To be submitted to: *Scientific Reports*

Running title: Plastid phylogenomics of the orchid family

Plastid phylogenomics of the orchid family: Solving phylogenetic ambiguities within Cymbidieae and Orchidoideae

Maria Alejandra Serna-Sánchez[a,b], Astrid Catalina Alvarez-Yela[c], Juliana Arcila[a], Oscar A. Pérez-Escobar [d], Steven Dodsworth[e] and Tatiana Arias[a*]

a Laboratorio de Biología Comparativa. Corporación para Investigaciones Biológicas (CIB), Cra. 72 A No. 78 B 141, Medellín, Colombia.

b Biodiversity, Evolution and Conservation. EAFIT University, Cra. 49, No. 7 sur 50, Medellín, Colombia

c Centro de Bioinformática y Biología Computacional (BIOS). Ecoparque Los Yarumos Edificio BIOS, Manizales, Colombia.

d Comparative Plant and Fungal Biology, Royal Botanic Gardens, Kew, TW9 3AE, London, UK.

e School of Life Sciences, University of Bedfordshire, University Square, Luton, LU1 3JU, UK.

* Corresponding Author: T.A.: Corporación para Investigaciones Biológicas, Cra. 72 A No. 78 B 141, Medellín, Colombia. E-mail: tatiana.tatianaarias@gmail.com

All data have been deposited in Bioproject (XXXXXXX) and SRA (XXXXXXX, Appendix 1).

ABSTRACT

Recent phylogenomic analyses have solved evolutionary relationships between most of the Orchidaceae subfamilies and tribes, yet phylogenetic relationships remain unclear within the hyperdiverse tribe Cymbidieae and within the Orchidoideae subfamily. Here we address these knowledge-gaps by focusing taxon sampling on the Cymbidieae subtribes Stanhopeinae, Maxillariinae, Zygopetalinae, Eulophiinae, Catasetinae, and Cyrtopodiinae. We further provide a more solid phylogenomic framework for the Codonorchideae subtribe within the Orchidoideae subfamily. Our global phylogenetic analysis includes 86 plastomes obtained from GenBank and 11

newly sequenced orchid plastomes genomes using a Genome Skimming approach. Whole genome phylogenies confirmed phylogenetic relationships in Orchidaceae as recovered in previous studies. Our results provide a more robust phylogenomic framework together with new hypotheses on the evolutionary relationships among subtribes within Cymbidieae, compared with previous phylogenies derived from plastome coding regions. Here, maximum statistical support in a maximum likelihood analysis was achieved for all the internal relationships in Cymbidieae, and Maxillariinae is recovered as sister to Oncidiinae for the first time. In Orchidoideae, we recovered Codonorchideae + Orchideae as a strongly supported clade. Our study provides an expanded plastid phylogenomic framework of the Orchidaceae and provides new insights on the relationships of one of the most species-rich orchid tribes.

Key words: Cymbidieae, High-throughput sequencing, Orchidaceae, Orchidoideae, Phylogenomics, Whole Plastid Genome

1. Introduction

The Orchidaceae, with *ca*. 25,000 species and ~800 genera[1,2] is one of the most diverse and widely distributed flowering plant families on earth and has captivated scientists for centuries[3]. The family has a striking floral morphological diversity and has evolved multiple interactions with fungi, animal and plants[4,5], and a diverse array of sexual systems[6,7]. Countless research efforts have been made to understand the natural history, evolution and phylogenetic relationships within the family[2,7–12]. To date, there are six nuclear genome sequences available, i.e., *Apostasia shenzhenica*[13], *Dendrobium catenatum*[14], *Dendrobium officinale*[15], *Gastrodia elata*[16], *Phalaenopsis* hybrid cultivar[17], *Phalaenopsis aphrodite*[18], *Vanilla planifolia*[19], 287 complete plastid genomes and 1,639 Sequence Read Archives for Orchidaceae in NCBI.

Phylogenomic approaches have been implemented to solve the main relationships between major orchids lineages in deep time[2,9,11,12], nevertheless extensive uncertainties remain regarding the phylogenetic placement of several subtribes and countless genera and species. This knowledge-gap stems from the large gaps in both taxon and genomic sampling efforts that would be required to comprehensively cover all orchid lineages at the subtribal and/or generic level. Givnish[2]

published the first well-supported phylogeny for the Orchidaceae based on plastid phylogenomic analyses. They used 75 genes from the plastid genome of 39 orchid species and performed a Maximum Likelihood (ML) analysis covering 22 subtribes, 18 tribes and five subfamilies. This robust but taxonomically-under-sampled study agrees with most of the phylogenetic relationships between and inside subfamilies and tribes, when compared with previous multilocus phylogenies[9–12].

Multiple relationships scattered across the orchid family remain unresolved, however, partly due to the limited phylogenetic information of plastid genes to resolve relationships in rapidly diversifying lineages[20,21] but also because of reduced taxon sampling[22]. This is particularly true for the Cymbidieae, one of the most species-rich tribes whose internal sub-tribal relationships are largely the product of rapid diversifications[23] that are often difficult to resolve using only a few loci[21,24]. The tribe Cymbidieae comprises 10 subtribes, ~145 genera and nearly 3,800 species[1], 90% of which occur in the Neotropical region[23]. Four of the subtribes within Cymbidieae are some of the most species-rich and abundant subclades in the Andean region (Maxillariinae, Oncidiinae, Stanhopeinae and Zygopetaliinae[25]).

Another group whose sub-tribal phylogenetic positions are largely unresolved is the Orchidoideae subfamily[1,26]. This group comprises four tribes, 25 subtribes and more than 3,600 species, the majority of which are terrestrial. The subfamily is distributed in all continents except the Antarctic and contains species with a single stamen (monandrous), with a fertile anther that is erect and basitonic[27]. Previous efforts to disentangle the phylogenetic relationships in the subfamily have mostly relied on a small set of nuclear and plastid markers[28], and more recently on extensive plastid coding sequence data[2].

The wide geographical range of these groups in the tropics and temperate regions, together with their striking vegetative and reproductive morphological variability place them as ideal model lineages for disentangling the contribution of abiotic and biotic drivers of orchid diversification across biomes. Occurring from alpine ecosystems to grasslands, they have conquered virtually all ecosystems available in any altitudinal gradient[29–31]. Moreover, they have evolved a diverse array of pollination systems[32–34], including male Euglossine-bee and pseudo-copulation[35,36]. Yet the absence of a solid phylogenetic framework has precluded the study of how such systems evolved, as well as the diversification dynamics of Cymbidieae and Orchidoideae more broadly.

Phylogenies are crucial to understanding the drivers of diversification in orchids, including the mode and tempo of morphological evolution[25,37]. High-throughput sequencing and modern

comparative methods have enabled the production of massive molecular datasets to reconstruct evolutionary histories, and thus provide unrivalled knowledge on plant phylogenetics[38]. Here we present the most densely sampled plastome phylogeny of the Orchidaceae, including eleven new plastid genomes, which expand the current generic representation for the Orchidaceae and clarify previously unresolved phylogenetic relations within the Cymbidieae and Orchidoideae. Two general approaches were used: a) phylogenetic analysis using whole plastome sequences, and b) phylogenetic analysis using 60 coding regions. The two different topologies reported here provide a robust phylogenomic framework of the orchid family and new insights into relationships at both deep and shallow phylogenetic levels.

2. Results

2.1 High-throughput sequencing of orchid plastid genomes

Eleven new orchid plastid genomes were sequenced. Supplementary table S1 shows the amount of sequencing data produced for each sample. From 4.9 Mb (*Gongora pleiochroma*) to 10.8 Mb (*Goodyera repens*) of raw reads were recovered from all samples (Table S1). The plastid genome with the highest average coverage was that of *Scaphosepalum antenniferum* (292X), and the one with the lowest average coverage was that of *Maxillaria sanderiana* (13X) (Table S1). The smallest plastid genome corresponds to *Maxillaria sanderiana* (132,712 bp) and the largest corresponds to *Sobralia mucronata* (161,827 bp) (Fig. S1 & Table 1). GC content was similar among all 11 plastomes and it ranges from 37 to 38.6%. The *M. sanderiana* plastome contains 123 different genes, of which 99 were single-copy and 24 were duplicated. Of these genes, 62 are protein-coding genes, four are rRNA genes and 33 are tRNA (Fig. S1 & Table 1). All new plastomes reported here have rRNA genes (rRNA4.5, rRNA5, rRNA16S, rRNA23S) and approximately 13 tRNA genes are located in the inverted repeat regions (Fig. S1).

2.2 Phylogenomic inferences from whole plastid genomes and coding regions

The ML tree derived from the complete plastid genome alignment is provided in Fig. 1. Virtually all nodes were recovered as strongly supported (i.e. LBS = 90-100), except for the relationship between Cymbidieae and Vandeae tribes (LBS = 71) and the MRCA of *Goodyera procera*, *G. repens* and *G. schlechteriana* (LBS = 57).

The analysis performed using 60 concatenated protein-coding regions further yielded a strongly supported phylogeny. Most of the nodes were recovered as strongly supported (LBS = 90-100, PP = 0.77-1.0), and only a few positions remained unresolved. Here, the relationship between Codonorchidae+Orchideae was moderately supported (LBS = 86) together with that of Cymbidiinae and the remaining Cymbidieae (LBS = 62). The monophyly of Nervilieae and Triphoreae was moderately supported (LBS = 79), as well as the phylogenetic relationships of Nervilieae+Triphoreae and the remainder of Epidendroideae (LBS = 75), and Epidendreae and *Coelia* + *Eria* (LBS = 52) (Fig. 2).

2.3 Molecular characterisation of plastid genomes

Whole plastome sequences belonging to 97 species (11 sequenced here and 86 reported in NCBI) were annotated for 75 protein-coding genes. Five additional genes were recovered when concatenating this data matrix with the protein coding regions matrix used by Givnish[2], giving a

total number of 80 genes for 124 orchid species and three outgroups.

On average, 30 transfer RNA (tRNA) genes, and four ribosomal RNA (rRNA) genes were also identified. Annotated genes belong to photosystems I and II, the cytochrome b/f complex, ATP synthase, NADH dehydrogenase, RubisCO large subunit, RNA polymerase, ribosomal proteins, clpP, matK, hypothetical plastome reading frames (ycf), transfer RNAs and ribosomal RNAs. It is common to find tRNA genes, ribosomal RNAs, ribosomal protein genes, *ndhB* and *ycf2* genes within the inverted repeated regions (IR) of orchid plastomes. Genes such as *ycf1*, ribosomal protein genes, photosystem genes and the majority of the *ndh* genes are commonly found within the short single copy region (SSC) (Fig. S1). Finally, the rest of the protein-coding genes are found in the long single copy region (LSC), as well as other tRNA genes (Table 1).

From these 80 genes, 20 were found to be problematic due to being out of reading frame or having multiple stop codons (*accD, ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK, petA, petB, petD, rpl16, rpoC1, rpoC2, rps12, ycf1*), and thus they were not included in the final alignment, which had a final sequence length of 41,942 bp.

Consistent losses of the *ndhF* gene were reported in 5 of the 11 new plastid genomes (*Gongora pleiochroma, Maxillaria nasuta, Maxillaria sanderiana, Otoglossum globuliferum* and *Telipogon glicensteinii*). The tRNA genes *trnT-UGU*, *trnI-AAU*, and *trnG-UCC* were also commonly lost in 7 plastid genomes. The plastome of *Sobralia mucronata* has all tRNA genes, but *Sobralia decora* and *Sobralia mandonii* lack *trnG-UCC*. Contrastingly, *Maxillaria sanderiana* lacks *trnT-UGU* and *trnI-AAU*. The gene *ndhK* is lost in *Gongora pleiochroma* and *Telipogon glicensteinii*. The plastome reported to have experienced the most genes losses is *Telipogon glicensteinii*, which lacks *ndhC, ndhF, ndhJ, ndhK, trnT-UGU, trnI-AAU, trnG-UCC* and *trnL-CAG*. The 11 plastomes have portions of the genes *rpl22* and *ycf1* duplicated, contributing to the expansion among inverted regions flanking the small single-copy region (Fig. S1).

3. Discussion

3.1 Orchid plastome evolution

Comparing orchid plastomes with the *Nicotiana tabacum* plastid genome reported at NCBI, some differences were identified. In terms of total gene content, *N. tabacum* plastome has 144 genes, whilst in orchids the gene content is around 120. Protein-coding genes are more abundant in *N. tabacum* than in orchids, being 98 and around 62 respectively. Two protein-coding genes found in orchid plastomes (*infA* and *pbf1*) were not found in *N. tabacum*, and six protein-coding genes (*ndhB, rpl2, rpl23, rps12, rps7 and ycf2*) were found as duplicated genes within the IR regions in both plastomes. Many studies have documented the movement of the *ndh* genes between the plastid genome and the nucleus. The *N. tabacum* plastome has 11 *ndh* genes (*ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*), in common with the plastid genome of *Apostasia wallichii*, which has been shown to transcribe all 11 *ndh* genes and these have been predicted to be translated into functional proteins[39]. These findings indicate that the common ancestor of orchids likely had a complete functional set of *ndh* genes. For some other orchids, not all those 11 genes are present, as in the case of *Gongora pleiochroma*, where just 8 *ndh* genes are present (*ndhA, ndhB, ndhC, ndhD, ndhE, ndhG, ndhH, ndhI*).

Diverse patterns of junctions between IR and SSC regions are seen in the 11 orchids sequenced here. Some plastomes have portions of the genes *rpl22* and *ycf1* within the IR region. Those genes seem to be repeated in some orchids, contributing to the expansion and contraction among the inverted regions, which flank the small single-copy of the plastomes. Studies regarding plastome content have also found both loss and retention of *ndh* genes among orchids[40,41]. Few *ndh* genes are thought to encode for functional *ndh* proteins in *Oncidium* and *Cymbidium*[42,43]. *ndh* gene function is thought to be related to land plant adaptation and photosynthesis[44]. However, Lin[41] found that no significant differences in biogeography or growth conditions (including light and water requirements) were observed between orchids where *ndh* genes were lost and orchids where the same *ndh* genes are present. Mechanisms leading to shifts in IR boundaries and the variable loss or retention of *ndh* genes are still unclear[12,40].

3.2 Extended support for major relationships in orchids

Previous phylogenomic studies of the orchid family included up to 74 species representing 18 tribes, 18 subtribes and 63 genera[22]. Our study sampled 94 species from all subfamilies, representing 15 tribes, 18 subtribes and 29 genera. In general, our phylogenomic frameworks are essentially in agreement with previously published family-wide orchid phylogenies either inferred

from dozens of markers[2,12] or from a handful of loci[24]. Here, representativeness within Cymbidieae has increased from 8[2] to 12 genera, whilst two new genera were included from the Pleurothallidinae subtribe (Epidendreae).

Our whole plastome analysis led to similar results as reported by Givnish et al. (2015) and Niu et al. (2017). Sampling within subtribes (Stanhopeinae, Maxillariinae, Oncidiinae, Eulophiinae and Cymbidiinae) resulted in the same topologies but with higher bootstrap values higher in all cases compared to previously published results (Figs. 3 and 4). Twenty protein-coding genes were identified as problematic due to multiple stop codons and uncertain ORFs. Few species could be aligned to the *ycf1* gene, which if included, may have caused noise in the phylogenetic analysis. Some of these genes have also been removed from other orchid phylogenies previously reported, for similar reasons[43,45,46].

3.3 Evolutionary relationships within Cymbidieae

Several phylogenies have been generated by morphological and molecular analyses in order to solve relationships within Cymbidieae[23,24]. Relationships among subtribes have recently been inferred using plastome coding genes *psaB, rbcL, matK, ycf1* combined with the low-copy nuclear gene *Xdh*[21]. In that study, the proposed phylogeny placed Cymbidiinae as sister to the rest of the Cymbidieae tribe. Poor support, however, and incongruent topologies were found among Catasetinae, Eulophiinae and Eriopsidinae subtribes with respect to the topologies obtained by Whitten et al. (2014), Freudenstein & Chase (2015) and Pérez-Escobar et al. (2017). In these phylogenies Eulophiinae and Catasetinae formed a clade. Also, Eriopsidinae was not clearly placed in the results obtained by Li et al. (2016), but it was strongly-supported as the sister group of (Maxillariinae(Stanhopeinae(Coeliopsidinae))) in Freudenstein & Chase (2015) and Pérez-Escobar et al., (2017). In Li et al. (2016), Cyrtopodiinae appears as the second outermost group differing from the topology obtained in Givnish et al. (2015), in which Cyrtopodiinae is clustered with Catasetinae.

Orchid phylogenomics using the most complete taxonomic sampling to date[2] included 8 of 10 subtribes belonging to Cymbidieae, but some subtribal relationships are still unresolved: Stanhopeinae (20 genera), Maxillariinae (12 genera), Zygopetalinae (36 genera), Oncidiinae (65 genera) and Eulophiinae (13 genera). A clade formed by Stanhopeinae and Maxillariinae had poor statistical support (BS=62) and their relationship with respect to Zygopetalineae had moderate support (BS=72). Relationship between sister clades Eulophiinae and a clade containing Stanhopeinae, Maxillariinae, Zygopetalinae, and Oncidiinae also had poor support (BS=42).

The outcome of our expanded sampling is the improvement of statistical support in

Cymbidieae, more specifically in the nodes of groups that arose from rapid diversifications and that historically have been problematic to resolve[2,24]. Our results provide resolution among Cymbidieae subtribes; however, we are still constrained by the lack of representatives for the subtribes Eriopsidiinae and Coleopsidinae. In our phylogeny, obtained using 60 plastome-coding regions, the relationships of Stanhopeinae with Zygopetalinae, and Oncidiinae with Maxillariinae differ from previous studies[2,23]. Also, our coding genes phylogeny disagrees with the whole plastome phylogeny presented here (Fig. 5). When using whole plastomes, Stanhopeinae remains as a sister group to Maxillariinae. However, when using only coding regions, Stanhopeinae is defined as sister to Zygopetalinae, and both are sister subtribes to the Maxillariinae + Oncidiinae clade (Fig. 5).

The Cymbidieae phylogenies proposed by Freudenstein & Chase (2015), Li et al., (2016), Pérez-Escobar et al., (2017) differ from the one presented here through coding regions analysis. Differences are found in the placement of the subtribes Maxillariinae (sister to Stanhopeinae), Zygopetalinae (sister to Maxillarinae and Stanhopeinae) and Eulophinae, which is sister to Catasetinae in studies reported by Freudenstein & Chase (2015) and Pérez-Escobar et al., (2017). Li et al., (2016) and Pérez-Escobar et al., (2017) found Dipodiinae (*Dipodium*) as the sister subtribe to the rest of Cymbidieae. However, the genus *Dipodium* has been previously included within Eulophiinae[1] and it is not represented in our phylogeny. Phylogenetic relationships within the tribe Cymbidieae have changed through the years according to the available data and approximations taken, either morphological and/or genetic. In Dressler, (1993), Cymbidieae contained seven subtribes (Goveniinae, Bromheadiinae, Eulophiinae, Theostelinae, Cyrtopodiinae, Acriopsidinae and Catasetinae), and circumscriptions were very different from what is currently accepted. A later study has shown that Cymbidieae could comprise up to 11 subtribes[21], but the latest study[23] reported 10 well-supported and circumscribed subtribes: (Cymbidiinae,((Cyrtopodiinae,(Catasetinae,Eulophiinae)),(Oncidiinae,(Zygopetalinae,(Eriopsidinae,(Maxillariinae,(Coeliopsidinae,Stanhopeinae))))))). Some topological differences can be identified with respect to our study. Here, relationships among most derived subtribes showed Stanhopeinae as a sister group to Zygopetalinae, and Maxillariinae as the sister subtribe of Oncidiinae. Also, the position of Eulophiinae within Catasetinae and Cyrtopodiinae, does not agree with our findings, because Eulophiinae was placed as sister group to the most derived Cymbidieae subtribes, and Catasetinae was clustered together with Cyrtopodiinae (Figs. 4 and 5).

Most of the Cymbidieae species are epiphytes, however almost all subtribes also have terrestrial species. Evolutionary transitions from terrestrial to epiphytic habit have played an important role in orchid diversification: gains of epiphytism habit are concomitant with increases

in diversification rates[10]. Those subtribes with greatest species richness (Oncidiinae = 1615, Maxillariinae = 819, Zygopetalinae = 437 and Catasetinae = 354) may be so partly due to the adoption of the epiphytic habit. This relationship could relate to movement into mountainous areas[2], and because of changes in the rate of uplift of the Andes[23]. Unlike other subtribes, most Eulophiinae species are terrestrial and widely distributed in the Old-World tropics of Africa, Asia and Australasia, with few taxa in the Neotropics. However, the Madagascan genera *Cymbidiella*, *Eulophiella*, *Grammangis* and *Paralophia* are all epiphytes[29]. Nevertheless, in Eulophiinae, more species-rich genera are terrestrial (*Eulophia*: 200 species and *Oeceoclades*: 38 species).

3.4 Evolutionary relationships amongst Orchidoideae

Here we present, for first time, a well-supported phylogeny for the backbone of Orchidoideae. The phylogeny obtained using complete plastomes yielded a strongly supported topology: Diurideae + Cranichideae and Orchideae as the outermost group, lacking a representative of Codonorchideae. Our approach using 60 coding regions, supports findings of Pridgeon et al., (2001), in which Diurideae and Cranichideae are sister groups, as well as Codonorchideae and Orchideae. Our findings differ from Givnish et al. (2015) and Salazar et al. (2003), in which Diurideae + Cranichideae form a clade – as here – but this clade is a sister group to Codonorchideae, with Orchideae placed as sister to the rest of Orchidoideae (Fig. 6). Givnish et al. (2015) included four (out of four) tribes and six of 21 subtribes for Orchidoideae, but the relationship between Diurideae and Cranichideae was still poorly supported (BS=34) with respect to Codonorchideae.

All Orchidoideae members have terrestrial habits and a cosmopolitan distribution. The most species-rich subtribe is Orchidinae (Orchideae) with 1,811 species. Records on pollination have shown that *Dactylorhiza* is pollinated by dipterans and beetles, which are attracted by scent[47]. At the same time, *Habenaria* is pollinated by moths[48]. Inflorescences within Orchidoideae are commonly terminal and racemose, but in the case of the monotypic tribe Codonorchideae (one genus = *Codonorchis*), those characters are not present. In fact, *Codonorchis* presents a single flower. This genus is only present in the south of the Andes and Paraná state. Rhizanthellinae and Thelymitrinae are grouped together within the Diurideae tribe. They share a geographical distribution, being common in Southeast Asia, Japan, New Zealand and Australia. The monotypic group Rhizanthellinae has a very particular inflorescence. It seems to be a solitary inflorescence but when it blooms under the leaf litter (which is also a unique character), tiny and densely grouped flowers can be observed. The inflorescences in Thelymitrinae are quite different from the rest of the subtribes within Orchidoideae; in this case, the size of the flowers is considerably bigger

(1 to 6 cm, compared to 1 cm or less in other subtribes).

In our analysis, Diurideae and Cranichideae are strongly supported as sister to one another (LBS=94), which was also recovered by Givnish et al. (2015). A synapomorphy shared by Diurideae and Cranichideae is the presence of binary/bilobed xylem in leaf midrib. The absence of tubers is only common in Cranichideae. Although these synapomorphies were identified against molecular phylogenies, authors have emphasized inadequate interpretations of the characters due to the discrepancies generated between the well-supported phylogenetic relationships and current classifications based on morphological characters[28]. Our results differ from those obtained in previous studies[2,28] in the categorization of Codonorchideae, where this tribe appeared as the sister group of Diurideae + Cranichideae. We recovered a strong sister relationship between Codonorchideae and Orchideae (LBS=86), although this could be due to branch effects by limited taxon sampling in Codonorchideae (consists of only two species of *Codonorchis*). Nevertheless, our results are in agreement with the phylogeny reported by Pridgeon et al., (2001), which used the *rbcL* gene and maximum parsimony to infer the Orchidoideae topology.

Conclusions

This study presents a well-resolved and better-supported phylogeny for the Orchidaceae family than any produced thus far by plastid DNA analyses. Here we report the complete plastid genome sequences of 11 orchid species: *G. pleiochroma, M. nasuta, M. sanderiana, O. globuliferum, T. glicensteinii, S. antenniferum, T. aliana, S. decora, S. mandonii, S. mucronata* and *G. repens*. These 11 plastomes differ in the IR boundaries and the loss/retention of *ndh* genes. For deep branches within the Cymbidieae subtribe, statistical support was improved. Similarly, our analyses provide the first well-supported phylogeny for Orchidoideae. Comparison of two approaches to infer phylogenies from plastome data showed different topologies most likely due to differences in taxon sampling. Although sampling was sufficient to resolve the relationships between the major clades in the family, sampling of several key genera (*Zygopetalum*, *Catasetum* and *Cyrtopodium*) and representatives for Eriopsidiinae and Coleopsidinae subtribes, would further enhance future work on orchid plastome phylogenetics.

Material and methods

Sampling, DNA extraction and sequencing

Eleven species representing Cymbidieae (subtribes Stanhopeinae, Maxillariinae and Oncidiinae), Epidendreae (Pleurothallidinae), Sobralieae (Sobraliinae), and Cranichideae (Goodyerinae) were sampled (Table 1). Fresh leaves were stored in silica gel for subsequent DNA extraction using a CTAB method[49]. Total DNA was purified with silica columns and then eluted in Tris-EDTA[50]. DNA samples were adjusted to 50 ng/uL to be sheared to fragments of approximately 500 bp. The library preparation, barcoding and sequencing on an Illumina HiSeqX were conducted at Rapid Genomics LLC (Gainesville, FL, USA). Pair end reads of 150 bp were obtained for fragments with insert size of 300-600 bp.

High-throughput sequencing

Rapid Genomics LLC first determined the concentration of DNA using a Qubit 3.0 (Life Technologies® Carlsbad, California, EE.UU.) and evaluated the integrity of the DNA using agarose gel electrophoresis. Purified genomic DNA (ratio OD260/280 between 1.8 to 2.0) was fragmented into smaller fragments of less than 800 bp using a Bioruptor 200 (Cosmo Bio Co. Ltd, Tokyo, Japan). Fragment size was checked by electrophoresis; qualified products were purified with a DNA purification kit (QIAGEN). A paired-end (PE) library with 150 bp insert size was constructed for each sample and sequencing was conducted on the Illumina HiSeq 4000 platform at Rapid Genomics LLC.

Overhangs were blunt ended using T4 DNA polymerase, Klenow fragment and T4 polynucleotide kinase. Subsequently, a base 'A' was added to the 3 'end of the phosphorylated blunt DNA fragments, and final products were purified. DNA fragments were ligated to adapters, which have the overhang of the base 'T'. Ligation products were gel-purified by electrophoresis to remove all unbound adapters or split adapters that were ligated together. Ligation products were then selectively enriched and amplified by PCR. For each sample, more than 10 million paired-end reads of 90 bp were generated.

Plastid genome assembly

Different bioinformatic tools were assessed for each of the steps of data processing in order to get the most efficient ones. Here we present the softwares that yielded better results when processing the data.

Sequence pre-processing

Raw sequences obtained by genome skimming were quality filtered using Trimmomatic[51] in order to eliminate sequencing artefacts, improve uniformity in the read length (>40 bp) and ensure quality (>20) for further analysis. Filtered sequences were processed with BBNorm[52] to normalize coverage by down-sampling reads over high-depth areas of the genomes (maximum depth coverage 900x and minimum depth 6x). This step creates a flat coverage distribution in order to improve read assembly. Subsequently, overlapping reads were merged into single reads using BBmerge[53] in order to accelerate the assembly process. Overlapping of paired reads was evaluated with Flash[54] to reduce redundancy. Merged reads were used to carry out the whole genome de novo assembly with SPAdes (Hash length 33,55,77)[55].

Plastome assembly

Assembler MIRA 4[56] was used to obtain whole plastid genomes. This program can map data against a consensus sequence of a reference assembly (simple mapping). MIRA has been useful for assembling complicated genomes with many repetitive sequences[57–59]. Additionally, the program improves assemblies with an iterative extension of the reads or contigs based on additional information obtained by overlap of paired reads or by automatic corrections. MIRA reduces the number of reads in the Illumina mapping without sacrificing coverage information. The program tracks coverage with respect to each base in the reference and creates a sequence of synthetic length, with the Coverage of Equivalent Reads (CER). Reads that do not map at a 100% remain as independent entities.

Consensus sequences were generated using SAMTOOLS[60], which provides a summary of coverage of reads mapped to a reference sequence. In theory, it can call variants by mapping reads to an appropriate reference. For each of the 11 plastomes, phylogenetically closed plastomes (available in the NCBI) were used as reference (*Masdevallia picturata, Masdevallia coccinea, Cattleya crispata, Goodyera fumata, Oncidium sphacelatum, Sobralia callosa*).

Plastome annotations

A search for other orchid plastomes was carried out through NCBI. Ninety-five plastomes from orchids and three from external groups (*Iris sanguinea, Agapanthus coddii* and *Asparagus officinalis*) were recovered. One hundred and six plastomes obtained (11 new plastomes and 95 from the NCBI) were annotated through the Chlorobox portal of the Max Planck Institute[61]. Sequences were uploaded as fasta files and running parameters were established as follow: BLAST

protein search identity=65%, BLAST rRNA, tRNA, DNA search identity=85%, genetic code = Bacterial/Plant plastid, max intron length=3,000, options= allow overlaps. The species *Oncidium sphacelatum* was set as the 'Server Reference' and *Masdevallia coccinea* was set as the 'Custom Reference' for CDS and tRNA, rRNA, primer, other DNA or RNA specifications.

Phylogenetic analysis

Whole plastome phylogenies

From the 106 plastids obtained, 97 (11 new plastomes and 86 from the NCBI) were used as phylogenetic markers. These were aligned to find the best hypothesis of homology[62] using MAFFT 7[63]. This step was performed at the supercomputing center APOLO, EAFIT University, Medellín, Colombia. Phylogenetic reconstruction based on Maximum Likelihood (ML) was implemented in RAxML v. 8.X[64], using 1,000 bootstrap replicates and the GTR+GAMMA model. Bayesian analysis was conducted in PhyloBayes MPI v. 1.5a (Lartillot, Lepage, & Blanquart, 2009) on the CIPRES server (Miller, Pfeiffer, & Schwartz, 2010), using the CAT model for site-specific equilibria and exchange rates defined by a Poisson distribution with 8 rate categories. Two independent chains were run until convergence was achieved (maxdiff <0.1 between chains), for 37,375 cycles each. Majority-rule consensus tree was built from the posterior of 7,419 trees post-burnin (1000), with trees sampled every 10 cycles.

Coding regions phylogenies

A set of 60 plastid genes was used to reconstruct phylogenetic relationships within Orchidaceae. One output of Chlorobox includes an alignment of each gene across all species. Additional alignments were made to include genes used by Givnish et al. (2015), thus obtaining alignments of up to 127 species (Table S2). Each gene was manually checked for start and stop codons.

We conducted an additional ML analysis using 60 coding regions of species belonging to Cymbidieae and six outgroups across external Orchidaceae tribes (*Apostasia wallichii, Habenaria radiata, Neottia ovata, Sobralia callosa, Sobralia mucronata* and *Vanilla planifolia*) using RAxML 8[64] program. In this analysis the taxon sampling for the coding region phylogeny was expanded in Cymbidieae (21 species) with respect to the whole plastid genome phylogeny (17 species). The same analysis was conducted for Orchidoideae subfamily, using 12 species of this group but removing *Chloreae gavilu* and *Dactylorhiza fuchsii* sequences from Givnish et al., (2015), due to their poor quality. Here we used four outgroups (*Cypripedium japonicum, Paphiopedilum niveum, Vanilla planifolia* and *Apostasia wallichii*). Concatenation of these 60

genes for both, the Cymbidieae tribe and Orchidoideae subfamily, was made using Geneious 9. Concatenated protein-coding sequences for all taxa were aligned using MAFFT[63] and polished.

Acknowledgments

We would like to thank Esteban Urrea for helping with bioinformatics pipelines. We thank Norris Williams and Mark Whitten from University of Florida for collecting and preparing the specimens. Kurt Neubig from Southern Illinois University provided the sequences of the 11 new samples. We also thank Janice Valencia for critical feedback on the paper, Juan David Pineda Cardenas for advising about computational resources used through EAFIT and Juan Carlos Correa for computational advices at BIOS. Finally, we would like to thank IDEA WILD for supporting with photographic equipment and Sociedad Colombiana de Orquideología for supporting M. A. Serna-Sánchez with a grant to conduct her undergraduate studies.

References

1. Chase, M. W. *et al.* An updated classification of Orchidaceae: Updated Classification of Orchidaceae. *Botanical Journal of the Linnean Society* **177**, 151–174 (2015).
2. Givnish, T. J. *et al.* Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proceedings of the Royal Society B: Biological Sciences* **282**, 20151553 (2015).
3. *darwin, c. r. 1862. on the various contrivances by which british and foreign orchids are fertilised by insects, and on the good effects of intercrossing. london: john murray. 1st ed., 1st issue.* (1862).
4. Fay, M. F. & Chase, M. W. Orchid biology: from Linnaeus via Darwin to the 21st century. *Annals of Botany* **104**, 359–364 (2009).
5. Ramírez, S. R. *et al.* Asynchronous diversification in a specialized plant-pollinator mutualism. *Science* **333**, 1742–1746 (2011).
6. Borba, E. L., Barbosa, A. R., Melo, M. C. de, Gontijo, S. L. & Oliveira, H. O. de. Mating systems in the Pleurothallidinae (Orchidaceae): evolutionary and systematic implications. *1* (2011). doi:10.15517/lank.v11i3.18275
7. Pérez-Escobar, O. A. *et al.* Multiple Geographical Origins of Environmental Sex Determination enhanced the diversification of Darwin's Favourite Orchids. *Scientific Reports* **7**, 12878 (2017).
8. Bateman, R. & Rudall, P. Evolutionary and Morphometric Implications of Morphological Variation Among Flowers Within an Inflorescence: A Case-Study Using European Orchids. *Annals of botany* **98**, 975–93 (2006).
9. Dong, W.-L. *et al.* Molecular Evolution of Chloroplast Genomes of Orchid Species: Insights into Phylogenetic Relationship and Adaptive Evolution. *International Journal of Molecular Sciences* **19**, 716 (2018).
10. Freudenstein, J. V. & Chase, M. W. Phylogenetic relationships in Epidendroideae (Orchidaceae), one of the great flowering plant radiations: progressive specialization and diversification. *Annals of Botany* **115**, 665–681 (2015).
11. Luo, J. *et al.* Comparative Chloroplast Genomes of Photosynthetic Orchids: Insights into Evolution of the Orchidaceae and Development of Molecular Markers for Phylogenetic Applications. *PLoS ONE* **9**, e99016 (2014).
12. Niu, Z. *et al.* The Complete Plastome Sequences of Four Orchid Species: Insights into the Evolution of the Orchidaceae and the Utility of Plastomic Mutational Hotspots. *Frontiers in Plant Science* **8**, (2017).
13. Zhang, G.-Q. *et al.* The Apostasia genome and the evolution of orchids. *Nature* **549**, 379–383 (2017).
14. Zhang, G.-Q. *et al.* The Dendrobium catenatum Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. *Sci Rep* **6**, 19029 (2016).
15. Yan, L. *et al.* The Genome of Dendrobium officinale Illuminates the Biology of the Important Traditional Chinese Orchid Herb. *Mol Plant* **8**, 922–934 (2015).
16. Yuan, Y. *et al.* The Gastrodia elata genome provides insights into plant adaptation to heterotrophy. *Nature Communications* **9**, (2018).
17. Huang, J.-Z. *et al.* The genome and transcriptome of *Phalaenopsis* yield insights into floral organ development and flowering regulation. *PeerJ* **4**, e2017 (2016).
18. Chao, Y.-T. *et al.* Chromosome-level assembly, genetic and physical mapping of Phalaenopsis aphrodite genome provides new insights into species adaptation and resources for orchid breeding. *Plant Biotechnol. J.* **16**, 2027–2041 (2018).
19. Hu, Y. *et al.* Genomics-based diversity analysis of Vanilla species using a Vanilla planifolia draft genome and Genotyping-By-Sequencing. *Sci Rep* **9**, 3416 (2019).
20. Jin, W.-T. *et al.* Phylogenetics of subtribe Orchidinae s.l. (Orchidaceae; Orchidoideae) based on seven

markers (plastid matK, psaB, rbcL, trnL-F, trnH-psba, and nuclear nrITS, Xdh): implications for generic delimitation. *BMC Plant Biology* **17**, (2017).
21. Li, M.-H., Zhang, G.-Q., Liu, Z.-J. & Lan, S.-R. Subtribal relationships in Cymbidieae (Epidendroideae, Orchidaceae) reveal a new subtribe, Dipodiinae, based on plastid and nuclear coding DNA. *Phytotaxa* **246**, 37 (2016).
22. Li, Y.-X. *et al.* Phylogenomics of Orchidaceae based on plastid and mitochondrial genomes. *Molecular Phylogenetics and Evolution* **139**, 106540 (2019).
23. Pérez-Escobar, O. A. *et al.* Recent origin and rapid speciation of Neotropical orchids in the world's richest plant biodiversity hotspot. *New Phytologist* **215**, 891–905 (2017).
24. Whitten, W. M., Neubig, K. M. & Williams, N. H. Generic and Subtribal relationShipS in neotropical cymbidieae (orchidaceae) baSed on matK/ycf1 plaStid data. *Lankesteriana* **13**, (2014).
25. Pridgeon, A. *Genera Orchidacearum Vol. 5, Vol. 5,*. (Oxford University Press, 2009).
26. Górniak, M., Paun, O. & Chase, M. W. Phylogenetic relationships within Orchidaceae based on a low-copy nuclear coding gene, Xdh: Congruence with organellar and nuclear ribosomal DNA results. *Mol. Phylogenet. Evol.* **56**, 784–795 (2010).
27. Pridgeon, A. M., Cribb, P. J. & Chase, M. W. *Genera Orchidacearum: Volume 2. Orchidoideae*. (OUP Oxford, 2001).
28. Salazar, G. A., Chase, M. W., Soto Arenas, M. A. & Ingrouille, M. Phylogenetics of Cranichideae with emphasis on Spiranthinae (Orchidaceae, Orchidoideae): evidence from plastid and nuclear DNA sequences. *American Journal of Botany* **90**, 777–795 (2003).
29. Bone, R. E., Cribb, P. J. & Buerki, S. Phylogenetics of Eulophiinae (Orchidaceae: Epidendroideae): evolutionary patterns and implications for generic delimitation: Evolutionary patterns in Eulophiinae. *Botanical Journal of the Linnean Society* **179**, 43–56 (2015).
30. Pérez-Escobar, O. A. *et al.* Andean Mountain Building Did not Preclude Dispersal of Lowland Epiphytic Orchids in the Neotropics. *Scientific Reports* **7**, 4919 (2017).
31. Salazar, G. *et al.* Phylogenetic systematics of subtribe Spiranthinae (Orchidaceae, Orchidoideae, Cranichideae) based on nuclear and plastid DNA sequences of a nearly complete generic sample. *Botanical Journal of the Linnean Society* **In press**, (2018).
32. Martins, A. *et al.* From tree tops to the ground: Reversals to terrestrial habit in Galeandra orchids (Epidendroideae: Catasetinae). *Molecular Phylogenetics and Evolution* **127**, (2018).
33. Nunes, C. *et al.* More than euglossines: the diverse pollinators and floral scents of Zygopetalinae orchids. *The Science of Nature* **104**, (2017).
34. Pansarin, L., Pansarin, E., Gerlach, G. & Sazima, M. The Natural History of Cirrhaea and the Pollination System of Stanhopeinae (Orchidaceae). *International Journal of Plant Sciences* 000–000 (2018). doi:10.1086/697997
35. Cisternas, M. A. *et al.* Phylogenetic analysis of Chloraeinae (Orchidaceae) based on plastid and nuclear DNA sequences. *Botanical Journal of the Linnean Society* (2012). doi:10.1111/j.1095-8339.2011.01200.x
36. Ramirez, S. R., Roubik, D. W., Skov, C. & Pierce, N. E. Phylogeny, diversification patterns and historical biogeography of euglossine orchid bees (Hymenoptera: Apidae). *Biological Journal of the Linnean Society* **100**, 552–572 (2010).
37. Cingel, N. A. V. D. *An Atlas of Orchid Pollination: European Orchids*. (CRC Press, 2001).
38. Weitemier, K. *et al.* Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* **2**, 1400042 (2014).
39. Givnish, T. J. *et al.* Assembling the Tree of the Monocotyledons: Plastome Sequence Phylogeny and Evolution of Poales [1]. *Annals of the Missouri Botanical Garden* **97**, 584–616 (2010).
40. Chris Blazier, J., Guisinger, M. M. & Jansen, R. K. Recent loss of plastid-encoded ndh genes within

Erodium (Geraniaceae). *Plant Molecular Biology* **76**, 263–272 (2011).
41. Lin, C.-S. *et al.* The location and translocation of ndh genes of chloroplast origin in the Orchidaceae family. *Scientific Reports* **5**, (2015).
42. Wu, F.-H. *et al.* Complete chloroplast genome of Oncidium Gower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiinae. 12 (2010).
43. Yang, J.-B., Tang, M., Li, H.-T., Zhang, Z.-R. & Li, D.-Z. Complete chloroplast genome of the genus Cymbidium: lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evolutionary Biology* **13**, 84 (2013).
44. Martín, M. & Sabater, B. Plastid ndh genes in plant evolution. *Plant Physiology and Biochemistry* **48**, 636–645 (2010).
45. Chang, C.-C. *et al.* The Chloroplast Genome of Phalaenopsis aphrodite (Orchidaceae): Comparative Analysis of Evolutionary Rate with that of Grasses and Its Phylogenetic Implications. *Molecular Biology and Evolution* **23**, 279–291 (2006).
46. Logacheva, M. D., Schelkunov, M. I. & Penin, A. A. Sequencing and Analysis of Plastid Genome in Mycoheterotrophic Orchid Neottia nidus-avis. *Genome Biol Evol* **3**, 1296–1303 (2011).
47. Gutowski, J. M. Pollination of the orchid Dactylorhiza fuchsii by longhorn beetles in primeval forests of Northeastern Poland. *Biological Conservation* **51**, 287–297 (1990).
48. Smith, G. R. & Snow, G. E. Pollination Ecology of Platanthera (Habenaria) Ciliaris and P. blephariglottis (Orchidaceae). *Botanical Gazette* **137**, 133–140 (1976).
49. Doyle, J. & Doyle, J. Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem Bull* **19**, 11–15 (1987).
50. Neubig, K. M. *et al.* Variables affecting DNA preservation in archival plant specimens. in *DNA Banking for the 21st Century: Proceedings of the US Workshop on DNA Banking* (eds. Applequist, W. & Campbell, L.) 81–112 (William L. Brown Center, Missouri Botanical Garden, 2014).
51. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
52. Bushnell. BBMap/BBTools. (2017). Available at: https://sourceforge.net/projects/bbmap/files/. (Accessed: 28th April 2019)
53. Bushnell, B., Rood, J. & Singer, E. BBMerge – Accurate paired shotgun read merging via overlap. *PLOS ONE* **12**, e0185056 (2017).
54. Magoc, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
55. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* **19**, 455–477 (2012).
56. Chevreux, B., Wetter, T. & Suhai, S. Genome sequence assembly using trace signals and additional sequence information. in *German conference on bioinformatics* **99**, 45–56 (Citeseer, 1999).
57. Cock, P. J. A., Grüning, B. A., Paszkiewicz, K. & Pritchard, L. Galaxy tools and workflows for sequence analysis with applications in molecular plant pathology. *PeerJ* **1**, e167 (2013).
58. Parakhia, M. V. *et al.* Draft Genome Sequence of the Endophytic Bacterium Enterobacter spp. MR1, Isolated from Drought Tolerant Plant (Butea monosperma). *Indian J Microbiol* **54**, 118–119 (2014).
59. Ward, J. A., Ponnala, L. & Weber, C. A. Strategies for transcriptome analysis in nonmodel plants. *American Journal of Botany* **99**, 267–276 (2012).
60. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
61. Tillich, M. *et al.* GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* **45**, W6–W11 (2017).
62. Chan, C. X. & Ragan, M. A. Next-generation phylogenomics. *Biology Direct* **8**, (2013).

63. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780 (2013).
64. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

Table 1. Comparison of major features of eleven orchid plastid genomes

Species	Accession number ****	Size (bp)	LSC* length (bp)	SSC** length (bp)	IR*** length (bp)	Number of different genes	Duplicated genes in IR	Protein-coding genes	tRNA genes	rRNA gene	GC content (%)
Gongora pleiochroma	XXXXXXXX	146,990	82,808	13,005	25,442	117	22	61	30	4	37.3
Maxillaria sanderiana	XXXXXXXX	132,712	74,195	8,638	24,807	123	24	62	33	4	38.6
Maxillaria nasuta	XXXXXXXX	144,213	81,128	12,357	25,251	121	22	64	31	4	37.7
Otoglossum globuliferum	XXXXXXXX	145,149	82,340	11,902	25,447	121	22	64	31	4	37.3
Telipogon glicensteinii	XXXXXXXX	143,414	80,462	11,785	25,559	113	22	57	30	4	37.0
Scaphosepalum antenniferum	XXXXXXXX	156,106	84,789	19,973	25,802	118	22	62	30	4	37.0
Teagueia aliana	XXXXXXXX	155,682	83,712	18,225	27,562	119	24	62	29	4	37.2
Sobralia decora	XXXXXXXX	160,230	87,540	20,449	26,282	120	24	61	31	4	37.3
Sobralia mandonii	XXXXXXXX	160,062	87,346	19,454	27,313	120	24	61	31	4	37.4
Sobralia mucronata	XXXXXXXX	161,827	88,602	19,845	27,311	122	24	64	30	4	37.1
Goodyera repens	XXXXXXXX	151,361	81,945	17,583	26,305	122	22	64	32	4	37.6

* Long Single Copy (LSC) section of the plastome

** Short Single Copy (SSC) section of the plastome

*** Inverted Repeats (IR) of the plastome

**** We are in the process of submitting the sequences to the GenBank.

Fig. 1. Whole plastome phylogeny for Orchidaceae based on ML analysis of sequence variation in 94 orchids under GTRGAMMA model and 3 Asparagales outgroups. Colored boxes correspond to new plastome sequences, the rest are plastid genomes found in NCBI. Bootstrap (1000 repetitions) support values are shown above each branch.

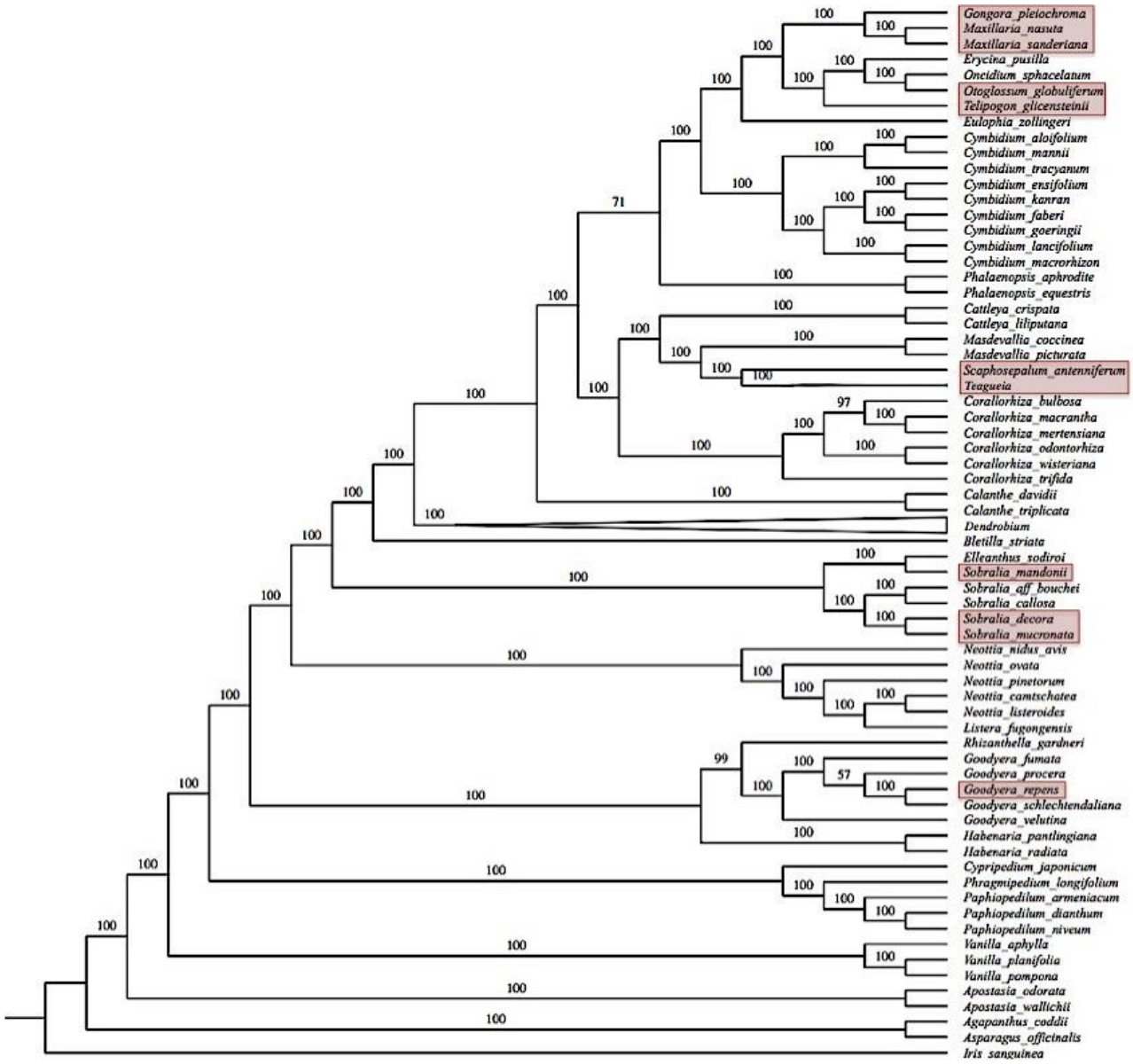


Fig. 2. Comparison between A) Givnish et al., 2015 phylogeny and B) best-scoring ML phylogeny presented here based on 60 coding regions with ML Bootstrap percentage above the branches. Terminals in *Eulophia*, *Cymbidium*, *Phalaenopsis*, *Cattleya*, *Masdevallia*, *Corallorhiza*, *Calanthe*, *Dendrobium*, *Bletilla*, *Sobralia*, *Neottia*, *Goodyera*, *Habenaria*, *Paphiopedilum*, *Vanilla* and *Apostasia* are collapsed. Colored boxes correspond to tribes, and bold words to subfamilies.

A) B)

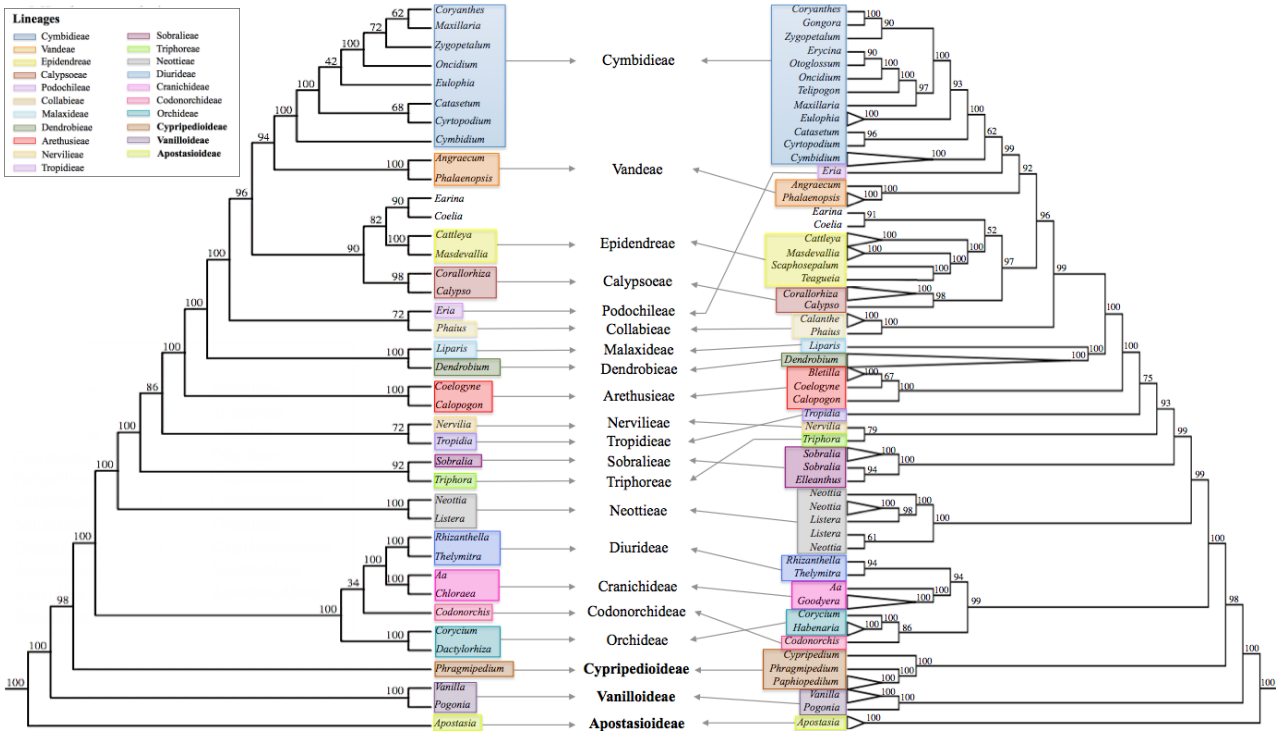


Fig. 3. Cymbidieae phylogeny based on ML analysis under GTRGAMMA model: 60 genes for 21 species and 6 outgroups. Bootstrap (1000 repetitions) support values are shown above each branch. The inset shows the phylogram of the Cymbidieae cladogram obtained here.

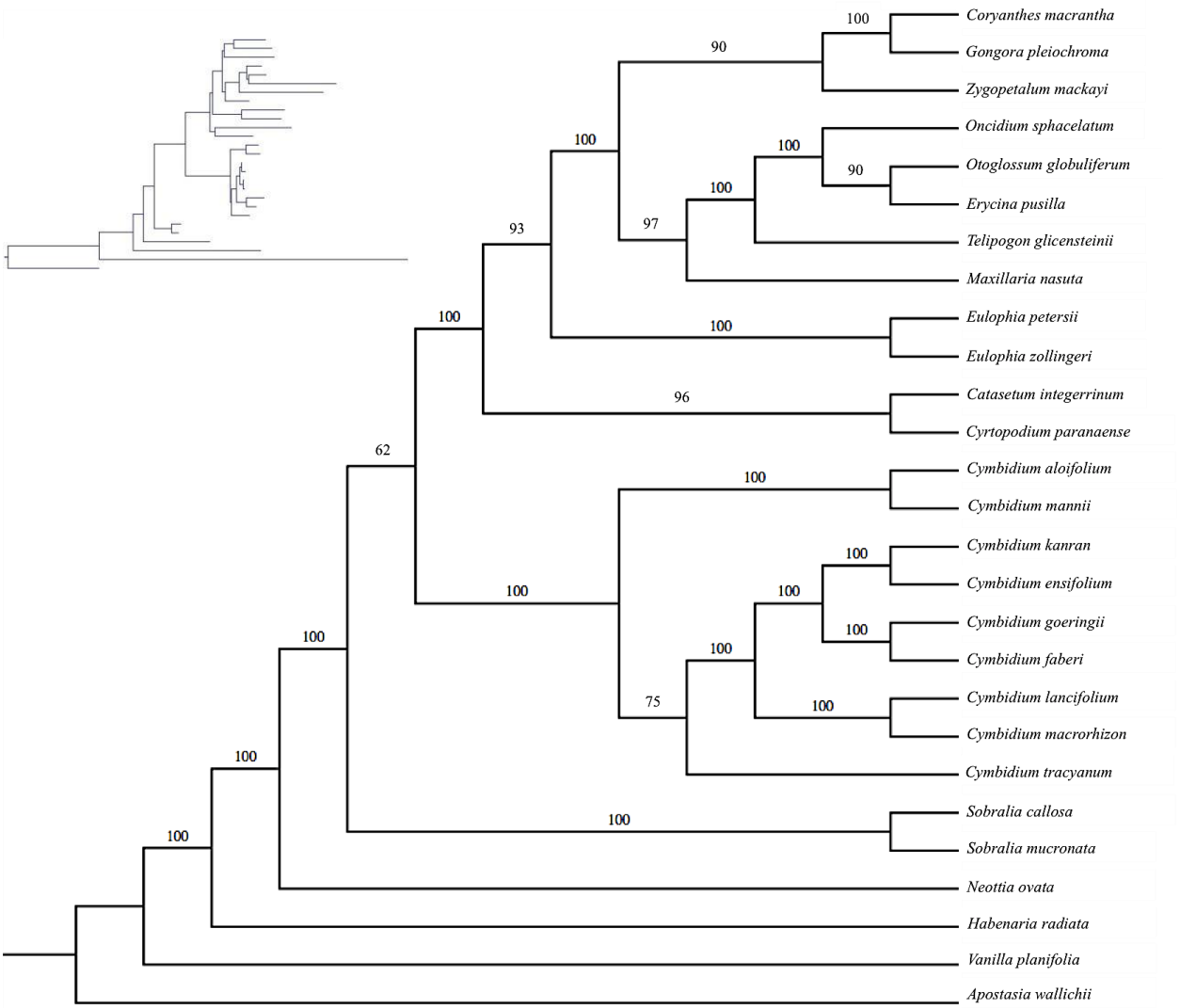


Fig. 4. Comparison between A) Cymbidieae phylogeny achieved by Givnish et al., 2015 and B) Zoom of Cymbidieae tribe from all Orchidaceae best-scoring ML phylogeny based on 60 genes. Colored boxes correspond to subtribes. Genera names in the photos from top to bottom: *Gongora*, *Zygopetalum*, *Maxillaria*, *Erycina*, *Eulophia*, *Catasetum*, *Cyrtopodium*, *Cymbidium*. Photos: LE. Mejía, M. Rincón and O. Pérez-Escobar.

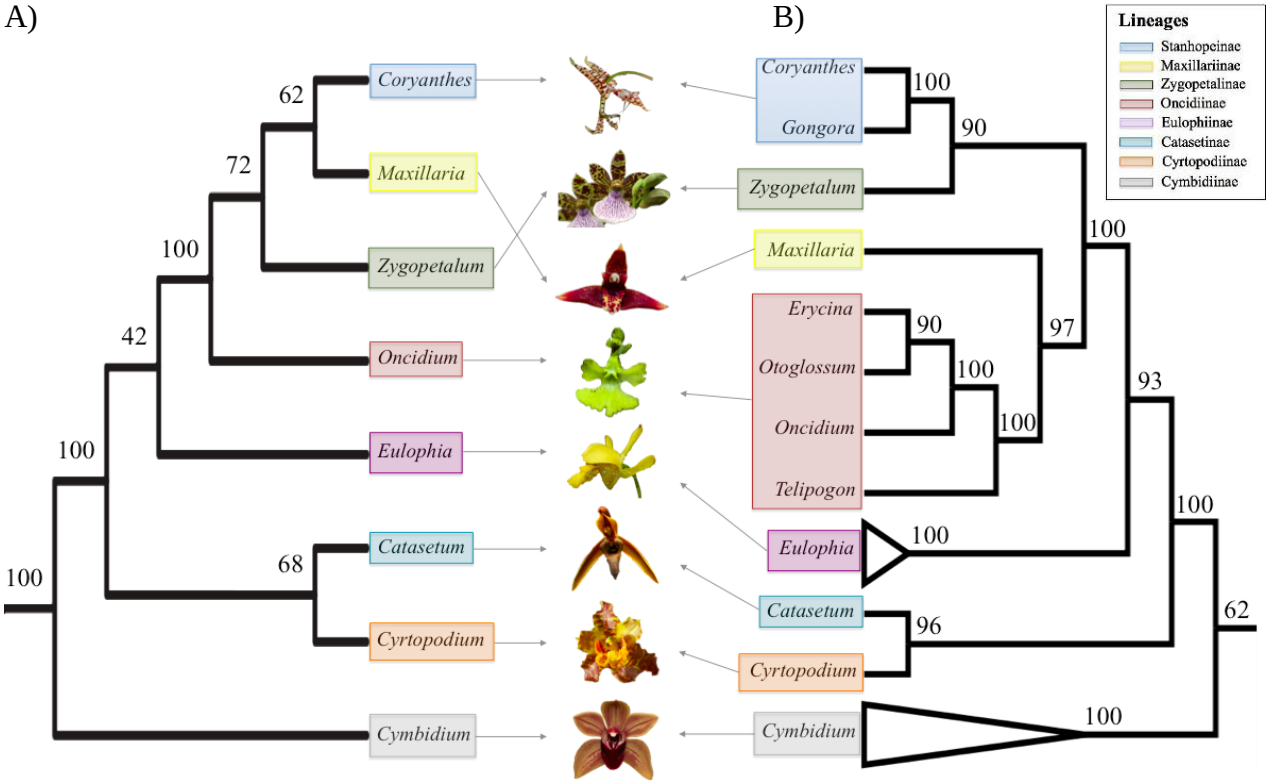


Fig. 5. Comparison between A) Whole plastome phylogeny and B) Zoom of Cymbidieae tribe from all Orchidaceae best-scoring ML phylogeny based on 60 genes. Colored boxes correspond to subtribes. Genera names in the photos from top to bottom: *Gongora*, *Zygopetalum*, *Maxillaria*, *Erycina*, *Eulophia*, *Catasetum*, *Cyrtopodium*, *Cymbidium*. Photos: LE. Mejía, M. Rincón and O. Pérez-Escobar.

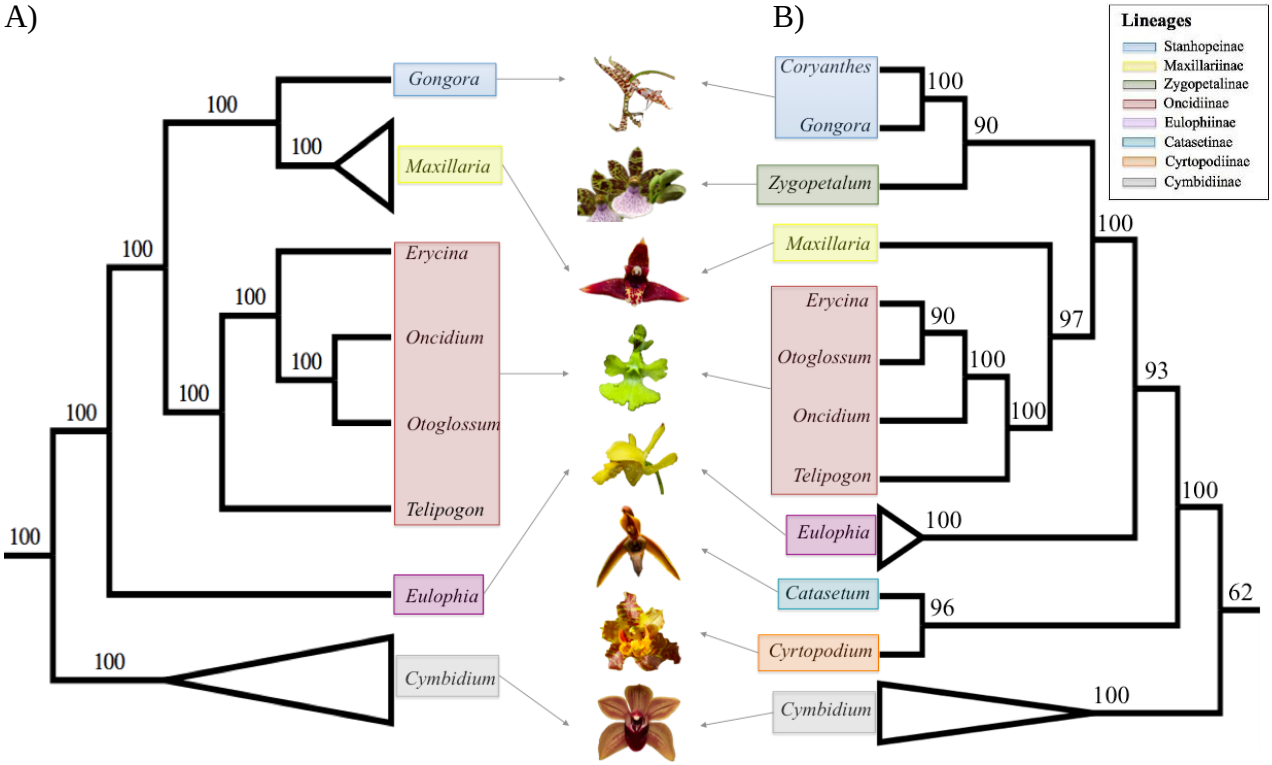
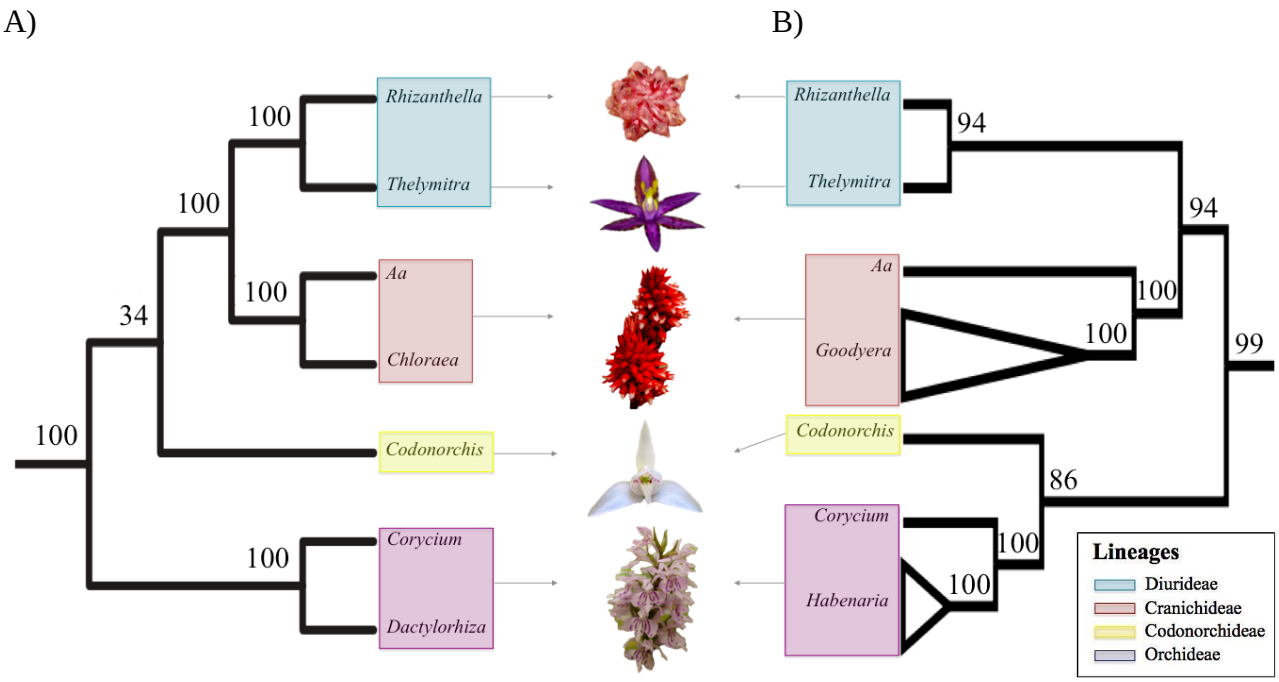
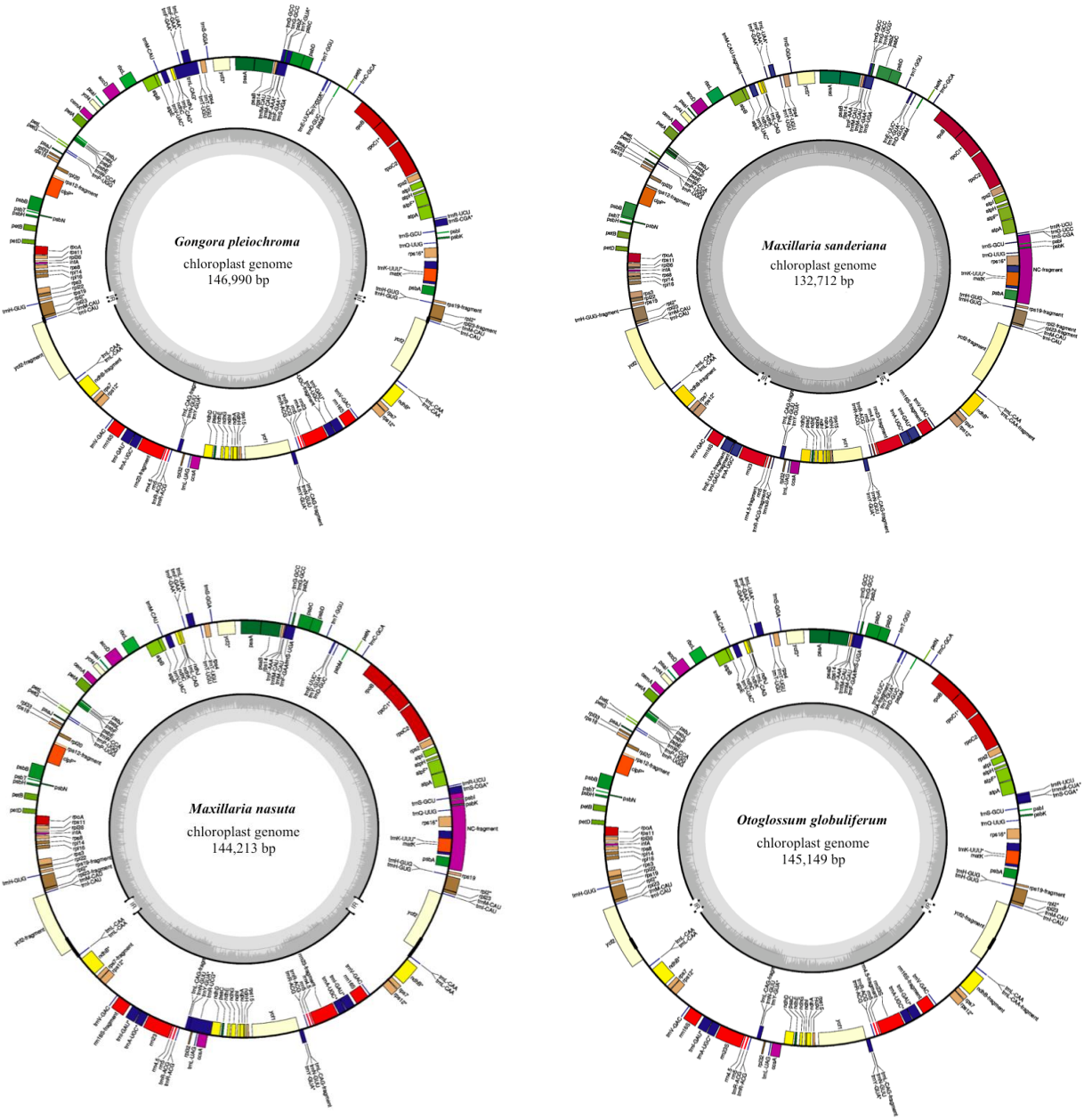


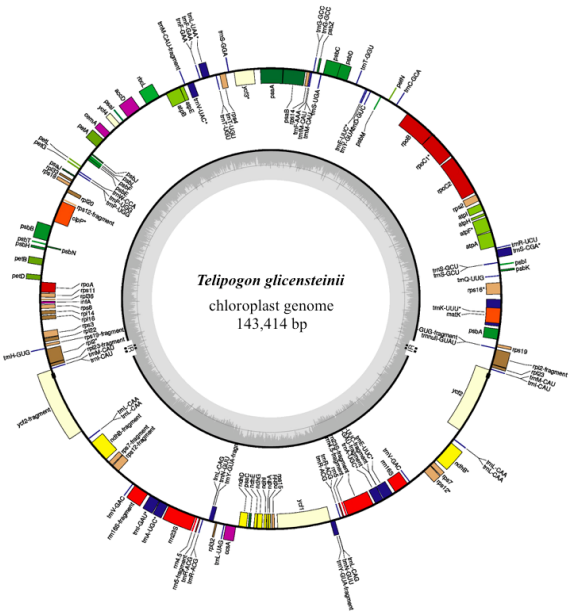
Fig. 6. Comparison between A) Orchidoideae phylogeny by Givnish et al., 2015 and B) zoom of best-scoring ML phylogeny based on 60 genes. Colored boxes correspond to tribes. Genera names in the photos from top to bottom: *Rhizanthella*, *Thelymitra*, *Stenorrhynchos*, *Codonorchis*, *Orchis*. Photos: M. Clements, C. Busby and O. Pérez-Escobar.

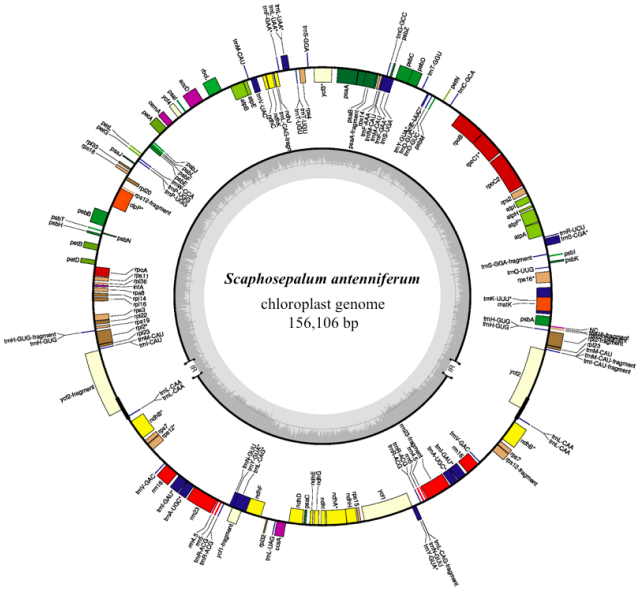


Supplementary materials

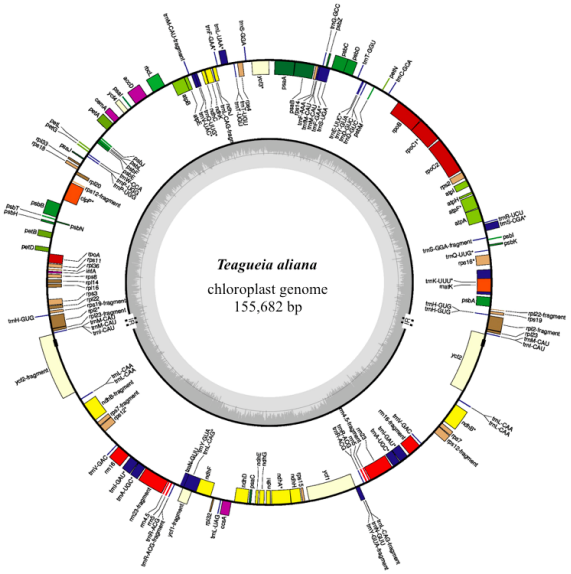
Fig. S1. Plastid genomes found in eleven orchids sequenced here. Genes shown inside the circle are transcribed clockwise, and those outside the circle are transcribed counter clockwise.

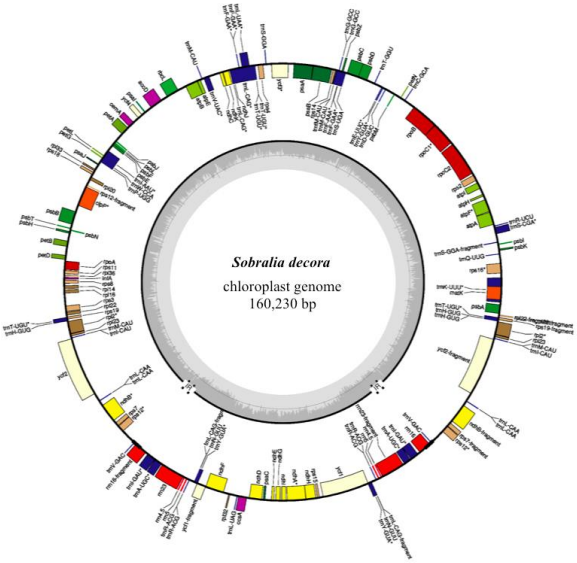



Telipogon glicensteinii
chloroplast genome
143,414 bp

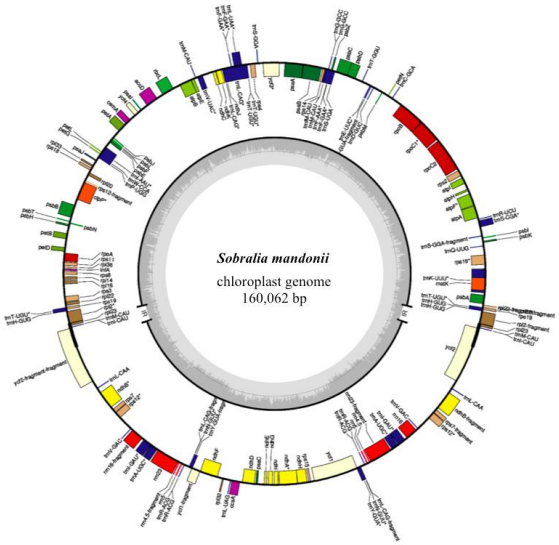

Scaphosepalum antenniferum
chloroplast genome
156,106 bp

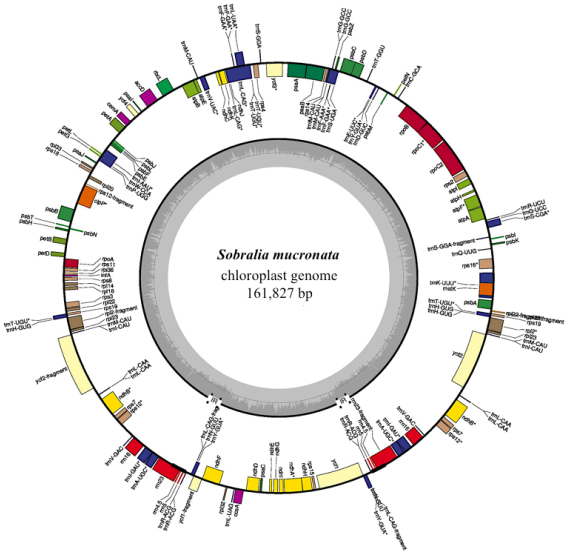
720

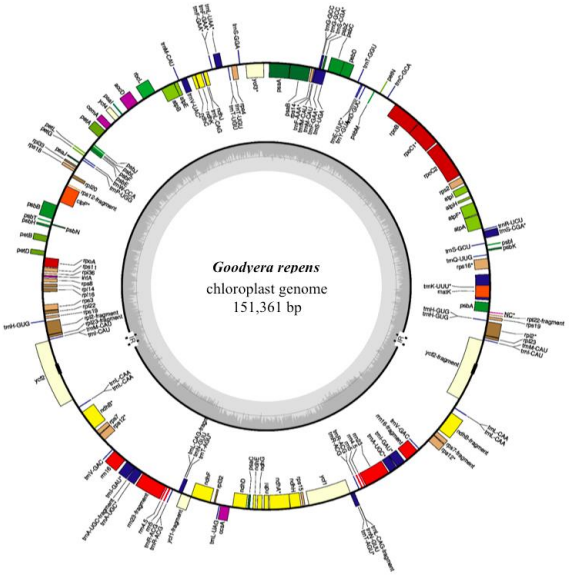

Teagueia aliana
chloroplast genome
155,682 bp


Sobralia decora
chloroplast genome
160,230 bp

721


Sobralia mandonii
chloroplast genome
160,062 bp


Sobralia mucronata
chloroplast genome
161,827 bp


Goodyera repens
chloroplast genome
151,361 bp

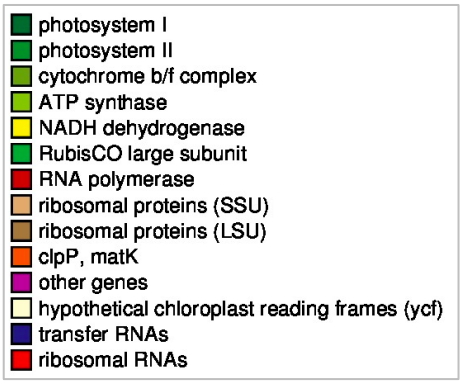

photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

Fig. S2. Coding regions phylogeny for Orchidaceae based on Bayesian analysis of sequence variation in 124 orchids and 3 Asparagales outgroups. PP values are shown above each branch. Terminals in *Eulophia*, *Cymbidium*, *Phalaenopsis*, *Masdevallia*, *Cattleya*, *Corallorhiza*, *Dendrobium*, *Bletilla*, *Sobralia*, *Neottia*, *Goodyera*, *Habenaria*, *Paphiopedilum*, *Vanilla* and *Apostasia* are collapsed.

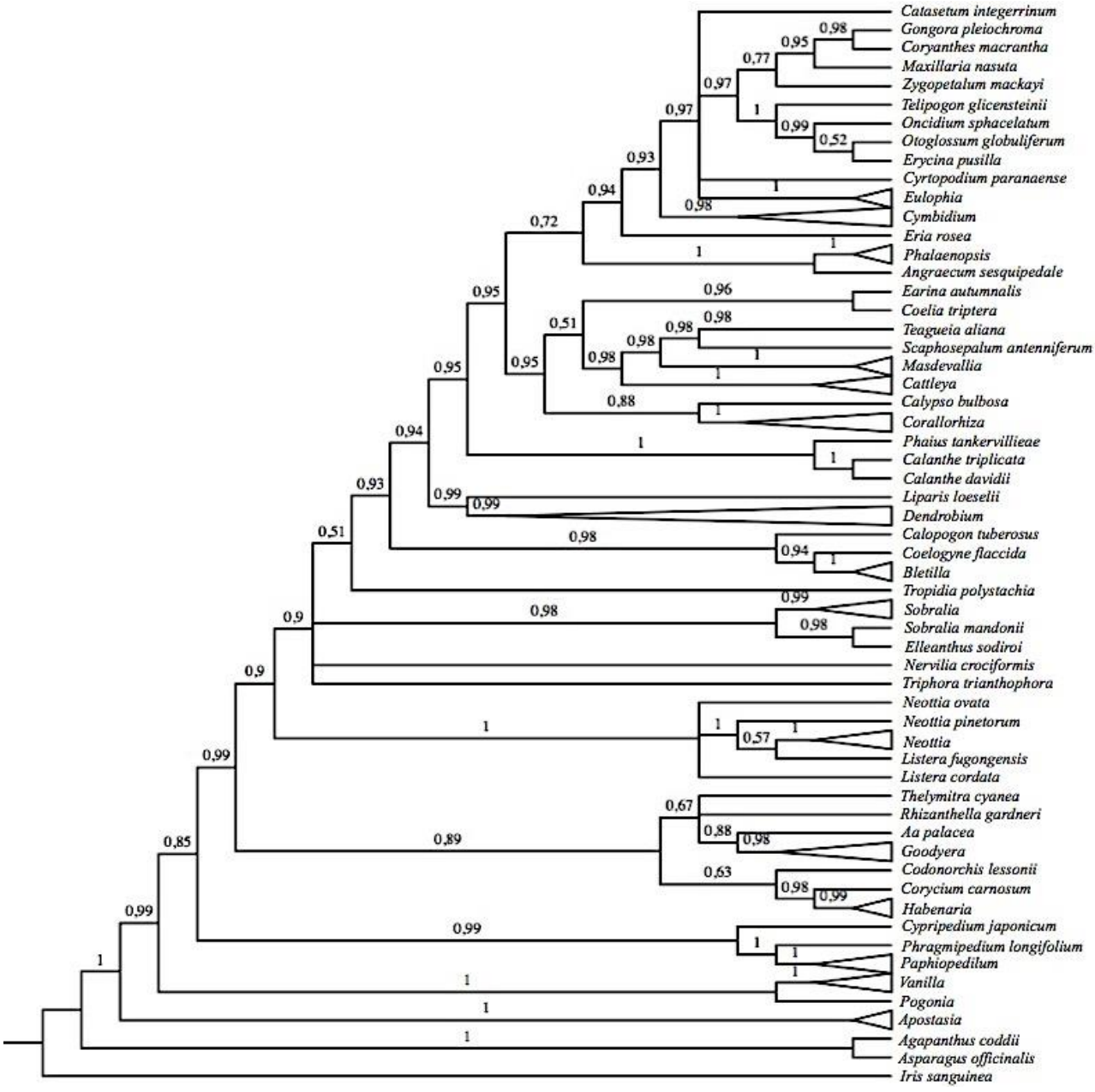


Table S1. List of eleven species included in this study and assembly data

Species	Raw reads (pairs)	Read pairs after pre-process	SPAdes (Whole genome)		MIRA (Plastome)		
			Contigs	Largest contig (bp)	Contig length (bp)	Reads	Average coverage
Gongora pleiochroma	4.955.260	4.362.598	2584	26.649	149.958	71.547	43
Maxillaria sanderiana	8.146.842	7.127.204	5134	13.572	148.730	22.004	13
Maxillaria nasuta	10.104.368	8.947.244	2907	20.135	149125	77.016	46
Otoglossum globuliferum	8.368.010	7.396.624	1993	34.030	149.411	101.195	61
Telipogon glicensteinii	5.488.188	4.671.888	3325	31.722	148.629	81.250	50
Scaphosepalum antenniferum	9.806.852	8.793.358	1740	56.374	158.607	510.750	292
Teagueia aliana	10.528.540	9.207.146	4820	40.822	160.875	303.586	168
Sobralia decora	10.404.770	9.538.426	2033	39.445	162.833	108.091	60
Sobralia mandonii	6.635.130	5.776.780	2316	27.089	165.531	132.225	92
Sobralia mucronata	7.105.254	6.388.396	2458	55.057	162.802	214.486	122
Goodyera repens	10.843.708	9.572.634	2235	23.053	157.822	346.014	197

Table S2. Comparison between the set of genes alignments (taxa per gene)

Gene	127 species*	97 species**
accD	54	54
atpA	111	85
atpB	114	87
atpE	115	90
atpF	113	86
atpH	116	88
atpI	117	90
ccsA	68	43
cemA	103	74
clpP	86	59
infA	117	90
matK	74	47
ndhA	48	35
ndhB	92	59
ndhC	59	41
ndhD	26	4
ndhE	88	70
ndhF	1	-
ndhG	50	29
ndhH	36	19

Gene	127 species*	97 species**
psbF	117	90
psbH	111	89
psbI	114	91
psbJ	117	92
psbK	114	88
psbL	117	90
psbM	106	83
psbN	110	90
psbT	109	91
psbZ	92	91
rbcL	105	77
rpl14	121	97
rpl16	29	-
rpl2	117	90
rpl20	83	62
rpl22	89	61
rpl23	115	87
rpl32	96	77
rpl33	116	89
rpl36	117	95

ndhI	17	8	rpoA	105	79
ndhJ	18	-	rpoB	111	87
ndhK	19	4	rpoC1	110	84
pbf1	91	90	rpoC2	66	67
petA	117	93	rps11	119	91
petB	27	-	rps12	28	-
petD	28	-	rps14	122	96
petG	117	91	rps15	109	88
petL	114	90	rps16	62	36
petN	113	91	rps18	110	87
psaA	109	85	rps19	118	92
psaB	114	88	rps2	103	74
psaC	115	91	rps3	112	85
psaI	109	87	rps4	117	94
psaJ	100	75	rps7	121	92
psbA	113	85	rps8	118	93
psbB	105	90	ycf1	10	7
psbC	115	90	ycf2	111	84
psbD	113	87	ycf3	109	83
psbE	119	92	ycf4	95	76

* Includes 86 whole plastomes from NCBI, 11 new plastids and 30 species sampled in Givnish et al, 2015.

** Includes 86 whole plastomes from NCBI and the 11 new whole plastomes.