Comparative Chloroplast Genomes of Photosynthetic Orchids: Insights into Evolution of the Orchidaceae and Development of Molecular Markers for Phylogenetic Applications

Jing Luo*, Bei-Wei Hou*, Zhi-Tao Niu, Wei Liu, Qing-Yun Xue, Xiao-Yu Ding*

College of Life Sciences, Nanjing Normal University, Nanjing, China

Abstract

The orchid family Orchidaceae is one of the largest angiosperm families, including many species of important economic value. While chloroplast genomes are very informative for systematics and species identification, there is very limited information available on chloroplast genomes in the Orchidaceae. Here, we report the complete chloroplast genomes of the medicinal plant Dendrobium officinale and the ornamental orchid Cypripedium macranthos, demonstrating their gene content and order and potential RNA editing sites. The chloroplast genomes of the above two species and five known photosynthetic orchids showed similarities in structure as well as gene order and content, but differences in the organization of the inverted repeat/small single-copy junction and ndh genes. The organization of the inverted repeat/small single-copy junctions in the chloroplast genomes of these orchids was classified into four types; we propose that inverted repeats flanking the small single-copy region underwent expansion or contraction among Orchidaceae. The AT-rich regions of the ycf1 gene in orchids could be linked to the recombination of inverted repeat/small single-copy junctions. Relative species in orchids displayed similar patterns of variation in ndh gene contents. Furthermore, fifteen highly divergent protein-coding genes were identified, which are useful for phylogenetic analyses in orchids. To test the efficiency of these genes serving as markers in phylogenetic analyses, coding regions of four genes (accD, ccsA, matK, and ycf1) were used as a case study to construct phylogenetic trees in the subfamily Epidendroideae. High support was obtained for placement of previously unlocated subtribes Collabiinae and Dendrobiinae in the subfamily Epidendroideae. Our findings expand our understanding of the diversity of orchid chloroplast genomes and provide a reference for study of the molecular systematics of this family.

Introduction

The orchid family Orchidaceae is one of the two largest families of flowering plants, with over 25,000 species [1] and five recognized subfamilies (Apostasioideae, Cypripedioideae, Epidendroideae, Orchidoideae, and Vanilloideae) [2]. A large number of orchids have significant economic value [3]. For example, some cultivars have been used as cut flowers or potted plants, while others can be utilized as food or medicine because of their nutritious or medical efficacy. Overexploitation and habitat destruction have threatened the survival of many wild orchid species. At the same time, numerous cultivated varieties and crossbreeds have been developed worldwide. Therefore, molecular information on orchids is of interest not only for the study of systematics, but also for species conservation and flower cultivation.

Epidendroideae is the largest of the five orchid subfamilies and includes approximately 20,000 species [1,2]. Several perspectives on its classification have long been debated [2,4–16]. Burns-Balogh and Funk (1986) have reviewed previous morphological classification systems of Orchidaceae. Freudenstein and Rasmussen (1999) pointed out that most of the previously established classifications have a highly developed tribal and subtribal classification within the Epidendroideae; they first performed a cladistic analysis of Orchidaceae. However, major groups of genera are equivalent to subfamilial groups, and the detailed classifications at the tribal level are not well supported by morphological and anatomical features in most cases [12,16]. In recent years, molecular data has been used in phylogenetic studies, but some relationships among subtribes or tribes remain questionable. Major disputes were focused on whether some tribes or subtribes were monophyletic, polyphyletic, or paraphyletic; which tribe or subtribe was the most basal; and the locations of Agrostophyllinae, Collabiinae and Dendrobiinae. Limited sampling with few variable loci in most of the common DNA regions has impeded reasonable and robust estimates of phylogenetic
patterns. Recent comparative chloroplast (cp) genomics has provided large quantities of data that are useful for selecting pertinent markers to resolve obscure phylogenetic relationships in seed plants [17–21]. However, cp genome information is still limited for the Epidendroideae.

In most land plants, the cp genome is a single circular molecule of 120–220 kb that consists of one large single-copy (LSC) region, one small single-copy (SSC) region, and a pair of inverted repeats (IRs). Although gene organization and content are conserved in cp genomes of higher plants, their genome sizes are diverse and depend largely on the extent of gene duplication, small repeats, and the size of intergenic spacers [22]. The information on sequence insertion or deletion, transition or transversion, and nucleotide repeats may help to clarify evolutionary relationships [23–27]. To date, cp genomes from seven orchid genera (Corallorhiza, Cymbidium, Erycina, Neotia, Oncidium, Phalaenopsis, and Rhizanthella) have been sequenced. The former six genera belong to the subfamily Epidendroideae, whereas the last one falls into the subfamily Orchidoideae. All species in these seven genera are photosynthetic orchid except Rhizanthella gardneri, Corallorhiza striata, and Neotia nihoa-acts being nonphotosynthetic plants [17,28–34]. However, the cp genome of the Dendrobiiinae, the largest and most economically important subtribe in the Epidendroideae, has not yet been sequenced.

Dendrobium officinale Kimura et Migo, a perennial epiphytic herb of the Dendrobiiinae, is endemic in moderately damp mountains in China [35]. The stems of D. officinale have been widely used as a traditional Chinese medicine (TCM) called “Teipi Fendou.” The efficacious compounds in D. officinale include phenols, alkaloids, coumarins, and polysaccharides [36]; and its medical benefits include stimulation of saliva, improvement in eyesight, warming of the stomach, enhancement of immuity, and inhibition of tumor growth [36]. As a result of its habitat shrinking and human overexploitation, natural populations of D. officinale have been progressively destroyed, and in 1992 it was classified as an endangered species in the Chinese Plant Red Book [37]. D. officinale has recently been rescued by tissue culture in southern China.

The subfamily Cypripedioideae comprises approximately 155 species in five genera [2]. All Cypripedioideae species have special flowers with a saccate lip, two fertile stamens, a shield-like staminode, and a synsepal composed of fused lateral sepals [38]. Because of its attractive morphological characteristics, this subfamily has been investigated widely in theoretical and applied research. Nonetheless, molecular information on this subfamily is still limited. Cypripedium macranthos Sw, is a terrestrial herbaceous plant in the subfamily and naturally distributed in East Asia [39]. Because of the commercial value of its pretty red or pink flowers, it has been cultivated as a potted and garden plant.

In this study, we sequenced the complete cp genomes of D. officinale and C. macranthos using a next-generation sequencing (NGS) approach. Our objectives were to deepen understanding of the structural diversity of orchid cp genomes and to provide information for resolving uncertain relationships within the Epidendroideae. The cp genomes of seven photosynthetic orchid species (C. macranthos, Cymbidium mannii, D. officinale, Erycina pusilla, Oncidium Gower Ramsey, Phalaenopsis aphrodite, and Phalaenopsis equestris) were compared to elucidate the diversity of gene order, gene content, and genome structure among them. Four regions were filtered according to the sequence divergence of protein-coding genes, and 56 taxa from 36 genera were used as a case study to determine phylogenetic relationships within the Epidendroideae.

**Materials and Methods**

Chloroplast DNA extraction and genome sequencing, assembly, and PCR-based validation

This study was approved by the Ethics Committee of Forestry Bureau of Zhejiang Province and Nanjing Normal University, China. We collected seeds of D. officinale from Yandang experimental base of Zhejiang Branch, College of Life Sciences, Nanjing Normal University.

Young leaves of D. officinale were taken from 6-month-old seedlings grown in a greenhouse. Intact chloroplasts were isolated using the Percoll gradient method [22–45%] [40]. Purified chloroplast DNA was extracted according to the 2× CTAB protocol [41]. Fresh leaves of C. macranthos were collected from Yunnan Province, China. Total DNA was extracted using a Qiagen DNeasy plant mini kit (Qiagen, Germany). DNA concentration and quality were determined using a NanoDrop 9000 Spectrophotometer (Thermo Scientific, Wilmington, DE). High quality DNA (concentration >300 ng/μl, A260/A280 ratio = 1.9–2.0 and A260/A230 ratio >1.7) was used for sequencing.

Purified DNA was fragmented and used to construct short-insert libraries (insert size~500 bp) according to the manufacturer’s instructions (Illumina). The short fragments were sequenced using an Illumina Hiseq 2000 sequencing system [42].

The raw reads for D. officinale were trimmed with error probability <0.001 and assembled using SOAPdenovo version 1.03 with default parameters [43]. The de Bruijn graph approach was applied to assembly with an optimal k-mer size of 79. The contigs shorter than 200 bp were removed. Then the paired-end information was used to join the contigs into scaffolds with the cp genome of P. aphrodite (Accession Number: AY916449) as a reference. Gaps among scaffolds were filled using paired-end extracted reads.

The short reads for C. macranthos were trimmed with error probability <0.05 and assembled using CLC Genomic Workbench 6.0.1 (CLC Bio, Aarhus, Denmark). The contigs shorter than 200 bp were discarded; others were compared with plant cp genomes in the National Center for Biotechnology Information (NCBI) using BLAST (http://blast.ncbi.nlm.nih.gov) searches. Contigs matching referenced genomes with E values <10−5 were selected for annotation.

Based on the reference genomes in Orchidaceae [28,29], gaps and four junction regions between LSC/SSC and IRs were confirmed by PCR amplification and Sanger sequencing using the primers listed in Table S1.

**Genome annotation**

Protein-coding and ribosomal RNA genes were annotated using DOGMA (http://dogma.cbb.utexas.edu/) [44]. The boundaries of each annotated gene were manually determined by comparison with orthologous genes from other orchid cp genomes. Genes of tRNA were predicted using tRNAscan (http://lowelab.ucsc.edu/tRNAscan-SE) [45] and ARAGORN version 1.2 (http://130.235.46.10/ARAGORN/) [46]. The circular genome maps were drawn using GenomeVx, followed by manual modification [47]. The sequencing data and gene annotation were submitted to GenBank with accession numbers KC771275 and KF925434.

**Analyses of RNA editing sites**

Thirty protein-coding genes of D. officinale and C. macranthos cp genomes were used to predict potential RNA editing sites using the online program Predictive RNA Editor for Plants (PREP) suite (http://prep.unl.edu/) [48] with a cutoff value of 0.8.
Phylogenetic analyses

Sixty-three common protein-coding genes were extracted from 10 cp genomes. Seven photosynthetic orchid species were involved in analyses with *Calamus caryotoides*, *Phorion dactylifera*, and *Typha latifolia* as outgroups. The GenBank accession numbers of all taxa are shown in Table S2. The *accD*, *infA*, *rps16*, *rps19*, *ycf1*, and *ndh* genes were not included in the data set because they were pseudogenized in some cp genomes. Alignments were performed using the MUSCLE program in Mega 5.03 [49], without including gaps, and start and stop codons. The aligned sequences were concatenated and used for phylogenetic reconstruction.

The ML tree was constructed by means of GTR+G model with rapidGAP version 1.2 (http://sourceforge.net/projects/raxmlgui/) [50] and a rapid bootstrap value of 1,000. A Bayesian inference (BI) tree was constructed using CAT model with PhyloBayes version 3.2 [51]. Two Independent MCMC chains were run. The first 25% of the cycles were removed as burn-in, and convergence of three chains was checked on the basis of maxdiff <0.3 by following the PhyloBayes manual.

Sequence divergence of protein-coding genes

To obtain suitable markers for phylogenetic analysis within subfamilies, complete cp genomes of six orchid species (C. macranthos, C. mannii, D. officinale, E. pusilla, O. Gower Ramsey, and P. aphrodite) were applied. The average pairwise distances of nucleotide and protein substitutions for 68 protein-coding genes were estimated using Kimura’s two-parameter model and p-distance, respectively, in Mega 5.03 [49].

Phylogenetic application of cp genomes, a case study on the Epidendroideae

We selected the Epidendroideae as an example for phylogenetic analysis. Data sets for four incomplete gene sequences (*npf*1, *mufB*, *ccsA*, and *accD*) were obtained for 36 taxa from 36 genera. The data matrix included 11 subtribes and one tribe in the Epidendroideae, with *C. caryotoides*, *P. dactylifera*, and *T. latifolia* as outgroups. Six taxa from two additional orchid subfamilies were used as internal checks. Sequences from 10 of the taxa were extracted from complete cp genomes; sequences from the other 46 taxa were obtained by PCR amplification and sequencing of PCR products with an ABI PRISM 3730XL DNA analyzer (Applied Biosystems). Primers for *accD* and *ccsA* were designed using Primer Premier version 6 [52] based on homologous sequences from orchid cp genomes (Table S3). All newly generated sequences were deposited in GenBank with accession numbers KF361524-KF361707. Sources of species and GenBank accession numbers are indicated in Table S4. These regions were aligned separately using Mega 5.03 [49] with manual modifications, and gaps were coded as “-.” Sequence information was analyzed using Mega 5.03 and DnaSP version 5.0 [53]. The combined matrix was utilized for phylogenetic analyses. Modeltest version 3.7 [54] was employed to select the best nucleotide substitution model under the Akaike Information Criterion (AIC); the GTR+I+G model was chosen as the best fit for our data set. The ML and BI analyses were performed according to the same protocol as that used for phylogenetic analyses.

Results

Sequencing and genome assembly

The raw Illumina paired-end sequencing of *D. officinale* produced 350 Mb of data. After quality trim, 210 Mb of data remained with an average read length of 80 bp. The subsequent de novo assembly produced 13 scaffolds, 12 of which were >2 Kb and the scaffold N50 size was 84,551 bp. The average coverage depth was 1,400×. These scaffolds were used for the following assembly.

We sequenced 2.5 Gb of Illumina paired-end reads for *C. macranthos* (average read length of 90 bp). The initial assembly included 12,148 contigs. After compared with plant cp genomes, 41 contigs were obtained with E values<10−5 and mean coverage depth >26×. Four of these contigs were larger than 10 kb with average depth coverage 129×, resulting in a nearly complete draft genome. After assembly and gap closure, two complete chloroplast genomes were obtained.

Characteristics of the chloroplast genomes of *Dendrobium officinale* and *Cypridium macranthos*

The complete cp genomes of *D. officinale* and *C. macranthos* were circular, having 132,221 and 157,050 bp, respectively. Similar to other angiosperms, both cp genomes were AT-rich (62.53% and 62.17%, respectively). The *D. officinale* plastome contained 110 different genes, of which 91 were single-copy genes and 19 were duplicated genes (Fig. 1). Its cp genome consisted of 76 protein-coding genes, 4 rRNA genes, and 30 tRNA genes. *C. macranthos* encoded 113 different genes (94 single-copy and 19 duplicated genes). The *C. macranthos* cp genome included 79 protein-coding genes, four rRNA genes, and 30 tRNA genes (Fig. 2). The gene content of the *D. officinale* cp genome was relatively conserved compared with other known orchid cp genomes. The gene content of the *C. macranthos* cp genome was also relatively conserved with the exception of the following. A coding sequence (CDS) of *infA* (coding for translation initiation factor) was interrupted because of a 5-bp deletion (53 bp downstream of the start codon). This gene was lost from *Discorea* in monocots [55]. At the N terminus of *rps19*, a surplus nucleotide A in the poly(A) tract interrupted the open reading frame (ORF), causing a frameshift. Furthermore, we recognized *rps16* as a pseudogene because a partial intron and the second exon were missing in it.

Potential RNA editing sites

In the present study, potential RNA editing sites were predicted for 30 genes; as a result, a total of 51 RNA editing sites were identified in genes of *Cypridium* and *Dendrobium* (Table S5). No potential editing sites were identified in seven genes (*petD*, *petE*, *petA*, *psbB*, *psbL*, *psbA*, and *rpl23*) in both cp genomes. Of the 51 editing sites, 9 (17.6%) and 42 (82.4%) were located at the first and second codon position, respectively, in *Cypridium*; 8 (15.7%) and 43 (84.3%) were located at the first codon and the second codon position, respectively, in *Dendrobium*; but no editing sites were found at the third codon position. Just as in other terrestrial plants, the editing types in *Cypridium* and *Dendrobium* were all G-to-U [56-58]. The amino acid conversion S to L occurred most frequently, while P to S and R to C occurred least. Thirty-four common RNA editing sites were shared in genes of the two species. We also observed RNA editing (C to U conversion) in the initiation codon of *rpl2* transcripts of *D. officinale*, which is a common phenomenon among angiosperms and has been verified in *P. aphrodite* and *R. gardneri* [28,30].

Phylogenetic analyses of the seven orchids

Our phylogenomic construction was based on 63 protein-coding genes of cp genomes, and the aligned data set comprised 47,736 bp. The BI and ML trees had the same topology (Fig. 3), demonstrating that *Cypridium* (Cypridioideae) was sister to the Epidendroideae. In the Epidendroideae, *Dendrobium* was sister to other species, and *Cymbidium* and *Oncidium-Erycina* were sister to *Phalaenopsis*.
Figure 1. Map of the chloroplast genome of *Dendrobium officinale*. Thick lines indicate inverted repeats (IRs). Genes shown inside the circle are transcribed clockwise, and those outside the circle are transcribed counterclockwise. doi:10.1371/journal.pone.0099016.g001
Comparison of chloroplast genomes of seven photosynthetic orchids

Six photosynthetic orchid species representing four subtribes of the subfamily Epidendroideae—Cymbidiinae (C. mannii), Aeridinae (P. aphrodite and P. equestris), Oncidiinae (O. Grower Ramsey and E. pusilla), and Dendrobiinae (D. officinale)—were compared with the reference species C. macranthos in the organization and gene content of their cp genomes. The seven cp genomes ranged from 146,484 to 157,050 bp (average length = 150,307.6489 bp) (Table 1). Compared with C. macranthos, the other taxa had reduced IR length. The organization of the cp genomes of the Epidendroideae was similar to that of the C. macranthos, except for three sequences: \Psi_{ycf1}-ndhF, ndhG-ndhJ, and ndhD-ndhH. Variations in \Psi_{ycf1}-ndhF sequence were due to reductions in the lengths of \Psi_{ycf1}, ndhF, and \Psi_{ycf1}-ndhF non-coding regions located at the IRB/SSC junction. Variations in ndhC-ndhJ and ndhD-ndhH sequences were caused by pseudogenization or loss of ndh genes.

Comparison of sequences flanking IR/SC junctions in the Orchidaceae

Sequences flanking IR/SC (single copy) junctions vary among cp genomes of different species [59]. Here, we compared

Figure 2. Map of the chloroplast genome of Cypripedium macranthos. Thick lines indicate inverted repeats (IRs). Genes shown inside the circle are transcribed clockwise, and those outside the circle are transcribed counterclockwise.

doi:10.1371/journal.pone.0099016.g002

Comparison and Application of Orchid Cp Genomes

PLOS ONE | www.plosone.org 5 June 2014 | Volume 9 | Issue 6 | e99016
sequences flanking IR/SC junctions among seven orchid cp genomes (Fig. 4); all of them were found to have similar structures at the IR/LSC junction. The trnL-psb19 cluster was duplicated and involved in IR. The IRB/LSC junction (JLB) was located within rpl22 in all seven orchid cp genomes. As a result, a duplicated rpl22 was nested within IRB.

On the other side, the orchid chloroplast genomes had distinct characteristics at the IR/SSC junction. In P. Aphrodite, the IRS/ SSC junction (JSB) was located upstream of ycf1, whereas in other species JSB was located within ycf1. Four types of junctions in the orchid cp genomes were characterized on the basis of the organization of genes flanking the IR/SSC junction (JSB). Cypripedium and Dendrobium shared type I structure in which JSB was located upstream of the ndhF-ycf32 cluster. Type II junction was found in Cymbidium and was characterized by an overlap between ycf1 and ndhF, resulting in JSB being located within these two genes. Type III was shown in Oncidium, Erycina, and P. equestris, in which JSB was located inside the ycf1-ycf32 cluster, with the loss of ndhF gene. The type IV structure was present in P. aphrodite and characterized by the entire incorporation of the entire ycf1 into the SSC, with JSB inside trnN-psb19.

Chloroplast-encoded ndh genes in seven orchid species

Chloroplast-encoded ndh genes were investigated in C. macranthos and the six photosynthetic Epidendroideae species (Fig. 3). The 11 ndh genes in Cypripedium cp genome were intact, but many ndh genes had either truncations or indels, resulting in frame shifts or pseudogenes in the six Epidendroideae cp genomes. The ndhD gene in all these Epidendroideae species contained indels or stop codons. The characteristics of other ndh genes differed among the genera. In Dendrobium, ndhB was intact; ndhC, I, and K were lost; and ndhF was truncated with two sequence inserts, creating two frame shifts. In the two Phalaenopsis species the ndhA and ndhF genes were absent and the remnants of seven ndh genes became pseudogenes. The ndhE genes in P. equestris and P. aphrodite were lost and incomplete, respectively. The two Oncidiinae species (Erycina and Oncidium) had similar patterns of diversity of ndh genes except ndhA, ndhE, and ndhI. Other varieties within Oncidinae shared major characteristics of ndh genes in Erycina and Oncidium [29]. In Cymbidium, most of the ndh genes were present in the ORF and remained intact [17].

Sequence divergence of protein-coding genes in the Orchidaceae

The pairwise distances of nucleotide and protein substitutions of 60 protein-coding genes were compared among six orchid species (Table 2). According to the average pairwise distance and numbers of nucleotide substitutions, three genes (psbA, psbB, and psbD) located in the IR regions had relatively low mean levels of sequence divergence. The ndhA and ndhF genes in the LSC and SSC regions showed higher evolutionary rates. Fifteen regions with relatively high divergence were identified in ndhE, matK, rpo15, rpl22, ccsA, psad, rps32, rpl16, rpl20, atpF, psbK, psbT, accD, and ndhB, located in the LSC, SSC, or SSC/IR junction regions. Similar patterns of divergence were also observed at the protein level, with the exception of psbT. Sequence divergence and gene length yielded a sufficient variety of loci (>600 bp); thus, the sequences of accD, ccsA, matK, and ycf1 were identified and used for phylogenetic analyses.

Molecular phylogeny within the Epidendroideae

To determine the availability of the accD, ccsA, matK, and ycf1 sequences for phylogenetic analyses, the Epidendroideae was used as a case study because of the disputes regarding its systematics. Sequences of the four genes were successfully amplified in all 46 taxa. The aligned combined dataset comprised 4,593 characters, of which 2,839 represented variable sites and 1,447 were parsimony-informative sites. The number of variable sites was highest in ycf1 and lowest in ccsA (Table S6).
Table 1. Comparison of major features of seven orchid chloroplast genomes.

| Species                      | Dendrobium officinale | Cymbidium mannii | Erycina pusilla | Oncidium Gower Ramsey | Phalaenopsis Aphrodite | Phalaenopsis equestris | Cypripedium macranthos |
|------------------------------|------------------------|------------------|-----------------|-----------------------|------------------------|------------------------|------------------------|
| Subfamily                    | Epidendroideae         | Cypripedioideae  |                 |                       |                        |                        |                        |
| Accession Number             | KC771275               | KC876129         | JF746994        | GQ324949              | AY916449               | JF719062               | KF925434               |
| Size (bp)                    | 152,221                | 153,308          | 143,164         | 146,484               | 148,964                | 148,959                | 157,050                |
| LSC length (bp)              | 85,109                 | 85,212           | 84,189          | 82,324                | 85,957                 | 85,967                 | 85,292                 |
| SSC length (bp)              | 14,516                 | 17,471           | 12,097          | 12,650                | 11,543                 | 11,300                 | 18,285                 |
| IR length (bp)               | 26,298                 | 26,304           | 23,439          | 25,755                | 25,732                 | 25,846                 | 26,777                 |
| Total number of different genes | 110                   | 113              | 108             | 111                   | 110                    | 109                    | 113                    |
| Duplicated genes in IR       | 19                     | 19               | 19              | 19                    | 19                     | 19                     | 19                     |
| AT content %                 | 62.53                  | 63.06            | 63.35           | 62.68                 | 64.4                   | 63.35                  | 62.17                  |
| Protein-coding gene          | 61.48                  | 63.05            | 61.38           | 61.34                 | 61.97                  | 61.45                  | 61.57                  |
| Ribosomal RNA genes          | 45.34                  | 45.37            | 45.52           | 45.43                 | 45.51                  | 45.53                  | 45.21                  |
| Untranslated regions         | 67.00                  | 67.53            | 68.51           | 67.40                 | 68.18                  | 68.41                  | 66.10                  |
| Gene with introns            | 18                     | 18               | 17              | 18                    | 17                     | 17                     | 18                     |

doi:10.1371/journal.pone.0099016.t001
Figure 4. Comparison of the regions flanking the junctions (JLB, JLA, JSB, and JSA) among seven orchid chloroplast genomes. Four types of junctions are present at the JSB in seven orchid species. Numbers in green indicate the length of \textit{ycf1}. Numbers in orange indicate the distance between \textit{ndhF} and JSB. Numbers in purple indicate the distance between 5\textsuperscript{\prime}--\textit{ycf1} and JSA. Numbers in blue indicate the distance between \textit{rps19} and JLA. This figure is not to scale.

doi:10.1371/journal.pone.0099016.g004
Table 2. Pairwise distances of nucleotide and protein substitutions among six orchid species.

| Order | Gene | Region | DNA (d) | DNA (S. E.) | Protein (d) | Protein (S.E.) | Miss data | Range of length |
|-------|------|--------|---------|-------------|-------------|---------------|-----------|-----------------|
| 1     | rps7 | IR     | 0.0083  | 0.0242      | 0.0065      | 0.0036        |           | 468            |
| 2     | rpl23 | IR     | 0.0085  | 0.0344      | 0.0105      | 0.0078        |           | 270–282        |
| 3     | rpl2 | IR     | 0.0140  | 0.0255      | 0.0162      | 0.0052        |           | 819–837        |
| 4     | petG | LSC    | 0.0150  | 0.0069      | 0.0090      | 0.0086        |           | 114            |
| 5     | rps12 | IR/LSC | 0.0157  | 0.0365      | 0.0108      | 0.0051        |           | 372–387        |
| 6     | psbL | LSC    | 0.0192  | 0.0078      | 0.0000      | 0.0000        |           | 117            |
| 7     | psbZ | LSC    | 0.0234  | 0.0179      | 0.0280      | 0.0140        |           | 189            |
| 8     | ycf2 | IR     | 0.0243  | 0.0013      | 0.0507      | 0.0029        |           | 6,666–6,876    |
| 9     | petL | LSC    | 0.0255  | 0.0099      | 0.0323      | 0.0179        |           | 96             |
| 10    | psbF | LSC    | 0.0255  | 0.0090      | 0.0000      | 0.0000        |           | 120            |
| 11    | psbD | LSC    | 0.0264  | 0.0030      | 0.0096      | 0.0033        |           | 1062           |
| 12    | atpH | LSC    | 0.0269  | 0.0069      | 0.0000      | 0.0000        |           | 246            |
| 13    | atpI | LSC    | 0.0269  | 0.0393      | 0.0181      | 0.0050        |           | 744            |
| 14    | psaC | SSC    | 0.0281  | 0.0067      | 0.0000      | 0.0000        |           | 246            |
| 15    | psbE | LSC    | 0.0291  | 0.0066      | 0.0280      | 0.0114        |           | 246–252        |
| 16    | psbN | LSC    | 0.0298  | 0.0092      | 0.0233      | 0.0141        |           | 132            |
| 17    | petN | LSC    | 0.0316  | 0.0136      | 0.0230      | 0.0158        |           | 90–96          |
| 18    | psaA | LSC    | 0.0319  | 0.0025      | 0.0104      | 0.0024        |           | 2223           |
| 19    | ycfS | LSC    | 0.0339  | 0.0051      | 0.0214      | 0.0068        |           | 507            |
| 20    | psbC | LSC    | 0.0339  | 0.0034      | 0.0038      | 0.0019        |           | 1,224–1,461    |
| 21    | rps19 | IR     | 0.0340  | 0.0068      | 0.0313      | 0.0135        | Cyripedium | 264–279        |
| 22    | psbD | LSC    | 0.0343  | 0.0026      | 0.0135      | 0.0028        |           | 2205           |
| 23    | psbA | LSC    | 0.0357  | 0.0034      | 0.0064      | 0.0029        |           | 1062           |
| 24    | petA | LSC    | 0.0360  | 0.0038      | 0.0307      | 0.0058        |           | 963            |
| 25    | clpP | LSC    | 0.0380  | 0.0049      | 0.0325      | 0.0070        |           | 594–615        |
| 26    | petD | LSC    | 0.0383  | 0.0056      | 0.0231      | 0.0084        |           | 531–564        |
| 27    | petB | LSC    | 0.0390  | 0.0046      | 0.0149      | 0.0046        |           | 648–654        |
| 28    | rpsC2 | LSC | 0.0396  | 0.0021      | 0.0798      | 0.0047        |           | 4,137–4,167    |
| 29    | rbcL | LSC    | 0.0398  | 0.0038      | 0.0243      | 0.0049        |           | 1,443–1,473    |
| 30    | atpA | LSC    | 0.0416  | 0.0031      | 0.0333      | 0.0048        |           | 1,524–1,530    |
| 31    | atpB | LSC    | 0.0419  | 0.0036      | 0.0311      | 0.0049        |           | 1,488–1,497    |
| 32    | infA | LSC    | 0.0424  | 0.0081      | 0.0052      | 0.0049        | Cyripedium | 264–279        |
| 33    | rps4 | LSC    | 0.0426  | 0.0047      | 0.0544      | 0.0102        |           | 606            |
| 34    | ycfA | LSC    | 0.0432  | 0.0056      | 0.0500      | 0.0102        |           | 555            |
| 35    | psbA | LSC    | 0.0451  | 0.0037      | 0.0143      | 0.0034        |           | 1,527          |
| Order | Gene | Region | DNA (d) | Protein (d) | Protein (S.E.) | Miss data | Range of length |
|-------|------|--------|---------|-------------|----------------|-----------|----------------|
| 36    | psbI | LSC    | 0.046   | 0.0125     | 0.0000         | 111       | 369            |
| 37    | rpl14| LSC    | 0.046    | 0.0073     | 0.0399         | 372       | 3213           |
| 38    | rpl18| LSC    | 0.047    | 0.0041     | 0.0474         | 305       | 305-315        |
| 39    | rpl14| LSC    | 0.047    | 0.0011     | 0.0049         | 303       | 303            |
| 40    | rplC1| LSC    | 0.048    | 0.0032     | 0.0559         | 303       | 203-306.1      |
| 41    | rpl36| LSC    | 0.048    | 0.0143     | 0.0270         | 114       | 123            |
| 42    | rpl1 | LSC    | 0.049    | 0.0046     | 0.0300         | 369       | 402-495        |
| 43    | rps18| LSC    | 0.051    | 0.0087     | 0.0474         | 3213      | 687-690        |
| 44    | rps14| LSC    | 0.052    | 0.0014     | 0.0104         | 711       | 129-335        |
| 45    | rpl14| LSC    | 0.052    | 0.0165     | 0.0196         | 105       | 105-114        |
| 46    | rpl1 | LSC    | 0.052    | 0.0149     | 0.0095         | 415       | 1014-1201      |
| 47    | rps18| LSC    | 0.053    | 0.0076     | 0.0599         | 303       | 201            |
| 48    | rps14| LSC    | 0.055    | 0.0113     | 0.0100         | 222       | 222            |
| 49    | rps14| LSC    | 0.055    | 0.0087     | 0.0081         | 687-690   | 687-690        |
| 50    | rps18| LSC    | 0.056    | 0.0080     | 0.0079         | 306-359   | 306-359        |
| 51    | rps14| LSC    | 0.056    | 0.0069     | 0.0064         | 3213      | 212-335        |
| 52    | rps14| LSC    | 0.056    | 0.0049     | 0.0027         | 711       | 711            |
| 53    | rpl14| LSC    | 0.058    | 0.0127     | 0.0016         | 105-114   | 105-114        |
| 54    | rpl14| LSC    | 0.058    | 0.0165     | 0.0035         | 415       | 1014-1201      |
| 55    | rpl14| LSC    | 0.058    | 0.0048     | 0.0088         | 303       | 201            |
| 56    | rpl14| LSC    | 0.058    | 0.0044     | 0.0135         | 711       | 129-335        |
| 57    | rpl14| LSC    | 0.058    | 0.0105     | 0.0100         | 222       | 222            |
| 58    | rpl14| LSC    | 0.058    | 0.0061     | 0.0101         | 687-690   | 687-690        |
| 59    | rpl14| LSC    | 0.058    | 0.0081     | 0.0091         | 306-359   | 306-359        |
| 60    | rpl14| LSC    | 0.058    | 0.0080     | 0.0081         | 306-359   | 306-359        |
| 61    | rpl14| LSC    | 0.058    | 0.0105     | 0.0100         | 222       | 222            |
| 62    | rpl14| LSC    | 0.058    | 0.0081     | 0.0091         | 306-359   | 306-359        |
| 63    | rpl14| LSC    | 0.058    | 0.0048     | 0.0088         | 303       | 201            |
| 64    | rpl14| LSC    | 0.058    | 0.0027     | 0.0088         | 711       | 711            |
| 65    | rpl14| LSC    | 0.058    | 0.0027     | 0.0088         | 711       | 711            |
| 66    | rpl14| LSC    | 0.058    | 0.0027     | 0.0088         | 711       | 711            |
| 67    | rpl14| LSC    | 0.058    | 0.0027     | 0.0088         | 711       | 711            |
| 68    | rpl14| LSC    | 0.058    | 0.0027     | 0.0088         | 711       | 711            |

Order and S.E. indicate overall mean distance and standard error, respectively. Nucleotide and amino acid distances were calculated using Kimura’s 2-parameter model distances and a p-distance model, respectively.}

**Table 2. Cont.**

The DOI for the original publication is [10.1371/journal.pone.0099016] and the link to the original publication is [www.plosone.org/doi/10.1371/journal.pone.0099016].
Phylogenetic analyses using BI and ML approaches resulted in the same topology (Fig. 5). Most nodes had high support among tribes and subtribes within the subfamily Epidendroideae. Within it, Coelogyinae was sister to all the other subtribes or tribes, with strong support (ML BP 100%, BI PP 1.00). The Balbophyllum group clustered with the Epigeneum and Dendrobiun-Flickgingeria to form a monophyletic clade of Dendrobiinae that was closely allied to Malaxideae (Laparis and Othonna). The Dendrobiinae-Malaxideae clade was sister to the rest of the subfamily. Podochilinae and Eriinae were not monophyletic clades; these two subtribes (both of tribe Podochilae) were sister to Collabinae.

Discussion

Comparison of RNA editing sites

Involved in plastid posttranscriptional regulation, RNA editing provides an effective way to create transcript and protein diversity [60,61]. Some chloroplast RNA editing sites are conserved in higher plants [62,63]. In Orchidaceae, RNA editing sites were identified in 24 protein-coding transcripts in higher plants [62,63]. In Orchidaceae, RNA editing sites were identified in 24 protein-coding transcripts in higher plants [62,63]. Potential editing sites also were identified in P. equestre and O. Gower Ramsey [33]. Of the examined 30 genes in above-mentioned seven orchids, 15 potential RNA editing sites out of 11 genes (atpA, atpF, ctpP, mnaK, petB, petF, rpo20, rpoA, rpoB, rpoC1 and ycf3) were shared; the number of shared editing sites increased in Epidendroideae species (28 sites out of 16 genes) (Table S5 and [33]). Therefore, RNA editing is more conserved from the same subfamily than which from different subfamily. However, orchids and other angiosperms have relatively less common editing sites. For example, 10 potential RNA editing sites were shared by orchids and Cocos nucifera; comparisons among Nicotiana tabacum, Arabidopsis thaliana, grasses and orchid RNA editing sites showed low conservation of editing sites (only one common editing sites in rpoB) (Table S5). These cases indicate that the evolutionary conservation of RNA editing is essential for only a few plastid-editing sites [64-66].

IR expansion or contraction in the Orchidaceae

The variability of genes flanking IR/SSC junctions results in IR expansion or contraction [59,67]. At the IR/LSC boundaries, most IRs of non-orchid monocots contained trnH-psbA19 gene clusters, excluding trnV-psb22 genes, leading to more-progressive expansion of IRs than that having occurred in non monocot angiosperms [17,20,55,59,68-71]. In all known photosynthetic orchid cp genomes, trnH-psbA19 clusters and trnV-psb22 genes were involved in IRs at the IR/LSC junctions. The IR/LSC junctions were the standard type III [71], and IRs experienced the largest expansion at the IR/LSC junction compared with other monocots.

The IR/SSC junction types II–IV in orchids differed from those in other monocots, while type I (in Cyripedium and Dendrobium cp genomes) was similar to that in Acorus (Fig. 4) with ycf1 extending over the JSA and ycf1 located within IR adjacent to the JSB. Although Yang et al. suggested most likely evolutionary routes of IRs in monocots [59], no studies have proposed a model about the evolutionary dynamics of the IR/SSC junctions within orchids. Here, we hypothesize two evolutionary routes to explain the expansion or contraction of IRs adjacent to IR/SSC junctions from an Acorus-like ancestor to the existing orchids. The first route proceeded from type I to type II; ycf1 further expanded into the IRA, resulting in an expansion of duplicated ycf1 in the IRA. During this period, an overlap occurred between ndhF remnant and ycf1. On the second route, ycf1 shifted continuously into the SSC, resulting in a shorter, duplicated ycf1 adjacent to the JSA.

Continually, ycf1 was embedded completely into the SSC, leading to the loss of duplicated ycf1. This contractive process of IR involved the structural change from type I to type IV via type III. Moreover, IRs expansion or contraction may not correlate with the taxonomic relationships. More molecular data need to be collected for intensifying our understanding of variations in sequences flanking IR/SSC junctions.

The loss of the border between the IR and SSC in orchids was associated with the ycf1 gene. Compared with the average AT content of protein-encoding genes, all known orchid ycf1 genes exhibited usage bias of AT base pairs (see Table 1 and Table S7). AT base pairs are bound by two hydrogen bonds, while GC base pairs are bound by three hydrogen bonds; therefore, DNA with high AT content is less stable than that with low AT content. Pol (A) tract sequences at IR/LSC boundaries might be closely linked with the dynamics of IR/LSC junctions and expansion of IR [67,71]. Similarly, the AT-rich nature of ycf1 gene may be linked to the recombination of IR/SSC junction.

The loss or pseudogenization of ndh genes in orchid chloroplast genomes

Instances of gene loss or pseudogenes have been elucidated in the cp genomes of monocots [21]. Chloroplast-encoded gene degeneration in photosynthetic orchids is mostly embodied in structural changes of ndh genes. There are 11 chloroplast-encoded ndh genes in the cp genomes of land plants, located in several transcriptional units and encoding for the thylakoid Ndh complex [72]. Non-functional chloroplast-encoded ndh genes have been found in CAM and C3 plants [32], including gymnosperms and grasses [66,73,74]. Sequence truncations and indels are common phenomena in orchid chloroplast-encoded ndh genes [17,28–33,75]. Pseudogenization or loss of the ndh gene did not correlate well with the divergent patterns of Epidendroideae lineages observed in the phylogenetic trees (Fig. 3). However, 10 common ndh pseudogenes of two Phalaenopsis species showed a high degree of similarities in sequence and indel patterns [33]. Both Erycina and the allied genus Oncidium lost two ndh genes (ndhF and ndhK) and had six pseudogenes (ndhB, C, D, G, and J); similar results were obtained from Oncidium and related Oncidiinae varieties [29]. Thus, we infer that relative species had similar patterns of variation in ndh gene content.

The loss of some chloroplast-encoded genes might not affect the plant life cycle. Gene transfer from chloroplast to nucleus is known to occur frequently during evolutionary processes [76]. The ancestral plastid ndh genes of orchids are presumed to have been transferred to the nucleus [28]. Moreover, fungal symbionts may contribute to the fate of ndh genes [77]. Therefore, the functions of lost chloroplast-encoded ndh genes could be performed by homologous genes from other resources; this hypothesis needs to be tested in the future study.

Phylogenetic relationships based on complete cp genomes

The cp sequences have been used in deep phylogenetic analyses because of their low substitution rates [20,78]. Phylogenetic analyses based on complete chloroplast genomes have resolved some bewildering relationships in angiosperms. Using two tree construction methods with different models, we obtained consistent results on the relationships among Phalaenopsis (Aeridinae), Cymbidium (Cymbidinae), Dendrobium (Dendrobiinae), Oncidium and Erycina (Oncidinae) within Epidendroideae, which are congruent with matK and rbcL analyses by Gustafsson et al. (2010) [8] and morphological cladistic analysis by Freudenstein and Rasmussen
Phylogenetic reconstruction of the Epidendroideae

Variability of genes in cp genomes has been calculated according to nucleotide diversity in previous studies [80,81]. If we considered sequence divergence at the nucleotide and protein levels, rpoB, rpoC2, and ycf2 were conserved with low evolutionary distance, with the exception of rpoC2, which exhibited medium divergence in the IR regions. These results are consistent with previous reports of slower divergence of sequences in the IR regions compared to other regions [80,82]. Although the ycf2 gene has been demonstrated to be one of the most rapidly evolved genes among 16 vascular plant species [80], the present study showed that it had relatively slow nucleotide divergence and moderate protein divergence within the Orchidaceae.

In this study, highly divergent genes were acquired according to pairwise distance of nucleotide substitutions. While ycf1 was located at the IR/SSC junction, 14 other genes bordered the LSC and SSC regions, four of which were selected to construct phylogenetic trees. Of these, matK and ycf1 have been used in previous studies [4,6], while accD and coxA were applied for the first time to the phylogenetic analysis of the subfamily Epidendroideae in the present study. These genes can be used as good phylogenetic markers at the subfamily level because of the following three reasons. First, these regions are variable, which highlights their unusual evolutionary properties. According to the pairwise distance of protein substitutions (Table 2), ycf1 and matK have high divergence, and accD and coxA have relatively moderate substitution rates that are higher than rbcL, which has been used in previous systematic analyses within the Epidendroideae [4,5,75]. Second, these regions are sufficiently long (>600 bp) to yield adequate loci for phylogenetic analysis. Third, the sequences are easily obtained by PCR amplification and relatively conservative for alignment.

Phylogenetic reconstruction of the Epidendroideae

The phylogeny of the Epidendroideae has long been debated. Here, eleven common subtribes and one tribe from Epidendroideae were used as a case study to identify the phylogenetic relationships within this subfamily using four cp sequences. Fig. S1 illustrates the relationships among these subtribes or tribes in previous studies based on molecular data. With polyphyletic and paraphyletic groups excluded from phylogenetic analyses, major debates were the placement of Dendrobiinae, Malaxideae, and Collabiinae, as well as identification of the basal subtribe or tribe. On the basis of a concatenated data set, we clarified several relationships that were previously poorly resolved, and the majority of nodes at the subtribe level in the gene trees had high support.

The placement of Coelogyninae (tribe Arethuseae) varied according to morphological and molecular proofs. Based on observing that Arethuseae has cormous and reed-stem habits, Dressler (1986) claimed that Arethuseae is the basal group in the “reed-stem” phylad [83]. Dressler (1990) divided advanced Epidendroideae into four major clades (Gastrodioidae, Nerviieae, Cymbidioid phylad, and Epidendroid phylad); and placed Arethuseae and Dendrobioid subclade in Epidendroid phylad, and Maxillarieae, Cymbidieae, Malaxideae in Cymbidioid phylad [13]. Dressler (1993) held that Arethuseae appeared to be paraphyletic due to their ever-shifting boundaries and tenuous morphological definitions [84]. However, Van den Berg et al. (2005) fixed subtribe Coelogyninae in distinct positions based on different methods using nrITS and four plastid sequences [4]. The results in this study strongly support that Coelogyninae was the most basal subtribe within the sampled subtribes, which was in line with the BI analysis by Van den Berg et al. (2005) [4], the MP analysis by Neubig et al. (2009) [6], and analyses by Gorniak et al. (2010) [7].

Previously, the placement of Collabiinae and Dendrobiinae was problematic, but their positions have been recovered. Collabiinae was polyphyletic based on matK and rbcL [15]. Van Den Berg (2005) proposed Collabiinae was in an unfixed position in MP and BI analyses, and Gorniak et al. (2010) posited that Collabiinae was sister to Aeridinae and Erinaceae with high support based on nuclear gene Xdh [4,7]. Our results suggest that Collabiinae was sister to the Podochilinae-Eriinae (tribe Podochileae) clade with moderate support; this is congruent with MP analysis of Van Den Berg (2005), which had weak support [4]. The positions of Dendrobiinae and Malaxideae were also confirmed. Dressler (1990) placed Malaxideae and Dendrobieae in two separated groups, Cymbidioid phylad and Epidendroid phylad, according to reed stem, upper lateral inflorescences and spherical silica bodies [13]. Chase (2003) recognized Dendrobiinae as a subtribe rather than tribe Dendrobieae [2]. By inferring from nrITS and four chloroplast sequences, Van den Berg et al. (2005) held that Dendrobiinae was beside Malaxideae [4]. Dendrobiinae is similar to Malaxideae in the synapomorphic state of the naked pollinium [84]. Like other analyses based on Xdh and rbcL, our analyses support that Dendrobiinae and Malaxideae were sister relatives [5,7], which was consistent with the morphological similarities between them. Controversially, the position of Dendrobiinae-Malaxideae clade was going up in the analysis of Xdh [7], and Podochileae was sister to this clade in the analysis of plastid gene rbcL (bootstrap support <50%) [5]; however, this clade was sister to other clades except Coelogyninae with high support in the present study. More extensive sampling and sequencing of mitochondrial and nuclear genomes should be conducted to resolve uncertain relationships within the Epidendroideae with confidence.

Conclusions

In summary, complete chloroplast genomes can provide abundant information for resolving evolutionary questions. The gene content, organization, and sequence of chloroplast genome have been used as important markers in systematic research. This study determined complete cp genomes of Dendrobium officinale and Cyrtopodium macranthos and compared cp genomes of seven photosynthetic orchids including the above two, which showed structural similarities but differences in IR/SSC junctions and ndh genes. We propose that the AT bias of ycf1 in the Epidendroideae may be related to recombination of the IR/SSC junction. In...
addition, relationships among subtribes and tribes in the subfamily Epidendroideae were resolved with high or moderate support in the present study. The highly divergent genes of cp genomes identified in this study can be used as markers in phylogenetic analyses. Further plastome sequencing of orchids will be necessary to clarify the diversity of chloroplast genomes and to improve our understanding of the relationships within this family.

Supporting Information

Figure S1 Phylogenetic relationships among 11 subtribes and one tribe within the subfamily Epidendroideae resulting from previous studies. All trees were drawn according to the cited references. Molecular markers and methods used in phylogenetic analyses are enclosed by parentheses below the cited studies. Subtribal and tribal delimitations refer to Chase et al. [2].

Table S1 Primers used for gap closure, assembly and junction verification.

Table S2 Accession numbers for taxa used in phylogenomic analysis and genome comparison.

Table S3 Primers for phylogenetic analyses of orchids.

References

1. Chase MW (2005) Classification of Orchidaceae in the age of DNA data. Curtis’s Bot Mag 22: 2–7.
2. Chase MW, Cameron KM, Barrett RL, Freudenstein JV (2003) DNA data and systematics of Orchidaceae: a new phylogenetic classification. In: Dixon KW, Kell SF, Barrett RL, Cribb PJ, editors. Orchid conservation. Kota Kinabalu, Sabah, Malaysia: Natural History Publications. pp. 69–89.
3. Raubeson LA, Jansen RK (2005) Chloroplast genomes of plants. In: Henry RJ, editor. Plant diversity and evolution: genotypic and phenotypic variation in higher plants. Cambridge: CAB International. pp. 45–68.
4. Van den Berg C, Goldman DH, Freudenstein JV, Prideon AM, Cameron KM, et al. (2005) An overview of the phylogenetic relationships within Epidendroideae inferred from multiple DNA regions and recircumscription of Epidendreae and Arethuseae (Orchidaceae). Am J Bot 92: 613–624.
5. Cameron KM, Chase MW, Whittem WM, Korov Pj, Jarrell DC, et al. (1999) A phylogenetic analysis of the Orchidaceae: evidence from dL nucleotide. Am J Bot 86: 208–224.
6. Neubig KM, Whittem WM, Carllward BS, Blanco MA, Endara L, et al. (2009) Phylogenetic utility of matK in orchids: a plastid gene more variable than atp6. Plant Syst Evol 227: 75–84.
7. Gorniak M, Paun O, Chase MW (2010) Phylogenetic relationships within Orchidaceae based on a low-copy nuclear coding gene, Xfrk: congruence with organellar and nuclear ribosomal DNA results. Mol Phylogenet Evol 36: 784–795.
8. Gustafsson AL, Verola CF, Antonelli A (2010) Reassessing the temporal evolution of orchids with a new fossil and a Bayesian relaxed clock, with implications for the diversification of the rare South American genus Hoffmannseggia (Orchidaceae: Epidendroideae). BMC Evol Biol 10: 177.
9. Freudenstein JV, Harris EM, Rasmussen FN (2002) The evolution of another morphology in orchids: incumbent anthers, superposed pollinia, and the vandoid complex. Am J Bot 89: 1747–1755.
10. Cameron KM (2004) Utility of plastid psbA gene sequences for investigating intratribal relationships within Orchidaceae. Mol Phylogenet Evol 31: 1157–1160.
11. Dresler RL (1981) The orchids: natural history and classification. Cambridge: Harvard University Press. 335p.
12. Freudenstein JV, Rasmussen FN (1999) What does morphology tell us about orchid relationships?—a cladistic analysis. Am J Bot 86: 225–248.
13. Dresler RL (1990) The major clades of the Orchidaceae – Epidendroideae. Lindleyana 5: 115–125.
14. Burns-Balogh P, Funk VA, editors (1986) A Phylogenetic Analysis of the Orchidaceae. Washington: Smithsonian Institution Press. 79 p.
15. Freudenstein JV, van den Berg C, Goldman DH, Kores PJ, Molvray M, et al. (2004) An expanded plastid DNA phylogeny of Orchidaceae and analysis of jackknife branch support strategy. Am J Bot 91: 149–157.
16. Burns-Balogh P, Funk VA (1986) A phylogenetic analysis of the Orchidaceae: a summary. Lindleyana 1: 131–139.
17. Yang JB, Tang M, Li HT, Zhang ZR, Li DZ (2013) Complete chloroplast genome of the genus Cymbidium: lights into the species identification, phylogenetic implications and population genetic analyses. BMC Evol Biol 13: 84.
18. Dong W, Xu C, Cheng T, Lin K, Zhou S (2013) Sequencing angiosperm plastid genomes made easy: a complete set of universal primers and a case study on the phylogeny of Saxifragales. Genome Biol Evol 5: 989–997.
19. Watanabe S, Kirakae T, Miyoshi-Akiyama T (2015) Complete genome sequence of Streptoscyos dysgalactia subsp. equisilius 167 carrying Lancefield group C antigen, and comparative genomics of S. dysgalactia subsp. equisilius strains. Genome Biol Evol 5: 1641–1651.
20. Wu QZ, Ge S (2012) The phylogeny of the BEP clade in grasses revisited: evidence from the whole-genome sequences of chloroplasts. Mol Phylogenet Evol 62: 573–578.
21. Liu J, Qi ZC, Zhao YP, Fu GX, Jenny Xiang QY (2012) Complete cpDNA genome sequence of Smilax chine and phylogenetic placement of Liladases-influences of gene partitions and taxon sampling. Mol Phylogenet Evol 64: 545–562.
22. Green BR (2011) Chloroplast genomes of photosynthetic eukaryotes. Plant J 66: 34–44.
23. Liu Y, Huo N, Dong L, Wang Y, Zhang S, et al. (2013) Complete Chloroplast Genome Sequences of Mongolia medicine Artemisia frigida and Phylogenetic Relationships with Other Plants. PLoS One 8: e57533.
24. Grewe F, Gao W, Gubbels EA, Hansen AK, Mower JP (2013) Complete plastid genomes from Ophioglossum vulgatum, Peltium nudum, and Equisetum hyemale reveal an ancestral land plant genome structure and resolve the position of Equisetales among monilophytes. BMC Evol Biol 13: 8.
25. Gao L, Yi X, Yang YX, Su YJ, Wang T (2009) Complete chloroplast genome sequence of a fern Allophylus spinulosa: insights into evolutionary changes in fern chloroplast genomes. BMC Evol Biol 9: 130.
26. Wolf PG, Roper JM, Duffy AM (2010) The evolution of chloroplast genome structure in ferns. Genome 53: 731–738.
27. Wu CS, Lin CP, Hu CY, Wang RJ, Chaw SM (2011) Comparative chloroplast genomes of Pinaceae: insights into the mechanism of diversified genomic organizations. Genome Biol Evol 3: 309–319.
28. Chang CC, Lin HC, Lin IP, Chow TY, Chen HH, et al. (2006) The chloroplast genome of Phalanopsis aphrodite (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. Mol Biol Evol 23: 279–291.
29. Wu FH, Chan MT, Liao DC, Hsu CT, Lee YW, et al. (2010) Complete chloroplast genome of Oncidium Gower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiumae. BMC Plant Biol 10: 60.
30. Delannoy E, Fuji S, Colas des Francs-Small C, Brundrett M, Small I (2011) Rampant gene loss in the underground orchid *Rhzanthella gardneri* highlights evolutionary constraints on plastid genomes. Mol Biol Evol 28: 2037–2056. 
31. Logacheva MD, Schelkunov MI, Pein AV (2011) Sequencing and analysis of plastid genome in mycoheterotrophic orchid *Notothyla nudicaulis*. Genome Biol Evol 3: 1293–1298. 
32. Pan JC, Liao DC, Wu FH, Dianhui H, Singh ND, et al. (2012) Complete chloroplast genome sequence of an orchid model plant candidate: *Eyraea paulii* apply in tropical Orchidum breeding. PLoS One 7: e34738. 
33. Jiang CF, Chen TG, Lin JY, Wu Wl, Chang CC (2012) The comparative chloroplast genomic analysis of photosynthetic orchids and developing DNA markers to distinguish *Phalaenopsis* orchids. Plant Sci 190: 62–73. 
34. Barrett CF, Davis JI (2012) The plastid genome of the mycoheterotrophic *Ceaefferia stiata* (Orchidaceae) is in the relatively early stages of degradation. Am J Bot 99: 1513–1523. 
35. Zhu GH, Ji ZH, Wood JF, Wood HP (2009) *Dendrotham* In: Wu CY, Raven PH, Hong DY, editors. Flora of China, 25. Beijing: Science Press. pp. 367–397. 
36. Committee SP (2010) Pharmacopoeia of the People's Republic of China. Beijing: People's Medical Publishing House. 2853 p. 
37. Fu L (1992) China plant red data book - rare and endangered plants, I. Beijing: Science Press: 714p. 
38. Cox AV, Pridgeon AM, Albert VA, Chase MW (1997) Phylogenetics of the slipper orchids (Cypripedioideae, Orchidaceae): nuclear rDNA ITS sequences. Pl Syst Ecol 208: 197–223. 
39. Chen XQ, Liu ZJ, Cribb PJ (2009) Subliman. Cypripedioideae. In: Wu CY, Raven PH, Hong DY, editors. Flora of China, 25. Beijing: Science Press. pp. 22–44. 
40. Robinson SP, Downotn WJ (1964) Potassium, sodium, and chloride content of isolated intact chloroplasts in relation to ionic compartmentation in leaves. Arch Biochem Biophys 228: 197–206. 
41. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19: 11–15. 
42. Borstegrom E, Lundin S, Lundeborg J (2011) Large scale library generation for high throughput sequencing. PLoS One 6: e19119. 
43. Li R, Fan W, Tian G, Zhu H, He L, et al. (2010) The sequence and de novo assembly of the giant panda genome. Nature 463: 311–317. 
44. Wisman SK, Jansen RK, Boore JL (2004) Automatic annotation of anglerfish genomes with DOGMA. Bioinformatics 20: 3252–3255. 
45. Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, muscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res 33: W667–689. 
46. Laedtet D, Canback B (2004) ARAGORN, a program to detect RNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32: 11–16. 
47. Conant GC, Wolfe KH (2008) GenomeVx: simple web-based creation of editable circular chromosome maps. Bioinformatics 24: 861–862. 
48. Mower JP (2009) The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. Nucleic Acids Res 37: W253–259. 
49. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGAS: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739. 
50. Silvestro D, Michalak I (2012) raxmlGUI: a graphical front-end for RAxML. Organisms Diversity & Evolution 12: 335–337. 
51. Lartillot N, Lepage T, Blanquart S (2004) PhyloBayes 3: a Bayesian software for molecular evolutionary genetics analysis using maximum likelihood, evolution- ary distance, and maximum parsimony methods. Mol Biol Evol 21: 1261–1269. 
52. Lewis GD, Chong CL, Chang CC, Wu CL, Su TM, et al. (2008) Dynamics and evolution of the inverted repeat-large single copy junctions in the plastid genomes of monocots. BMC Evol Biol 8: 36. 
53. Martin M, Sabater B (2010) Plastid snf genes in plant evolution. Plant Physiol 48: 636–645. 
54. McCoy SR, Kuehl JV, Boore JL, Raubeson LA (2008) The complete plastid genome sequence of *Goic lacryma-jobi* and a comparative molecular evolutionary analysis of plastomes in cereals. J Mol Evol 69: 311–318. 
55. Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, et al. (2008) Dynamics and evolution of the inverted repeat-large single copy junctions in the plastid genomes of monocots. BMC Evol Biol 8: 130. 
56. Wu CS, Lai YT, Lin CP, Wang YN, Chaw SM (2009) Evolution of reduced and compact chloroplast genomes (cpDNAs) in guineophytes: selection toward a lower-cost strategy. Mol Phylogenet Evol 52: 113–124. 
57. Barrett CF, Feinstein JV (2008) Molecular evolution of snfL in the mycoheterotrophic coralroot orchids (Ceaafferia Gageana, Orchidaceae). Mol Phylogenet Evol 47: 665–679. 
58. Huang CY, Ayliffe MA, Timmis JN (2005) Direct measurement of the transfer rate of chloroplast DNA into the nucleus. Nature 422: 72–76. 
59. Wang B, Qiu YL (2006) Phylogenetic distribution and evolution of mycorrhizas in land plants. Mycorrhiza 16: 299–363. 
60. Jansen RK, Cai Z, Raubeson LA, Daniel H, Deparmelas CW, et al. (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci 104: 19369–19374. 
61. Sohn DE, Albert VA, Savilevich Y, Hsia K, Qiu YL, et al. (2004) Genome-scale data, angiosperm relationships, and "ending incongruence": a cautionary tale in phylogenetics. Trends Plant Sci 9: 477–483. 
62. Kim KJ, Lee HL (2004) Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Neay) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res 11: 247–261. 
63. Qian J, Song J, Gao H, Zhu Y, Xu J, et al. (2013) The Complete Chloroplast Genome Sequence of the Medicinal Plant *Salvia miltiorrhiza*. PLoS One 8(6): e67907. 
64. Maier RM, Neerckermann K, Iglói GL, Kossel H (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. J Mol Biol 251: 614–628. 
65. Dresler RL (1986) Recent advances in orchid phylogeny. Lindleyana 1: 5–20. 
66. Dresler RL (1993) Phylogeny and Classification of the Orchid Family. Portland, Oregon, USA: Timber Press. 514p.