperm family, with approximately 25 000 species (Cribb and Govaerts, 2005); of these, 10 000 are estimated to be CAM or C3–CAM orchids (based on data from Silvera *et al.*, 2009). In this sense, the Orchidaceae is the largest CAM clade. Therefore, sequencing and comparing genomes and transcriptomes of representative types of orchids will provide an unprecedented opportunity to understand the evolution and molecular regulation of CAM. There is still no comparative omics study of CAM in orchids, however, so the knowledge of how CAM evolved to become an important type of photosynthesis in the Orchidaceae is lacking. Here, we performed a systematic analysis of the gene families encoding carbon fixation enzymes in 10 orchids, along

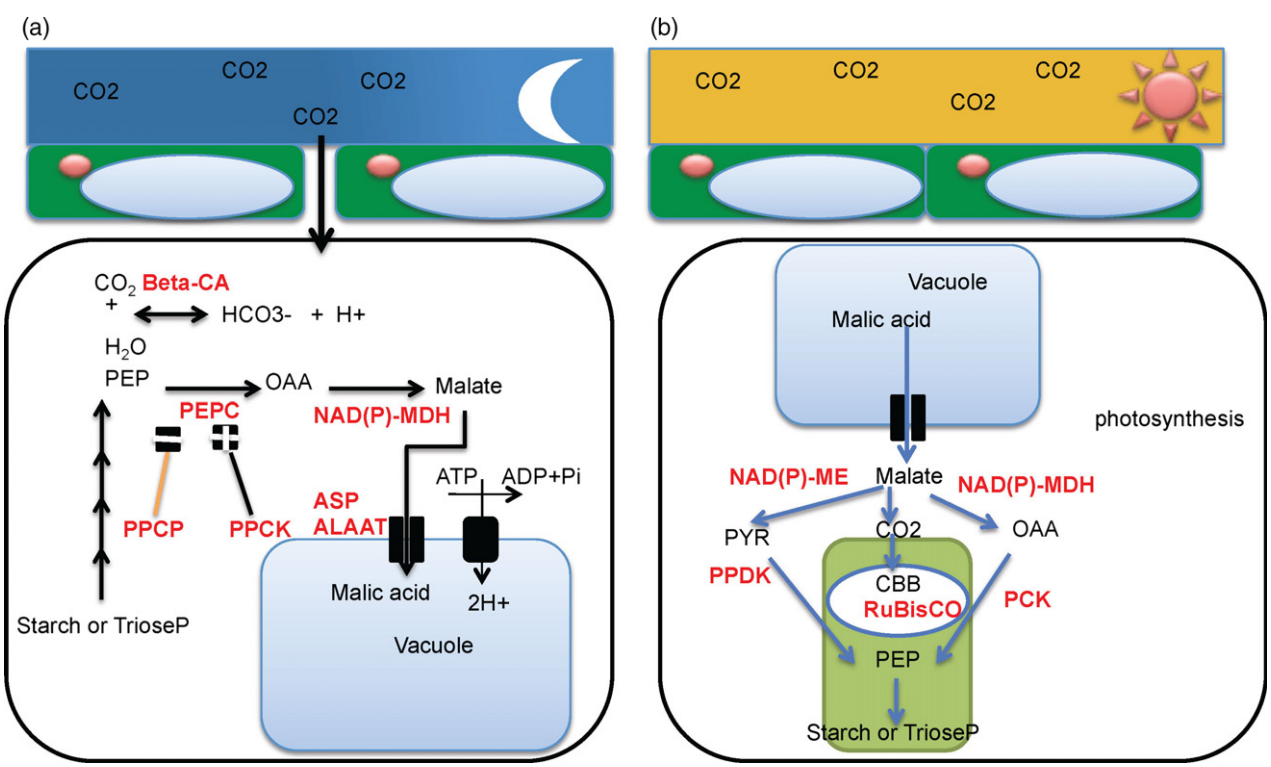

Figure 1. Carbon fixation metabolism pathways of crassulacean acid metabolism (CAM) involving core genes/enzymes. (a) Nocturnal reactions (black lines) are shown when stomata are open. (b) Diurnal reactions (blue lines) are shown when stomata are closed. Genes/enzymes are labeled in red. The yellow line indicates inhibition during the day.

with those of other C3 and C4 land plants. Our results suggest that the CAM pathway in Orchidaceae may be primarily controlled by the transcriptional regulation of key genes in spatial and temporal manners, rather than by the gene dosage effect, but with moderate-sized photosynthesis-related genes.

RESULTS

Transcriptome sequencing, assembly and annotation of 13 orchids

The genomes and transcriptomes of 13 orchids were used to compare the genetic background of CAM plants with C3 and C4 plants. The genomes of *Phalaenopsis equestris* (Cai *et al.*, 2014) and *Dendrobium catenatum* (Zhang *et al.*, 2015) were available. We also labeled the photosynthetic types according to the $\delta^{13}C$ (‰) value (Table S1), a broadly accepted indicator for identifying CAM, and *P. equestris*, *Dendrobium terminale*, *Cymbidium atropurpureum* and *Cymbidium mannii* were indicated as CAM plants (Table S1). The transcriptomes of three representative CAM orchids, *P. equestris*, *D. terminale* and *C. mannii*, were sampled under both light and dark conditions. All transcriptomes of the orchids were assembled and annotated (see Experimental procedures), and show high levels of overlapping genes compared with sequenced genomes of *P. equestris* and *D. catenatum*, and other model plants such as *Arabidopsis thaliana* and *Oryza sativa* (rice). This suggests that these orchid transcriptomes have similar high-level, nearly genome-wide coverage (Figure S1), and are therefore suitable for phylogenetic tree construction and the comparison of CAM genes. Based on the annotated genes, we performed a phylogenetic analysis of the 13 orchids, and other land plants, relying on 82 nuclear orthologous genes (Figure 2).

Phylogenetic trees of the 13 carbon fixation related gene families

The carbon fixation pathway is one of the most remarkable innovations in CAM plants, and it is a dominant trait that clearly distinguishes these plants from C3 and C4 plants. Because its unique primary assimilation and secondary fixation of carbon are controlled by a circadian clock, CAM plants allow the opening of the stomatal aperture under dark conditions, but not under light conditions, thus allowing the plant to save water and to survive in arid environments. We selected genes that might be involved in the carbon fixation pathway and constructed a phylogenetic tree for each of the 13 gene families involved in carbon fixation, using lineages of the Orchidaceae and other land plants (Figures 3 and S2–S14). Among the 13 gene families, the *PPCK*, *PPDK*, *PCK*, *RBCS* and *PPCP* genes have a single origin in land plants, whereas the remaining genes have origins in two or three clusters. The *PPC* genes are the key genes that bind CO_2 in CAM plants and can be divided into the fol-

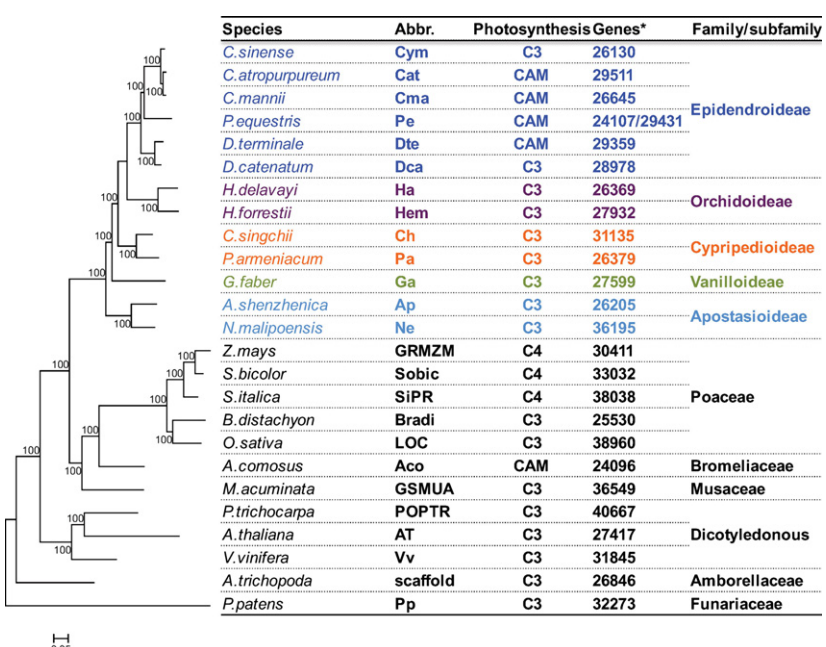

Figure 2. The phylogenetic tree of the 13 orchids in the context of land plants. The maximum-likelihood (ML) tree of the 13 orchids and other land plants used 82 single-copy genes. Photosynthetic types are shown in the second column. The five colors represent the five subfamilies of orchids: red for the crown subfamily Epidendroideae; green for Orchidoideae; magenta for Cypripedioideae; blue for Vanilloideae; and light blue for Apostasioideae. The stars are the number of genes that come from the annotation of transcriptomes, except for *Phalaenopsis equestris*, which was calculated based on both transcriptomes and the genome.

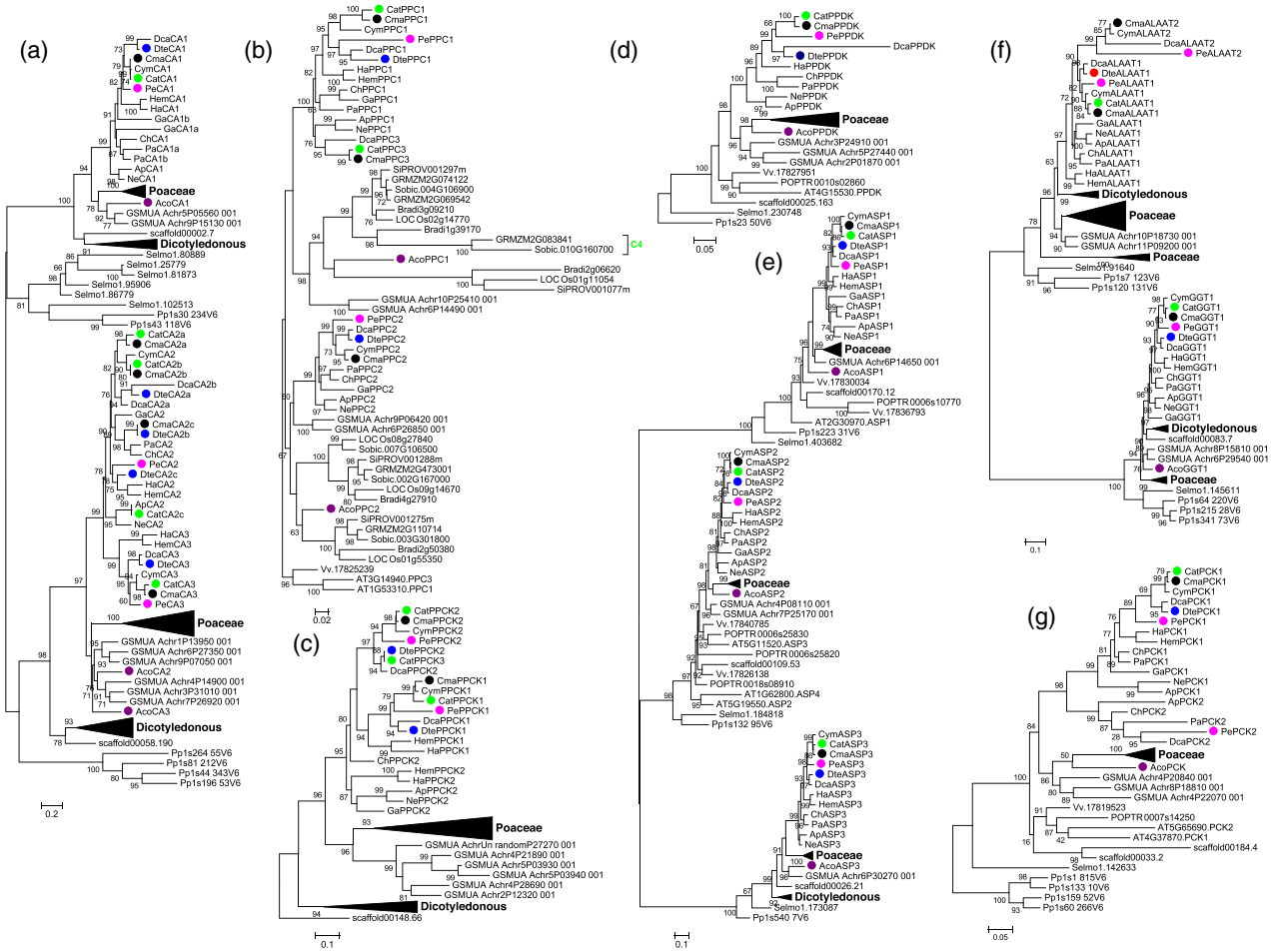

Figure 3. Phylogenetic trees of the 13 gene families involved in carbon fixation from the 13 orchids and other representative land plants: (a) *b-CA*; (b) *PPC*; (c) *PPCK*; (d) *PPDK*; (e) *ASP*; (f) *ALAAT*; (g) *PCK*; (h) *NAD-ME*; (j) *NADP-ME*; (k) *NADH-MDH*; (i) *NADP-MDH*; (m) *RBCS*; and (n) *PPCP*. Names of orchid genes are abbreviated as two or three letters, whereas those of other sequenced genomes are shown in the original gene ID. Detailed species information is provided in Figure 2. Different colors of dots represent different groups of orchids: magenta for *Phalaenopsis equestris*; green for *Cymbidium atropurpureum*; navy blue for *Dendrobium terminale*; black for *Cymbidium mannii*; and purple for *Ananas comosus*.

lowing two groups in each orchid: *PPC-1* and *PPC-2* (Figure 3b). The monocot *PPC*s have two clusters, whereas eudicots have only one cluster as the out-group, indicating that gene duplication occurred in an ancestor of monocots. In all, the 13 gene families have a total of 23 clusters and each cluster has orthologs from our annotated orchids (Tables S2 and S3), indicative of the sufficient coverage of the transcriptomes.

Among the 23 gene clusters, six clusters, including β-*CA*, *PPCK*, *NADP-ME*, *ALAAT*, *RBCS* and *PCK*, have undergone one duplication within the orchid lineage, whereas the other families or subfamilies have a single copy in the orchid lineage (Figures S2–S14). Except for *RuBisCO*, all genes have duplications shared by the CAM orchids *P. equestris*, *D. terminale* and *C. mannii*. The duplication of *RuBisCO* in *P. equestris* is species or genus specific, whereas *D. terminale* and *C. mannii* shared the same duplication (Figure 3). Noticeably, among all duplications, both C3 and CAM orchids share a recent duplication. Considering all of the enrolled gene families (details in Tables S2 and S3), the CAM orchid *P. equestris* has smaller or equal numbers of genes compared with the C4 plants *Zea mays* (maize) and sorghum, but no greater than those in C3 land plants (Table 1). Therefore, the copy numbers of these genes are conserved, with small variations among C3, C4 and CAM plants, and are especially conserved in C3 and CAM orchids. Compared with eudicot and Poaceae species, orchids have fewer core CAM genes because eudicot and Poaceae species have undergone successive whole-genome duplications (WGDs), and the rapid evolution in orchids have led to large-scale gene loss (Hsiao *et al.*, 2011) or fewer WGDs (Cai *et al.*, 2014). These results strongly suggest that CAM is not regulated by gene dosage or gene duplication.

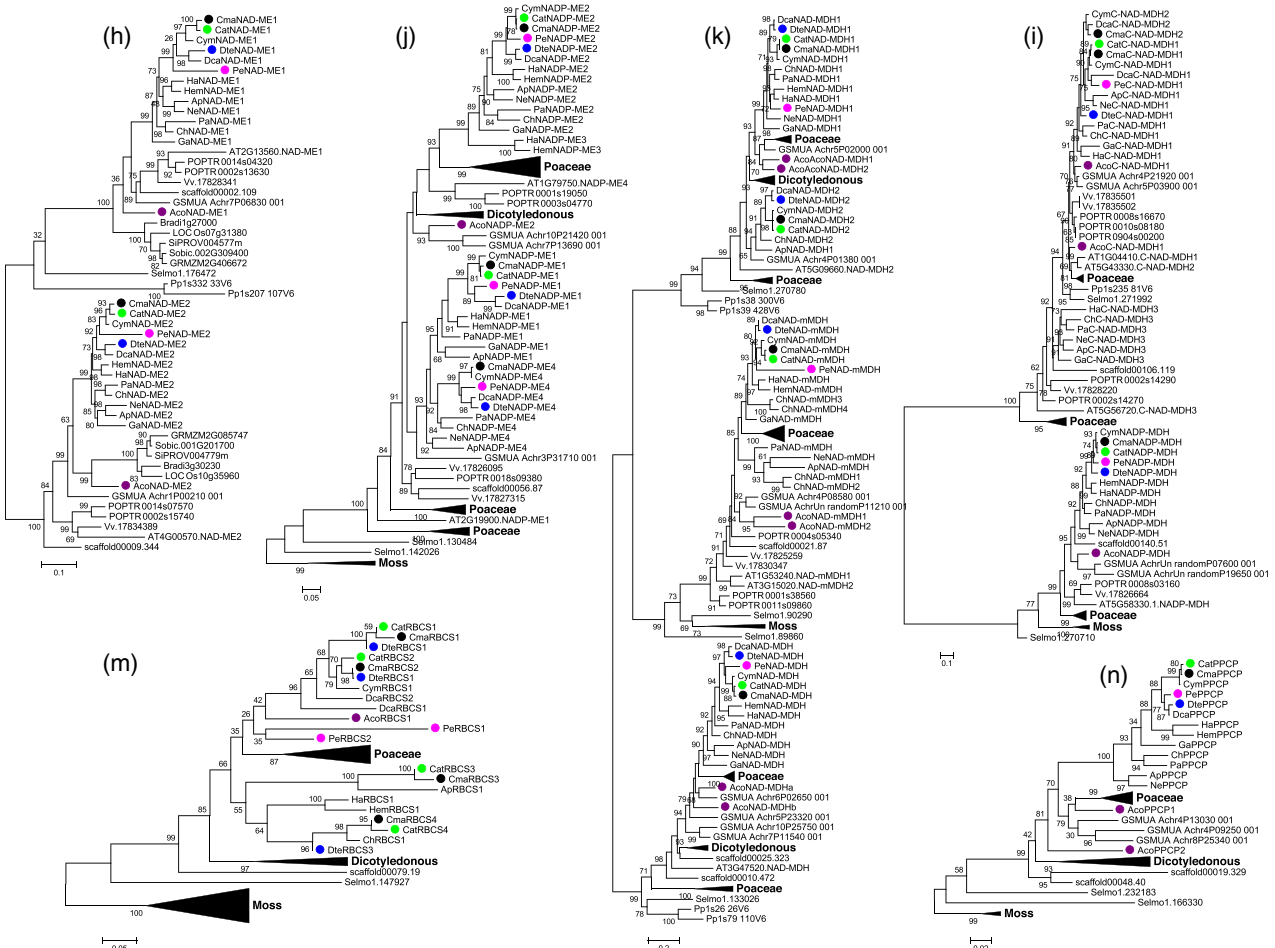

Figure 3. (Continued).

Expressional profiling of the identified genes from the 13 gene families involved in the carbon fixation pathway in CAM

As the gene copy number of carbon fixation-related genes is unlikely to be involved in modulating CAM in Orchidaceae, we examined the expression of genes potentially involved in carbon fixation in CAM orchids in various tissues to test whether differential transcriptional regulation could occur (Figures 4 and S15). We measured the expression of genes involved in carbon fixation with biochemical validation in *P. equestris* and *Apostasia shenzhenica*, and compared them with the available data from rice, maize, and Arabidopsis, in leaf, root, stem (siliques in *Arabidopsis*) and flower tissues (Tables S4–S8). We also analyzed the expression of these genes in response to the circadian rhythm of *P. equestris*, *C. mannii* and *D. terminale* (Figure 4; Tables S9–11). The CAM-related genes show higher expressions in leaves and at night than during the day in *P. equestris, C. mannii* and *D. terminale*. *RBCS2&1*, *NADP-MDH2*, *ASP3*, *NAD-ME1*, *PPDK*, *BCA2&1* and *ASP1* showed threefold higher expression in leaf tissue than in other tissues (Figure S15a). Because the orthologs of *BCA*, *GGT1* and *RBCS* also have high expression in the leaves of the C3 orchid *A. shenzhenica*, and the C3 eudicot Arabidopsis and monocot rice, and C4 maize (Figure S15b–e), they are not involved specifically in the CAM pathway. They may function in other developmental processes of leaf tissue, however. *PPC1*, *PCK*, *PPDK* and *NAD-MDH* also exhibited high expression in the leaves of C4 maize, but not in C3 leaves, suggesting that these four genes are likely to be involved in the carbon fixation pathways in both C4 and CAM plants.

CAM is regulated by controlling key gene expression

Interestingly, we also found key regulated genes in CAM. *PPCK1*, *PPCK2*, *PPC1* and *PPDK* had a threefold greater expression in the dark than in the light in leaf tissues (Figure 4a) in *P. equestris*. Likewise, in *C. mannii* and *D. terminale*, *PPCK1*, *PPCK2*, *PPC1* and *PPDK* genes exhibited higher expression in the dark than in the light, suggesting an important role in CAM. Taken together, these results strongly suggest that the evolution of CAM occurred

Table 1 The numbers of each CAM-related genes (13 gene families) in each taxon.

Species	Pho[a]	CA	PPC	PPCK[b]	PPDK	NAD-ME	NADP-ME	NAD-MDH	NADP-MDH	ALAAT	ASP	RBCS	PCK[b]	PPCP
Cymbidium sinense	C3	3	2	2	0	2	3	4	3	3	3	1	1	1
Cymbidium atropurpureum	CAM	4	3	2	1	2	3	4	2	2	3	4	1	1
Cymbidium mannii	CAM	6	2	3	1	2	3	4	3	3	3	4	1	1
Phalaenopsis equestris	CAM	3	2	2/0	1	2	3	3	2	3	3	2	1/2	1
Dendrobium terminale	CAM	4	3	2	1	2	3	4	2	2	3	3	1	1
Dendrobium catenatum	C3	4	3	2	1	2	3	4	2	3	3	2	2	1
Habenaria delavayi	C3	3	1	2	1	2	3	3	3	2	3	1	1	1
Hemipilia forrestii	C3	3	2	2	0	2	3	3	2	2	3	1	1	1
Cypripedium singchii	C3	2	2	1	1	2	2	7	3	2	3	1	2	1
Paphiopedilum armeniacum	C3	3	2	0	1	2	3	3	3	2	3	0	2	1
Galeola faberi	C3	3	2	1	0	2	2	3	3	2	2	0	1	1
Neuwiedia malipoensis	C3	2	2	1	1	2	2	3	3	2	3	0	1	1
Apostasia shenzhenica	C3	2	2	1	1	2	3	3	3	2	3	1	2	1
Zea mays	C4	6	5	4	2	2	5	6	4	6	3	2	2	2
Sorghum bicolor	C4	5	5	3	2	2	6	6	4	5	3	1	1	2
Setaria italica	C4	11	4	4	3	2	8	10	5	7	8	3	5	2
Brachypodium distachyon	C3	3	5	3	1	2	5	6	3	5	4	4	1	2
Oryza sativa	C3	2	5	2	2	2	4	7	3	6	4	3	1	1
Ananas comosus	CAM	3	2	0	1	2	1	6	2	2	3	1	1	2
Musa acuminata	C3	8	5	6	3	2	3	8	5	5	4	0	3	3
Populus trichocarpa	C3	9	3	4	1	4	5	9	6	4	6	2	2	2
Arabidopsis thaliana	C3	6	3	2	1	2	4	5	4	4	5	4	2	3
Vitis vinifera	C3	6	2	2	1	2	3	5	4	2	5	0	1	2
Amborella trichopoda	C3	2	0	1	1	2	1	4	2	1	3	1	2	2
Physcomitrella patens	C3	6	0	0	1	2	3	7	4	5	3	24	4	3

[a]Photosynthetic types are shown in second column.
[b]Sequences from genome data shown after the slash, while number before slash stands for sequence from transcriptome.

through the regulation of the expression of core carbon fixation genes, such as *PPCK1*, *PPC1* and *PPDK* (Figure 4). Notably, we found that *PPC1* and *PPC2* in *P. equestris* have undergone rapid selection. Among the genes examined, *PPC1* had an Reads Per Kilobase of transcript per Million mapped reads (RPKM) of 10739.9 in the dark, whereas *PPC2* had an RPKM of only 91.646. *PPC1* has low expression in light in both leaf and other tissues (Tables S4 and S9) in *P. equestris*, demonstrating that *PPC1* could have been specialized to function within CAM. *PPC* orthologs, however, show low expression in the C3 plants *A. shenzhenica* and rice (Figure S15). Both the *PPC1* of orchids and the C4 *PPC* genes are located in the same subfamily (Wang *et al.*, 2009). Unlike the single amino acid mutations of A774S and R884G that led to the greatly enhanced enzyme activity of C4 *PPC*s (Paulus *et al.*, 2013), we did not detect these mutations in *P. equestris PPC1* (Figure 5). In C4 plants and CAM plants, *PPC1* might have evolved differentially, with the switch between C4 and CAM attained by enabling the *PPC1* enzyme to participate in the CAM photosynthetic pathway. Furthermore, in *P. equestris PPCK1*, *PPCK2* and *PPDK* exhibited dramatic expressional fold changes in circadian rhythms, and high levels of co-expression with *PPC1* during the night. This suggests that *P. equestris* may achieve CAM by enhancing *PPC1* enzymatic activity through the high expression of kinases *PPCK* and *PPDK*, which consequently phosphorylate *PPC1* (Figure 6). Thus, *PPCK1*, *PPC1* and *PPDK* are strongly co-expressed (Figure 4), and have been co-recruited to the CAM pathway. *PPCK* is known to phosphorylate *PPC1* in the dark (Silvera *et al.*, 2014), and *PPDK* may also contribute to the conversion of PEP in the dark because this reversible reaction takes place in the light. Therefore, this reaction could be triggered in the dark, and thus we propose that in the dark, CO_2 is primarily fixed through two metabolic pathways via *PPDK* and *PPCK*, respectively (Figure 6). Additionally, we also show that in orchids, especially in *D. terminale*, the conversion from PEP to malate,

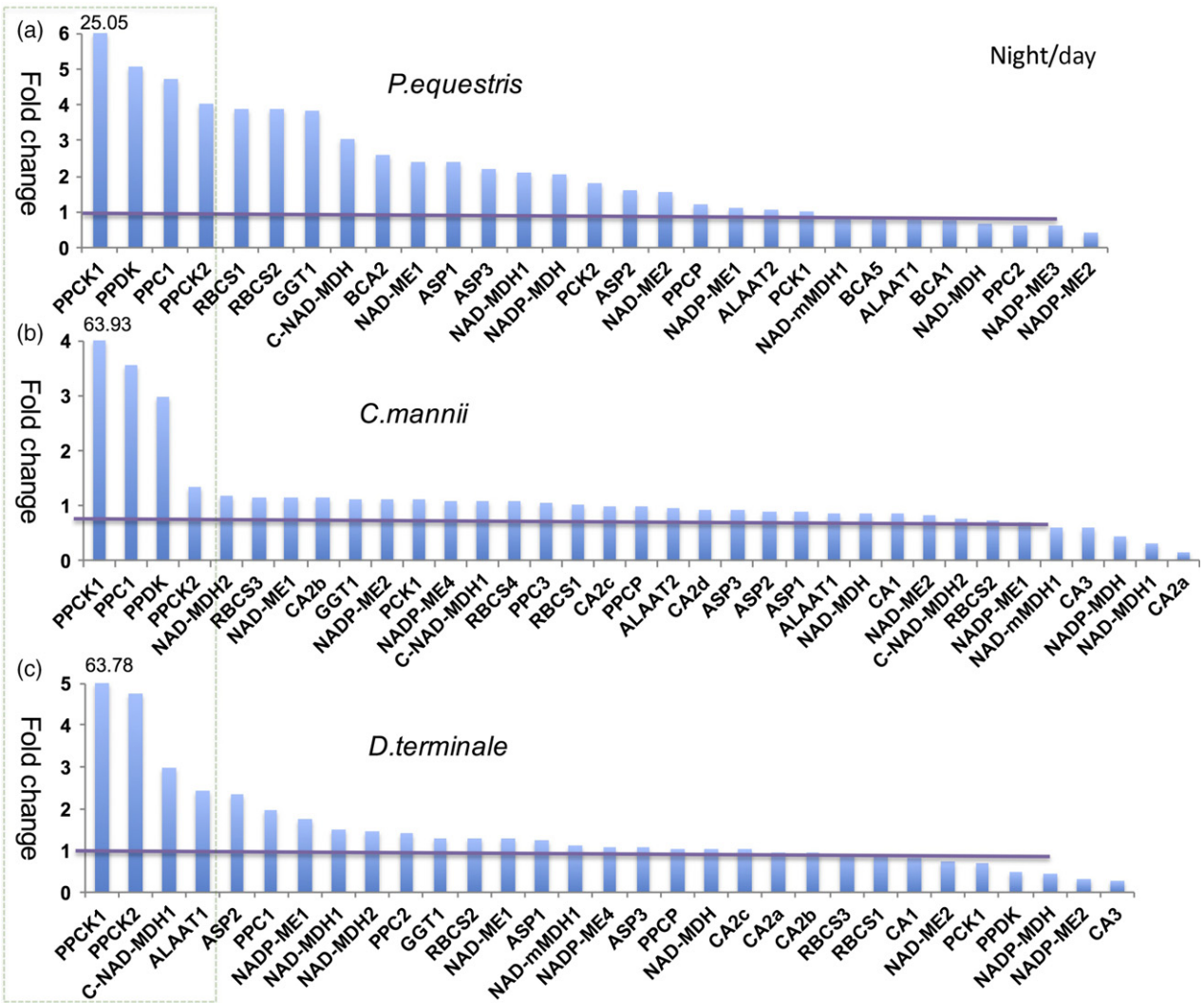

Figure 4. Expression of the 13 gene families in *Phalaenopsis equestris*, *Cymbidium mannii* and *Dendrobium terminale* during the day and night: (a) *P. equestris*; (b) *C. mannii*; and (c) *D. terminale*. Bars indicate the expressional fold change of night/day, which is calculated as: (leaf + 10)/[average (flower, stem, root) + 10] for normalization.

which usually happens under dark conditions, takes place under light conditions because of the reversibility of this reaction, and that *DteNADP-MDH* and *DtePCK1* show higher expression under light than under dark conditions.

DISCUSSION

Molecular evolution of gene families in orchids

To date, little information is available on orchids except for the analysis of the MADS-box, b-CA (Cai *et al.*, 2014), and PEPC family (Silvera *et al.*, 2014; Deng *et al.*, 2016). In this study, we provided a phylogenetic analysis of the complete carbon fixation pathway in orchids, which greatly complements previous genetic research of Orchidaceae. Moreover, the five main branches of the Orchidaceae are inferred using the phylogeny of 82 single-copy genes from 13 orchids representing the five subfamilies. This tree provides a more robust and reliable orchid phylogeny than previous studies of orchids based solely on morphology or on a single plastid/nuclear gene, such as *MatK* and *ITS* rDNA (van den Berg *et al.*, 2009; Givnish *et al.*, 2015). Because a single-copy gene has a more conserved evolutionary trajectory, whereas multi-copy genes have stronger selective pressure, this tree could serve as the most accurate orchid phylogeny to date (Figure 2), and is consistent with the widely accepted phylogeny of monocots (Chase, 2004). The phylogeny also suggests that the nuclear single-copy gene is a powerful tool for constructing the most complex orchid phylogeny. The data set presented also advances comparative genomics in orchid biology.

Our comparative analyses of 13 gene families in 13 orchids uncovered gene duplications among β-*CA*, *PPCK*, *NADP-ME*, *ALAAT*, *RBCS* and *PCK* genes in the orchid lineage, which suggests that there was a single WGD event within

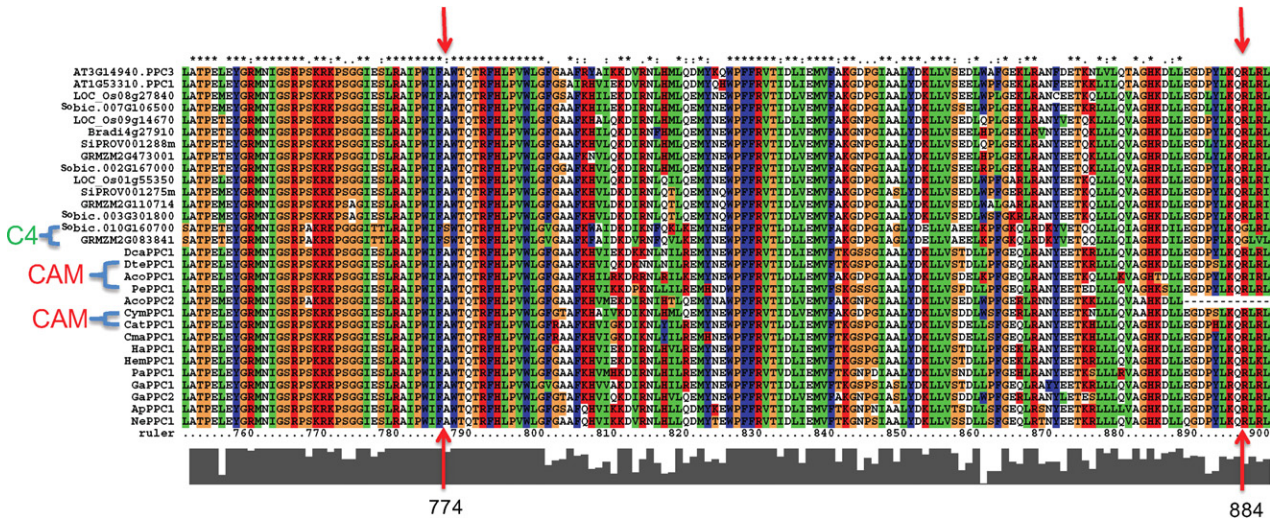

Figure 5. Sequence alignment of C3-, C4- and crassulacean acid metabolism (CAM)-determining regions in various *PPC*s. Arrows indicate the two mutations from C4 sequences that enhanced the photosynthetic efficiency.

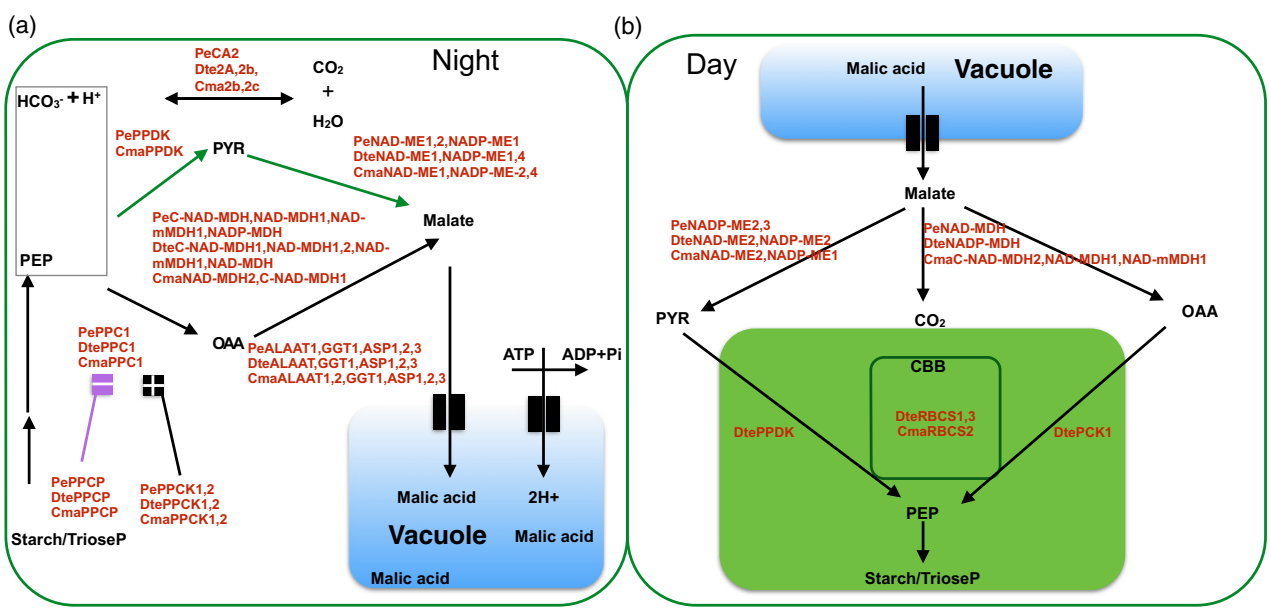

Figure 6. Carbon fixation metabolism pathways of crassulacean acid metabolism (CAM) involving core genes/enzymes: (a) nocturnal reactions are shown when stomata are open at night; (b) diurnal reactions are shown when stomata are closed. Genes/enzymes are labeled in red and blue. The purple line indicates the inhibition during the day.

the Orchidaceae. Furthermore, ALAAT and PPCK from Epidendroideae have two copies, whereas *A. shenzhenica* and *N. malipoensis* only have one copy. There are also gene families in *Galeola faberi* that have only one copy, suggesting that the WGD occurred in the ancestor of Apostasioideae or Vanilloideae. Genomic data for *P. equestris* suggest a WGD (Cai *et al.*, 2014), and our work suggests the existence of a WGD in Orchidaceae based on the phylogenetic analysis of gene families, and another WGD within the subfamilies of Apostasioideae or Vanilloideae, which has not yet been reported. Analyses of orchid gene families and genome duplications will promote additional studies of orchid genomics, and help to understand the mechanism of the rapid evolution and ecology of orchids.

Evolution of CAM genes in orchids

The major characteristics that distinguish CAM from C3 and C4 are the temporal regulation of CO_2 absorption and fixation. In this study, we identified the carbon fixation genes involved in CAM via comparisons of circadian

expression. The higher expression in the dark of *PPC1*, *PPDK*, *PPCP*, *PPCK1*, *PPCK2*, *NAD-MD* and *C-NAD-MDH* suggest that they have been recruited to the CAM photosynthetic pathway because compared with their C3/C4 orthologs, these genes have the opposite expression pattern. We also hypothesize that future studies could unveil the involvement of a series of other genes such as the circadian clock, light-regulated and temperature-regulated genes, and possibly nutrition-related genes, which could have roles in the switch to CAM. Our transcriptome data from orchids will be a great reference for comparative studies on other economically important CAM plants, such as agave, kalanchoe, and ice plants (Yang *et al.*, 2015). It is clear that the breeding of crops with enhanced photosynthetic performance and low WUE in the future relies on the better understanding of CAM-related genes.

Origin of CAM in orchids

A simple hypothesis is that CAM is the result of the duplication of multiple gene families, including *PPC*s, such as with the origin of C4 photosynthesis (Wang *et al.*, 2009; Christin *et al.*, 2013); however, our findings do not support this hypothesis in CAM because the genes involved in C3 and CAM orchids bear the same gene copy number and evolutionary history, especially for *PPC*. There are no correlations between the appearance of CAM and gene duplication.

An alternative hypothesis is that CAM resulted from the adaptation of regulating gene expression. Considering the fast radiation of orchids, the diversity of CAM or non-CAM in orchids, and the tissue-specific and diel-rhythmic expression of key CO_2 fixation genes in *P. equestris*, our findings strongly suggest that CAM probably resulted from the regulation of the transcription levels of some key genes involved in CO_2 fixation (Figure 6), and not because of copy number. Therefore, *PPC1*, *PPCK1* and *PPDK* may be the primary targets of the CAM pathway for future studies on their detailed transcriptional regulation and precise manipulation. We propose here that in the dark, CO_2 is primarily fixed not only via *PPCK*, but also via *PPDK*, as an alternative pathway in CAM orchids, which is the complementation to the recent genome analysis of the CAM plant pineapple (Ming *et al.*, 2015). The importance of this alternative pathway in orchids should also be validated by future biochemical studies, and whether or not this pathway is orchid specific or present in other CAM plants. The future focus on CAM should be directed towards the potential regulation of these key genes at a transcriptional level, rather than at the gene dosage level, to enhance water-efficient photosynthesis.

EXPERIMENTAL PROCEDURES

Carbon isotope determinations

The carbon isotope ratio ($\delta^{13}C$) is generally regarded as a rapid and accurate method to determine the type of photosynthesis of a plant, including C3, C3-CAM and CAM (Borland *et al.*, 2009). The $\delta^{13}C$ value was derived from 3-mg samples of dried mature whole leaves. The $^{13}C/^{12}C$ ratio was analyzed according to Holtum and colleagues (Holtum *et al.*, 2004).

Transcriptome sequencing, assembly, annotation and gene expression

The transcriptomes of *Cymbidium sinense*, *Cypripedium singchii*, *Galeola faber*, *Habenaria delavayi*, *Hemipilia forrestii*, *Neuwiedia malipoensis* and *Paphiopedilum armeniacum* were made from flower samples and sequenced in our previous study (Fu *et al.*, 2011). The sequencing reads were reassembled in this work. Samples from leaves, stem, roots and flowers of *Phalaenopsis equestris*, and leaves, stem, roots and flowers of *Apostasia shenzhenica* were sequenced to compare the tissue/organ expressional profiles. To study the diel expressional differences of CAM, leaves of CAM orchids *P. equestris*, *Dendrobium terminale* and *Cymbidium mannii* were sampled at 11:00 h and 23:00 h, according to the methode described by Silvera *et al.* (2014).

Total RNA was extracted using the Sigma Spectrum™ Plant Total RNA Kit. Library construction and sequencing were performed following the protocols of Peng *et al.* (2012) with an Illumina HiSeq2000 instrument that produced 100-base pair (bp) reads. Reads were *de novo* assembled and annotated with TRINITY (Grabherr *et al.*, 2013). The commands and parameters used for running TRINITY were as follows: Trinity –seqType fq –JM 200G –min_contig_length 200 –left sample_1.fq –right sample_2.fq. The sequences from genomic studies of *P. equestris* (Cai *et al.*, 2014) and *D. catenatum* (Zhang *et al.*, 2015) were compared and curated to confirm that the sequences were sufficient, and that the gene family members were complete.

Homolog gene identification for comparative analysis

Orthologs from the model plant *A. thaliana*, obtained from TAIR 10 (http://www.arabidopsis.org) were used as seeds to search against the annotated genome and transcriptomes. Hits with e-values less than $1e^{-10}$ were obtained for phylogenetic comparison with gene models from model species, such as Arabidopsis and rice. The genes obtained were phylogenetically identified as reliable orthologs from each orchid. The 13 gene families were listed in Table S2.

Multiple sequence alignment and phylogenetic tree construction

Multiple sequences were aligned with MUSCLE integrated in MEGA 5.2.2 (Tamura *et al.*, 2011). A maximum-likelihood phylogenetic tree was constructed using FASTTREE 2.1 (Price *et al.*, 2009, 2010). For the best protein substitution model, we have tried JTT+CAT, WAG+CAT and GTR+CAT models in FASTTREE, and all produced the same topology and similar local support values. We adopted the JTT+CAT model as it produced the highest supporting values.

Single copy gene retrieval and species tree construction

Protein domains were identified using HMMER (Eddy, 2011) against the Pfam-A data set (Punta *et al.*, 2012) for annotation. Single copy nuclear ortholog genes were determined with single copy ortholog domain-containing genes (Torruella *et al.*, 2012). To infer the phylogeny of the 13 orchids and other plants, 82 single copy nuclear genes were selected based on annotations of transcriptomes (eight orchids in this study) and genomes (*P. equestris* and other land plants), based on the methods of Zeng *et al.* (2014).

ACCESSION NUMBERS

The accession numbers for *P. equestris* leaves, stem, roots and flowers were SRR2080202, SRR2080200, SRR2080194 and SRR2080204, respectively. The accession numbers for *A. shenzhenica* leaves, stem, roots and flowers were SRR3183249, SRR3183244, SRR3183239 and SRR3183251, respectively.

Leaf transcriptomes of CAM *P. equestris* (accessions SRR2962594 and SRR292604), *Dendrobium terminale* (accessions SRR2967016 and SRR2968862), *Cymbidium mannii* (accessions SRR2976385 and SRR2976410), and *Ananas comosus* (accession SRR2989012 and SRR2989014) were sampled at 11:00 h and 23:00 h. The leaf transcriptome from *Cymbidium atropurpureum* was sequenced using accession SRR2976606.

ACKNOWLEDGEMENTS

We would like to thank Israel Ausin for helpful comments on the article. This work was supported by the National Natural Science Foundation of China (81502437), a Graduate Innovation Programme of Jiangsu Province (no. KYLX_0552), and a scholarship from the China Scholarship Council (CSC, no. 201406850018) to Fei Chen. This work was supported by Fujian–Taiwan Joint Innovative Center for Germplasm Resources and Cultivation of Crop [Fujian 2011 Program, (2015)75].

AUTHOR CONTRIBUTIONS

LSZ, ZMC, ZJL, FC designed the project; GQZ, FC, SN, JSX carried out the experiments; LSZ, FC, YQZ analyzed the data; and LSZ, FC, ZMC, ZJL, ZGL wrote and revised the manuscript.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Pfam domains shared among the 13 orchids and other sequenced genomes.

Figure S2. Phylogeny of *CA* genes of orchids and other representative land plants.

Figure S3. Phylogeny of *PPC* genes of orchids and other representative land plants.

Figure S4. Phylogeny of *PPCK* genes of orchids and other representative land plants.

Figure S5. Phylogeny of *PPDK* genes of orchids and other representative land plants.

Figure S6. Phylogeny of *NAD-ME* genes of orchids and other representative land plants.

Figure S7. Phylogeny of *NADP-ME* genes of orchids and other representative land plants.

Figure S8. Phylogeny of *NAD-MDH* genes of orchids and other representative land plants.

Figure S9. Phylogeny of *NADP-MDH* genes of orchids and other representative land plants.

Figure S10. Phylogeny of *AAT* (including *ALAAT* and *GGT*) genes of orchids and other representative land plants.

Figure S11. Phylogeny of *ASP* genes of orchids and other representative land plants.

Figure S12. Phylogeny of *RBCS* genes of orchids and other representative land plants.

Figure S13. Phylogeny of *PCK* genes of orchids and other representative land plants.

Figure S14. Phylogeny of *PPCP* genes of orchids and other representative land plants.

Figure S15. Expression of core genes in representative plants.

Table S1. The $\delta^{13}C$ (‰) of leaves in each species.

Table S2. The gene names and ID numbers of core genes in Arabidopsis, maize, rice, *Phalaenopsis equestris*, *Dendrobium catenatum* and *Apostasia shenzhenica.*

Table S3. The gene names and ID numbers of core genes in the 13 orchids.

Tables S4–S8. The expressions of core genes in *Phalaenopsis equestris*, *Apostasia shenzhenica*, maize, rice, and Arabidopsis, with leaf and other tissues/stages.

Table S9. The expressions of core genes in *Phalaenopsis equestris* during the day and night.

Table S10. The expressions of core genes in *Cymbidium mannii* during the day and night.

Table S11. The expressions of core genes in *Dendrobium terminale* during the day and night.

REFERENCES

van den Berg, C., Higgins, W.E., Dressler, R.L., Whitten, W.M., Soto-Arenas, M.A. and Chase, M.W. (2009) A phylogenetic study of Laeliinae (Orchidaceae) based on combined nuclear and plastid DNA sequences. *Ann. Bot.* **104**, 417–430.

Borland, A.M. and Yang, X. (2013) Commentary informing the improvement and biodesign of crassulacean acid metabolism via system dynamics modelling. *New Phytol.* **200**, 946–949.

Borland, A.M., Griffiths, H., Hartwell, J. and Smith, J.A.C. (2009) Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands. *J. Exp. Bot.* **60**, 2879–2896.

Cai, J., Liu, X., Vanneste, K. *et al.* (2014) The genome sequence of the orchid Phalaenopsis equestris. *Nat. Genet.* **47**, 65–72.

Chase, M.W. (2004) Monocot relationships: an overview. *Am. J. Bot.* **91**, 1645–1655.

Christin, P., Boxall, S.F., Gregory, R., Edwards, E.J., Hartwell, J. and Osborne, C.P. (2013) Parallel recruitment of multiple genes into C4 photosynthesis. *Genome Biol. Evol.* **5**, 2174–2187.

Cribb, P. and Govaerts, R. (2005) Just how many orchids are there? In *Proceedings of the 18th World Orchid Conference* (Raynal-Roques, A., Roguenant, A. and Prat, D., eds). Naturalia, Dijon, France: Turriers, pp. 161–172.

Deng, H., Zhang, L., Zhang, G., Zheng, B. and Liu, Z. (2016) Molecular Phylogenetics and Evolution Evolutionary history of PEPC genes in green plants: implications for the evolution of CAM in orchids. *Mol. Phylogenet. Evol.* **94**, 559–564.

Depaoli, H.C., Borland, A.M., Tuskan, G.A., Cushman, J.C. and Yang, X. (2014) Synthetic biology as it relates to CAM photosynthesis: challenges and opportunities. *J. Exp. Bot.* **65**, 3381–3393.

Eddy, S.R. (2011) Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195.

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

Givnish, T.J., Spalink, D., Ames, M. *et al.* (2015) Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proc. R. Soc. B*, **282**, 20151553.

Grabherr, M.G., Haas, B.J., Yassour, M. *et al.* (2013) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **29**, 644–652.

Holtum, J.A.M., Aranda, J., Virgo, A., Gehrig, H.H. and Winter, K. (2004) δ13C values and crassulacean acid metabolism in Clusia species from Panama. *Trees*, **18**, 658–668.

Hsiao, Y., Pan, Z., Hsu, C. *et al.* (2011) Research on orchid biology and biotechnology. *Plant Cell Physiol.* **52**, 1467–1486.

Kluge, M. and Ting, I.P. (1978) *Crassulacean Acid Metabolism: Analysis of An Ccological Adaptation*. Heidelberg: Springer-Verlag.

Ming, R., VanBuren, R. and Wai, C.M. (2015) The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* **47**, 1435–1442.

Paulus, J.K., Schlieper, D. and Groth, G. (2013) Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution. *Nat. Commun.* **4**, 4–10.

Peng, Z., Cheng, Y., Tan, B.C.-M. *et al.* (2012) Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nat. Biotechnol.* **30**, 253–260.

Price, M.N., Dehal, P.S. and Arkin, A.P. (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650.

Price, M.N., Dehal, P.S. and Arkin, A.P. (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE*, **5**, e9490.

Punta, M., Coggill, P.C., Eberhardt, R.Y. *et al.* (2012) The Pfam protein families database. *Nucleic Acids Res.* **40**, 290–301.

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

Silvera, K., Neubig, K.M., Whitten, W.M., Williams, N.H., Winter, K. and Cushman, J.C. (2010) Evolution along the crassulacean acid metabolism continuum. *Funct. Plant Biol.* **37**, 995.

Silvera, K., Winter, K., Rodriguez, B.L., Albion, R.L. and Cushman, J.C. (2014) Multiple isoforms of phosphoenol pyruvate carboxylase in the Orchidaceae (subtribe Oncidiinae): implications for the evolution of crassulacean acid metabolism. *J. Exp. Bot.* **65**, 3623–3636.

Smith, J.A.C. and Winter, K. (1996) Taxonomic distribution of crassulacean acid metabolism. In *Crassulacean Acid Metabolism: Biochemistry, Ecophysiology and Evolution* (Winter, K. and Smith, J. A. C., eds). Berlin/Heidelberg: Springer-Verlag, pp. 427–436.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.

Torruella, G., Derelle, R., Paps, J., Lang, B.F., Roger, A.J., Shalchian-Tabrizi, K. and Ruiz-Trillo, I. (2012) Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. *Mol. Biol. Evol.* **29**, 531–544.

Wang, X., Gowik, U., Tang, H., Bowers, J.E., Westhoff, P. and Paterson, A.H. (2009) Comparative genomic analysis of C4 photosynthetic pathway evolution in grasses. *Genome Biol.* **10**, R68.

Yang, X., Cushman, J.C., Borland, A.M. *et al.* (2015) A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. *New Phytol.* **207**, 491–504.

Zeng, L., Zhang, Q., Sun, R., Kong, H., Zhang, N. and Ma, H. (2014) Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat. Commun.* **5**, 4956.

Zhang, G.-Q., Xu, Q. and Bian, C. *et al.* (2015) The Dendrobium catenatum Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. *Sci. Rep.* **5**, 19029.

