



## INVITED SPECIAL ARTICLE

For the Special Issue: Exploring Angiosperms353: a Universal Toolkit for Flowering Plant Phylogenomics

# Hundreds of nuclear and plastid loci yield novel insights into orchid relationships

Oscar Alejandro Pérez-Escobar<sup>1,12</sup> , Steven Dodsworth<sup>2,12</sup>, Diego Bogarín<sup>3</sup>, Sidonie Bellot<sup>1</sup>, Juan A. Balbuena<sup>4</sup>, Rowan J. Schley<sup>1</sup>, Izai A. Kikuchi<sup>5</sup>, Sarah K. Morris<sup>1</sup>, Niroshini Epitawalage<sup>1</sup>, Robyn Cowan<sup>1</sup>, Olivier Maurin<sup>1</sup>, Alexandre Zuntini<sup>1</sup>, Tatiana Arias<sup>6</sup>, Alejandra Serna-Sánchez<sup>6</sup>, Barbara Gravendeel<sup>5</sup>, Maria Fernanda Torres Jimenez<sup>7</sup>, Katharina Nargar<sup>8,9</sup>, Guillaume Chomicki<sup>10</sup>, Mark W. Chase<sup>1,11,\*</sup>, Ilia J. Leitch<sup>1,\*</sup>, Félix Forest<sup>1,\*</sup>, and William J. Baker<sup>1,\*</sup> 

Manuscript received 17 November 2020; revision accepted 10 June 2021.

<sup>1</sup> Royal Botanic Gardens Kew, Richmond TW9 3AE, UK

<sup>2</sup> School of Biological Sciences, University of Portsmouth, Portsmouth PO1 2UP, UK

<sup>3</sup> Lankester Botanic Garden, University of Costa Rica, Cartago, Costa Rica

<sup>4</sup> University of Valencia, Valencia, Spain

<sup>5</sup> Naturalis Biodiversity Center, Leiden, The Netherlands

<sup>6</sup> Colombian Orchid Society, Medellin, Colombia

<sup>7</sup> University of Gothenburg, Gothenburg, Sweden

<sup>8</sup> Australian Tropical Herbarium, James Cook University, Australia

<sup>9</sup> National Research Collections, Commonwealth Industrial and Scientific Research Organization, Australia

<sup>10</sup> Department of Animal and Plant Sciences, Alfred Denny Building, University of Sheffield, Western Bank, Sheffield S10 2TN, UK

<sup>11</sup> Department of Environment and Agriculture, Curtin University, Bentley, Western Australia 6102, Australia

<sup>12</sup> Authors for correspondence (e-mails: o.perez-escobar@kew.org; steven.dodsworth@beds.ac.uk)

\*Joint senior authors.

**Citation:** Pérez-Escobar, O. A., S. Dodsworth, D. Bogarín, S. Bellot, J. A. Balbuena, R. J. Schley, I. A. Kikuchi, S. K. Morris, et al. 2021. Hundreds of nuclear and plastid loci yield novel insights into orchid relationships. *American Journal of Botany* 108(7): 1–15.

doi:10.1002/ajb2.1702

**PREMISE:** The inference of evolutionary relationships in the species-rich family Orchidaceae has hitherto relied heavily on plastid DNA sequences and limited taxon sampling. Previous studies have provided a robust plastid phylogenetic framework, which was used to classify orchids and investigate the drivers of orchid diversification. However, the extent to which phylogenetic inference based on the plastid genome is congruent with the nuclear genome has been only poorly assessed.

**METHODS:** We inferred higher-level phylogenetic relationships of orchids based on likelihood and ASTRAL analyses of 294 low-copy nuclear genes sequenced using the Angiosperms353 universal probe set for 75 species (representing 69 genera, 16 tribes, 24 subtribes) and a concatenated analysis of 78 plastid genes for 264 species (117 genera, 18 tribes, 28 subtribes). We compared phylogenetic informativeness and support for the nuclear and plastid phylogenetic hypotheses.

**RESULTS:** Phylogenetic inference using nuclear data sets provides well-supported orchid relationships that are highly congruent between analyses. Comparisons of nuclear gene trees and a plastid supermatrix tree showed that the trees are mostly congruent, but revealed instances of strongly supported phylogenetic incongruence in both shallow and deep time. The phylogenetic informativeness of individual Angiosperms353 genes is in general better than that of most plastid genes.

**CONCLUSIONS:** Our study provides the first robust nuclear phylogenomic framework for Orchidaceae and an assessment of intragenomic nuclear discordance, plastid-nuclear tree incongruence, and phylogenetic informativeness across the family. Our results also demonstrate what has long been known but rarely thoroughly documented: nuclear and plastid phylogenetic trees can contain strongly supported discordances, and this incongruence must be reconciled prior to interpretation in evolutionary studies, such as taxonomy, biogeography, and character evolution.

**KEY WORDS** Angiosperms353; incongruence; multilocus phylogenetic trees; nuclear-plastid discordance; Orchidaceae; recombination.

Phylogenetic inference based on nuclear and organellar DNA sequences has revolutionized plant systematics and evolution (e.g., Cameron et al., 1999; Eiserhardt et al., 2018; Soltis et al., 2000). From species complexes (Bogarín et al., 2018; Fernández-Mazuecos et al., 2018; Pérez-Escobar et al., 2020) to families and beyond (Bateman et al., 2018; Nauheimer et al., 2018; Wan et al., 2018; Wong et al., 2020), molecular phylogenetics has radically shaped our understanding of plant evolution at widely varying scales and has driven substantial changes to classifications to better reflect monophyly (e.g., APG IV, 2016). Historically, Sanger-sequenced DNA markers have been widely used to infer phylogenies at various taxonomic levels (e.g., Baldwin, 1992; Baldwin and Markos, 1998; Cameron et al., 1999; Soltis et al., 2000). In particular, plastid DNA loci such as *rbcl*, *matK*, and the *trnL-trnF* region, the nuclear ribosomal internal transcribed spacers (nrITS), and the low-copy nuclear gene *Xdh* have been used frequently (e.g., Chase et al., 1993; Baldwin et al., 1995; Soltis et al., 1999; Hilu et al., 2003; Górnjak et al., 2010; Schley et al., 2018), reflecting the relative ease of sequencing these DNA regions, particularly those present in high copy numbers in the cell and genome (plastid regions and nrITS, but see Górnjak et al. 2010).

Recently, implementation of high-throughput sequencing methods has expanded the number of DNA regions available for phylogenetic inference, but there has been a continuing focus on plastid genes (Edger et al., 2018; Li et al., 2019b) and plastomes (Ross et al., 2016; Guo et al., 2017; Li et al., 2019a). The rapid growth of organellar phylogenomic studies is an evident trend in many plant families, including Araceae (Henriquez et al., 2014), Berberidaceae (Sun et al., 2018), Lauraceae (Song et al., 2020), Leguminosae (Zhang et al., 2020), and Orchidaceae (Givnish et al., 2015) due to the relative ease with which they can be generated. Parallel sequencing at shallow coverage (less than 1×) of genomic library preparations derived from recent or historical plant material can yield millions of organellar DNA sequencing reads for hundreds of individuals, enabling the sequencing of dozens of plastid loci at a lower cost per sample than Sanger sequencing (Straub et al., 2012; Dodsworth, 2015).

With at least 25,000 species and 700 genera, Orchidaceae are one of the two largest angiosperm families and are distributed across most terrestrial biomes. Orchids display a wide range of vegetative and reproductive traits that have long captivated biologists. They exhibit unusual relationships with animal pollinators (Darwin, 1877; Jersáková and Malinová, 2004; Ramirez et al., 2011; Martins et al., 2018; Bogarín et al., 2019; Balbuena et al., 2020) and mycorrhizal fungi (Dearnaley, 2007; Rasmussen, 2015; Fochi et al., 2017). Other unusual characters of interest include the velamen, a tissue that fosters water uptake and protects roots in epiphytic and some terrestrial orchids (Zotz and Winkler, 2013; Chomicki et al., 2015) and seeds that are mostly wind-dispersed seeds (Arditti and Ghani, 2000; Barthlott et al., 2014) or occasionally animal-dispersed (ants, bats, bees, crickets and frugivorous birds; Suetsugu et al., 2015; Morales-Linares et al., 2018). Their global distribution, high species-richness in the tropics, and wide variety of functional traits and ecological interactions make them an excellent model group for studying how biotic and abiotic factors affect plant diversification (Givnish et al., 2015, 2016; Pérez-Escobar et al., 2017a, 2017b).

Understanding plant relationships is essential to enable the interpretation of their extraordinary diversity (e.g., Chase et al., 1993; Eiserhardt et al., 2018; Smith and Brown, 2018; Grace et al., 2021). By including representatives of nearly all major taxa, phylogenetic trees inferred from the analysis of mostly organellar loci have provided a

robust set of relationships for a multitude of higher-order lineages. In the particular case of orchids, their relationships have been investigated intensively over the last century (e.g., Schlechter, 1926; Dressler, 1993; Cameron et al., 1999; Freudenstein and Chase, 2015; Givnish et al., 2015). As a result, well-supported organellar phylogenetic frameworks have provided a good understanding of relationships for most orchid tribes and subtribes (reviewed in Chase et al., 2015). They have further been widely employed to investigate the historical biogeography and evolution of selected traits for over 20 years (Neyland and Urbatsch, 1996; Cameron et al., 1999). However, the extent to which the evolutionary history of the maternally inherited organellar genomes (Chang et al., 2000; Cafasso et al., 2005) tracks that of the nuclear genome in the orchid family has not been properly tested in a phylogenomic framework. Across the angiosperms, there is mounting evidence of phylogenetic incongruence between plastid and nuclear genomes and among nuclear genes, driven by phenomena including long-branch attraction (Straub et al., 2014), hybridization, and incomplete lineage sorting (Soltis and Kuzoff, 1995; Smith et al., 2015; Pérez-Escobar et al., 2016; Vargas et al., 2017; Schley et al., 2020; Renner et al., 2021). The extent to which phylogenetic incongruence occurs in orchids (Sramkó et al., 2014; Bateman et al., 2021), or more generally in plants, remains to be assessed.

Here we present a phylogenomic nuclear data set for Orchidaceae relying on the universal Angiosperms353 target capture probe set (Johnson et al., 2019), which is increasingly being used in phylogenomic studies of flowering plants (Baker et al., 2021). We sampled 75 species representing all five recognized subfamilies and the majority of the tribes (Chase et al., 2015). To assess nuclear-plastid topological conflict, we also produced a plastid phylogenomic tree inferred from the published sequences of 78 genes for 264 species, also representing all the same sets of higher taxa. Lastly, we compared the phylogenetic informativeness of these nuclear and plastid loci. We addressed the following topics: (1) the extent to which the orchid plastid tree is congruent with that of the nuclear genome and (2) how well nuclear genes perform in recovering strongly supported relationships compared to plastid genes.

## MATERIALS AND METHODS

### Taxon sampling

We included 75 orchid species representing 69 genera, 24 subtribes (of 46 sensu Chase et al., 2015), 16 tribes (of 21), and all five subfamilies. In addition, 14 species from non-orchid monocot families were included as outgroups (Appendix S1). Newly generated nuclear DNA data were produced from 62 vouchered accessions stored in the DNA and tissue bank of the Royal Botanic Gardens, Kew (<https://dnabank.science.kew.org/homepage.html>). These DNA samples had been previously extracted from silica-dried leaves using a modified CTAB method (Doyle and Doyle, 1987). Each sample has a voucher herbarium specimen in the Kew Herbarium (K; Appendix S1). Nuclear short Illumina sequencing reads were data-mined from the Sequence Read Archive (SRA) using the fastq-dump software of the SRAtool-kit package (available at [https://ncbi.github.io/sra-tools/install\\_config.html](https://ncbi.github.io/sra-tools/install_config.html)) and the 1KP data repository (Wong et al., 2020; <https://sites.google.com/a/ualberta.ca/onekp/>) to expand our taxon sampling with 13 additional orchid species.

### Library preparation, targeted enrichment, and sequencing

The quality of the DNA extractions, including concentration and distribution of fragment lengths, were checked using a Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and a TapeStation 42000 system (Agilent Technologies, Santa Clara, CA, USA). Genomic library preparation and enrichment were conducted following the protocols of Johnson et al. (2019). Dual-indexed Illumina genomic libraries were prepared for each DNA sample using an insert size of ~350 bp and the Ultra II Library Prep Kit (New England BioLabs, Ipswich, MA, USA) following the manufacturer's protocol. We used the Angiosperms353 probe set to enrich each genomic library in 353 low-copy nuclear genes (Johnson et al., 2019; <https://arborbiosci.com/genomics/targeted-sequencing/mybaits/mybaits-expert/mybaits-expert-angiosperms-353/>). Sixty-two genomic libraries were pooled in equimolar quantities to make a 1 µg (total DNA) hybridization reaction, which was subsequently cleaned and sequenced on an Illumina MiSeq v3 (600 cycles, 300-bp paired-end reads) at the Royal Botanic Gardens, Kew to produce ~14.8 Gbp (~30 million paired-end reads). Raw sequence data generated in this study are accessible via the European Nucleotide Archive (<https://www.ebi.ac.uk/ena/browser/home>) under project PRJEB35285 and via the Kew Tree of Life Explorer (<https://treeoflife.kew.org/>; Baker et al., 2021).

### Plastome phylogenomics

To assess the performance of the Angiosperms353 nuclear loci for resolving orchid relationships and investigating nuclear/plastid gene tree discordance at different taxonomic levels, we utilized the 78-plastid-gene data set of Serna-Sánchez et al. (2021; available at <https://doi.org/10.6084/m9.figshare.14068892>) for 264 species representing 117 genera, 28 subtribes, and 18 tribes. The taxon sampling of nuclear and plastid datasets overlaps for 34 genera, 14 tribes, and 22 subtribes (Appendix S1). Detailed information on the completeness of this plastid DNA data set is provided in Serna-Sánchez et al. (2021).

### DNA sequence data analysis and phylogenomic inference

The quality of the newly generated sequencing data was assessed using the FastQC software (available at <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Paired-end DNA sequencing reads were adapter-trimmed and quality-filtered with the pipeline TrimGalore! v.0.6.5 (available at [https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)) using a Phred score quality threshold of 30 (flag *-q*), a minimum read length value of 20 (flag *--length*), and only read pairs that passed all quality-filtering thresholds. Data obtained from the SRA were already adapter- and quality-filtered. For each sample, the Angiosperms353 loci were retrieved using the HybPiper v.1.3.1 pipeline (Johnson et al., 2016) by mapping the clean reads against template sequences of the 353 low copy nuclear genes (available at <https://github.com/mossmatters/Angiosperms353>) using the Burrows-Wheeler Alignment (BWA) program v.0.7 (Li and Durbin, 2009) and then de novo assembling mapped reads for each gene separately using the software SPAdes v. 3.13 (Bankevich et al., 2012), with a minimum coverage threshold of 8×. Given that our data set included samples produced by the 1KP initiative (i.e., read sequencing data produced for CDS DNA only) and the risk of introducing noise in our phylogenetic analyses due to potential homoplasy in noncoding regions (Bellot

et al., 2020), we decided to harvest and utilize exonic regions only for all our phylogenetic analyses to reduce the proportion of missing data in individual gene and supermatrix data sets. For each gene, homologous sequences from each species were combined and aligned with the software MAFFT v. 7.4 (Katoh and Standley, 2013) using the FFT-NS-i strategy. Gene alignments were trimmed for spurious sequences (flags *-resoverlap* 0.75 and *-seqoverlap* 0.90) and positions with 90% missing data (flag *-gt* 0.90) in the software trimAL v.1.2 (Capella-Gutiérrez et al., 2009).

For each nuclear or plastid gene alignment, a maximum likelihood (ML) gene tree was computed using RAxML v8.0 (Stamatakis, 2014) with the GTR+GAMMA nucleotide substitution model, and branch support was estimated using 500 rapid bootstrap replicates (flags *-x* and *-#* 500). The same ML approach was used to estimate two species trees based on a supermatrix made of the 353 concatenated nuclear gene alignments and a supermatrix of the 78 concatenated plastid gene alignments. We treated both concatenated matrices as a single partition and assigned the GTR+GAMMA substitution model. To account for possible topological incongruence between nuclear gene trees, we also inferred a species tree by analyzing the nuclear gene trees together under the multispecies-coalescent (MSC) framework implemented in the software ASTRAL-III v5.6 (Zhang et al., 2018). Given their nonrecombining nature, plastid genomes are often regarded as a single frozen linkage group, yet trees inferred from different plastid genes might still produce incongruent topologies (e.g., Walker et al., 2019; Zhang et al., 2020). Such intragenomic conflict could be driven by variation in substitution rates across loci, changes in organelle inheritance, heteroplasmy, and even incomplete lineage sorting (ILS) or horizontal gene transfer events, although evidence for the last two phenomena is scant (Gonçalves et al., 2019, 2020; Straub et al., 2013). The use of coalescence-based approaches such as ASTRAL on genes with the same evolutionary trajectory is inappropriate to infer species trees, but we nevertheless analyzed individual plastid gene trees in ASTRAL as a way to visualize potential topological conflicts among these loci (see the next section, “Quantification of intragenomic and nuclear-plastid discordance”). Branches with likelihood bootstrap support (LBS) < 20 in the gene trees were first collapsed using the Newick Utilities toolkit (Junier and Zdonov, 2010), as recommended by Sayyari and Mirarab (2016). The resulting ASTRAL topologies were annotated with quartet support values (Sayyari and Mirarab, 2016) for the main topology (*q1*), the first alternative topology (*q2*), and the second alternative topology (*q3*; flag *-t* 2). In addition, we used the software SplitsTree4 (Huson, 1998) to infer neighbor-net networks based on uncorrected *p*-distances calculated from the nuclear supermatrix. Neighbor-nets (also known as splits graphs; Dress and Huson, 2004) are suitable diagrams to represent evolutionary relationships in groups that have experienced reticulation (Rutherford et al., 2018) and are useful to identify relationships that exhibit some ambiguity (Solís-Lemus et al., 2017).

### Quantification of intragenomic and nuclear-plastid discordance

The proportion of intragenomic discordance in the nuclear and plastid data sets was evaluated by looking at the normalized quartet scores produced by ASTRAL when inferring plastid and nuclear MSC species trees. The quartet score indicates the proportion of gene tree quartets that is in agreement with the species tree, and the magnitude is inversely proportional to incongruence, where a value of 1 indicates potential absence of gene tree discordance.



To assess the degree to which the evolutionary history reflected by the nucleus tracked that of the plastid genome, we compared Euclidean distances among terminals between each bootstrap replicate of each nuclear ML gene tree and each bootstrap replicate of the plastid ML species tree. The plastid ML species tree and associated bootstrap replicates were chosen over individual plastid ML gene trees because of the high degree of intragenomic congruence observed between the topologies derived from individual plastid loci (see Results). The comparisons were conducted with the Procrustean Approach to Cophylogenetics (PACo) pipeline implemented in R (Balbuena et al., 2013). The pipeline, originally designed to investigate cophylogenetic patterns between host and parasites, assesses the similarities between any two given trees by comparing the Euclidean distances separating terminals in both via Procrustean superimposition (i.e., the rotation and scaling of ordinations derived from the organelle distance matrix to fit those produced from the nuclear distance matrix; Balbuena et al., 2013). The efficiency of PACo to assess tree incongruence between nuclear–plastid associations (i.e., any given pair of nuclear–plastid terminals) was previously evaluated by Pérez-Escobar et al. (2016), and the pipeline has been widely used in other plant groups, including Asteraceae (Vargas et al., 2017), Fagaceae (Yang et al., 2018), Orchidaceae (Pérez-Escobar, 2016; Pérez-Escobar et al., 2016) and Rosaceae (Morales-Briones, Romoleroux, et al., 2018). The pipeline provides the sum of squared residuals (i.e., the disparity between an observed and a fitted value derived from a model) for each association and each pair of topologies evaluated (Balbuena et al., 2013). This sum of squared residuals can be interpreted as a concordance score because it is directly proportional to the magnitude of the topological conflict for the pair of terminals considered.

Because extremely long branches can bias the comparison of phylogenetic distances between terminals (De Vienne et al., 2011), we conducted PACo analyses on cladograms by assigning a value of 1 to each branch length in each tree, using the function `br.length` in the R package `ape` (Paradis et al., 2004). Differences in the position of terminals between nuclear and plastid trees were summarized in barplots using the R package `ggplot2` (Wickham, 2016), for which the sum of squared residuals for each pair of terminals across nuclear genes was classified in quartiles. Here, the magnitude of the discordance was assessed by the proportion of genes binned in quartiles 3 and 4 (50% and 75%) in each terminal; the more genes binned in quartiles 3 and 4, the more discordant the terminals.

To test for intergenomic conflicts occurring across different taxonomic levels, we conducted the same analysis but with trees including one representative of each genus sampled in our nuclear and plastid species trees. For this analysis, we conducted sequence alignments and ML inference on each nuclear gene alignment and on the matrix of 78 plastid genes as before (see earlier section “DNA sequence data analysis and phylogenomic inference of methods”). We then produced subtribe- and tribe-level trees by keeping one representative of each clade in our nuclear gene trees and plastid ML species tree using the approach of Matzke (2013b) as implemented in the R package `BioGeoBEARS` (Matzke, 2013a; script available at [http://phylo.wikidot.com/example-biogeobears-scripts#pruning\\_a\\_tree](http://phylo.wikidot.com/example-biogeobears-scripts#pruning_a_tree)). Specifically, all tips belonging to the same taxon were pruned except for the first species in the list of taxa representing each clade. Terminals identified as potentially conflicting were depicted in tanglegrams using nuclear and plastid ML species trees derived from matrices sampled to genus level as implemented in the function `tanglegram` of the R package `dendextend` (Galili, 2015). For subtribe- and tribe-level

analyses, we relied on pruned trees originally derived from the nuclear and plastid supermatrices sampled to genus level.

### Assessment of phylogenetic informativeness and support for nuclear and plastid data sets

We compared the performance of the Angiosperms353 low-copy nuclear genes and the 78 plastid genes for resolving orchid relationships in terms of support and phylogenetic informativeness (PI). The latter term refers to the potential of a given locus to resolve a four-terminal polytomy at a given time depth and is estimated from the substitution rates observed at all nucleotide sites of the locus (Townsend, 2007; Bellot et al., 2020). We first calculated the proportion of branches across genus-level nuclear and plastid gene trees that fell into nine discrete LBS categories defined by an interval length of 10 (excluding the interval LBS [81–100]). Secondly, we estimated the PI of each single nuclear and plastid gene alignment with regards to all species relationships recovered in the nuclear and plastid ML species trees, respectively. The trees were made ultrametric by assigning an arbitrary age of 1 to their root and of 0 to their tips (Townsend, 2007), using the software `PATHd8` (<https://www2.math.su.se/PATHd8/>; Britton et al., 2007). Phylogenetic informativeness of the nuclear and plastid genes was computed in `PhyDesign` (López-Giráldez and Townsend, 2011; <http://phydesign.townsend.yale.edu/>) using the `HyPhy` algorithm recommended for DNA sequences (Kosakovsky Pond et al., 2005) and the ultrametric ML species trees and nuclear and plastid supermatrices with gene partition information as input.

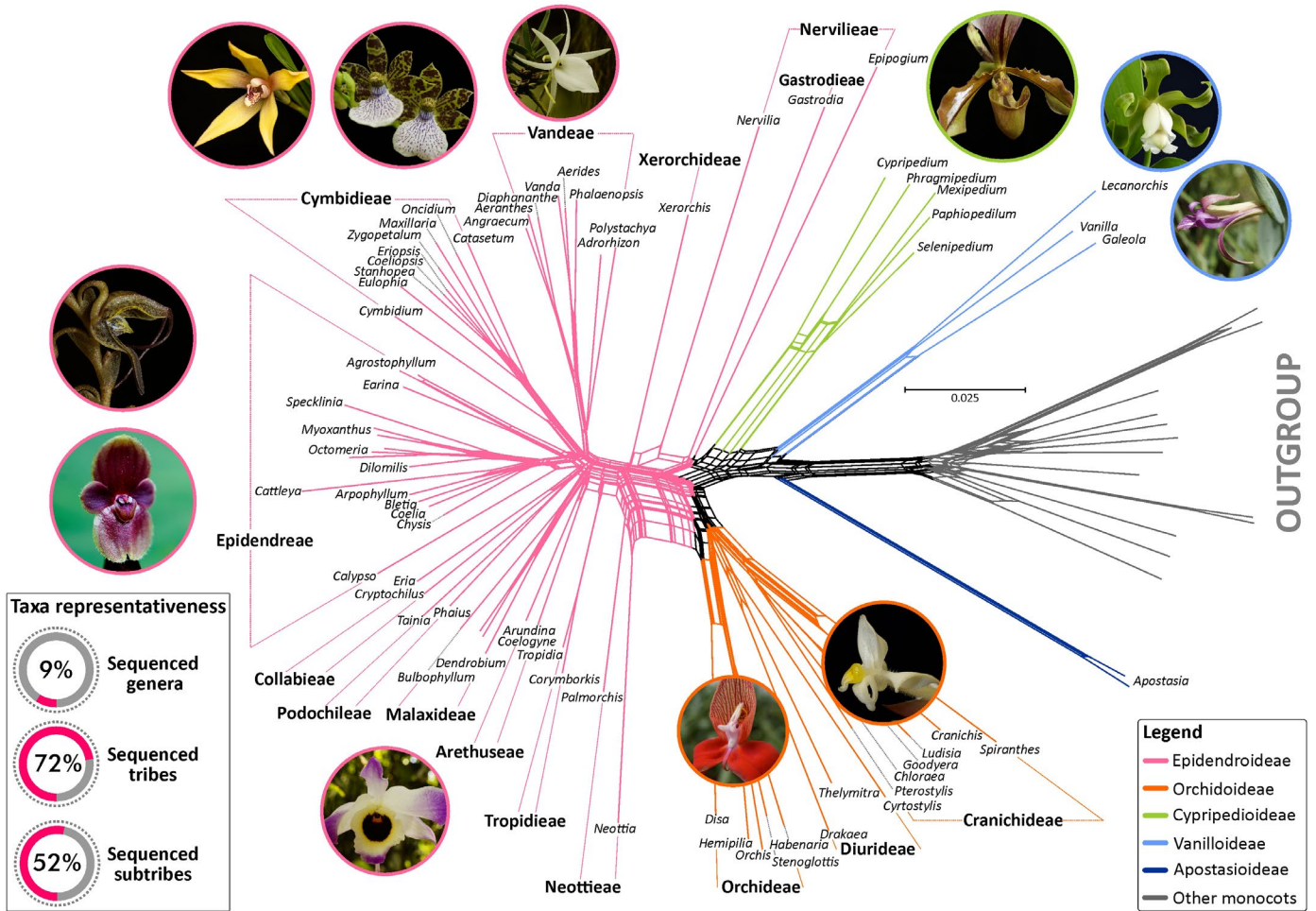
## RESULTS

### Nuclear phylogenomics of Orchidaceae

Enrichment success of the target genes ranged from 5% (*Neottia nidus-avis*) to 87% (*Dendrobium ellipsophyllum*) gene recovery for samples sequenced in this study (Appendices S2, S3). The proportion of nuclear genes recovered from SRA and 1KP data-mined accessions ranged from 16% (*Phalaenopsis equestris*) to 75% (*Mexipedium xerophyticum* and *Paphiopedilum malipoense*). After excluding samples with fewer than 15 genes retrieved and gene alignments including fewer than 20 sequences (i.e., ~80% missing data), the final nuclear data set consisted of 294 genes and 89 species (Appendix S3).

Splits graphs derived from the nuclear supermatrix revealed clear clustering between members of each orchid subfamily, tribe, and subtribe (Fig. 1). However, uncertainty regarding the phylogenetic placement of Neottieae, Nervilieae, and Xerorchideae representatives was reflected in an increased number of alternative splits connecting these groups to representatives of other subfamilies. Maximum likelihood inference of the nuclear supermatrix and MSC analyses converged on similar, strongly supported topologies (Fig. 2; Appendix S4). However, we found important differences between these analyses regarding the placement of Gastrodieae (represented by the mycoheterotrophic *Gastrodia elata*), which was placed as (Neottieae (Gastrodieae (Xerorchideae/other epidendroids))) in the coalescent tree, but as (Neottieae (Xerorchideae (Gastrodieae/other epidendroids))) in the ML results.

An overview of the quartet support obtained across the nuclear MSC indicated that for most branches between 40–96% of the gene trees agreed with the species tree topology (Fig. 2). For a few



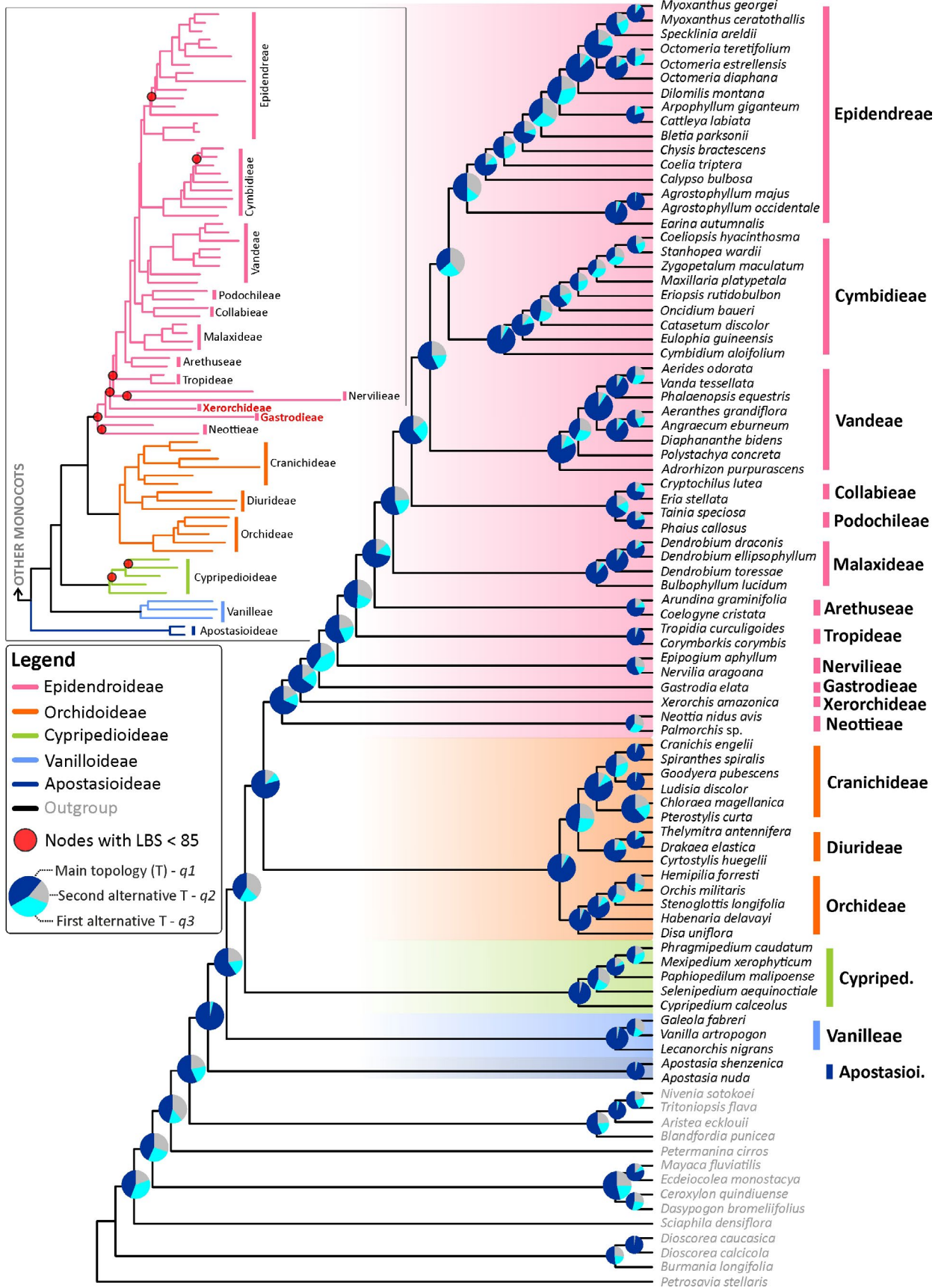
**FIGURE 1.** Split network of Orchidaceae computed from uncorrected  $p$ -distances and a supermatrix of 292 low-copy nuclear genes and 75 species. Splits are color-coded by orchid subfamily (see legend). Box on left: Taxonomic representativeness of orchid genera, tribes and subtribes sampled by nuclear gene datasets in this study. (Inset): Iconic representatives of different tribes in the orchid family (Vanilloideae: *Vanilla inodora*, *Cleistes rosea*; Cyripedioideae: *Paphiopedilum insigne*; Orchidoideae–Cranichideae: *Ludisia discolor*, *Disa uniflora*; Orchidoideae–Orchideae: *Disa uniflora*; Malaxidae: *Dendrobium nobile*; Epidendreae: *Pleurothallis perryi* and *Scaphosepalum verrucosum*; Cymbidieae: *Zygopetalum crinitum* and *Maxillaria pereziana*; Vandeeae: *Angraecum rutenbergianum*). Photo credits: O. A. Pérez-Escobar and Sebastian Vieira.

branches, the proportion of gene trees supporting the quartet displayed in the species tree was below 40%. This included the most recent common ancestors (MRCAs) of *Maxillaria*, *Zygopetalum*, *Stanhopea*, and *Coeliopsis* (36.27), *Bletia*/Epidendreae (34.52), and Cymbidieae/Epidendreae (38.76; Fig. 2). In particular, support for the *Neottia*/*Palmorchis* pair was flagged by ASTRAL as unreliable due to low numbers of gene trees recovered for *Neottia* (i.e., 15). Likelihood bootstrap support (LBS) percentages for the ML tree were high, with only 15 branches displaying LBS <100, of which nine had LBS <85 (Appendix S4). The majority of branches with low support were located near the base of Epidendroideae (the pink clade), in line with our findings of branches with quartet support values <40 as inferred by the MSC estimation.

#### Intragenomic conflict in nuclear and plastid data sets

Estimation of the proportion of gene tree quartets that agree with the species tree through normalized quartet scores indicated that

intragenomic incongruence was low. Here, the proportion of gene tree quartets in agreement with the species tree was 89% for the nuclear genome and 95% for the plastid genome. The proportion of gene quartets supporting topologies other than the species tree (i.e., quartet support) revealed that the majority of branches obtained values between 40 and 95% for the main topology ( $q1$ ) in MSC analyses conducted on nuclear and plastid data sets (Appendix S5). Exceptions to this pattern in the nuclear data set were a few branches with quartet values supporting alternative bipartitions linked to the MRCAs of Epidendreae/Cymbidieae ( $q1 = 38.6$ ;  $q2 = 38.6$ ;  $q3 = 22.69$ ) and *Gastrodia*/remainder of epidendroids ( $q1 = 40.3$ ;  $q2 = 45.2$ ;  $q3 = 14$ ). In the plastid data set, only four branches obtained quartet values robustly supporting multiple quartets. Three other exceptions were in Pleurothallidinae (represented by *Dilomilis*, *Octomeria*, *Myoxanthus*, and *Specklinia*), and one represented the MRCA of Podochilieae/Collabieae plus the remainder of Epidendroideae (Appendix S5).





**FIGURE 2.** Species-coalescence tree of the orchid family inferred from 292 maximum likelihood (ML) gene trees. Pie diagrams at nodes represent quartet support values, with  $q1$  (deep blue portion) representing the proportion of gene tree quartets that support the main (depicted) branch,  $q2$  (light blue portion) representing the proportion of quartets supporting the first alternative branch, and  $q3$  (gray portion) representing the proportion of quartets supporting the second alternative branch (an explanation of how the quartet values are computed is available at <https://github.com/smira/rab/ASTRAL/blob/master/astral-tutorial.md>). (Inset): ML tree derived from a supermatrix of 292 low copy nuclear genes. Terminal names with alternative positions to those obtained by the species-tree coalescence analysis are highlighted in bold and red. A detailed version of the tree is presented in Appendix S4.

A similar trend was found in the proportion of gene trees supporting the species tree or alternative topologies in the nuclear data set sampled to the genus level (Appendix S6), in which the proportion of gene tree topologies congruent with the MSC species-tree branches ranged from 15% to 75% and was often dominant over the proportion of gene trees supporting alternative topologies. Three notable exceptions to this pattern were the MRCA of Cymbidieae/Epidendreae and two early-divergent branches in Epidendroideae: MRCA of Nervilieae/remainder of Epidendroideae and Malaxideae/remainder of Epidendroideae. Here, gene trees supporting a second most common topology were dominant over the species-tree topology (Appendix S6). Overall, the proportion of gene trees supporting the species-tree topology in basal branches in Epidendroideae was low (from 21 to 50) but dominant over the proportion of gene trees supporting the second most-common topology and any other.

### Nuclear–plastid phylogenetic discordance

The incongruence analysis conducted in PACo on nuclear and plastid trees suggested 10 terminals as potentially conflicting (i.e., terminals with ~50% of their squared residual values falling into quartiles 3 and 4; see Methods; Fig. 3; Appendices S7, S8). The squared residual values of these terminals computed individually for each nuclear gene tree assigned to quartiles 3 and 4 were overall higher compared with nonconflicting terminals (Fig. 3; Appendix S8). After inspecting the position of these terminals in nuclear and plastid ML and MSC trees, *Cattleya*, *Coelia*, *Calypso*, and *Earina* appeared to be conflicting with moderate-strong (LBS 85–100) to weak-strong support (LBS 40–98) in the nuclear and plastid ML trees, respectively (Appendix S8, S9). These terminals were further linked to branches with quartet 1 ( $q1$ ) values  $>40$  and  $<76$  in the coalescent tree inferred from the nuclear dataset (Fig. 3, Appendix S5, S8). In the plastid dataset, however, the ASTRAL tree did not recover the same topology as ML, placing *Earina* and *Coelia* as the sister terminals to Epidendreae with  $q1 > 50$  (Appendix S5), *Calypso* as sister to *Changnienia* ( $q1 = 60$ ), and *Cattleya* as sister to Pleurothallidinae ( $q1 = 73$ ). Five other terminals identified as potentially conflicting (i.e., *Angraecum*, *Catasetum*, *Eulophia*, *Phalaenopsis*, and *Vanda*) were found in discordant positions with strong support in nuclear and plastid ML and coalescent trees (LBS  $> 88$ ;  $q1 > 48 < 76$ ; Fig. 3; Appendices S5, S8, S9).

Comparisons of nuclear and plastid trees (subtribe level) revealed 13 terminals were potentially conflicting (Appendix S10). As in the genus level analysis, the squared residual values of terminals computed for each nuclear tree were overall higher for these taxa than the remainder of the tips (Appendix S10). However, only the placements of Catasetinae, Eulophiinae, Angraecinae, Aeridinae, Tropicidiae, and Nervilieae were found to be conflicting with strong support in both nuclear and plastid ML and coalescent trees (Fig. 3; Appendix S11). Lastly, incongruence assessments conducted

between trees sampled to tribe level revealed that Cymbidieae, Epidendreae, Nervilieae, Tropicidiae, and Vandaeae were potentially conflicting (Appendix S12). The phylogenetic positions of Cymbidieae, Epidendreae, Nervilieae, Tropicidiae, and Vandaeae were found to be conflicting with moderate to strong support (Fig. 3, Appendix S12, S13) in nuclear and plastid ML trees. Quartet support for the MRCA of Epidendreae/Cymbidieae in the nuclear data set revealed almost equal support for three alternative bipartitions ( $q1 = 38.6$ ;  $q2 = 38.6$ ;  $q3 = 22.69$ ; Appendix S5).

### Phylogenetic informativeness and support of plastid and nuclear relationships

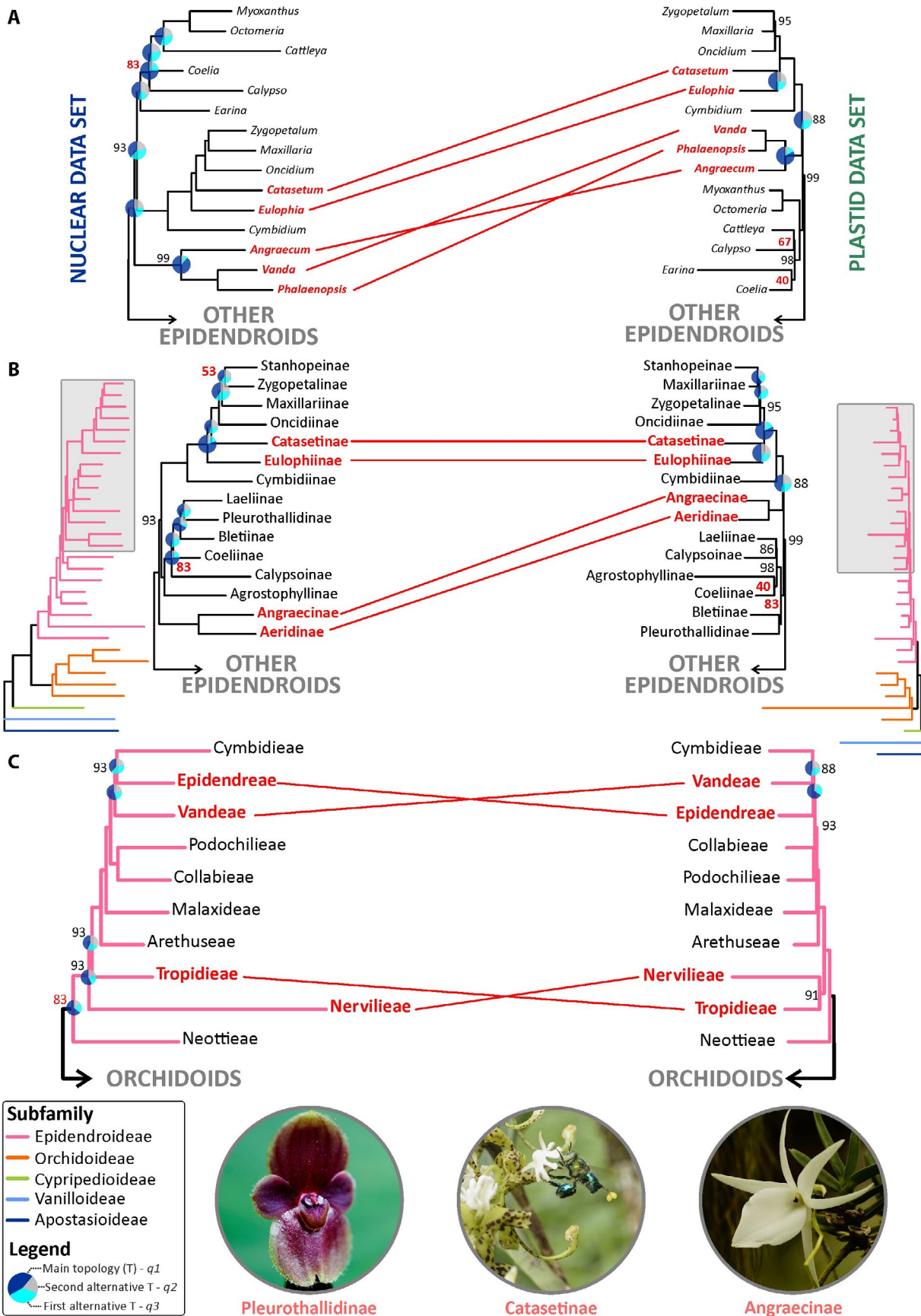
Profiles of phylogenetic informativeness (PI) for the nuclear and plastid data sets are shown in Appendix S14. Net and per-site PIs were in general higher in nuclear than plastid data sets, with average values of 82 and 0.192 versus 56 and 0.058 for nuclear and plastid data sets, respectively. The highest net PI values in nuclear alignments were attained between 0.2 and 0.6 in an arbitrary time scale  $t$  of 0 (tips) to 1 (root; see Materials and Methods), broadly corresponding to the initial diversification of Vanilloideae, Orchidoideae, and Epidendroideae. In contrast, the highest per-site PI occurred at  $t[0.4-0.1]$ , coinciding with the relative divergence times of most generic clades. Net PI values of plastid data sets overall attained uniform values from root to tips, with a notable decrease between  $t[0.4-0]$  and  $t[0.2-0]$  with plastid *ycf1* as an exception to this pattern (Appendix S14). Per-site PI of plastid data sets presented a broadly similar distribution pattern to the nuclear PI per-site values, attaining their highest values at relative time intervals of  $t[0.4-0.01]$ .

The distribution of LBS across gene tree branches strongly contrasted between nuclear and plastid data sets (Appendix S4). Here, a similar distribution pattern of LBS across the interval  $t20$  (LBS  $> 10 \leq 20$ ) and  $t80$  (LBS  $> 70 \leq 80$ ) was observed in nuclear and plastid data sets, with  $t10$  (LBS  $\leq 10$ ) and  $t100$  (LBS  $> 80$ ) scoring the largest values. However, the number of gene tree branches receiving LBS  $\geq 80$  ( $t100$ ) in the nuclear data set was 1600 (vs. 800 in the plastid data set), whereas the interval  $t10$  scored 1400 branches (vs. 3100 in the plastid data set).

## DISCUSSION

### Limitations of plastid-only analyses

With few exceptions (Bateman et al., 2018; Bogarín et al., 2018; Unruh et al., 2018; Brandrud et al., 2020; Pérez-Escobar et al., 2020), previous orchid studies (Cameron et al., 1999; Salazar et al., 2003; Neubig et al., 2012; Givnish et al., 2015; Y. Li et al., 2019; Serna-Sánchez et al., 2021) have relied on plastid data sets, although some have employed the low-copy nuclear gene *Xdh* (Górniak et al., 2010) and the nuclear ribosomal ITS (nrITS) region (Freudenstein and





**FIGURE 3.** Summary of nuclear–plastid phylogenomic incongruence based on PACo analysis across different taxonomic levels between clades of Epidendroideae based on nuclear and plastid ML trees. Conflicting positions between orchid (A) genera, (B) subtribes, and (C) tribes. Association of terminals found to be incongruent and placed with robust support in nuclear and plastid trees are highlighted in bold and red. Pie diagrams at nodes represent quartet support values. Likelihood bootstrap support (LBS) percentages at nodes are 100 unless shown otherwise (LBS < 85 are highlighted in red). Photos: Representatives of three strongly conflicting groups in the orchid family (Epidendreae: *Pleurothallis perryi*.; Catasetinae: *Cycnoches guttulatum* with its potential orchid bee pollinator (*Euglossa* aff. *cybelia*); Angraecinae: *Angraecum rutenbergianum*). Photo credits: O. A. Pérez-Escobar.

Chase, 2015). Historically, nuclear genes have been difficult to sequence due to a combination of factors including inefficient amplification (due to degradation of DNA samples and/or large intronic regions), the need to clone amplified products due to paralogy or allelic diversity, the difficulty of sorting out paralogous sequences from orthologous ones, and lack of universal polymerase chain reaction (PCR) primers (Bailey et al., 2003; Feliner and Rosello, 2007; Sramkó et al., 2014).

High-throughput sequencing has revolutionized our ability to sequence DNA regions at a genomic scale and thus promises to enhance our understanding of phylogenetic relationships by utilizing various approaches to gain more sequence data per taxon sampled. One of the most commonly used methods is genome skimming (Straub et al., 2012; Dodsworth, 2015), which focuses on sequencing high-copy genomic partitions that are present even in low-coverage genomic sequencing. Most notably, such regions include the plastid genome, which has been used extensively for phylogenetics. The ease of this method has meant that it has surged in popularity over recent years in many organisms, including monocots (Givnish et al., 2018) and orchids specifically (Parks et al., 2009; Edger et al., 2018; H. Li et al., 2019a; Kim et al., 2020; Zavala-Páez et al., 2020). Genome skimming thus has perpetuated the bias toward the use of plastid markers in phylogenetic studies of orchids, but at the same time has improved our understanding of their relationships. Nevertheless, it has long been acknowledged that without an assessment of nuclear genes, it would be challenging to evaluate a number of important biological phenomena that have shaped angiosperm evolution and diversification, including hybridization and polyploidization, population structure, gene flow, and introgression (e.g., Rieseberg et al., 1996; Vargas et al., 2017; Schley et al., 2020; Pérez-Escobar et al., 2021). Given the current reliance on plastid analyses for classification/taxonomy and for studying diversification rates, biogeography, and trait evolution, it is important to ask how well these reflect organismal evolution. Considering a nuclear framework to investigate characters across time will refine and may also radically affect our understanding of the mode and tempo of their evolution. An epidendroid example of this phenomenon involves the gains and losses of deceit pollination, a trait thought to be linked with increased speciation and extinction rates in orchids (Givnish et al., 2015). Using a plastid tree of the orchid subtribes as a framework, deceit pollination appeared to have evolved in Laeliinae and Pleurothallidinae independently (Givnish et al., 2015). However, if the same character is optimized on our nuclear tree (Fig. 2), deceit pollination would be inferred to have evolved in the MRCA of these two subtribes; this result could admittedly be a reflection of the sparse taxonomic sampling of our study, but it nevertheless illustrates the problem.

### Are nuclear and plastid evolutionary histories broadly congruent in orchids?

Our comparative phylogenomic analyses provide a solid evolutionary framework for the orchid family inferred from hundreds

of low-copy nuclear genes and a detailed assessment of how much these relationships depart from those previously estimated with plastid DNA. An overview of our current understanding of the phylogenetic relationships of orchids was produced by Chase et al. (2015). The topology provided in this review is highly congruent with studies on entire coding plastid and mitochondrial genomes (Givnish et al., 2015; Li et al., 2019b; Serna-Sánchez et al., 2021) and with that of the low-copy nuclear gene *Xdh* (Górniak et al., 2010). Overall, our quantitative comparisons of our much larger plastid and nuclear data sets support this view, revealing that there is a high degree of congruence between nuclear and organellar phylogenetic trees in orchids, including the monophyly of the five subfamilies and many of the tribal, subtribal, and generic relationships (Fig. 3).

Nevertheless, the topological test for quantification of incongruence conducted here reveals that the positions of several Epidendroideae groups are potentially in conflict. One particular case is the incongruence of *Coelia* and *Earina*, which fall into moderately to well-supported conflicting positions in nuclear ML and MSC analyses (Appendix S5). The plastid tree of Givnish et al. (2015) placed these genera as sister (LBS 90), a pattern also recovered in our ML plastid tree, albeit with lower support (LBS 40). The nuclear MSC analysis, in contrast, robustly places both genera as successively sister to the rest of Epidendreae, but with *Earina* recovering support for an alternative position supported by 25% of the gene quartets (vs. 65% of gene quartets supporting the species tree depicted in Appendix S5). *Earina* is often used for calibration in molecular clock analyses, as it is one of the three orchid macrofossils that has been unambiguously identified (Conran et al., 2009), and thus, its correct placement is important for estimating ages within the orchid tree of life.

Another notable exception to the general plastid–nuclear congruence includes the inter-relationships of Epidendreae, Cymbidieae, and Vandaeae. These three clades account for nearly a third of the known orchid species diversity worldwide and are thought to be derived from multiple rapid diversifications (Givnish et al., 2015; Pérez-Escobar et al., 2017a). Previous plastid phylogenomic studies and our own analyses place Epidendreae as sister to Cymbidieae/Vandaeae with maximum (Givnish et al., 2015; Y. Li et al., 2019b) to moderate support by (Fig. 3, Appendices S5, S12, S13), respectively. In contrast, our nuclear tree places Vandaeae as sister to Cymbidieae/Epidendreae with strong LBS (Fig. 3; Appendix S13). However, we must be cautious about the extent to which the entire nuclear genome agrees with this topology. The quartet support generated for the MRCA of Cymbidieae and Epidendreae indicates that an equal proportion of gene quartets support this and an alternative topology (Fig. 3, Appendix S5), which could suggest that nuclear gene tree discordance might be responsible for an equally plausible arrangement, i.e., (Epidendreae(Cymbidieae/Vandaeae)). Biological phenomena responsible for gene tree incongruence include ILS and hybridization/reticulation (e.g., Rieseberg et al., 1996; van der Niet and Peter Linder, 2008; Joly, 2011; Schley et al., 2020). Teasing apart the signal that these phenomena leave on

a topology is methodologically challenging (e.g., Morales-Briones et al., 2018a, 2018b). Given the overall low proportion of genes supporting the main and alternative topologies (Appendices S5, S6), we suggest that increasing the number of genes representing this particular branch could help determine whether there is a dominant branching pattern overall, as well as the relative influence of ILS and reticulation (Nute et al., 2018).

### Potential of conserved low-copy nuclear genes for orchid phylogenetics

Previously, it has been difficult to sequence more than a handful of low-copy nuclear genes in any plant group (Dodsworth et al., 2018), but genome complexity-reduction methods such as target enrichment have proven to be efficient for sequencing many nuclear loci (Bogarín et al., 2018; Brewer et al., 2019; Dodsworth et al., 2019; Johnson et al., 2019). This approach enables the capture of hundreds of nuclear genes simultaneously, and these loci are highly variable in their levels of conservation, making them useful as a set across all phylogenetic levels.

Our comparative analyses conducted on the per-gene informativeness and support offered by plastid coding data sets and the Angiosperms353 nuclear bait kit (Johnson et al., 2019) strongly point toward the higher performance of low-copy nuclear genes for resolving phylogenetic relationships in both shallow and deep time. These include the recent divergence of epidendroids and clades that experienced rapid radiations, such as Epidendreae/Laeliinae and Arethuseae/remainder of the epidendroids. Groups that remain problematic are those located toward the epidendroid MRCA and those that have been historically difficult to place, including Gastrodieae, Neottieae, and Nervilieae (Chase et al., 2015), which also proved difficult to sequence in our study (gene recovery was poor; Appendix S2). Finding a robust result for such clades will require an increase in their taxonomic representation as well as the number of genes included.

### CONCLUSIONS

We present the first large-scale assessment of the congruence of nuclear and plastid evolutionary histories of the orchid family and provide a generally robust nuclear phylogenomic framework for the family. Comparative analyses of the performance of hundreds of low-copy nuclear genes versus plastid genes reliably demonstrated that the Angiosperms353 genes, in general, are more informative and resolve more relationships with higher support as expected given the usually slower substitution rates of individual plastid genes. We also discovered that, although the plastid genome largely tracks the evolution of the orchid family reflected by the similarity of its phylogenetic reconstruction to that produced by our analysis of the nuclear genome, there are several instances of incongruence at varying taxonomic levels that require further study. These incongruences are a clear indication that in spite of the overall congruence between nuclear and plastid data, they are not interchangeable, particularly when it comes to the study of character and trait evolution through time. Both approaches provide insights into orchid phylogeny, and the task before us is not to eliminate one as “flawed”, but rather to seek to integrate them to provide the best possible inferences about the evolution of this vast family. Our study also highlights the potential benefits of nuclear data sets for assessing

the influence of hybridization and incomplete lineage sorting on patterns of diversification. Here, we found that nuclear gene tree discordance is limited, but nonetheless exists, and will be important to resolve in key groups for understanding the diversification of the orchids and the timing of this. We predict that our study will lead to more in-depth studies of the extent of orchid topological discordance both within the nuclear genome and between genomic compartments—and the phenomena driving this incongruence. By providing a well-supported nuclear phylogenomic backbone for most orchid tribes and subtribes, our study paves the way for future research to produce a densely sampled and strongly supported orchid family tree, building on and merging with decades of work that produced phylogenetic trees derived from Sanger-sequencing data with our phylogenomic framework.

### ACKNOWLEDGMENTS

This work was supported by grants from the Calleva Foundation and the Sackler Trust to the Plant and Fungal Trees of Life project at the Royal Botanic Gardens, Kew. The Associate Editor (Sean Graham) and two anonymous reviewers provided excellent feedback that helped increasing the quality of our study. Sebastian Vieira kindly provided a photograph of *Maxillaria pereziana* (Rodr.-Mart. & M. A. Blanco) Schuit. & M. W. Chase, displayed on Figure 1. O.A.P.E. acknowledges support from the Sainsbury Orchid Fellowship at the Royal Botanic Gardens, Kew and the Swiss Orchid Foundation. O.A.P.E. dedicates this study in loving memory of Professor Jochen Heinrichs (14 March 1969–22 April 2018), a bryologist by profession (yet orchid lover at heart) and a source of tremendous inspiration to O.A.P.E.

### AUTHOR CONTRIBUTIONS

O.A.P.E., D.B., W.J.B., and M.W.C. designed research; O.A.P.E., D.B., O.M., and N.E. collected samples; O.A.P.E., S.B., R.C., R.S., I.K., and S.K.M. performed all the lab work; O.A.P.E., D.B., S.B., M.F.T., and J.A.B. performed all analyses; O.A.P.E., S.D., G.C., S.B., and M.W.C. wrote the manuscript, with contributions from all coauthors.

### DATA AVAILABILITY

Raw sequence data generated in this study are accessible via the European Nucleotide Archive (<https://www.ebi.ac.uk/ena/browser/home>) under project PRJEB35285 and via the Kew Tree of Life Explorer (<https://treeoflife.kew.org/>). High-resolution figures presented in the supplementary material, together with the raw per-site and net PI values are freely available at <https://doi.org/10.6084/m9.figshare.14287538>.

### SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**APPENDIX S1.** Species names, taxonomic and voucher information (including herbarium specimens whenever available) for material used in this study. The number of Angiosperms353 low-copy

nuclear genes sampled in phylogenetic estimations and missing data per sample are also provided.

**APPENDIX S2.** Target enrichment success of de novo sequenced accessions. Heatmap denoting the proportion of the gene sequence length recovered for each sequenced accession.

**APPENDIX S3.** The number of reads produced for de novo sequenced accessions. The number of reads mapped and proportion of target gene sequenced are also provided.

**APPENDIX S4.** Maximum likelihood tree derived from a supermatrix of 292 low-copy nuclear genes. Likelihood bootstrap support (LBS) values is 100 unless shown otherwise (LBS <85 in red). (A) Number of branches per LBS interval derived from 292 ML nuclear gene trees. (B) Number of branches per LBS interval derived from 78 ML plastid gene trees.

**APPENDIX S5.** Multispecies coalescent (MSC) trees inferred from 292 ML nuclear gene trees (A) and 78 ML plastid gene trees (B). Pie diagrams at nodes represent quartet support, with  $q1$  (deep blue portion) representing the proportion of gene tree quartets that support the depicted branch,  $q2$  (blue portion) representing the proportion of quartets supporting the first alternative, and  $q3$  (grey portion) representing the proportion of quartets supporting the second alternative. Tribes follow the classification of Chase et al. (2015).

**APPENDIX S6.** MSC trees inferred from 292 ML nuclear gene trees. Pie diagrams at nodes represent the proportion of gene trees supporting the species tree. Here, the blue portion represents the proportion of concordant gene trees with the species tree, the green portion represents the proportion of gene trees supporting the most common alternative topology, the red portion represents the proportion of gene trees supporting any alternative topology, and the grey portion represents gene trees that are uninformative. Tribes follow the classification of Chase et al. (2015).

**APPENDIX S7.** Experimental design of nuclear-plastid incongruence analyses. The maximum number of terminals sampled in genus, subtribe, and tribe-level analyses of incongruence are provided together with the corresponding number of nuclear-plastid terminal associations analyzed and those deemed potentially conflicting.

**APPENDIX S8.** (A) Conflicting phylogenetic positions between nuclear and plastid trees, depicted on ML trees. Tip names and their corresponding connections are taxa flagged as conflicting. Terminals highlighted in red are placed with strong branch support; those in gray are weakly supported in either the nuclear or plastid tree. Pie diagrams at nodes are provided only for branches with conflicting terminals. They represent quartet support, with  $q1$  (deep blue portion) representing the proportion of gene tree quartets that support the depicted branch,  $q2$  (blue portion) representing the proportion of quartets supporting the first alternative and  $q3$  (gray portion) representing the proportion of quartets supporting the second alternative. All quartet support values are provided in Appendix S5. LBS values are 100 unless shown otherwise (LBS <85 in red). (B) Proportion of genes with summary of squared residual values binned in quartiles 1–4 for each nuclear-plastid association. Terminals with elevated numbers of genes binned in quartiles 3 and 4 were deemed potentially conflicting (labeled with an asterisk). Terminal names highlighted in bold and red denote those

conflicting with strong support in nuclear and plastid trees whereas terminals in bold and connected with grey lines were deemed potentially conflicting albeit with low statistical support in either phylogeny.

**APPENDIX S9.** (A) Maximum likelihood tree derived from a supermatrix of 292 low-copy nuclear genes and (B) 78-coding plastid genes. LBS are 100 unless shown otherwise (LBS <85 in red). The color of branches connected to internal nodes denote LBSs. Tree tips are color coded by subfamilies (pink: Epidendroideae; orange: Orchidoideae; light green: Cyprapedioideae; light blue: Vanilloideae; deep blue: Apostasioideae).

**APPENDIX S10.** (A) Conflicting phylogenetic positions between nuclear and plastid phylogenies sampled to subtribe level, depicted on trimmed ML trees derived from concatenated supermatrices sampled to genus level. Tip names and their corresponding connections are taxa flagged as conflicting. Terminals highlighted in red are placed with strong branch support; those in gray are weakly supported in either the nuclear or plastid tree. Pie diagrams at nodes are provided only for branches interacting with conflicting terminals. They represent quartet support values, with  $q1$  (deep blue portion) representing the proportion of gene tree quartets that support the main (depicted) branch,  $q2$  (blue portion) representing the proportion of quartets supporting the first alternative branch, and  $q3$  (gray portion) representing the proportion of quartets supporting the second alternative branch. All quartet support values are provided in Appendix S7. Experimental design of nuclear-plastid incongruence analyses. The maximum number of terminals sampled in genus, subtribe and tribe-level analyses of incongruence are provided together with the corresponding number of nuclear-plastid terminal associations analyzed and those deemed potentially conflicting. LBS values at nodes are 100 unless shown otherwise (LBS <85 are highlighted in red). (B) Proportion of genes with summary of squared residual values binned in quartiles 1–4 for each nuclear-plastid association. Terminals with elevated numbers of genes binned in quartiles 3 and 4 were deemed potentially conflicting (labelled with a black asterisk). Terminal names highlighted in bold and red denote tips found to be conflicting with strong support in nuclear and plastid phylogenies; terminals in bold and connected with gray lines were deemed potentially conflicting albeit with low statistical support in either phylogeny.

**APPENDIX S11.** (A) A subtribe-level trimmed maximum likelihood tree derived from a supermatrix of 292 low-copy nuclear genes and (B) 78-coding plastid genes sampled to genus level. LBS values at nodes are 100 unless shown otherwise (LBS < 85 is highlighted in red). The color of branches connected to internal nodes denote LBS values. Tree tips are color coded by subfamilies (pink: Epidendroideae, orange: Orchidoideae, light green: Cyprapedioideae, light blue: Vanilloideae, deep blue: Apostasioideae).

**APPENDIX S12.** (A) Conflicting phylogenetic positions between nuclear and plastid trees sampled to tribe level, depicted on trimmed maximum likelihood trees derived from supermatrices sampled to genus level. Tip names and their corresponding connections are taxa flagged as conflicting. Terminals highlighted in red are placed with strong branch support; those in gray are weakly supported in either the nuclear or plastid tree. Pie diagrams at nodes are provided only for branches interacting with conflicting terminals. They represent quartet support values, with  $q1$  (deep



blue portion) representing the proportion of gene tree quartets that support the main (depicted) branch,  $q_2$  (blue portion) representing the proportion of quartets supporting the first alternative branch and  $q_3$  (grey portion) representing the proportion of quartets supporting the second alternative branch. All quartet support values are provided in Appendix S7. LBS values at nodes are 100 unless shown otherwise (LBS <85 highlighted in red). (B) Proportion of genes with summary of squared residual values binned in quartiles 1–4 for each nuclear–plastid association. Terminals with elevated numbers of genes binned in quartiles 3 and 4 were deemed potentially conflicting (labelled with a black asterisk). Terminal names highlighted in bold and red denote tips found to be conflicting with strong support in nuclear and plastid phylogenies; terminals in bold and connected with gray lines were deemed potentially conflicting albeit with low statistical support in either phylogeny.

**APPENDIX S13.** (A) A tribe-level trimmed maximum likelihood tree derived from a supermatrix of 292 low-copy nuclear genes and (B) 78-coding plastid genes sampled to genus level. LBS values at nodes are 100 unless shown otherwise (LBS <85 highlighted in red). The color of branches connected to internal nodes denotes LBS values. Tree tips are color coded by subfamilies (pink: Epidendroideae, orange: Orchidoideae, light green: Cypridioideae, light blue: Vanilloideae, deep blue: Apostasioideae).

**APPENDIX S14.** Net and per-site phylogenetic informativeness analyses of nuclear and plastid data sets. The informativeness (either net or per site) of each gene tree (i.e., the potential of a given locus to resolve a four-terminal polytomy at a given time on the  $x$ -axis) is plotted for each of the 292 nuclear (A) and 79 plastid gene trees (B), respectively. The corresponding species ML trees with branch lengths equivalent to absolute time ( $t$ ) units, where  $t_0$  is present, and  $t_1$  is the tree root height. The informativeness of each gene tree is color-coded (per-site and net PI values are available at Figshare: <https://doi.org/10.6084/m9.figshare.14287538>). Tree tips are color coded by subfamilies (pink: Epidendroideae, orange: Orchidoideae, light green: Cypridioideae, light blue: Vanilloideae, deep blue: Apostasioideae).

## LITERATURE CITED

- APG IV. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* 181: 1–20.
- Arditti, J., and A. K. A. Ghani. 2000. Numerical and physical properties of orchid seeds and their biological implications. *New Phytologist* 145: 367–421.
- Bailey, C. D., T. G. Carr, S. A. Harris, and C. E. Hughes. 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Molecular Phylogenetics and Evolution* 29: 435–455.
- Baker, W. J., P. Bailey, V. Barber, A. Barker, S. Bellot, D. Bishop, L. Botigué, et al. 2021. A comprehensive platform for exploring the angiosperm tree of life. *Systematic Biology* syab035. <https://doi.org/10.1093/sysbio/syab035>.
- Balbuena, J. A., R. Míguez-Lozano, and I. Blasco-Costa. 2013. PACo: a novel procrustes application to cophylogenetic analysis. *PLoS One* 8: e61048.
- Balbuena, J. A., O. A. Pérez-Escobar, C. Llopis-Belenguier, and I. Blasco-Costa. 2020. Random Tanglegram Partitions (Random TaPas): an Alexandrian approach to the cophylogenetic Gordian knot. *Systematic Biology* 69: 1212.
- Baldwin, B. G. 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the compositae. *Molecular Phylogenetics and Evolution* 1: 3–16.
- Baldwin, B. G., and S. Markos. 1998. Phylogenetic utility of the external transcribed spacer (ETS) of 18S–26S rDNA: congruence of ETS and ITS trees of *Calycadenia* (Compositae). *Molecular Phylogenetics and Evolution* 10: 449–463.
- Baldwin, B. G., M. J. Sanderson, J. M. Porter, M. F. Wojciechowski, C. S. Campbell, and M. J. Donoghue. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82: 247–277.
- Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, et al. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455–477.
- Barthlott, W., B. Große-Veldmann, and N. Korotkova. 2014. Orchid seed diversity: a scanning electron microscopy survey. *Englera* 32. Botanic Garden and Botanical Museum Berlin, Berlin, Germany.
- Bateman, R. M., P. J. Rudall, A. R. Murphy, R. S. Cowie, D. S. Devey, and O. Pérez-Escobar. 2021. Whole plastomes are not enough: phylogenomic and morphometric exploration at multiple demographic levels of the bee orchid clade *Ophrys* sect. *Sphegodes*. *Journal of Experimental Botany* 72: 654–681.
- Bateman, R. M., G. Sramkó, and O. Paun. 2018. Integrating restriction site-associated DNA sequencing (RAD-seq) with morphological cladistic analysis clarifies evolutionary relationships among major species groups of bee orchids. *Annals of Botany* 121: 85–105.
- Bellot, S., T. C. Mitchell, and H. Schäfer. 2020. Phylogenetic informativeness analyses to clarify past diversification processes in Cucurbitaceae. *Scientific Reports* 10: 488.
- Bogarín, D., O. A. Pérez-Escobar, D. Groenewald, S. D. Holland, A. K. P. Karremans, E. R. Lemmon, A. R. Lemmon, et al. 2018. Anchored hybrid enrichment generated nuclear, plastid and mitochondrial markers resolve the *Lepanthes horrida* (Orchidaceae: Pleurothallidinae) species complex. *Molecular Phylogenetics and Evolution* 129: 27–47.
- Bogarín, D., O. A. Pérez-Escobar, A. K. P. Karremans, M. Fernández, J. Kruizinga, F. Pupulin, E. Smets, and B. Gravendeel. 2019. Phylogenetic comparative methods improve the selection of characters for generic delimitations in a hyperdiverse Neotropical orchid clade. *Scientific Reports* 9: 1–17.
- Brandrud, M. K., J. Baar, M. T. Lorenzo, A. Athanasiadis, R. M. Bateman, M. W. Chase, M. Hedrén, and O. Paun. 2020. Phylogenomic relationships of diploids and the origins of allotetraploids in *Dactylorhiza* (Orchidaceae). *Systematic Biology* 69: 91–109.
- Brewer, G. E., J. J. Clarkson, O. Maurin, A. R. Zuntini, V. Barber, S. Bellot, N. Biggs, et al. 2019. Factors affecting targeted sequencing of 353 nuclear genes from herbarium specimens spanning the diversity of angiosperms. *Frontiers in Plant Science* 10: 1102.
- Britton, T., C. L. Anderson, D. Jacquet, S. Lundqvist, and K. Bremer. 2007. Estimating divergence times in large phylogenetic trees. *Systematic Biology* 56: 741–752.
- Cafasso, D., A. Widmer, and S. Cozzolino. 2005. Chloroplast DNA inheritance in the orchid *Anacamptis palustris* using single-seed polymerase chain reaction. *Journal of Heredity* 96: 66–70.
- Cameron, K. M., M. W. Chase, W. M. Whitten, P. J. Kores, D. C. Jarrell, V. A. Albert, T. Yukawa, et al. 1999. A phylogenetic analysis of the Orchidaceae: evidence from *rbcL* nucleotide sequences. *American Journal of Botany* 86: 208–224.
- Capella-Gutiérrez, S., J. M. Silla-Martínez, and T. Gabaldón. 2009. trimAL: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973.
- Chang, S.-B., W.-H. Chen, H.-H. Chen, Y.-M. Fu, and Y.-S. Lin. 2000. RFLP and inheritance patterns of chloroplast DNA in intergenic hybrids of *Phalaenopsis* and *Doritis*. *Botanical Studies* 41: 219–223.
- Chase, M. W., K. M. Cameron, J. V. Freudenstein, A. M. Pridgeon, G. Salazar, C. van den Berg, and A. Schuiteman. 2015. An updated classification of Orchidaceae: updated classification of Orchidaceae. *Botanical Journal of the Linnean Society* 177: 151–174.
- Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, et al. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* 80: 528–580.
- Chomicki, G., L. P. R. Bidel, F. Ming, M. Coiro, X. Zhang, Y. Wang, Y. Baissac, et al. 2015. The velamen protects photosynthetic orchid roots against UV-B

- damage, and a large dated phylogeny implies multiple gains and losses of this function during the Cenozoic. *New Phytologist* 205: 1330–1341.
- Conran, J. G., J. M. Bannister, and D. E. Lee. 2009. Earliest orchid macrofossils: early Miocene *Dendrobium* and *Earina* (Orchidaceae: Epidendroideae) from New Zealand. *American Journal of Botany* 96: 466–474.
- Darwin, C. 1877. On the various contrivances by which British and foreign orchids are fertilised by insects. Appleton and Co., NY, NY, USA.
- De Vienne, D. M., G. Aguilera, and S. Ollier. 2011. Euclidean nature of phylogenetic distance matrices. *Systematic Biology* 60: 826–832.
- Dearnaley, J. D. W. 2007. Further advances in orchid mycorrhizal research. *Mycorrhiza* 17: 475–486.
- Dodsworth, S. 2015. Genome skimming for next-generation biodiversity analysis. *Trends in Plant Science* 20: 525–527.
- Dodsworth, S., M. S. Guignard, M. J. M. Christenhusz, R. S. Cowan, S. Knapp, O. Maurin, M. Struebig, et al. 2018. Potential of herbariomics for studying repetitive DNA in angiosperms. *Frontiers in Ecology and Evolution* 6: 174.
- Dodsworth, S., L. Pokorny, M. G. Johnson, J. T. Kim, O. Maurin, N. J. Wickett, F. Forest, et al. 2019. Hyb-Seq for flowering plant systematics. *Trends in Plant Science* 24: 887–891.
- Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.
- Dress, A. W., and D. H. Huson. 2004. Constructing splits graphs. *IEEE/ACM Transactions on Computational and Biological Bioinformatics* 1: 109–115.
- Dressler, R. L. 1993. Phylogeny and classification of the orchid family. Dioscorides Press, Portland, OR, USA.
- Edger, P. P., J. C. Hall, A. Harkess, M. Tang, J. Coombs, S. Mohammadin, M. E. Schranz, et al. 2018. Brassicales phylogeny inferred from 72 plastid genes: a reanalysis of the phylogenetic localization of two paleopolyploid events and origin of novel chemical defenses. *American Journal of Botany* 105: 463–469.
- Eiserhardt, W. L., A. Antonelli, D. J. Bennett, L. R. Botigué, J. G. Burleigh, S. Dodsworth, B. J. Enquist, et al. 2018. A roadmap for global synthesis of the plant tree of life. *American Journal of Botany* 105: 614–622.
- Feliner, G. N., and J. A. Rosello. 2007. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution* 44: 911–919.
- Fernández-Mazuecos, M., G. Mellers, B. Vigalondo, L. Sáez, P. Vargas, and B. J. Glover. 2018. Resolving recent plant radiations: power and robustness of Genotyping-by-Sequencing. *Systematic Biology* 67: 250–268.
- Fochi, V., W. Chitarra, A. Kohler, S. Voyron, V. R. Singan, E. A. Lindquist, K. W. Barry, et al. 2017. Fungal and plant gene expression in the *Tulasnella calospora*–*Serapias vomeracea* symbiosis provides clues about nitrogen pathways in orchid mycorrhizas. *New Phytologist* 213: 365–379.
- Freudenstein, J. V., and M. W. Chase. 2015. Phylogenetic relationships in Epidendroideae (Orchidaceae), one of the great flowering plant radiations: progressive specialization and diversification. *Annals of Botany* 115: 665–681.
- Galili, T. 2015. dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* 31: 3718–3720.
- Givnish, T. J., D. Spalink, M. Ames, S. P. Lyon, S. J. Hunter, A. Zuluaga, A. Doucette, et al. 2016. Orchid historical biogeography, diversification, Antarctica and the paradox of orchid dispersal. *Journal of Biogeography* 43: 1905–1916.
- Givnish, T. J., D. Spalink, M. Ames, S. P. Lyon, S. J. Hunter, A. Zuluaga, W. J. D. Iles, et al. 2015. Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proceedings of the Royal Society, B, Biological Sciences* 282: 1553.
- Givnish, T. J., A. Zuluaga, D. Spalink, M. Gomez, V. K. Lam, J. M. S. Saarela, Ch Saas, et al. 2018. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and functional model for the origin of monocots. *American Journal of Botany* 105: 1888–1910.
- Gonçalves, D. J. P., R. K. Jansen, T. A. Ruhlman, and J. R. Mandel. 2020. Under the rug: abandoning persistent misconceptions that obfuscate organellar variation. *Molecular Phylogenetics and Evolution* 151: 106903.
- Gonçalves, D. J. P., B. B. Simpson, E. M. Ortiz, G. H. Shimizu, and R. K. Jansen. 2019. Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. *Molecular Phylogenetics and Evolution* 138: 219–232.
- Górniak, M., O. Paun, and M. W. Chase. 2010. Phylogenetic relationships within Orchidaceae based on a low-copy nuclear coding gene, *Xdh*: congruence with organellar and nuclear ribosomal DNA results. *Molecular Phylogenetics and Evolution* 56: 784–795.
- Grace, O., O. A. Pérez-Escobar, E. J. Lucas, M. Vorontsova, G. P. Lewis, B. Walker, et al. 2021. Botanical monograph in the Anthropocene. *Trends in Plant Science* 26: 433–441.
- Guo, X., J. Liu, G. Hao, L. Zhang, K. Mao, X. Wang, D. Zhang, et al. 2017. Plastome phylogeny and early diversification of Brassicaceae. *BMC Genomics* 18: 176.
- Henriquez, C. L., T. Arias, J. C. Pires, T. B. Croat, and B. A. Schaal. 2014. Phylogenomics of the plant family Araceae. *Molecular Phylogenetics and Evolution* 75: 91–102.
- Hilu, K. W., T. Borsch, K. Müller, D. E. Soltis, P. S. Soltis, V. Savolainen, M. W. Chase, et al. 2003. Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany* 90: 1758–1776.
- Huson, D. H. 1998. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14: 68–73.
- Jersáková, J., and T. Malinová. 2004. Spatial aspects of seed dispersal and seedling recruitment in orchids. *New Phytologist* 176: 237–241.
- Johnson, M. G., E. M. Gardner, Y. Liu, R. Medina, B. Goffinet, A. J. Shaw, N. J. C. Zerega, and N. J. Wickett. 2016. HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* 4: 1600016.
- Johnson, M., L. Pokorny, S. Dodsworth, L. Botigue, S. Cowan, A. Devault, W. L. Eiserhardt, et al. 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using *k*-medoids clustering. *Systematic Biology* 7: 594–606.
- Joly, S. 2011. JML: testing hybridization from species trees. *Molecular Ecology Resources* 12: 179–184.
- Junier, T., and E. M. Zdobnov. 2010. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* 26: 1669–1670.
- Katoh, K., and D. M. Standley. 2013. MAFFT: multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.
- Kim, Y. K., S. Jo, S. H. Cheon, M. J. Joo, J. R. Hong, M. Kwak, and K. J. Kim. 2020. Plastome evolution and phylogeny of Orchidaceae, with 24 new sequences. *Frontiers in Plant Science* 11: 22.
- Kosakovsky Pond, S. L., S. D. W. Frost, and S. V. Muse. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21: 676–679.
- Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.
- Li, H. T., T. S. Yi, L. M. Gao, P. F. Ma, T. Zhang, J. B. Yang, M. A. Gitzendanner, et al. 2019a. Origin of angiosperms and the puzzle of the Jurassic gap. *Nature Plants* 5: 461–470.
- Li, Y.-X., Z.-H. Li, A. Schuiteman, M. W. Chase, J.-W. Li, W.-C. Huang, A. Hidayat, et al. 2019b. Phylogenomics of Orchidaceae based on plastid and mitochondrial genomes. *Molecular Phylogenetics and Evolution* 139: 106540.
- López-Giráldez, F., and J. P. Townsend. 2011. PhyDesign: an online application for profiling phylogenetic informativeness. *BMC Evolutionary Biology* 11: 152.
- Martins, A., T. Bochorny, O. A. Pérez-Escobar, G. Chomicki, S. Monteiro, and E. Smidt. 2018. From tree tops to the ground: reversals to terrestrial habit in *Galeandra* orchids (Epidendroideae: Catasetinae). *Molecular Phylogenetics and Evolution* 127: 952–960.
- Matzke, N. 2013a. BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. University of California, Berkeley, Berkeley, CA, USA.
- Matzke, N. J. 2013b. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossil allow improved accuracy and model-testing. *Frontiers of Biogeography* 5: 243–248.
- Morales-Briones, D. F., A. Liston, and D. C. Tank. 2018a. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). *New Phytologist* 218: 1668–1684.
- Morales-Briones, D. F., K. Romoleroux, F. Kolář, and D. C. Tank. 2018b. Phylogeny and evolution of the neotropical radiation of *Lachemilla* (Rosaceae):

- uncovering a history of reticulate evolution and implications for infrageneric classification. *Systematic Botany* 43: 17–34.
- Morales-Linares, J., J. G. García-Franco, A. Flores-Palacios, J. E. Valenzuela-González, M. Mata-Rosas, and C. Díaz-Castelazo. 2018. Orchid seed removal by ants in Neotropical ant-gardens. *Plant Biology* 20: 525–530.
- Nauheimer, L., R. J. Schley, M. A. Clements, C. Micheneau, and K. Nargar. 2018. Australasian orchid biogeography at continental scale: molecular phylogenetic insights from the Sun Orchids (*Thelymitra*, Orchidaceae). *Molecular Phylogenetics and Evolution* 127: 304–319.
- Neubig, K. M., W. M. Whitten, N. H. Williams, M. A. Blanco, L. Endara, J. G. Burleigh, K. Silvera, et al. 2012. Generic recircumscriptions of Oncidiinae (Orchidaceae: Cymbidieae) based on maximum likelihood analysis of combined DNA datasets. *Botanical Journal of the Linnean Society* 168: 117–146.
- Neyland, R., and L. E. Urbatsch. 1996. Phylogeny of subfamily Epidendroideae (Orchidaceae) inferred from *ndhF* chloroplast gene sequences. *American Journal of Botany* 83: 1195–1206.
- Nute, M., J. Chou, E. K. Molloy, and T. Warnow. 2018. The performance of coalescent-based species tree estimation methods under models of missing data. *BMC Genomics* 19: 286.
- Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20: 289–290.
- Parks, M., R. Cronn, and A. Liston. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology* 7: 84.
- Pérez-Escobar, O. A. 2016. Molecular phylogenetics, evolution of sexual systems and historical biogeography of Darwin's favorite orchids (Catasetinae) and Swan orchids (*Cynoches* Lindl.). Ph.D. dissertation, Ludwig-Maximilians Universität, Munich, Germany.
- Pérez-Escobar, O. A., J. A. Balbuena, and M. Gottschling. 2016. Rumbling orchids: how to assess divergent evolution between chloroplast endosymbionts and the nuclear host. *Systematic Biology* 65: 51–65.
- Pérez-Escobar, O. A., D. Bogarín, R. Schley, R. M. Bateman, G. Gerlach, D. Harpke, J. Brassac, et al. 2020. Resolving relationships in an exceedingly young Neotropical orchid lineage using Genotyping-by-sequencing data. *Molecular Phylogenetics and Evolution* 144: 106672.
- Pérez-Escobar, O. A., G. Chomicki, F. L. Condamine, A. P. Karremans, D. Bogarín, N. J. Matzke, D. Silvestro, and A. Antonelli. 2017a. Recent origin and rapid speciation of Neotropical orchids in the world's richest plant biodiversity hotspot. *New Phytologist* 215: 891–905.
- Pérez-Escobar, O. A., G. Chomicki, F. L. Condamine, J. M. de Vos, A. C. Martins, E. C. Smidt, B. Klitgård, et al. 2017b. Multiple geographical origins of environmental sex determination enhanced the diversification of Darwin's favourite orchids. *Scientific Reports* 7: 12878.
- Pérez-Escobar, O. A., S. Bellot, N. A. S. Przelomska, J. M. Flowers, M. Nesbitt, P. Ryan, R. M. Gutaker, et al. 2021. Molecular clocks and archaeogenomics of a Late Period Egyptian date palm leaf reveal introgression from wild relatives and add timestamps on the domestication. *Molecular Biology and Evolution*: in press.
- Ramirez, S. R., T. Eltz, M. K. Fujiwara, G. Gerlach, B. Goldman-Huertas, N. D. Tsutsui, and N. E. Pierce. 2011. Asynchronous diversification in a specialized plant-pollinator mutualism. *Science* 333: 1742–1746.
- Rasmussen, F. N. 2015. Trophic relationships in orchid mycorrhiza—diversity and implications for conservation and implications for conservation. *Lankesteriana* 7: 334–341.
- Renner, S. S., S. Wu, O. A. Pérez-Escobar, M. V. Silber, Z. Fei, and G. Chomicki. 2021. A chromosome-level genome of a Kordofan melon illuminates the origin of domesticated watermelons. *Proceedings of the National Academy of Sciences, USA* 118: e2101486118.
- Rieseberg, L. H., J. Whitton, and C. R. Linder. 1996. Molecular marker incongruence in plant hybrid zones and phylogenetic trees. *Acta Botanica Neerlandica* 45: 243–262.
- Ross, T. G., C. F. Barrett, M. Soto Gomez, V. K. Y. Lam, C. L. Henriquez, D. H. Les, J. I. Davis, et al. 2016. Plastid phylogenomics and molecular evolution of Alismatales. *Cladistics* 32: 160–178.
- Rutherford, S., M. Rossetto, J. G. Bragg, H. McPherson, D. Benson, S. P. Bonser, and P. G. Wilson. 2018. Speciation in the presence of gene flow: population genomics of closely related and diverging *Eucalyptus* species. *Heredity* 121: 1–16.
- Salazar, G. A., M. W. Chase, M. A. Soto Arenas, and M. Ingrouille. 2003. Phylogenetics of Cranichideae with emphasis on Spiranthinae (Orchidaceae, Orchidoideae): evidence from plastid and nuclear DNA sequences. *American Journal of Botany* 90: 777–795.
- Sayyari, E., and S. Mirarab. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution* 33: 1654–1668.
- Schlechter, R. 1926. Das System der Orchidaceen. *Notizblatt des Königlichen botanischen Gartens und Museums zu Berlin* 9: 563–591.
- Schley, R. J., M. de la Estrella, O. A. Pérez-Escobar, A. Bruneau, T. Barraclough, F. Forest, and B. Klitgaard. 2018. Is Amazona a 'museum' for Neotropical trees? The evolution of the *Brownea* clade (Detarioideae, Leguminosae). *Molecular Phylogenetics and Evolution* 126: 279–292.
- Schley, R. J., R. T. Pennington, O. A. Pérez-Escobar, A. J. Helmstetter, M. de la Estrella, I. Larridon, I. A. B. Sabino Kikuchi, et al. 2020. Introgression across evolutionary scales suggests reticulation contributes to Amazonian tree diversity. *Molecular Ecology* 29: 4170–4185.
- Serna-Sánchez, M. A., O. A. Pérez-Escobar, D. Bogarín, M. F. Torres, A. Alvarez-Yela, J. Arcila, C. Hall, et al. 2021. Plastid phylogenomics resolves ambiguous relationships within the orchid family and provides a solid timeframe for biogeography and macroevolution. *Scientific Reports* 11: 6858.
- Smith, S. A., and J. W. Brown. 2018. Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany* 105: 302–314.
- Smith, S. A., M. J. Moore, J. W. Brown, and Y. Yang. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology* 15: 150.
- Solis-Lemus, C., P. Bastide, and C. Ané. 2017. PhyloNetworks: a package for phylogenetic networks. *Molecular Biology and Evolution* 34: 3292–3298.
- Soltis, D. E., and R. K. Kuzoff. 1995. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution* 49: 727–742.
- Soltis, D. E., P. S. Soltis, M. W. Chase, M. E. Mort, D. C. Albach, M. Zanis, V. Savolainen, et al. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnean Society* 133: 381–461.
- Soltis, P. S., D. E. Soltis, and M. W. Chase. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402: 402–404.
- Song, Y., W. Bin Yu, Y. H. Tan, J. J. Jin, B. Wang, J. B. Yang, B. Liu, and R. T. Corlett. 2020. Plastid phylogenomics improve phylogenetic resolution in the Lauraceae. *Journal of Systematics and Evolution* 58: 423–439.
- Sramkó, G., M. V. Attila, J. A. Hawkins, and R. M. Bateman. 2014. Molecular phylogeny and evolutionary history of the Eurasian orchid genus *Himantoglossum* s.l. (Orchidaceae). *Annals of Botany* 114: 1609–1626.
- Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.
- Straub, S. C. K., R. C. Cronn, C. Edwards, M. Fishbein, and A. Liston. 2013. Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (Apocynaceae). *Genome Biology and Evolution* 5: 1872–1885.
- Straub, S. C. K., M. J. Moore, P. S. Soltis, D. E. Soltis, A. Liston, and T. Livshultz. 2014. Phylogenetic signal detection from an ancient radiation: effects of noise reduction, long-branch attraction, and model selection in crown clade Apocynaceae. *Molecular Phylogenetics and Evolution* 80: 169–185.
- Straub, S. C. K., M. Parks, K. Weitemier, M. Fishbein, R. C. Cronn, and A. Liston. 2012. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *American Journal of Botany* 99: 349–364.
- Suetsugu, K., A. Kawakita, and M. Kato. 2015. Avian seed dispersal in a mycoheterotrophic orchid *Cyrtosia septentrionalis*. *Nature Plants* 1: 15052.
- Sun, Y., M. J. Moore, J. B. Landis, N. Lin, L. Chen, T. Deng, J. Zhang, et al. 2018. Plastome phylogenomics of the early-diverging eudicot family Berberidaceae. *Molecular Phylogenetics and Evolution* 128: 203–211.
- Townsend, J. P. 2007. Profiling phylogenetic informativeness. *Systematic Biology* 56: 222–231.
- Unruh, S. A., M. R. McKain, Y.-I. Lee, T. Yukawa, M. K. McCormick, R. P. Shefferson, A. Smithson, et al. 2018. Phylotranscriptomic analysis and genome evolution of the Cyripedioideae (Orchidaceae). *American Journal of Botany* 105: 631–640.



- van der Niet, T., and H. Peter Linder. 2008. Dealing with incongruence in the quest for the species tree: a case study from the orchid genus *Satyrium*. *Molecular Phylogenetics and Evolution* 47: 154–174.
- Vargas, O. M., E. M. Ortiz, and B. B. Simpson. 2017. Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). *New Phytologist* 14: 1736–1750.
- Walker, J. F., N. Walker-Hale, O. M. Vargas, D. A. Larson, and G. W. Stull. 2019. Characterizing gene tree conflict in plastome-inferred phylogenies. *PeerJ* 7: e7747.
- Wan, T., Z. M. Liu, L. F. Li, A. R. Leitch, I. J. Leitch, R. Lohaus, Z. J. Liu, et al. 2018. A genome for gnetophytes and early evolution of seed plants. *Nature Plants* 4: 82–89.
- Wickham, H. 2016. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag, NY, NY, USA.
- Wong, G. K. S., D. E. Soltis, J. Leebens-Mack, N. J. Wickett, M. S. Barker, Y. Van De Peer, S. W. Graham, and M. Melkonian. 2020. Sequencing and analyzing the transcriptomes of a thousand species across the tree of life for green plants. *Annual Review of Plant Biology* 71: 741–765.
- Yang, Y., J. Zhu, L. Feng, T. Zhou, G. Bai, J. Yang, and G. Zhao. 2018. Plastid genome comparative and phylogenetic analyses of the key genera in Fagaceae: Highlighting the effect of codon composition bias in phylogenetic inference. *Frontiers in Plant Science* 9: 1–13.
- Zavala-Páez, M., L. do N. Vieira, V. A. de Baura, E. Balsanelli, E. M. de Souza, M. C. Cevallos, M. W. Chase, and E. de C. Smidt. 2020. Comparative plastid genomics of neotropical *Bulbophyllum* (Orchidaceae; Epidendroideae). *Frontiers in Plant Science* 11: 799.
- Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BCM Bioinformatics* 19: 153.
- Zhang, R., Y. H. Wang, J. J. Jin, G. W. Stull, A. Bruneau, D. Cardoso, L. P. De Queiroz, et al. 2020. Exploration of plastid phylogenomic conflict yields new insights into the deep relationships of Leguminosae. *Systematic Biology* 69: 613–622.
- Zotz, G., and U. Winkler. 2013. Aerial roots of epiphytic orchids: the velamen radicum and its role in water and nutrient uptake. *Oecologia* 171: 733–741.