To be submitted to: *Scientific Reports*

Running title: Plastid phylogenomics of the orchid family

Plastid phylogenomics resolves ambiguous relationships within the orchid family and provides a solid timeframe for biogeography and macroevolution

Maria Alejandra Serna-Sánchez[1,2+], Oscar A. Pérez-Escobar[3*+], Diego Bogarín[4,5+], María Fernanda Torres[6], Astrid Catalina Alvarez-Yela[7], Juliana E. Arcila[1], Climbie F. Hall[8], Fábio de Barros[8], Fábio Pinheiro[9], Steven Dodsworth[10], Mark W. Chase[3], Alexandre Antonelli[3,6,11], Tatiana Arias[1,7,12*]

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

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

[3] Royal Botanic Gardens, Kew, TW9 3AE, London, UK.

[4] Jardín Botánico Lankester, Universidad de Costa Rica, P. O. Box 302-7050, Cartago, Costa Rica.

[5] Naturalis Biodiversity Center, Endless Forms group, P.O. Box 9517, 2300 RA Leiden, The Netherlands; Herbario UCH, Universidad Autónoma de Chiriquí, David, Panamá.

[6] Gothenburg Global Biodiversity Centre, Department of Biological and Environmental Sciences, University of Gothenburg, 405 30 Gothenburg, Sweden.

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

[8] Instituto de Botânica, Núcleo de Pesquisa Orquídario do Estado, 68041, São Paulo.

[9] Universidade Estadual de Campinas, Instituto de Biologia, Departamento de Biologia Vegetal, 13083-862, Campinas-SP, Brazil

[10] School of Life Sciences, University of Bedfordshire, University Square, Luton, LU1 3JU, UK.

[11] Department of Plant Sciences, University of Oxford, South Parks Road, OX1 3RB Oxford, United Kingdom

[12] Tecnológico de Antioquia, Calle 78B NO. 72A - 220 Medellín- Colombia

[*] Corresponding Authors: T.A. (tatiana.arias48@tdea.edu.co) & O.A.P.E. (o.perez-escobar@kew.org)

[+] These authors contributed equally to the study

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

ABSTRACT [291 words]

Recent phylogenomic analyses based on the maternally inherited plastid organelle have enlightened evolutionary relationships between the subfamilies of Orchidaceae and most of the tribes. However, uncertainty remains within several subtribes and genera for which phylogenetic relationships have not ever been tested in a phylogenomic context. To address these knowledge-gaps, we here provide the most extensively sampled analysis of the orchid family to date, based on 78 plastid coding genes representing 264 species, 117 genera, 18 tribes and 28 subtribes. Divergence times are also provided as inferred from strict and relaxed molecular clocks and birth-death tree models. Our taxon sampling includes 51 newly sequenced plastid genomes produced by a genome skimming approach. We focus our sampling efforts on previously unplaced clades within tribes Cymbidieae and Epidendreae. Our results confirmed phylogenetic relationships in Orchidaceae as recovered in previous studies, most of which were recovered with maximum support (209 of the 262 tree nodes). We provide for the first time a clear phylogenetic placement for Codonorchideae within subfamily Orchidoideae, and Podochilieae and Collabieae within subfamily Epidendroideae. We also identify relationships that have been persistently problematic across multiple studies, regardless of the different details of sampling and genomic datasets used for phylogenetic reconstructions. Our study provides an expanded, robust temporal phylogenomic framework of the Orchidaceae that paves the way for biogeographical and macroevolutionary studies.

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

1. Introduction

Orchidaceae, with *ca*. 25,000 species and ~800 genera[1,2] are one of two of the most diverse and widely distributed flowering plant families on Earth and have captivated the attention of scientists for centuries[3]. The family has a striking morphological and ecological diversity and

evolved complicated interactions with fungi, animal and other plants[4,5] and a diverse array of sexual systems[6–8]. Numerous efforts have been made to understand the natural history, evolution and phylogenetic relationships within the family[2,7,9–13]. To date, there are seven nuclear genome sequences available, i.e., *Apostasia shenzhenica*[14], *Dendrobium catenatum*[15], *D. officinale*[16], *Gastrodia elata*[17], *Phalaenopsis equestris*[18], a *Phalaenopsis* hybrid cultivar[19], *P. aphrodite*[20], *Vanilla planifolia*[21], 221 complete plastid genomes and 2,678 sequence read archives for Orchidaceae in NCBI (accessed 22 August 2020) .

Phylogenomic approaches have been implemented to infer relationships between major orchids clades in deep and recent time[2,10,12,13,22,23], but extensive uncertainties remain regarding the phylogenetic placement of several subtribes. This knowledge-gap stems from the large gaps in both taxonomic and genomic sampling efforts that would be required to comprehensively cover all major orchid clades (subtribes/groups of genera). Givnish et al.[2] published the first well-supported analysis of Orchidaceae based on plastid phylogenomics. They performed a maximum likelihood (ML) analysis of 75 genes from the plastid genome of 39 orchid species, covering 22 subtribes, 18 tribes and five subfamilies. This robust but taxonomically under-sampled study agreed corroborated relationships of the subfamilies and tribes, observed in previous studies[10–13].

Previous orchid studies have failed to resolve relationships in rapidly diversifying clades[24–26] because of reduced taxon and data sampling[27]. This is particularly true for Cymbidieae and Pleurothallidinae, the two most species-rich groups in which generic relationships are largely the product of rapid diversification[28] that is difficult to resolve using only a few loci[25,29]. Cymbidieae comprise 10 subtribes, ~145 genera and nearly 3,800 species[1], 90% of which occur in the Neotropics [28]. Four of these subtribes are among the most species-rich in the Andean and Chocoan region (Maxillariinae, Oncidiinae, Stanhopeinae and Zygopetaliinae[30,31]). Pleurothallidinae include ~ 5,500 exclusively Neotropical species in 47 genera. Pleurothallid orchids are one of the most prominent components of the cloud forest flora in the northern and central Andes and Central America[32].

Another group in which phylogenetic relationships are unresolved is Orchidoideae [1,33]. This group comprises four mostly terrestrial tribes, 25 subtribes and over 3,600 species. The subfamily occurs on all continents except the Antarctic. Previous efforts to disentangle the phylogenetic relationships in the subfamily have mostly relied on a small set of nuclear and plastid markers[34], and more recently on extensive plastid coding sequence data[2].

The wide geographical range of these groups in the tropics and temperate regions and their

98 striking vegetative and reproductive morphological variability make them ideal model clades for
99 disentangling the contribution of abiotic and biotic drivers of orchid diversification across biomes.
100 Occurring from alpine ecosystems to grasslands, they have conquered virtually all ecosystems
101 available in any elevational gradient[35–37], showing independent transitions to terrestrial, rupicolous
102 and epiphytic habit. Moreover, they have evolved a diverse array of pollination systems[38–40],
103 rewarding species offering scent, oil and nectar, and even food- and sexual deceptive species[41,42].
104 However, the absence of a solid phylogenetic framework has precluded the study of how such
105 systems evolved and the diversification dynamics of Cymbidieae, Pleurothallidinae and
106 Orchidoideae more broadly.

107 Phylogenetic analyses are crucial to understanding the drivers of diversification in orchids,
108 including the mode and tempo of morphological evolution[30,43]. High-throughput sequencing and
109 modern comparative methods have enabled the production of massive molecular datasets to
110 reconstruct evolutionary histories and thus provide unrivalled knowledge on plant phylogenetics[44].
111 Here, we present the most densely sampled plastid analysis of Orchidaceae, including data from 51
112 newly sequenced plastid genomes,. We apply two general approaches: a) maximum likelihood
113 phylogenetic analysis conducted on 78 plastid coding regions to inform relationships; b) Bayesian
114 inference in combination with strict and relaxed molecular clocks and a birth-death model applied
115 to a subset of the plastid coding regions to produce a temporal framework of the orchid family. Our
116 study expands the current generic representation for the Orchidaceae and clarifies previously
117 unresolved phylogenetic relations within the Cymbidieae, Pleurothallidinae and Orchidoideae. The
118 results reported here provide a robust framework for the orchid family and new insights into
119 relationships at both deep and shallow phylogenetic levels.

2. Results

2.1 Phylogenetic relationships and divergence times in the orchid family

The ML tree derived from the 78 plastid genes is provided in Fig. 1. Two hundred-and-thirty-one nodes were recovered as strongly supported (i.e. likelihood bootstrap percentage [LBP] = 85-100), of which 209 attained maximum support. Only 26 nodes recovered LBPs between 25 and 84 (Fig. 1, inset). Unsupported relationships were restricted to Epidendroideae and Orchidoideae but were more frequent in Epidendroideae and often linked to low levels of sequence variation. Here, poorly supported relationships occurred mostly towards the backbone of the tribes Arethuseae, Cymbidieae, Epidendreae and Neottieae and Tropidieae + Nervilieae and the most recent common ancestor (MRCA) of Arethuseae, Malaxideae, Podochilieae, Collabieae, Epidendreae, Vandeae and Cymbidieae. Intrageneric relationships were robustly supported, with only two instances for which few nodes were recovered as poorly supported (*Dendrobium*: 3; *Cymbidium:* 1; Fig. S1).

Absolute times of divergence under strict and relaxed clocks for Orchidaceae, subfamilies and most tribes are provided in Table 1 (phylogenetic trees with mean ages and intervals of confidence produced under both clock models are provided on Figs S2, S3). Strict and relaxed molecular clocks revealed similar ages of divergence for the majority of the MRCAs of main orchid clades, although we found stark differences in the length of the 95% highest posterior density intervals (HPD) derived from both models are obvious, with the relaxed clock producing larger HPDs (Tab. 1; Fig. S2, S3; Tab. S1, S2). Under the strict and relaxed clocks, Orchidaceae diversified first during the late Cretaceous (88.1 my ± 3; 89.1 my ±9, respectively). The largest differences on the MRCA ages occurred in Epidendroideae (44 my ± 2 vs 60 my ± 10 under a strict and relaxed clock models, respectively) and Vanilloideae (80 my ± 4 vs 67 my ± 9). A complete account of mean and median ages, HPDs, branch lengths and rate values estimated for all nodes of chronograms estimated strict and relaxed molecular clock models are provided on Tab. S1, S2.

2.2. Phylogenetic informativeness of plastid genes

Phylogenetic informativeness plots are provided on Fig. S4 (see Tab. S3,S4 for a detailed account of PI per-site and net values for each assessed locus). Per-site and net phylogenetic informativeness (PI) analyses recovered both *ycf*1 as the most informative locus, which attained the highest values at a reference time (phylogenetic depth) of 0.51. On average, plastid loci attained their highest PI value at a reference time of 0.85 (SD=0.16). In contrast, the highest PI values of

152 the 10 most informative loci occurred at an average reference time of 0.63 (SD=0.11) and 0.80
153 (SD=0.17) for per-site and net PI calculations.
154
155 **3. Discussion**
156 *3.1 A robust temporal phylogenomic framework for the orchid family*
157 Previous phylogenomic studies of the orchid family included up to 74 species representing
158 18 tribes, 19 subtribes and 66 genera[27]. Our study sampled 264 species from all subfamilies,
159 representing 18 tribes (out of 22), 28 subtribes (out of 46) and 74 genera (~10% of the currently
160 recognised genera; Fig. 2). In general, our phylogenomic frameworks are in agreement with
161 previously published family-wide orchid analyses either inferred from dozens of markers[2,13] or
162 from a handful of loci[29]. Here, representativeness of Cymbidieae and Epidendreae, two of the most
163 prominent tropical Epidendroideae[45]clades, have increased from eight to 32 genera and six to 30,
164 respectively[2,27]. In particular, relationships inferred from extensive plastid data within
165 Zygopetaliinae (Cymbidieae) and Pleurothallidinae (Epidendreae) are presented for the first time.
166 Our 78-coding sequence plastid ML analysis led to similar results as reported by Givnish et al.[2],
167 Niu et al.[13] and Li et al.[27] but with an overall clear increase in support: 22% of nodes with LBS <
168 85 in Givnish et al.[2] and 21% in Li et al.[27] *vs* 11.5% in this study. This is particularly evident in
169 relationships inferred within Orchidoideae (see section 3.5 of *Discussion*) and Cymbidieae,
170 Epidendreae (see sections 3.3 and 3.4 of *Discussion,* respectively) and Collabieae. For the last, for
171 the first time we provide high support for the previously unresolved relationship of
172 Podochilieae+Collabieae[2,27].
173 The absolute age estimates derived from our strict and relaxed molecular clocks and five of
174 the most informative plastid loci are in line with previous nuclear-plastid multi-locus and
175 phylogenomic plastid-only chronograms[2,46,47]. Nonetheless, our ML tree also identifies intricate
176 relationships that have been consistently recovered as unsupported in several studies. These include
177 unsupported basal nodes in Epidendroideae representing Sobralieae, Nervilieae and Tiphoreae[27,48],
178 *Arundina*+remainder of Arethuseae[27], and the position of Eulophiinae in the Cymbidieae[25,28,49]
179 (Fig. 2). Uncertainty around the phylogenetic position of these clades might be due to limited taxon
180 sampling in this and previous studies. Alternatively, intragenomic conflict[50–52] and lack of
181 phylogenetic informativeness required to sort out relationships derived from rapid
182 diversifications[22,53,54] in plastid DNA sequences (regardless of whether whole plastid genome
183 datasets are employed[55]) might hamper the phylogenetic placement of clades with robust support.

184

185 *3.3 Improved support of phylogenetic relationships within Cymbidieae*

186 Multiple studies have inferred evolutionary relationships in Cymbidieae from
187 morphological and molecular characters[28,29]. Relationships among subtribes have recently been
188 estimated using the plastid genes *psaB, rbcL, matK* and *ycf1* combined with the low-copy nuclear
189 gene *Xdh*[25]. Here, Cymbidiinae was sister to the remainder of Cymbidieae. Poorly supported and
190 incongruent relationships were found among Catasetinae, Eulophiinae and Eriopsidinae, however,
191 when compared with the topologies obtained by Whitten et al.[29], Freudenstein & Chase[48] and
192 Pérez-Escobar et al.[7]

193 The most complete taxonomic sampling conducted to date under a plastid phylogenomic
194 framework[2] included 8 of 11 subtribes of Cymbidieae, but some inter-subtribal relationships were
195 unresolved: Stanhopeinae (20 genera), Maxillariinae (12 genera), Zygopetalinae (36 genera),
196 Oncidiinae (65 genera) and Eulophiinae (13 genera). A clade formed by Stanhopeinae and
197 Maxillariinae had poor support (LBP=62) and their relationship to Zygopetaliinae also had low
198 support (LBP=72). The relationship between Eulophiinae and a clade of Stanhopeinae,
199 Maxillariinae, Zygopetalinae and Oncidiinae also had poor support (LBP=42). One of the
200 outcomes of our expanded sampling (nine subtribes) is the improvement of support in Cymbidieae,
201 more specifically for nodes of some groups involved in rapid diversifications that historically have
202 been problematic to resolve[2,29]. In particular, Maxillariinae+Stanhopeinae and
203 Catasetinae+Cyrtopodiinae are now both strongly supported (LBP=100). In addition, our results
204 also support the placement of *Dipodium* (Dipodiinae) is supported as sister to the rest of
205 Cymbidieae, a relationship which was previously recovered from a few loci[25]. However, our
206 plastid phylogenomic framework is still incomplete due to absence of representatives of
207 Eriopsidiinae and Coleopsidinae.

208 One other novelty of our study is the inference of relationships in Zygopetalinae, a subtribe
209 in which relationships have previously been poorly understood[56]. The most extensively sampled
210 analysis of Zygopetalinae inferred from plastid markers (*mat*K-*ycf1*) [29] included 60 species and 27
211 genera, but relationships between most genera attained only low support. Our expanded molecular,
212 but taxonomically reduced, matrix (i.e. 20 genera and 21 species) produced greater support for the
213 backbone relationships in the subtribe, including the radiation of the *Huntleya* clade (*Dichaea*,
214 *Huntleya*, *Chaubardia* and the *Chondrorhyncha* complex[56,57]). Nonetheless, relationships between
215 the *Huntleya* grade (i.e. *Huntleya* clade + *Cryptarrhena*) and the remainder of Zygopetalinae still

216 remains unresolved.

217 Our phylogenetic analyses further place for the first time in the orchid tree of life
218 *Cheiradenia* and *Hoehneella* with moderate to strong support (Fig. 1, S1). *Cheiradenia* is a
219 monospecific genus restricted to the lowland wet forests of Venezuela and Guyana whereas
220 *Hoehneella* includes two species exclusively distributed in the Brazilian evergreen wet forests of
221 the Brazilian states of Espirito Santo and São Paulo[58]. Referring to the similarity of both vegetative
222 and floral reproductive characters, Pupulin [58] hypothesised that *Cheiradenia* should be closely
223 related to members of the *Zygopetalum* clade (e.g. *Koellensteina*, *Paradisanthus*), with *Hoehneella*
224 being related to the *Huntleya* clade (i.e. *Huntleya* and *Chaubardia*). Our ML tree supports both
225 assumptions, placing *Cheiradenia* as sister to *Paradisanthus* with maximum support and
226 *Hoehneella* as sister to *Chaubardia* in a moderately supported clade (83 LBP: Fig. 1, S1).
227 *Koellensteina kellneriana* (the taxonomic type of the genus) clustered with *Acacallis* and not with
228 *Otostylis* and *Paradisanthus*, and therefore we confirm that *Koellensteina* in the strict sense is
229 related to *Acacallis*. In addition, *Otostylis* is recovered as sister to *Warrea* and not to
230 *Paradisanthus* as previously suggested by Williams et al.[56] based on a weakly supported
231 placement. Our results also highlight the extensive and independent terrestrial and epiphytic habit
232 transitions occurring in this clade, as most sister genera shows different habit types.
233

234 *3.5. Novel and robust relationships in the most rapidly diversifying subtribe Pleurothallidinae*

235 One of the most spectacular Neotropical plant diversifications is perhaps that of the
236 Pleurothallidinae, for it involves the evolution of ~5,000 species that have conquered virtually all
237 biogeographical regions in the American tropics[32,45]. The rapid radiation of Pleurothallidinae
238 occurring in the last ~20 Myrs[28] is associated with the evolution of a diverse suite of pollination
239 systems ranging from food deception[59] to pseudocopulation[60] linked to dipterans[61,62] and a complex
240 array of reproductive and vegetative morphologies[22,32]. Understanding of relationships in the
241 subtribe has relied mostly on relatively small number of markers[63–65], which have informed with
242 some confidence the phylogenetic placement and monophyly of genera in Pleurothallidinae, yet
243 basal nodes in these trees have often lacked good support.

244 Several attempts have been conducted to estimate generic relationships in the subtribe,
245 most of which have relied on nuclear rITS and plastid *matK* markers[66]. A synthesis of the
246 phylogenetic relationships in the subtribe based on such studies was conducted by Karremans[67].
247 Here, a cladogram depicting the commonest topologies of relationships between genera was

248 provided and nine clades were defined (termed "affinities" by the author) but without considering
249 the magnitude of the support for these (see Figure 2 in Karremans[67]). Our plastid phylogenomic
250 analysis recovered well-supported relationships in Pleurothallidinae that are mostly in line with
251 previously published studies [28,63,68]. However, these previous trees based on a handful of DNA
252 nuclear and plastid markers yielded poor resolution and low support for backbone nodes as well as
253 infrageneric relationships. In contrast, our plastid phylogenomic inferences recovered high support
254 along the backbone, thus recovering novel placements. Some of these noteworthy well-supported
255 relationships are the position of *Acianthera* as sister to *Myoxanthus* and *Dresslerella* as sister to
256 *Barbosella*+*Restrepia* (Fig. 1, S1).

257 *Acianthera* includes over 300 species distributed throughout the American tropics and
258 subtropics[64,69,70], is often retrieved as sister to the remainder of Pleurothallidinae with moderate
259 support[68]. Karremans[67] used a series of "affinities" to describe to groups of genera affiliated with a
260 core genus of these group and thus described the "*Acianthera* affinity" as the frequent clustering of
261 several Central American genera with *Acianthera*[64]. Our study contradicts Karreman's[67] concept of
262 the *Acianthera* affinity by placing with high support *Acianthera* in the *Restrepia* affinity as sister to
263 *Myoxanthus. Dresslerella* was previously recovered with low support as sister to the remaining
264 genera in the *Restrepia* affinity (*Barbosella*, *Echinosepala*, *Myoxanthus*, *Restrepia*, *Restrepiella*
265 and *Restrepiopsis*). In contrast, our analysis robustly places *Dreslerella* as sister to *Restrepia* and
266 *Barbosella*, a result that does not support the monophyly of the *Restrepia* affinity.

267 Although estimates of the ancestral distribution of the Pleurothallidinae are still uncertain,
268 most of the early divergent Pleurothallidinae and their sister groups are found in the Antilles or
269 Brazil[28]. The remarkable relationship recovered here for *Acianthera*+*Myoxanthus* could yield more
270 clues about the biogeographic history and evolution of the subtribe because Brazil harbours a high
271 species diversity of *Acianthera* and some of the early divergent clades in *Myoxanthus* (particularly
272 the species close to *M. lonchophyllus*), whereas *Myoxanthus* is notably absent in the Antilles. In
273 addition, other early divergent clades such as *Octomeria* and *Barbosella* are more diverse in Brazil.
274 These early diverging clades share the lack of stem annulus as a morphological symplesiomorphy,
275 a character that later appears in more diverse clades such as *Masdevallia*+*Dracula*, *Lepanthes*, and
276 *Pleurothallis*+*Stelis*[71]. Members of these clades probably diversified after a migration to the
277 mountainous areas of the northern Andes ca 16 ± 5 Ma and together account for almost 80% of the
278 species in the subtribe[28]. The modern range extends mostly along the Andean and Central
279 American mountain ranges. Here, another noteworthy relationship is that the less diverse

Specklinia clade (*Scaphosepalum*+*Platystele*) was recovered as sister to the most species-rich clades of the subtribe (*Masdevallia*, *Lepanthes*, and *Pleurothallis*). In previous phylogenetic analyses *Specklinia* clade was recovered as sister to just *Pleurothallis*[28].

Likewise, relationships between early divergent members in the *Lepanthes* affinity (*Anathallis*, *Draconanthes*, *Epibator*, *Lepanthes*, *Opilionanthe*, *Trichosalpinx* and *Tubella*) were largely weakly supported, demonstrating the need for increased taxon sampling, principally in *Lepanthopsis* and *Tubella*[32]. In particular, the early diversification of the *Lepanthes* affinity (>1500 spp.), inferred to have occurred around 8 Ma, has been linked to colonisation of newly formed environments in the Andean Cordillera, a product of accelerated mountain uplift and specific pollination systems (pseudocopulation and food mimicry[60]).

Another novel placement concerns *Teagueia* (diverse in Colombia, Ecuador and Peru[72–74]), which resembles *Platystele*[75]. Karremans[76] had suggested a close relationship between *Teagueia* and *Scaphosepalum*, but our results place *Teagueia* as sister to *Platystele* with high support, thus corroborating the long-standing hypotheses of their sister relationship based on the of their reproductive structures[74,75].

3.5. Evolutionary relationships in Orchidoideae

Our study provides a well-supported tree for Orchidoideae. Our ML inference supports the findings of Pridgeon et al.[35] in which Diurideae is sister to Cranichideae and Codonorchideae to Orchideae. Our findings differ from Givnish et al.[2] and Salazar et al.[34], in which Diurideae/Cranichideae are sister to Codonorchideae, with Orchideae sister to all these (Fig. 2). Givnish et al.[2] included all four tribes but only six of 21 subtribes of Orchidoideae, and the relationship of Diurideae to Cranichideae was poorly supported.

Conclusions

This study presents a well-resolved, more densely sampled and strongly supported analysis of Orchidaceae and their absolute times of divergence than all previous such studies. For deep branches and recent diversifications in Cymbidieae and Epidendreae, support is improved, yet several recalcitrant nodes that historically have been challenging to resolve were also found (e.g. early divergent taxa in the Epidendroideae, initial radiation of the *Lepanthes* affinity in Pleurothallidinae). Similarly, our analyses provide the a well-supported result for Orchidoideae. Although taxon sampling was sufficient to resolve the relationships between the major clades in

312 the family, sampling of unrepresented genera and representatives of Eriopsidiinae, Goodyerinae,
313 and Coleopsidinae would further enhance our understanding of phylogenetic relationships.

314 **Material and methods**

315

316 *2.1 Sampling, DNA extraction and sequencing*

317 Two-hundred and sixty-four species representing 117 genera, 28 subtribes and 18 tribes
318 were sampled in this study. For 51 species plastid genomes were sequenced. Table S5 provides
319 accession numbers of plastid genomes sourced from NCBI and GenBank numbers of those newly
320 generated. Fresh leaves were stored in silica gel for subsequent DNA extraction using a CTAB
321 method[77]. Total DNA was purified with silica columns and then eluted in Tris-EDTA[78]. DNA
322 samples were adjusted to 50 ng/uL and sheared to fragments of approximately 500 bp.

323

324 *High-throughput sequencing*

325 The library preparation, barcoding and sequencing (Illumina HiSeqX) were conducted at
326 Rapid Genomics LLC (Gainesville, FL, USA) and Genewiz GmbH (Leipzig, Germany). Pair-end
327 reads of 150 bp were obtained for fragments with insert size of 300-600 bp. Overhangs were blunt
328 ended using T4 DNA polymerase, Klenow fragment and T4 polynucleotide kinase. Subsequently,
329 a base 'A' was added to the 3' end of the phosphorylated blunt DNA fragments. DNA fragments
330 were ligated to adapters, which have a thymine (T) overhang. Ligation products were gel-purified
331 by electrophoresis to remove all unbound adapters or split adapters that were ligated together.
332 Ligation products were then selectively enriched and amplified by PCR. For each sample, between
333 one and 10 million paired-end reads were generated.

334

335 *Plastid genome assembly and annotation*

336 Raw sequences were quality filtered using Trimmomatic[79] in order to eliminate sequencing
337 artefacts, improve uniformity in the read length (>40 bp) and ensure quality (>20) for further
338 analysis. Filtered sequences were processed with BBNorm[80] to normalize coverage by down-
339 sampling reads over high-depth areas of the genomes (maximum depth coverage 900x and
340 minimum depth 6x). This step creates a flat coverage distribution in order to improve read
341 assembly. Subsequently, overlapping reads were merged into single reads using BBmerge[81] in
342 order to accelerate the assembly process. Overlapping of paired reads was evaluated with Flash[82] to
343 reduce redundancy. Merged reads were used to carry out the whole genome de novo assembly with
344 SPAdes (hash length 33,55,77)[83].

345 To produce contiguous, linear plastid genome sequences we relied on a refence-based and

346 *de-novo* approaches. The reference based approach was conducted on MIRA v. 4[84], a software that
347 maps read data against a consensus sequence of a reference assembly (simple mapping). MIRA has
348 been useful for assembling complicated genomes with many repetitive sequences[85–87]. MIRA
349 produces BAM files as output, which were subsequently used to generate consensus sequences in
350 SAMTOOLS[88]. We sourced 11 reference plastomes from the NCBI repository that represent
351 related species, namely: *Cattleya crispata, Goodyera fumata, Masdevallia picturata*, *M. coccinea*,
352 *Oncidium sphacelatum* and *Sobralia callosa*. The *de-novo* assembly approach relied on
353 GetOrganelle[89], using the recommended default settings for assemblies of green-plant plastid
354 genomes.
355 Newly sequenced and datamined plastid genomes were annotated through the Chlorobox
356 portal of the Max Planck Institute[90]. Sequences were uploaded as fasta files, and running
357 parameters were established as follow: BLAST protein search identity=65%, BLAST rRNA,
358 tRNA, DNA search identity=85%, genetic code = bacterial/plant plastid, max intron length=3,000,
359 options= allow overlaps. *Apostasia wallichii, Masdevallia picturata, Oncidium sphacelatum,*
360 *Sobralia callosa* and *Goodyera fumata* were set as the 'Server Reference' and *Cattleya liliputana*
361 was set as the 'Custom Reference' for CDS and tRNA, rRNA, primer, other DNA or RNA
362 specifications.
363
364 *Phylogenetic analysis*
365 A set of 78 plastid genes was used to reconstruct phylogenetic relationships in Orchidaceae.
366 These were aligned[91] using MAFFT 7[92] and subsequently concatenated (proportions of missing
367 data per species is provided on Tab. S5). This step was performed at the supercomputing centre
368 APOLO, EAFIT University, Medellín, Colombia. Phylogenetic reconstruction based on maximum
369 likelihood (ML) was implemented in RAxML v. 8.0[93], using 1,000 bootstrap replicates and the
370 GTR+GAMMA model. Absolute age estimation analyses relied on fossil and secondary calibration
371 points, strict and molecular clocks and a birth/death model implemented in BEAST v. 1.8[94]. The
372 fossil constraint was added to the MRCA of *Dendrobium* following Xiang et al.[95] using a normal
373 distribution with mean value of 21.07 and a standard deviation (SD) of 3.0. Following Givnish et
374 al.[2], the two secondary calibration points were added to the root of the tree and MRCA of the
375 Orchidaceae, using a normal distribution and mean values of 123.48 (SD=2.0) and 90 (SD=2.0).
376 Because dating analyses conducted on dozens of gene alignments and hundreds of terminals are
377 extremely computationally greedy, we estimated absolute ages on the five most phylogenetically

informative genes (see below) and by constraining the tree topology to the ML tree derived from RAxML. For each clock model, we conducted two MCMC analyses with 250 million generations each with a sampling frequency of 10000 generations. The convergence of the strict and relaxed molecular clocks parameters was confirmed on the software TRACER v1.6. (http://tree.bio.ed.ac.uk/software/tracer/). Maximum clade credibility trees were summarised from the MCMC trees in the program TreeAnnotator v.1.8. of the software BEAST.

Phylogenetic informativeness profiles

To estimate the phylogenetic informativeness (PI) of plastid genes we calculated the per-site and net values for each assessed locus with the HyPhy substitution rates algorithm for DNA sequences[96] using in the web application PhyDesign http://phydesign. townsend.yale.edu/). The input files were the consensus ML ultrametric tree converted with the function *chronos* of the R-package APE (http://ape-package.ird.fr/) using an smoothing rate of 1 and a relaxed clock model, and the partitioned concatenated gene alignments.

Acknowledgments

We would like to thank Esteban Urrea for helping with the bioinformatics pipelines. We thank Norris Williams and the late Mark Whitten (University of Florida) for collecting and preparing the specimens. Kurt Neubig from Southern Illinois University provided the sequences of 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. The University of Costa Rica provided access to the genetic material for the projects B8257 and B6140. 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. O.A.P.E. is supported by the Swiss Orchid Foundation and the Lady Sainsbury Orchid Fellowship. A.A. acknowledges financial support from the Swedish Research Council (2019-05191), the Swedish Foundation for Strategic Research (FFL15-0196) and the Royal Botanic Gardens, Kew.

408 **References**

409 1. Chase, M. W. *et al.* An updated classification of Orchidaceae: Updated Classification of
410 Orchidaceae. *Bot. J. Linn. Soc.* **177,** 151–174 (2015).
411 2. Givnish, T. J. *et al.* Orchid phylogenomics and multiple drivers of their extraordinary
412 diversification. *Proc. R. Soc. B Biol. Sci.* **282,** 1553 (2015).
413 3. Darwin, Ch., *On the various contrivances by which british and foreign orchids are fertilised*
414 *by insects, and on the good effects of intercrossing. London: John Murray. 1st ed., 1st issue.*
415 (1862).
416 4. Fay, M. F. & Chase, M. W. Orchid biology: from Linnaeus via Darwin to the 21st century.
417 *Ann. Bot.* **104,** 359–364 (2009).
418 5. Ramírez, S. R. *et al.* Asynchronous diversification in a specialized plant-pollinator
419 mutualism. *Science* **333,** 1742–1746 (2011).
420 6. Borba, E. L., Barbosa, A. R., Melo, M. C. de, Gontijo, S. L. & Oliveira, H. O. de. Mating
421 systems in the Pleurothallidinae (Orchidaceae): evolutionary and systematic implications.
422 *Lankesteriana* (2011). doi:10.15517/lank.v11i3.18275
423 7. Pérez-Escobar, O. A. *et al.* Multiple geographical origins of environmental sex
424 determination enhanced the diversification of Darwin's favourite orchids. *Sci. Rep.* **7,** 12878
425 (2017).
426 8. Pérez-Escobar, O. A., Gottschling, M., Whitten, W. M., Salazar, G. & Gerlach, G. Sex and
427 the Catasetinae (Darwin's favourite orchids). *Mol. Phylogenet. Evol.* **97,** 1–10 (2016).
428 9. Bateman, R. & Rudall, P. Evolutionary and Morphometric Implications of Morphological
429 Variation Among Flowers Within an Inflorescence: A Case-Study Using European Orchids.
430 *Ann. Bot.* **98,** 975–93 (2006).
431 10. Dong, W.-L. *et al.* Molecular Evolution of Chloroplast Genomes of Orchid Species: Insights
432 into Phylogenetic Relationship and Adaptive Evolution. *Int. J. Mol. Sci.* **19,** 716 (2018).
433 11. Freudenstein, J. V. & Chase, M. W. Phylogenetic relationships in Epidendroideae
434 (Orchidaceae), one of the great flowering plant radiations: progressive specialization and
435 diversification. *Ann. Bot.* **115,** 665–681 (2015).
436 12. Luo, J. *et al.* Comparative Chloroplast Genomes of Photosynthetic Orchids: Insights into
437 Evolution of the Orchidaceae and Development of Molecular Markers for Phylogenetic
438 Applications. *PLoS One* **9,** e99016 (2014).
439 13. Niu, Z. *et al.* The Complete Plastome Sequences of Four Orchid Species: Insights into the

440 Evolution of the Orchidaceae and the Utility of Plastomic Mutational Hotspots. *Front. Plant*
441 *Sci.* **8,** (2017).
442 14. Zhang, G.-Q. *et al.* The *Apostasia* genome and the evolution of orchids. *Nature* **549,** 379–
443 383 (2017).
444 15. Zhang, G.-Q. *et al.* The *Dendrobium catenatum* Lindl. genome sequence provides insights
445 into polysaccharide synthase, floral development and adaptive evolution. *Sci. Rep.* **6,** 19029
446 (2016).
447 16. Yan, L. *et al.* The Genome of *Dendrobium officinale* Illuminates the Biology of the
448 Important Traditional Chinese Orchid Herb. *Mol. Plant* **8,** 922–934 (2015).
449 17. Yuan, Y. *et al.* The *Gastrodia elata* genome provides insights into plant adaptation to
450 heterotrophy. *Nat. Commun.* **9,** (2018).
451 18. Cai, J. *et al.* The genome sequence of the orchid *Phalaenopsis equestris*. *Nat. Genet.* **47,** 65–
452 72 (2014).
453 19. Huang, J.-Z. *et al.* The genome and transcriptome of *Phalaenopsis* yield insights into floral
454 organ development and flowering regulation. *PeerJ* **4,** e2017 (2016).
455 20. Chao, Y.-T. *et al.* Chromosome-level assembly, genetic and physical mapping of
456 *Phalaenopsis aphrodite* genome provides new insights into species adaptation and resources
457 for orchid breeding. *Plant Biotechnol. J.* **16,** 2027–2041 (2018).
458 21. Hu, Y. *et al.* Genomics-based diversity analysis of Vanilla species using a *Vanilla planifolia*
459 draft genome and Genotyping-By-Sequencing. *Sci. Rep.* **9,** 3416 (2019).
460 22. Bogarín, D. *et al.* Anchored Hybrid Enrichment generated nuclear, plastid and mitochondrial
461 markers resolve the *Lepanthes horrida* (Orchidaceae: Pleurothallidinae) species complex.
462 *Mol. Phylogenet. Evol.* (2018). doi:10.1016/j.ympev.2018.07.014
463 23. Pérez-Escobar, O. A. *et al.* Resolving relationships in an exceedingly young Neotropical
464 orchid lineage using Genotyping-by-sequencing data. *Mol. Phylogenet. Evol.* **144,** 106672
465 (2020).
466 24. Jin, W.-T. *et al.* Phylogenetics of subtribe Orchidinae s.l. (Orchidaceae; Orchidoideae) based
467 on seven markers (plastid matK, psaB, rbcL, trnL-F, trnH-psba, and nuclear nrITS, Xdh):
468 implications for generic delimitation. *BMC Plant Biol.* **17,** (2017).
469 25. Li, M.-H., Zhang, G.-Q., Liu, Z.-J. & Lan, S.-R. Subtribal relationships in Cymbidieae
470 (Epidendroideae, Orchidaceae) reveal a new subtribe, Dipodiinae, based on plastid and
471 nuclear coding DNA. *Phytotaxa* **246,** 37 (2016).

472 26. Nauheimer, L., Schley, R. J., Clements, M. A., Micheneau, C. & Nargar, K. Australasian
473 orchid biogeography at continental scale: Molecular phylogenetic insights from the Sun
474 Orchids (*Thelymitra*, Orchidaceae). *Mol. Phylogenet. Evol.* **127,** 304–319 (2018).
475 27. Li, Y.-X. *et al.* Phylogenomics of Orchidaceae based on plastid and mitochondrial genomes.
476 *Mol. Phylogenet. Evol.* **139,** 106540 (2019).
477 28. Pérez-Escobar, O. A. *et al.* Recent origin and rapid speciation of Neotropical orchids in the
478 world's richest plant biodiversity hotspot. *New Phytol.* **215,** 891–905 (2017).
479 29. Whitten, W. M., Neubig, K. M. & Williams, N. H. Generic and subtribal relationships in
480 neotropical Cymbidieae (Orchidaceae) based on matK/ycf1 plastid data. *Lankesteriana* **13,**
481 375–392 (2014).
482 30. Pridgeon, A. *Genera Orchidacearum Vol. 5, Vol. 5,*. (Oxford University Press, 2009).
483 31. Pérez-Escobar, O. A. *et al.* The origin and diversification of the hyperdiverse flora in the
484 Chocó biogeographic region. *Front. Plant Sci.* **10,** 1328 (2019).
485 32. Bogarín, D. *et al.* Phylogenetic comparative methods improve the selection of characters for
486 generic delimitations in a hyperdiverse Neotropical orchid clade. *Sci. Rep.* **9,** 1–17 (2019).
487 33. Górniak, M., Paun, O. & Chase, M. W. Phylogenetic relationships within Orchidaceae based
488 on a low-copy nuclear coding gene, Xdh: Congruence with organellar and nuclear ribosomal
489 DNA results. *Mol. Phylogenet. Evol.* **56,** 784–795 (2010).
490 34. Salazar, G. A., Chase, M. W., Soto Arenas, M. A. & Ingrouille, M. Phylogenetics of
491 Cranichideae with emphasis on Spiranthinae (Orchidaceae, Orchidoideae): evidence from
492 plastid and nuclear DNA sequences. *Am. J. Bot.* **90,** 777–795 (2003).
493 35. Bone, R. E., Cribb, P. J. & Buerki, S. Phylogenetics of Eulophiinae (Orchidaceae:
494 Epidendroideae): evolutionary patterns and implications for generic delimitation:
495 Evolutionary patterns in Eulophiinae. *Bot. J. Linn. Soc.* **179,** 43–56 (2015).
496 36. Pérez-Escobar, O. A. *et al.* Andean mountain building did not preclude dispersal of lowland
497 epiphytic orchids in the Neotropics. *Sci. Rep.* **7,** 4919 (2017).
498 37. Salazar, G. A. *et al.* Phylogenetic systematics of subtribe spiranthinae (Orchidaceae:
499 Orchidoideae: Cranichideae) based on nuclear and plastid DNA sequences of a nearly
500 complete generic sample. *Bot. J. Linn. Soc.* **186,** 273–303 (2018).
501 38. Martins, A. *et al.* From tree tops to the ground: Reversals to terrestrial habit in *Galeandra*
502 orchids (Epidendroideae: Catasetinae). *Mol. Phylogenet. Evol.* **127,** (2018).
503 39. Nunes, C. *et al.* More than euglossines: the diverse pollinators and floral scents of

504 Zygopetalinae orchids. *Sci. Nat.* **104,** (2017).
505 40. Pansarin, L., Pansarin, E., Gerlach, G. & Sazima, M. The Natural History of *Cirrhaea* and
506 the Pollination System of Stanhopeinae (Orchidaceae). *Int. J. Plant Sci.* 000–000 (2018).
507 doi:10.1086/697997
508 41. Cisternas, M. A. *et al.* Phylogenetic analysis of Chloraeinae (Orchidaceae) based on plastid
509 and nuclear DNA sequences. *Botanical Journal of the Linnean Society* (2012).
510 42. Ramirez, S. R., Roubik, D. W., Skov, C. & Pierce, N. E. Phylogeny, diversification patterns
511 and historical biogeography of euglossine orchid bees (Hymenoptera: Apidae). *Biol. J. Linn.*
512 *Soc.* **100,** 552–572 (2010).
513 43. Cingel, N. A. Van Der. *An Atlas of Orchid Pollination: European Orchids*. (CRC Press,
514 2001).
515 44. Weitemier, K. *et al.* Hyb-Seq: Combining target enrichment and genome skimming for plant
516 phylogenomics. *Appl. Plant Sci.* **2,** 1400042 (2014).
517 45. Pridgeon, A. M., Cribb, P. J., Chase, M. W. & Rasmussen, F. N. *Genera Orchidacearum:*
518 *Vol. 5. Epidendroideae (part two)*. (Oxford University Press, 2009).
519 46. Chomicki, G. *et al.* The velamen protects photosynthetic orchid roots against UV-B damage,
520 and a large dated phylogeny implies multiple gains and losses of this function during the
521 Cenozoic. *New Phytol.* **205,** 1330–1341 (2015).
522 47. Ramírez, S. R., Gravendeel, B., Singer, R. B., Marshall, C. R. & Pierce, N. E. Dating the
523 origin of the Orchidaceae from a fossil orchid with its pollinator. *Nature* **448,** 1042–1045
524 (2007).
525 48. Freudenstein, J. V. & Chase, M. W. Phylogenetic relationships in Epidendroideae
526 (Orchidaceae), one of the great flowering plant radiations: progressive specialization and
527 diversification. *Ann. Bot.* (2015). doi:10.1093/aob/mcu253
528 49. Pérez-Escobar, O. A. Molecular phylogenetics, evolution of sexual systems and historical
529 biogeography of Darwin's favorite orchids (Catasetinae) and Swan orchids (Cycnoches
530 Lindl.). (Ludwig-Maximilians Universität, 2016).
531 50. Soltis, D. E. & Kuzoff, R. K. Discordance between Nuclear and Chloroplast Phylogenies in
532 the Heuchera Group Author (Saxifragaceae). *Evolution (N. Y).* **49,** 727–742 (1995).
533 51. van der Niet, T. & Peter Linder, H. Dealing with incongruence in the quest for the species
534 tree: a case study from the orchid genus *Satyrium*. *Mol. Phylogenet. Evol.* **47,** 154–74
535 (2008).

536 52. Pérez-Escobar, O. A., Balbuena, J. A. & Gottschling, M. Rumbling Orchids: How to Assess
537 Divergent Evolution between Chloroplast Endosymbionts and the Nuclear Host. *Syst. Biol.*
538 **65,** (2016).
539 53. Fragoso-Martínez, I. *et al.* A pilot study applying the plant Anchored Hybrid Enrichment
540 method to New World sages (Salvia subgenus Calosphace; Lamiaceae). *Mol. Phylogenet.*
541 *Evol.* **117,** 124–134 (2017).
542 54. Léveillé-Bourret, É., Starr, J. R., Ford, B. A., Moriarty Lemmon, E. & Lemmon, A. R.
543 Resolving Rapid Radiations within Angiosperm Families Using Anchored Phylogenomics.
544 *Syst. Biol.* **67,** 94–112 (2018).
545 55. Zhang, R. *et al.* Exploration of Plastid Phylogenomic Conflict Yields New Insights into the
546 Deep Relationships of Leguminosae. *Syst. Biol.* **69,** 613–622 (2020).
547 56. Mark Whitten, W., Williams, N. H., Dressler, R. L., Gerlach, G. & Pupulin, F. Generic
548 relationships of Zygopetalinae (Orchidaceae: Cymbidieae): Combined molecular evidence.
549 *Lankesteriana* **5,** 87–107 (2005).
550 57. Neubig, K. M., Williams, N. H., Whitten, W. M. & Pupulin, F. Molecular phylogenetics and
551 the evolution of fruit and leaf morphology of *Dichaea* (Orchidaceae: Zygopetalinae). *Ann.*
552 *Bot.* **104,** 457–467 (2009).
553 58. Pupulin, F. in *Genera Orchidacearum. Vol. 5. Epidendroideae (Part Two)* (eds. Pridgeon,
554 A. M., Cribb, P. J., Chase, M. W. & Rasmussen, F. N.) **5,** 456–546 (Oxford University
555 Press, Oxford., 2009).
556 59. Bogarín, D. *et al.* Pollination of *Trichosalpinx* (Orchidaceae: Pleurothallidinae) by biting
557 midges (Diptera: Ceratopogonidae). *Bot. J. Linn. Soc.* (2018).
558 doi:10.1093/botlinnean/box087
559 60. Blanco, M. A. & Barboza, G. Pseudocopulatory pollination in *Lepanthes* (Orchidaceae:
560 Pleurothallidinae) by fungus gnats. *Ann. Bot.* **95,** 763–772 (2005).
561 61. Damon, A. *et al.* A survey of pollination in remnant orchid populations in Soconusco ,
562 Chiapas , Mexico. **48,** 1–14 (2007).
563 62. Policha, T. *et al.* Disentangling visual and olfactory signals in mushroom-mimicking
564 *Dracula* orchids using realistic three-dimensional printed flowers. *New Phytol.* (2016).
565 doi:10.1111/nph.13855
566 63. Pridgeon, A. M., Solano, R. & Chase, M. W. Phylogenetic relationships in Pleurothallidinae
567 (Orchidaceae): combined evidence from nuclear and plastid DNA sequences. *Am. J. Bot.* **88,**

2286–2308 (2001).

64. Karremans, A. P. *et al.* Phylogenetic reassessment of *Acianthera* (Orchidaceae: Pleurothallidinae). *Harvard Pap. Bot.* **21,** 171–187 (2016).

65. Bogarín, D., Karremans, A. P. & Fernández, M. Genus-level taxonomical changes in the *Lepanthes* affinity (Orchidaceae, pleurothallidinae). *Phytotaxa* **340,** 128–136 (2018).

67. Karremans, A. P. Genera Pleurothallidinarum: an updated phylogenetic overview of Pleurothallidinae. *Lankesteriana* **16,** 219–241 (2016).

68. Pridgeon, A. in *Genera Orchidacearum Volume 4: Epidendroideae (Part 1)* 319–422 (2005).

69. Damián, A., Mitidieri, N. & Chiron, G. A taxonomic synopsis of *Acianthera* (Orchidaceae: Pleurothallidinae) in Peru, including two new species. *An. del Jard. Bot. Madrid* **75,** 1–21 (2018).

70. Luer, C. A. A Systematic Method of Classification of the Pleurothallidinae Versus a Strictly Phylogenetic Method. *Selbyana* **23,** 57–110 (2002).

71. Stern, W. W. L., Pridgeon, A. & Luer, C. Stem structure and its bearing on the systematics of Pleurothallidinae (Orchidaceae). *Bot. J. Linn. Soc.* **91,** 457–471 (1985).

72. Jost, L. & Shepard, A. Two new species of Teagueia (Orchidaceae: Pleurothallidinae) from east-central Ecuador. *Lankesteriana* **11,** 9–14 (2011).

73. Luer, C. A. Icones Pleurothallidinarum VIII. Systematics of *Lepanthopsis*, *Octomeria* subgenus Pleurothallopsis, *Restrepiella*, *Restrepiopsis*, *Salpistele*, and *Teagueia*. Addenda to *Platystele*, *Porroglossum*, and *Scaphosepalum*. *Monogr. Syst. Bot. from Missouri Bot. Gard.* **39,** 1–161 (1991).

74. Luer, C. A. Icones Pleurothallidinarum XXIV. A First Century Of New Species of *Stelis* of Ecuador part one. Addenda to *Lepanthes* of Ecuador. Addenda to *Barbosella*, *Dracula*, *Dresslerella*, *Lepanthopsis*, *Platystele*, *Pleurothallis*, *Restrepia*, *Scaphosepalum*, *Teagueia* and . *Monogr. Syst. Bot. from Missouri Bot. Gard.* **88,** 1–122 (2002).

75. Luer, C. A. Icones Pleurothallidinarum VII. Systematics of *Platystele* (Orchidaceae). *Monogr. Syst. Bot. from Missouri Bot. Gard.* **38,** 1–115 (1990).

76. Karremans, A. P. *et al.* Phylogenetic reassessment of *Specklinia* and its allied genera in the pleurothallidinae (Orchidaceae). *Phytotaxa* **272,** 1–36 (2016).

77. Doyle, J. & Doyle, J. Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem Bull* **19,** 11–15 (1987).

600 78. Neubig, K. M. *et al.* in *DNA Banking for the 21st Century: Proceedings of the US Workshop*
601 *on DNA Banking* (eds. Applequist, W. & Campbell, L.) 81–112 (William L. Brown Center,
602 Missouri Botanical Garden, 2014).
603 79. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
604 sequence data. *Bioinformatics* **30,** 2114–2120 (2014).
605 80. Bushnell. BBMap/BBTools. (2017).
606 81. Bushnell, B., Rood, J. & Singer, E. BBMerge – Accurate paired shotgun read merging via
607 overlap. *PLoS One* **12,** e0185056 (2017).
608 82. Magoc, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve
609 genome assemblies. *Bioinformatics* **27,** 2957–2963 (2011).
610 83. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to
611 Single-Cell Sequencing. *J. Comput. Biol.* **19,** 455–477 (2012).
612 84. Chevreux, B., Wetter, T. & Suhai, S. Genome sequence assembly using trace signals and
613 additional sequence information. in *German conference on bioinformatics* **99,** 45–56
614 (Citeseer, 1999).
615 85. Cock, P. J. A., Grüning, B. A., Paszkiewicz, K. & Pritchard, L. Galaxy tools and workflows
616 for sequence analysis with applications in molecular plant pathology. *PeerJ* **1,** e167 (2013).
617 86. Parakhia, M. V. *et al.* Draft Genome Sequence of the Endophytic Bacterium *Enterobacter*
618 spp. MR1, Isolated from Drought Tolerant Plant (Butea monosperma). *Indian J. Microbiol.*
619 **54,** 118–119 (2014).
620 87. Ward, J. A., Ponnala, L. & Weber, C. A. Strategies for transcriptome analysis in nonmodel
621 plants. *Am. J. Bot.* **99,** 267–276 (2012).
622 88. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–
623 2079 (2009).
624 89. Jin, J. J. *et al.* GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of
625 organelle genomes. *Genome Biol.* (2020). doi:10.1186/s13059-020-02154-5
626 90. Tillich, M. *et al.* GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic*
627 *Acids Res.* **45,** W6–W11 (2017).
628 91. Chan, C. X. & Ragan, M. A. Next-generation phylogenomics. *Biol. Direct* **8,** (2013).
629 92. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7:
630 Improvements in Performance and Usability. *Mol. Biol. Evol.* **30,** 772–780 (2013).
631 93. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large

632 phylogenies. *Bioinformatics* **30,** 1312–1313 (2014).
633 94. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling
634 trees. *BMC Evol. Biol.* **7,** 214 (2007).
635 95. Xiang, X. G. *et al.* Biogeographical diversification of mainland Asian *Dendrobium*
636 (Orchidaceae) and its implications for the historical dynamics of evergreen broad-leaved
637 forests. *J. Biogeogr.* **43,** 1310–1323 (2016).
638 96. Kosakovsky Pond, S. L., Frost, S. D. W. & Muse, S. V. HyPhy: Hypothesis testing using
639 phylogenies. *Bioinformatics* (2005). doi:10.1093/bioinformatics/bti079
640

641 **Author Contributions Statement**

642 M.A.S.S., O.A.P.E., and T.A. designed research; O.A.P.E., M.A.S.S., T.A., C.H. and A.A.
643 generated new data; M.A.S.S., O.A.P.E., M.F.T., A.C.A.Y., J.A. and S.D. performed all analyses;
644 O.A.P.E., D.B., M.W.C., M.A.S.S., S.D. and T.A. wrote the manuscript, with contributions from
645 all authors.
646

647 **Legends**

648 **Table 1.** Absolute ages and confidence intervals of main orchid lineages as inferred under a strict
649 and relaxed molecular clocks and a Birth-Death model.
650 **Figure 1.** Maximum Likelihood phylogeny of the orchid family inferred from 78 coding plastid
651 genes. Likelihood bootstrap support values (LBS) < 85% at nodes are highlighted in red together
652 with their corresponding subtending branches. Orchid genera, tribes and subfamilies are indicated
653 in the phylogeny together with photographs of selected representative species per subfamily.
654 (Inset): Bar plot showing the frequency of LBS values at nodes as computed by bin intervals of 5
655 units.
656 **Figure 2.** A comparison of the main plastid topologies of the orchid family published to date. A)
657 Givnish et al.[2]' inference based on 75 plastid genes and 39 orchid species; B) Li et al.[27]' inference
658 based on 76 plastid genes and 76 orchid species; C) This study: 78 plastid and 264 orchid species.
659 LBP at nodes are highlighted in red together with their corresponding subtending branches. (Inset):
660 trees witih branch lengths proportional to substitutions/site. Photos: O. Pérez-Escobar.
661

662 **Supplementary materials**

663 **Table S1.** Detailed absolute ages, confidence intervals and rates for the orchid family as inferred
664 under a strict molecular clock and a birth-death model. The table contains 528 rows and 16
665 columns and is available at https://doi.org/10.6084/m9.figshare.13185008.v1.
666 **Table S2.** Detailed absolute ages, confidence intervals and rates for the orchid family as inferred
667 under a relaxed molecular clock and a birth-death model. The table contains 528 rows and 16
668 columns and is available at https://doi.org/10.6084/m9.figshare.13185008.v1.
669 **Table S3.** Phylogenetic informativeness per-site.
670 **Table S4.** Phylogenetic net informativeness.
671 **Table S5.** Voucher information and proportion of missing data in gene alignments.
672
673 **Figure S1.** Detailed maximum likelihood tree of the orchid family inferred from 78 plastid genes.
674 LBP <100 are shown at nodes, with LBP <85 highlighted in red together with their corresponding
675 subtending branches.
676 **Figure S2.** Chronogram of the orchid family as inferred from a strict molecular clock and a birth-
677 death model. LBP at nodes <85 are highlighted in red together with their corresponding subtending
678 branches. Blue bars at nodes denote 95% high density probability (HDP) absolute age intervals.
679 **Figure S3.** Chronogram of the orchid family as inferred from a relaxed molecular clock and a
680 birth-death model. LBP at nodes <85 are highlighted in red together with their corresponding
681 subtending branches. Blue bars at nodes denote 95% high density probability (HDP) absolute age
682 intervals.
683 **Figure S4.** Phylogenetic informativeness (PI) of 78 plastid gene alignments used in this study to
684 infer orchid relationships. A) Chronogram of Orchidaceae as inferred by PATH8 from the ML tree
685 derived from RAxML; B) Per-site PI; C) Net PI.
686

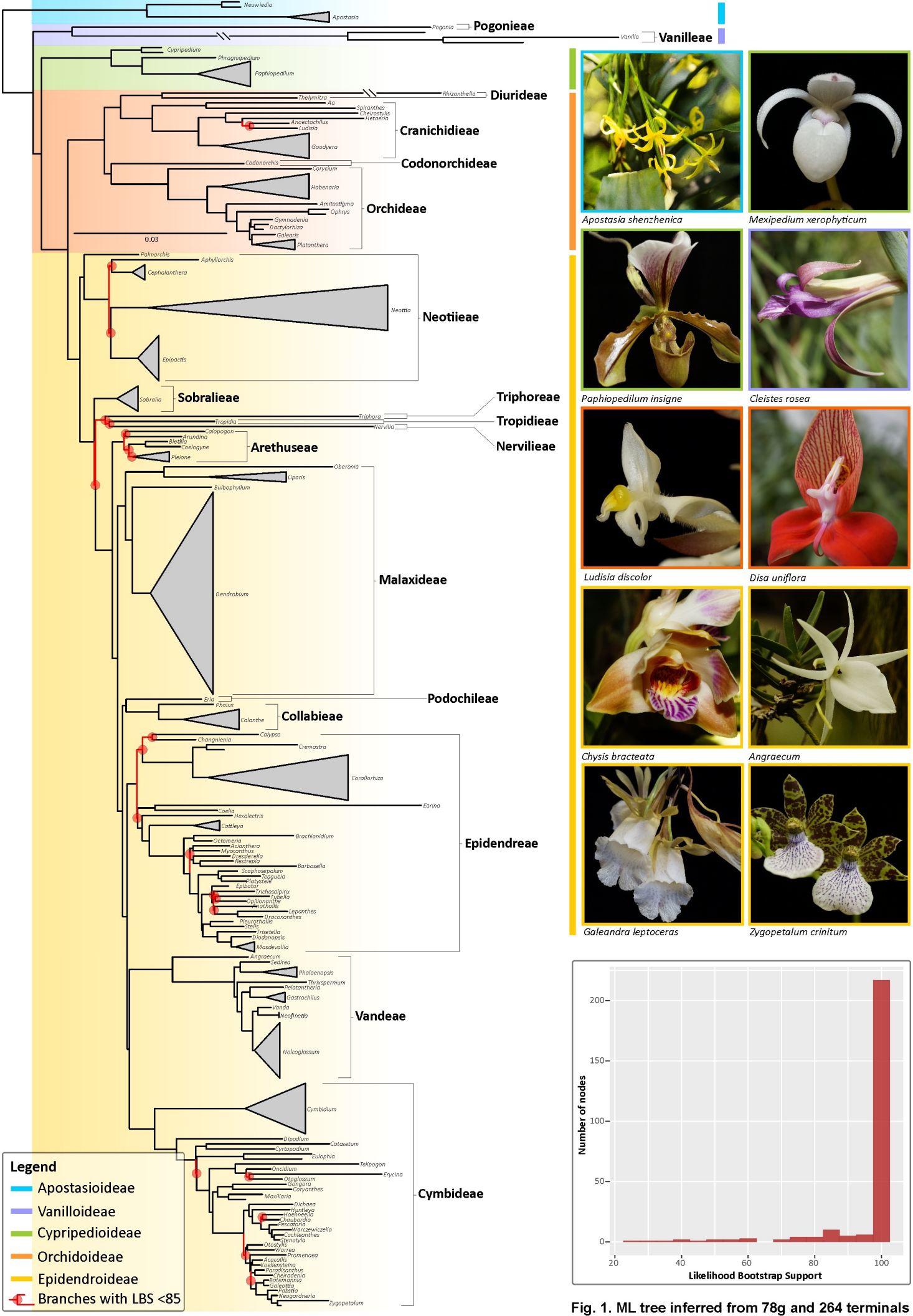
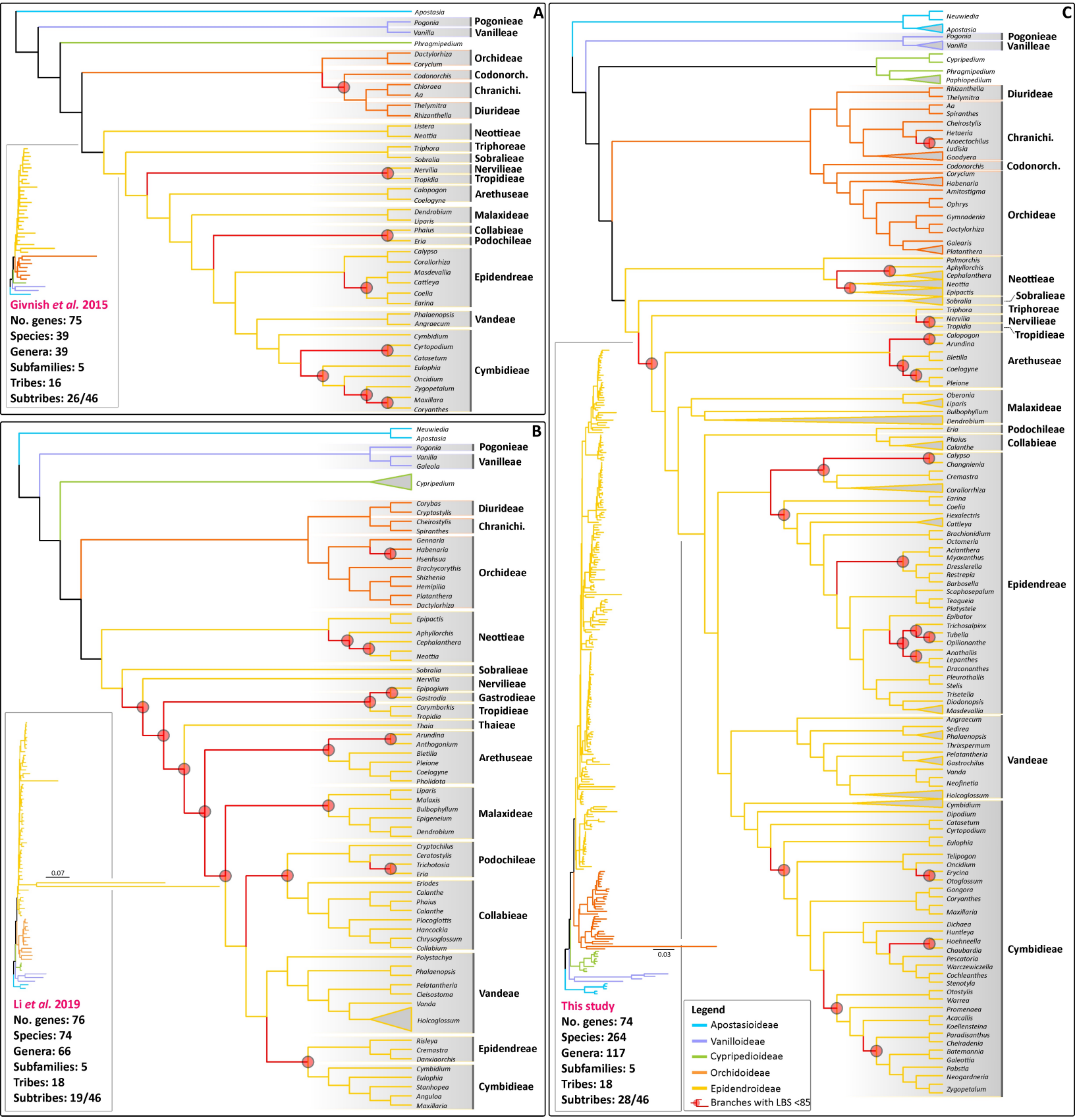


Fig. 1. ML tree inferred from 78g and 264 terminals


A
Apostasia
Pogonia
Vanilla
Pogonieae
Vanilleae
Phragmipedium
Dactylorhiza
Corycium
Orchideae
Codonorchis
Codonorch.
Chloraea
Aa
Chranichi.
Thelymitra
Rhizanthella
Diurideae
Listera
Neottia
Neottieae
Triphora
Triphoreae
Sobralia
Sobralieae
Nervilia
Nervilieae
Tropidia
Tropidieae
Calopogon
Coelogyne
Arethuseae
Dendrobium
Liparis
Malaxideae
Phaius
Collabieae
Eria
Podochileae
Calypso
Corallorhiza
Masdevallia
Cattleya
Coelia
Earina
Epidendreae
Phalaenopsis
Angraecum
Vandeae
Cymbidium
Cyrtopodium
Catasetum
Eulophia
Oncidium
Zygopetalum
Maxillara
Coryanthes
Cymbidieae
Givnish et al. 2015
No. genes: 75
Species: 39
Genera: 39
Subfamilies: 5
Tribes: 16
Subtribes: 26/46
B
Neuwiedia
Apostasia
Pogonia
Vanilla
Galeola
Pogonieae
Vanilleae
Cypripedium
Corybas
Cryptostylis
Diurideae
Cheirostylis
Spiranthes
Chranichi.
Gennaria
Habenaria
Hsenhsua
Brachycorythis
Shizhenia
Hemipilia
Platanthera
Dactylorhiza
Orchideae
Epipactis
Aphyllorchis
Cephalanthera
Neottia
Neottieae
Sobralia
Sobralieae
Nervilia
Epipogium
Nervilieae
Gastrodia
Gastrodieae
Corymborkis
Tropidia
Tropidieae
Thaia
Thaieae
Arundina
Anthogonium
Bletilla
Pleione
Coelogyne
Pholidota
Arethuseae
Liparis
Malaxis
Bulbophyllum
Epigeneium
Dendrobium
Malaxideae
Cryptochilus
Ceratostylis
Trichotosia
Eria
Podochileae
Eriodes
Calanthe
Phaius
Calanthe
Placoglottis
Hancockia
Chrysoglossum
Collabium
Collabieae
Polystachya
Phalaenopsis
Pelatantheria
Cleisostoma
Vanda
Holcoglossum
Vandeae
Risleya
Cremastra
Danxiaorchis
Epidendreae
Cymbidium
Eulophia
Stanhopea
Anguloa
Maxillaria
Cymbidieae
0.07
Li et al. 2019
No. genes: 76
Species: 74
Genera: 66
Subfamilies: 5
Tribes: 18
Subtribes: 19/46
C
Neuwiedia
Apostasia
Pogonia
Vanilla
Pogonieae
Vanilleae
Cypripedium
Phragmipedium
Paphiopedilum
Rhizanthella
Thelymitra
Diurideae
Aa
Spiranthes
Cheirostylis
Hetaeria
Anoectochilus
Ludisia
Goodyera
Chranichi.
Codonorchis
Codonorch.
Corycium
Habenaria
Amitostigma
Ophrys
Gymnadenia
Dactylorhiza
Galearis
Platanthera
Orchideae
Palmorchis
Aphyllorchis
Cephalanthera
Neottia
Epipactis
Neottieae
Sobralia
Sobralieae
Triphora
Triphoreae
Nervilia
Tropidia
Nervilieae
Tropidieae
Calopogon
Arundina
Bletilla
Coelogyne
Pleione
Arethuseae
Oberonia
Liparis
Bulbophyllum
Dendrobium
Malaxideae
Eria
Podochileae
Phaius
Calanthe
Collabieae
Calypso
Changnienia
Cremastra
Corallorrhiza
Earina
Coelia
Hexalectris
Cattleya
Brachionidium
Octomeria
Acianthera
Myoxanthus
Dresslerella
Restrepia
Barbosella
Scaphosepalum
Teagueia
Platystele
Epibator
Trichosalpinx
Tubella
Opilionanthe
Anathallis
Lepanthes
Draconanthes
Pleurothallis
Stelis
Trisetella
Diodonopsis
Masdevallia
Epidendreae
Angraecum
Sedirea
Phalaenopsis
Thrixspermum
Pelatantheria
Gastrochilus
Vanda
Neofinetia
Holcoglossum
Vandeae
Cymbidium
Dipodium
Catasetum
Cyrtopodium
Eulophia
Telipogon
Oncidium
Erycina
Otoglossum
Gongora
Coryanthes
Maxillaria
Dichaea
Huntleya
Hoehnaella
Chaubardia
Pescatoria
Warczewiczella
Cochleanthes
Stenotyla
Otostylis
Warrea
Promenaea
Acacallis
Koellensteina
Paradisanthus
Cheiradenia
Batemannia
Galeottia
Pabstia
Neogardneria
Zygopetalum
Cymbidieae
0.03
This study
No. genes: 74
Species: 264
Genera: 117
Subfamilies: 5
Tribes: 18
Subtribes: 28/46
Legend
Apostasioideae
Vanilloideae
Cypripedioideae
Orchidoideae
Epidendroideae
Branches with LBS <85