frontiers
in Plant Science

Target Nuclear and Off-Target Plastid Hybrid Enrichment Data Inform a Range of Evolutionary Depths in the Orchid Genus *Epidendrum*

Carolina Granados Mendoza[1*], Matthias Jost[2], Eric Hágsater[3], Susana Magallón[1], Cássio van den Berg[4], Emily Moriarty Lemmon[5], Alan R. Lemmon[6], Gerardo A. Salazar[1*] and Stefan Wanke[2]

[1] Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Mexico City, Mexico, [2] Institut für Botanik, Technische Universität Dresden, Dresden, Germany, [3] Herbario AMO, Instituto Chinoin, A.C., Mexico City, Mexico, [4] Departamento de Ciências Biológicas, Universidade Estadual de Feira de Santana, Feira de Santana, Brazil, [5] Department of Biological Science, Florida State University, Tallahassee, FL, United States, [6] Department of Scientific Computing, Florida State University, Tallahassee, FL, United States

Universal angiosperm enrichment probe sets designed to enrich hundreds of putatively orthologous nuclear single-copy loci are increasingly being applied to infer phylogenetic relationships of different lineages of angiosperms at a range of evolutionary depths. Studies applying such probe sets have focused on testing the universality and performance of the target nuclear loci, but they have not taken advantage of off-target data from other genome compartments generated alongside the nuclear loci. Here we do so to infer phylogenetic relationships in the orchid genus *Epidendrum* and closely related genera of subtribe Laeliinae. Our aims are to: 1) test the technical viability of applying the plant anchored hybrid enrichment (AHE) method (Angiosperm v.1 probe kit) to our focal group, 2) mine plastid protein coding genes from off-target reads; and 3) evaluate the performance of the target nuclear and off-target plastid loci in resolving and supporting phylogenetic relationships along a range of taxonomical depths. Phylogenetic relationships were inferred from the nuclear data set through coalescent summary and site-based methods, whereas plastid loci were analyzed in a concatenated partitioned matrix under maximum likelihood. The usefulness of target and flanking non-target nuclear regions and plastid loci was assessed through the estimation of their phylogenetic informativeness. Our study successfully applied the plant AHE probe kit to *Epidendrum*, supporting the universality of this kit in angiosperms. Moreover, it demonstrated the feasibility of mining plastome loci from off-target reads generated with the Angiosperm v.1 probe kit to obtain additional, uniparentally inherited sequence data at no extra sequencing cost. Our analyses detected some strongly supported incongruences between nuclear and plastid data sets at shallow divergences, an indication of potential lineage sorting, hybridization, or introgression events in the group. Lastly, we found that the per site phylogenetic informativeness of the *ycf1* plastid gene surpasses that of all

other plastid genes and several nuclear loci, making it an excellent candidate for assessing phylogenetic relationships at medium to low taxonomic levels in orchids.

Keywords: Orchidaceae, anchored hybrid enrichment, universal probe set, off-target data, coalescent methods, phylogenomics

INTRODUCTION

Powerful hybrid enrichment strategies (HES), a toolset for selectively capturing genomic regions of interest prior to sequencing (Summerer, 2009; Mamanova et al., 2010; Lemmon and Lemmon, 2013), are increasingly being applied to plant phylogenomics, boosting generation of massive sequence data and, therefore, opening exciting new possibilities for plant evolutionary studies. Previously HES applied to angiosperm phylogenomics include an assortment of nuclear-exon and organelle (plastid) enrichment methods targeting a range of taxonomic levels and lineages. Some differences among these techniques include 1) how capture probes are designed, e.g., whether they target dozens (de Sousa et al., 2014) to thousands (Mandel et al., 2014; Weitemier et al., 2014) of more or less conserved genomic regions; 2) if the focus is taxon-specific (e.g., *Gossypium*, Grover et al., 2015; *Sabal*, Heyduk et al., 2016; *Helianthus*, Stephens et al., 2015) or has a wider taxonomic scope (e.g., angiosperms, Johnson et al., 2019; Zingiberales, Sass et al., 2016; eudicots, Stull et al., 2013); 3) the targeted genome i.e., nuclear (de Sousa et al., 2014; Mandel et al., 2014; Grover et al., 2015) or plastid (Stull et al., 2013). The mitochondrial genome has only begun to be targeted for angiosperm phylogenomic studies (Li et al., 2019). Due to the relatively small size of the plastid genome and its relatively high copy number per cell, assembling complete plastomes (Stull et al., 2013; Sass et al., 2016) or large portions of them (Heyduk et al., 2016) can easily be achieved. Additionally, off-target plastid reads can potentially be recovered after nuclear enrichment, providing a valuable added source of orthologous, uniparentally inherited sequence data (Weitemier et al., 2014; Stephens et al., 2015; Nikolov et al., 2019) to complement nuclear data.

Plant anchored hybrid enrichment (AHE; Buddenhagen et al., 2016) is a highly efficient HES that has rapidly been applied to several angiosperm lineages and an ample range of taxonomic levels (Cardillo et al., 2017; Fragoso-Martínez et al., 2017; Mitchell et al., 2017; Shahi Shavvon et al., 2017; Wanke et al., 2017; Léveillé-Bourret et al., 2018; Crowl et al., 2019; Kriebel et al., 2019). The design of plant AHE probes was based on complete and low-coverage genomes from a variety of species representing all main angiosperm lineages, being potentially applicable to any flowering plant lineage. Since less conserved flanking regions can be captured along with the conserved target regions, retrieved loci can also inform at a range of evolutionary depths.

In this work, we explore the utility of the plant AHE method to resolve phylogenetic relationships in the orchid genus *Epidendrum* (subtribe Laeliinae, tribe Epidendreae, subfamily Epidendroideae), which currently includes over 1,500 Neotropical species exhibiting a great degree of vegetative and reproductive variation, habitat preferences, and ecological interactions (Pérez García, 1993; van den Berg, 2000; Hágsater and Soto-Arenas, 2005; Pinheiro and Cozzolino, 2013). Previous molecular phylogenetic studies of this genus have been based on a single [i.e., nuclear ribosomal internal transcribed spacer (ITS) region; van den Berg et al., 2000] or a few markers (ITS plus plastid *atpI-atpH*, *rpl32-trnL*, *rps16*, *trnH-psbA*, *trnL-trnF*, *trnS-trnfM*, and *matK-trnK* regions: Hágsater and Soto-Arenas, 2005; van den Berg et al., 2009; Quiroga-González, 2017; Pagnussat Klein et al., 2019) obtained *via* Sanger sequencing. Those studies indicated that *Epidendrum* (currently including: *Oerstedella*, *Amblostoma*, *Lanium*, and *Nanodes*; Hágsater and Soto-Arenas, 2005) forms a clade, known as the *Epidendrum* Alliance, together with the genera *Orleanesia*, *Barkeria*, *Caularthron* (van den Berg et al., 2000; van den Berg et al., 2009), and *Microepidendrum* (Hágsater and Soto-Arenas, 2005).

The study of Hágsater and Soto-Arenas (2005) provided a general phylogenetic framework recognizing some major clades of *Epidendrum*. However, support for these major clades and, in general, for backbone relationships remains low due to the lack of informative data. Several fundamental questions persist, such as whether or not the genus is monophyletic, and the identity of its sister lineage (Hágsater and Soto-Arenas, 2005). A well-resolved and strongly supported phylogenetic framework will also facilitate the establishment of a formal infrageneric classification.

For these reasons, *Epidendrum*, along with a careful selection of outgroup taxa, conforms an ideal system to test the technical implementation of the AHE and its performance in resolving and supporting phylogenetic relationships at intermediate to shallow taxonomic levels. Because no orchid representatives were originally included in the set of reference species used to design the plant AHE capture probes, *Epidendrum* represents an excellent group for testing the general applicability of this technique in angiosperms. Additionally, we here explore the feasibility of mining off-target plastid genes from targeted nuclear enrichment data to increase the amount of potentially informative data, as well as to generate an uniparentally inherited dataset without additional sequencing effort. Our aims are to 1) test the technical viability of applying the plant AHE probe set (Angiosperm v.1 probe kit) to *Epidendrum* and outgroup species of subtribe Laeliinae, 2) mine plastid protein coding genes from off-target reads; and 3) evaluate the performance of the target nuclear AHE and off-target plastid loci in resolving and supporting phylogenetic relationships at a range of taxonomical depths within both Laeliinae and *Epidendrum*. In order to account for potential nuclear gene tree discordance, coalescent summary (Zhang et al., 2018) and site-based

(Chifman and Kubatko, 2014) methods were performed for phylogenetic inference. Plastid data was analyzed under a concatenated partitioned approach under maximum likelihood. The phylogenetic utility of loci was assessed through the phylogenetic informativeness method of Townsend (2007).

MATERIALS AND METHODS

Taxon Sampling and Deoxyribonucleic Acid Extraction

Our taxon sampling comprised 18 *Epidendrum* species representing the two main clades previously recognized by Hágsater and Soto-Arenas (2005), including five species previously shown to be closely related to each other as members of the "*Epidendrum anisatum* group" (Quiroga-González, 2017), as well as one species each of the genera *Arpophyllum*, *Barkeria*, *Broughtonia*, and *Caularthron* of subtribe Laeliinae. Such sampling strategy permitted us to assess the phylogenetic utility of our nuclear and plastid data both among major clades, as well as closely-related species of *Epidendrum*. Phylogenetic trees were rooted with *Pleurothallis cardiothallis* of subtribe Pleurothallidinae, because this subtribe was recovered in previous phylogenetic analyzes of Orchidaceae as sister to subtribe Laeliinae (**Table 1**; Chase et al., 2015). Genomic DNA of one individual per species was extracted from fresh or silica-gel dried leaf tissue with the cetyl trimethylammonium bromide (CTAB) method of Doyle and Doyle (1987) modified to include RNase A (Qiagen, 100 mg/ml) and proteinase K (Thermo Scientific, 1 mg/ml) during incubation phases. A NanoDrop 2000/2000c spectrophotometer (Thermo Scientific) was used to ensure a minimum amount of 2.3 μg of DNA per sample with 260/280 and 230/260 purity ratios ≥0.84. Agarose (2.0%) test gels were run for 90 min at 120 V to confirm the presence of bands of high molecular weight and visual assessment of DNA fragmentation.

Plant Anchored Hybrid Enrichment

The Angiosperm v.1 probe kit (Buddenhagen et al., 2016) was used for enrichment. Details of how this probe kit was designed have been extensively explained elsewhere (Buddenhagen et al., 2016; Fragoso-Martínez et al., 2017; Wanke et al., 2017). A recent study in the orchid genus *Lepanthes* (subtribe Pleurothallidinae; Bogarín et al., 2018) applied a modified version of this kit targeting longer and potentially more variable loci so that the retrieved markers were more suitable for population level studies. Since the aim of the present study is to investigate phylogenetic relationships at higher taxonomic ranks (i.e., above the species level) among species of *Epidendrum* and to other genera of Laeliinae and Pleurothallidinae, we applied the original version of the plant AHE probe set of Buddenhagen et al. (2016).

Library preparation and enrichment was performed in the Center for Anchored Phylogenomics at Florida State University (www.anchoredphylogeny.com) as described in Fragoso-Martínez et al. (2017). Up to 16 samples were pooled per enrichment reaction. Enriched libraries were sequenced on one PE150 Illumina HiSeq2000 lane at the Translational Science Laboratory in the College of Medicine at Florida State University.

TABLE 1 | Taxon sampling and voucher information including collector, and collection number (herbarium code as in http://sweetgum.nybg.org/science/ih/).

Function	Subtribe	Taxon	Voucher information	Lab #
Outgroup	Pleurothallidinae	*Pleurothallis cardiothallis*	Soto Arenas, 3679 (AMO)	E100
	Laeliinae	*Arpophyllum giganteum*	Soto Arenas, 6858 (AMO)	E106
		Barkeria melanocaulon	Jiménez Machorro, 2445 (AMO)	E094
		Broughtonia sanguinea	Brieger, 14440 (ESA)	E218
		Caularthron bicornutum	van den Berg, 2997 HUEFS)	E217
Ingroup		*Epidendrum anisatum*	Hágsater, 14559 (AMO)	E026
		E. ciliare	Hágsater, 9468 (AMO)	E058
		E. cusii	Salazar Chávez, 7467 (AMO)	E432
		E. gasteriferum	Salazar Chávez, 7566 (AMO)	E147
		E. juergensenii	Salazar Chávez, 7867 (AMO)	E145
		E. lacertinum	Soto Arenas, 9087 (AMO)	E063
		E. longicaule	Jiménez Machorro, 2763-B (AMO)	E013
		E. magnoliae	Hágsater, 14573 (AMO)	E079
		E. mathewsii	Hágsater, 13559 (AMO)	E070
		E. matudae	Salazar Chávez, 7468 (AMO)	E008
		E. nocturnum	Hágsater, 13963 (AMO)	E065
		E. octomerioides	Hágsater, 13804 (AMO)	E017
		E. parkinsonianum	Hágsater, 14582 (AMO)	E042
		E. propinquum	Hágsater, 14591 (AMO)	E015
		E. sophronitoides	Hágsater, 14552 (AMO)	E048
		E. succulentum	Salazar Chávez, 6723 (AMO)	E435
		E. summerhayesii	Hágsater, 14587 (AMO)	E044
		E. trialatum	Hágsater, 14558 (AMO)	E022

Read Processing, Assembly, Orthology Assessment, and Alignment of Nuclear Loci

The CASAVA v. 1.8 pipeline was used to filter low-quality raw reads applying a high-chastity setting. Filtered reads were demultiplexed and those failing to match exactly any of 13 in-house developed indexes were discarded. The code and parameter settings used for read merging, assembly, orthology assessment, and alignment of nuclear loci is available as **Supplementary Material S1**.

A conservative method, designed to prevent read merging at highly repetitive regions, was performed following Rokyta et al. (2012). Merged and unmerged reads were assembled with the quasi-*de novo* assembler described by Prum et al. (2015). In the first step of this assembler, reads are mapped to conserved regions of the target loci with three distant species (*Arabidopsis thaliana*, *Billbergia nutans*, and *Carex lurida*) being selected as references from the set of species used by Buddenhagen et al. (2016) in the probe set design. The assembler uses a library of spaced kmers (k = 20), derived from the conserved sites of the alignments of the three reference species,

to determine which target locus a particular read is derived from. Preliminary candidate locus matches are identified if a minimum of 17 (out of 20) matches are found between a spaced kmer and the read. Then the read is compared to the reference sequence from which the kmer was derived and if 55 bases out of 100 bases surrounding the kmer match between the read and the kmer, the read is said to be a confirmed match. Approximate alignment position of reads mapped this way was estimated using the position of the spaced 20-mer. In the second step reads assembled in the first step are used to create a hash table of 60-mers that serve as references to extend the assembly into more variable flanking regions. The two assembly steps are used to traverse repeatedly the read files until no additional reads are mapped.

One consensus sequence with heterozygous sites coded as IUPAC ambiguity codes was produced for each species per orthologous locus. Unambiguous bases were called if no polymorphism was observed or if polymorphisms could be attributed to sequencing errors. Bases were called as N if coverage was below 10. In order to prevent cross contamination and potential sequencing errors in index reads, assembly contigs with less than 30 reads were discarded.

For orthology assessment, consensus sequences were grouped by locus and a distance matrix was generated with the pairwise distance between two sequences as the percent of 20-mers found in both sequences. Based on these distance matrices, sequences were clustered using the neighbor-joining algorithm (Saitou and Nei, 1987). If gene duplication, low coverage and contamination are absent, each locus should produce a single cluster with a single sequence per species. If more than one cluster was retrieved per locus, each cluster of orthologs was considered as a different locus and separated from the other(s). Clusters with less than 50% of the target species were discarded in order to reduce the effect of missing data. Pre-alignments were produced using MAFFT v.7.023b (Katoh and Standley, 2013) and subsequently trimmed following Prum et al. (2015) and Hamilton et al. (2016) criteria to generate the final nuclear alignments. All methods described in this section were performed by the Center for Anchored Phylogenomics.

Read Processing, Assembly, and Alignment of Plastid Loci

The raw data was assembled using CLC Genomics Workbench v.11.0 (https://www.qiagenbioinformatics.com/). A *de novo* assembly for each of the accessions was created, allowing for automatic word and bubble size, as well as an auto-detection of paired distances. To solve potential mis-assemblies or inconsistencies in the assemblies, readmapping files were extracted after mapping the reads back to the contigs. Plastid protein coding genes and ribosomal RNAs (rRNAs) were identified using BLAST and automatically aligned to a reference using the Python Workflow for 1kp Assemblies (written by Wesley K. Gerelle, University of British Columbia; https://github.com/wesleykg/1kp_workflow), combining a BLAST search (e-value = 1e−20) and an alignment of the hits back to the reference (MUSCLE, standard settings). The reference file was prepared by extracting sequences of 79

protein coding and four rRNA genes (a total of 83 loci) from the plastome of *Masdevallia coccinea* (NC_026541.1). The BLAST results in combination with the readmapping files were then used to extract the genes from the target species. Translocation between the plastid and mitochondrial genomes has been reported for genes of the *ndh* family in some Epidendroideae representatives (Lin et al., 2015); therefore, the complete *ndh* gene family (11 genes) was excluded from further analyses to prevent a possible mixture of mitochondrial and plastid copies for these genes. After exclusion of the *ndh* genes, 72 plastid loci were the targeted in our mining strategy. Sequences and BLAST results were visualized using AliView v.1.18.1 (http://ormbunkar.se/aliview/; Larsson, 2014), and readmapping files using Tablet v.1.17.08.17 (https://ics.hutton. ac.uk/tablet/). Single gene alignments were manually created using AliView v.1.18.1 and alignments were concatenated in Geneious v.11.1.5 (https://www.geneious.com).

Tree Reconstruction

Coalescent summary and site-based methods were selected for phylogenetic inference of the nuclear data set to accommodate potential gene-tree heterogeneity, for instance, caused by incomplete lineage sorting (Maddison, 1997; Degnan and Rosenberg, 2009), and because these methods are known to handle appropriately large numbers of loci and/or taxa (Chifman and Kubatko, 2014; Molloy and Warnow, 2017). Summary methods use gene trees as input for species tree estimation, so that gene trees are summarized into a species tree (Liu et al., 2009; Zhang et al., 2018). However, summary methods are sensitive to gene-tree estimation errors, for instance, those resulting from missing data or low number of informative sites per individual locus (Roch and Warnow, 2015). Due to these known caveats, and to confirm summary methods results, we also applied site-based methods which directly infer the species tree from site patterns in the alignments of the different loci (Chifman and Kubatko, 2014), thus circumventing the difficulties associated to gene tree estimation. Coalescence summary analyses were performed in ASTRAL-III (Zhang et al., 2018), estimating branch support through local posterior probabilities (Sayyari and Mirarab, 2016). Input nuclear gene trees for ASTRAL-III were generated using a workflow created in Geneious that runs in series maximum likelihood (ML) analyses on all loci within a folder with the implemented RAxML (Stamatakis, 2006) plugin, applying a GTR + Γ model as recommended in the RAxML manual. Twenty searches for the best ML tree were performed per locus and clade support was assessed with 1,000 bootstrap replicates. Node support was indicated on the best ML trees using a modification of the "applyRAxML2AllFilesInDirectory.pl" Perl script (https:// github.com/stamatak/standard-RAxML/tree/master/ usefulScripts). Gene trees were further manipulated to collapse nodes with bootstrap support (BS) < 33 (**Supplementary Material S2**), since this has been shown to increase accuracy of the species tree estimation in ASTRAL (Sayyari and Mirarab, 2016). Nodes were collapsed with the program TreeCollapseCL 4 (Emma Hodcroft, http://emmahodcroft.com/ TreeCollapseCL.html).

Coalescent site-based analyses were performed in SVDquartets (Chifman and Kubatko, 2014) as implemented in PAUP 4.0a 165 (Swofford, 2003). The input was a NEXUS file containing the concatenated nuclear matrix partitioned by locus. Up to 100,000 randomly sampled quartets were evaluated and 1,000 bootstrap replicates were performed.

Phylogenetic inference of plastid data was performed under a concatenated approach with maximum likelihood (ML). PartitionFinder2 v.2.1.1 (Lanfear et al., 2017) was used to determine the best-fit subset partitioning scheme of the concatenated matrix of all selected plastid loci (after excluding the *ndh* genes as explained above). One hundred independent searches for the best tree were run on the concatenated partitioned plastid matrix and node support was estimated through 1,000 bootstrap replicates on RAxML v.8.2.10 (Stamatakis, 2014).

Estimation of Phylogenetic Informativeness

Performance of plastid and nuclear loci, the latter divided into target and flaking non-target regions, in resolving a range of evolutionary depths within *Epidendrum* and subtribe Laeliinae was estimated with the phylogenetic informativeness method of Townsend (2007). Position of nuclear target and flaking non-target regions was determined on final alignments, after trimming ends and ambiguously aligned regions. Flanking regions falling outside the target region were considered non-target regions. Branch lengths of the topology obtained from the nuclear SVDQuartets analysis were optimized in RAxML, applying a GTR + Γ model and the combined data matrix partitioned by locus. The resulting phylogenetic tree was then made ultrametric, assigning time 0 for tree tips and time 1 to the root, in R v.3.5.0 with the function "chronopl" of the APE package (Paradis et al., 2004), setting lambda to 0.0. This modified ultrametric tree, along with a combined partitioned matrix, was uploaded to the PhyDesign web application (López-Giráldez and Townsend, 2011). Input partitions for this combined matrix corresponded to each of the plastid loci, nuclear target regions and nuclear non-target flaking regions. Substitution rates were estimated with the HyPhy program (Pond et al., 2005) applying a generalized time reversible (GTR) evolutionary model by inputting base frequencies and substitution rate matrix obtained from the analysis of the combined data set with Phyml (Guindon and Gascuel, 2003) in JModelTest2 v. 2.1.6 (Darriba et al., 2012). Net phylogenetic informativeness profiles were plotted for each individual plastid and nuclear locus and contrasted against the reference ultrametric tree. Additionally, maximum net phylogenetic informativeness (PImax) was recorded for each locus.

RESULTS

Nuclear Gene Capture

From the 517 loci included in the plant AHE kit, 316 were recovered with a single copy after the entire nuclear pipeline for our taxonomic sample. Additionally, eight loci had two copies and one had three copies, therefore the total nuclear data set was composed of 335 orthologs (**Supplementary Material S3**). Alignment length varied from 163 to 1,495 bp, with a mean of 581 bp (**Figure 1A**). Complete taxon sampling was achieved for 223 (67%) of the nuclear loci, 58 (17%) had one missing species, 17 (5%) had two and 37 (11%) had three or more missing species (**Figure 1B**). Within Laeliinae, the number of recovered loci per species ranged from 333 in *Arpophyllum giganteum* to 285 in *Epidendrum magnoliae*. As a result of our clustering step for orthology assessment, species can recover no copies, a single copy, or more than one copy. When considering the average number of recovered copies per species across all nuclear loci, representatives of Laeliinae ranged from 0.88 in *Broughtonia sanguinea* to 0.66 in *E. magnoliae* (**Figure 1C**).

The concatenated nuclear dataset had 194,841 sites of which 13.86% were variable (**Table 2**). In this matrix *P. cardiothallis* and *E. magnoliae* showed the highest proportion of missing data (i.e., proportion of bases called as N plus missing flanking regions; 12.5 and 9.5%, respectively), whereas all other species ranged from 6.8% in *Epidendrum mathewsii* to 1% in *Caularthron bicornutum*. Percentage of reads on target ranged from 1.8 in *E. magnoliae* to 12 in *Epidendrum nocturnum* (**Figure 1D**).

Mined Plastid Regions

A total of 68 plastid protein coding genes and four rRNAs (72 loci in total) were mined successfully for our taxon sampling. Length of individual plastid alignments varied widely, from 90 bp in the *petN* gene to 6,990 bp in the *ycf2* gene (**Figure 2A**). We were able to mine most plastid loci for the complete sample of taxa (66 out of 72); four loci were recovered for 22 species and the remaining two in 20 and 19 species, respectively (**Figure 2B**). The number of recovered loci per species range from 72 (in 16 of the 23 spp.) to 68 in *Epidendrum parkinsonianum* (**Figure 2C**).

The aligned length of the concatenated plastid dataset (**Supplementary Material S4**) was 63,421 bp, from which 2,610 bp (4.11%) were variable (**Table 2**). *E. magnoliae* and *Epidendrum octomerioides* had the highest proportion (13.5 and 12.5%, respectively) of missing data and the remaining species ranged from 7.1% (*Epidendrum summerhayesii*) to 0.1% (*Epidendrum longicaule*) (**Figure 2C**). In general, percentage of missing data increased with the percentage of reads on target (**Figure 2D** and **Supplementary Material S5**).

Phylogenetic Relationships Within Laeliinae and *Epidendrum*

No strongly supported incongruences were detected between the inference methods (SVDQuartets *vs.* ASTRAL) applied to the nuclear dataset (**Figure 4**). In contrast, six strongly supported incongruences were retrieved between the nuclear ASTRAL (**Supplementary Material S6**) and the plastid RAxML (**Supplementary Material S7**) trees involving the position of *Epidendrum sophronitoides*, *E. nocturnum*, *Epidendrum lacertinum*, *Epidendrum juergensenii*, *E. anisatum*, and *Epidendrum cusii*. When comparing the nuclear SVDQuartets
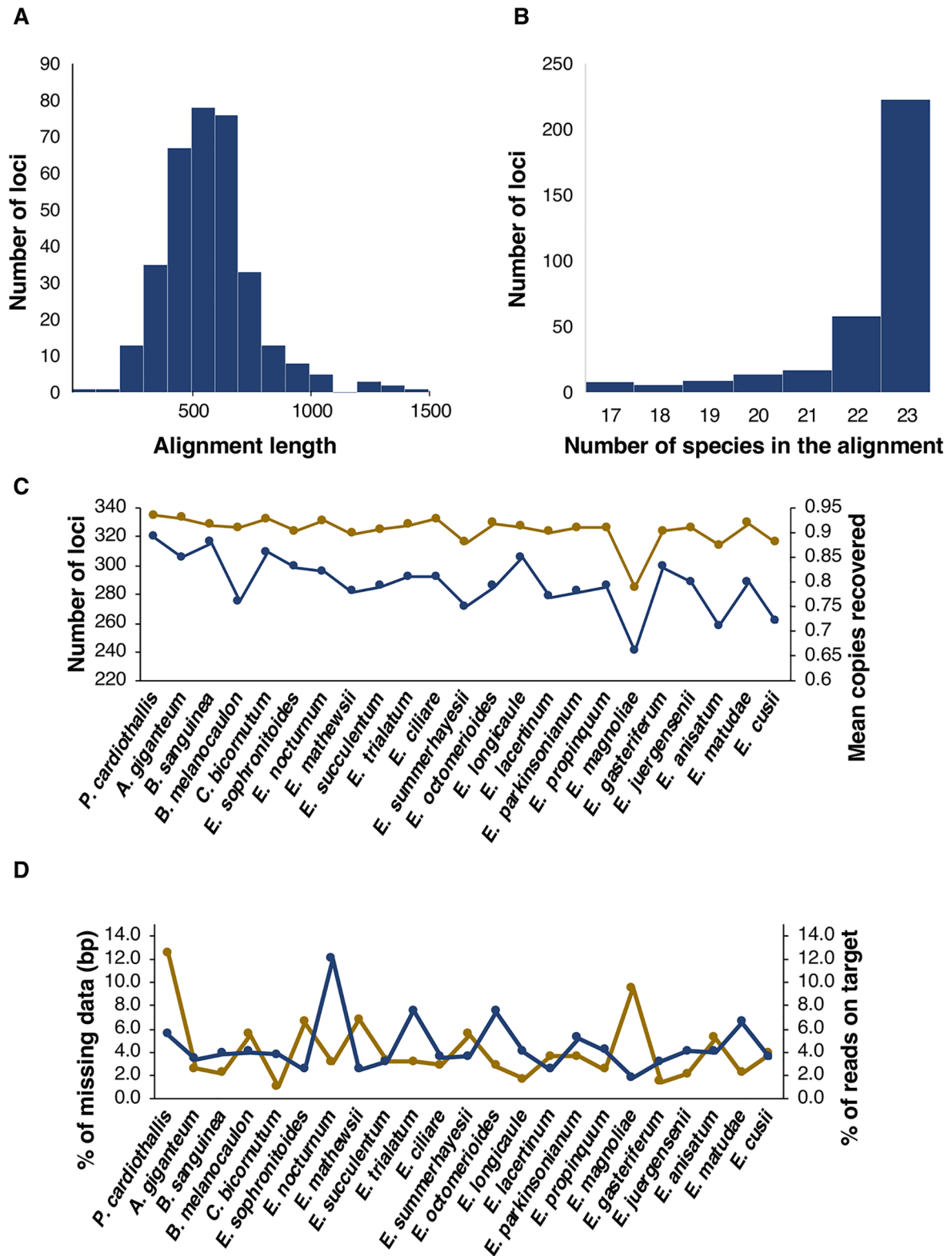
FIGURE 1 | Attributes of retrieved nuclear loci. **(A, B)** Histograms showing length and number of species in the alignments, respectively. **(C)** Number of loci (yellow points and lines) and mean copies recovered (blue points and lines) per species. **(D)** % of missing data per species, including bases called as N plus missing flanking regions of loci, in terms of number of base pairs (bp, yellow points and lines) and % of reads on target per species (blue points and lines).

TABLE 2 | Analyzed dataset characteristics.

	Nuclear	**Plastid**	**Combined**
# of retrieved loci	335	72	407
Alignments length range (bp)	163-1495	90-6990	163-6990
# (%) alignments with 23 spp.	223 (66.56)	66 (91.66)	289 (71)
# total sites	194,841	63,421	258,262
# (%) variable sites	27,006 (13.86)	2610 (4.11)	29,616 (11.46)

(**Supplementary Material S8**) and the plastid RAxML tree, only three of the previously mentioned incongruences were maintained as strongly supported, including the relationships of *E. lacertinum*, *E. anisatum*, and *E. cusii* (**Figure 4**).

Due to the higher congruence between the nuclear SVDQuartets and plastid RAxML analyses (**Figure 4**), phylogenetic relationships will be described based on the nuclear SVDQuartets tree (**Figure 3A**), where all but the already indicated relationships received BS > 85. Within subtribe Laeliinae, *A. giganteum* was recovered as sister to all other species, followed by a grade consisting of *B. sanguinea*, *Barkeria melanocaulon*, and *C. bicornutum*. The genus *Epidendrum* was found to be monophyletic and, within it, two main clades were recovered. One consists of a sister pair in which one clade includes *E. sophronitoides* sister to *E. nocturnum* (BS = 23) and the other includes *E. mathewsii*, *Epidendrum succulentum*, and *Epidendrum trialatum* as successive sisters
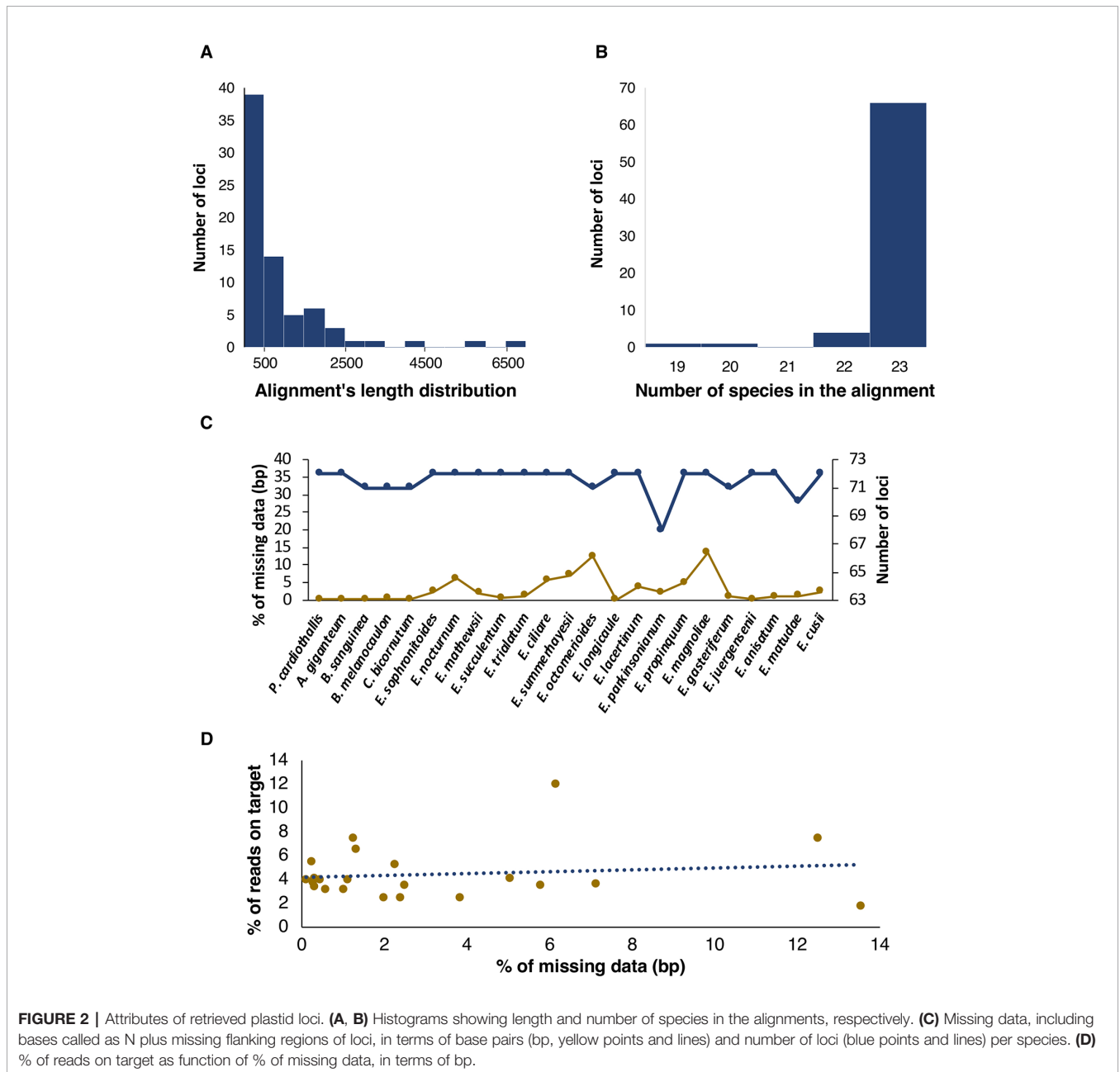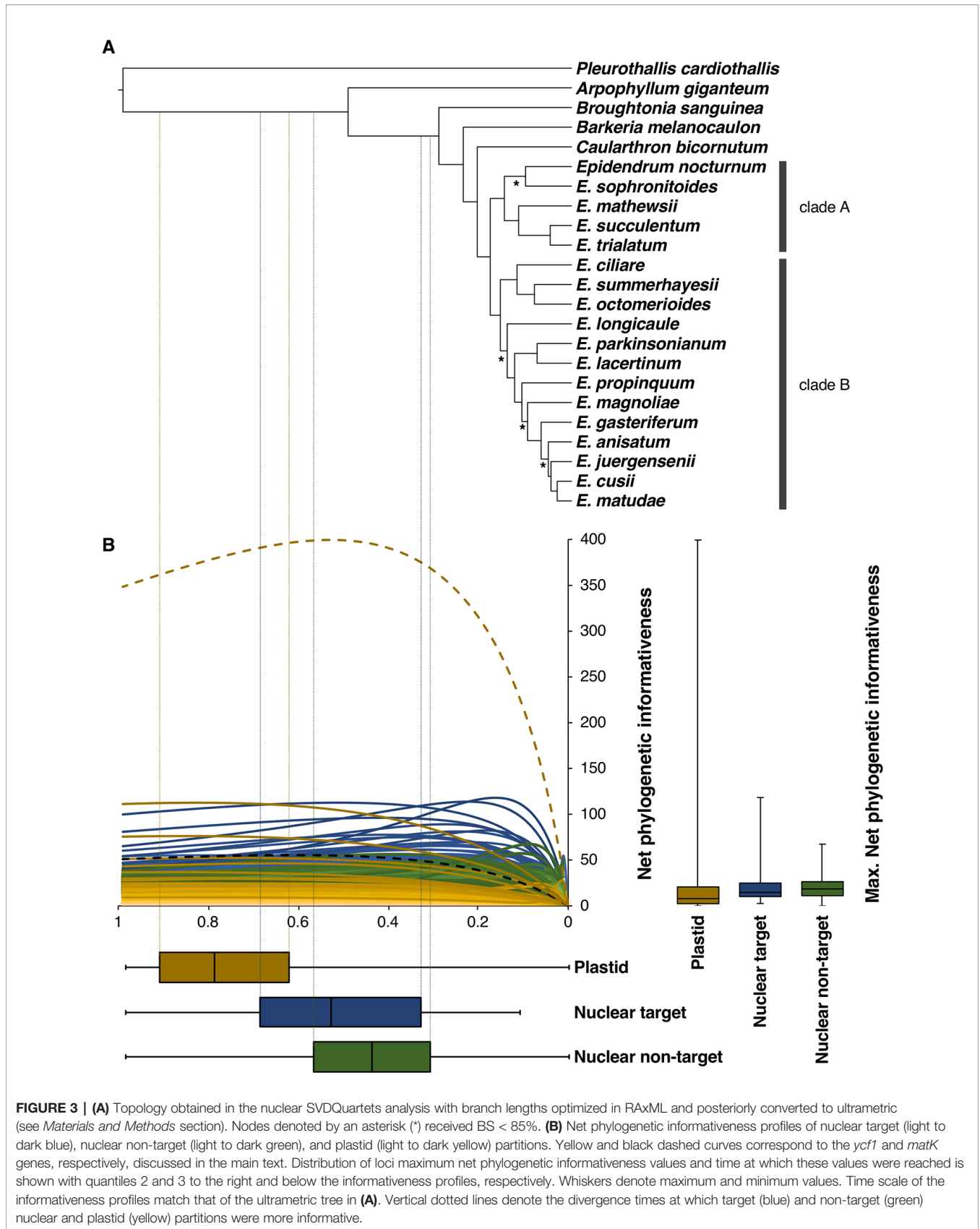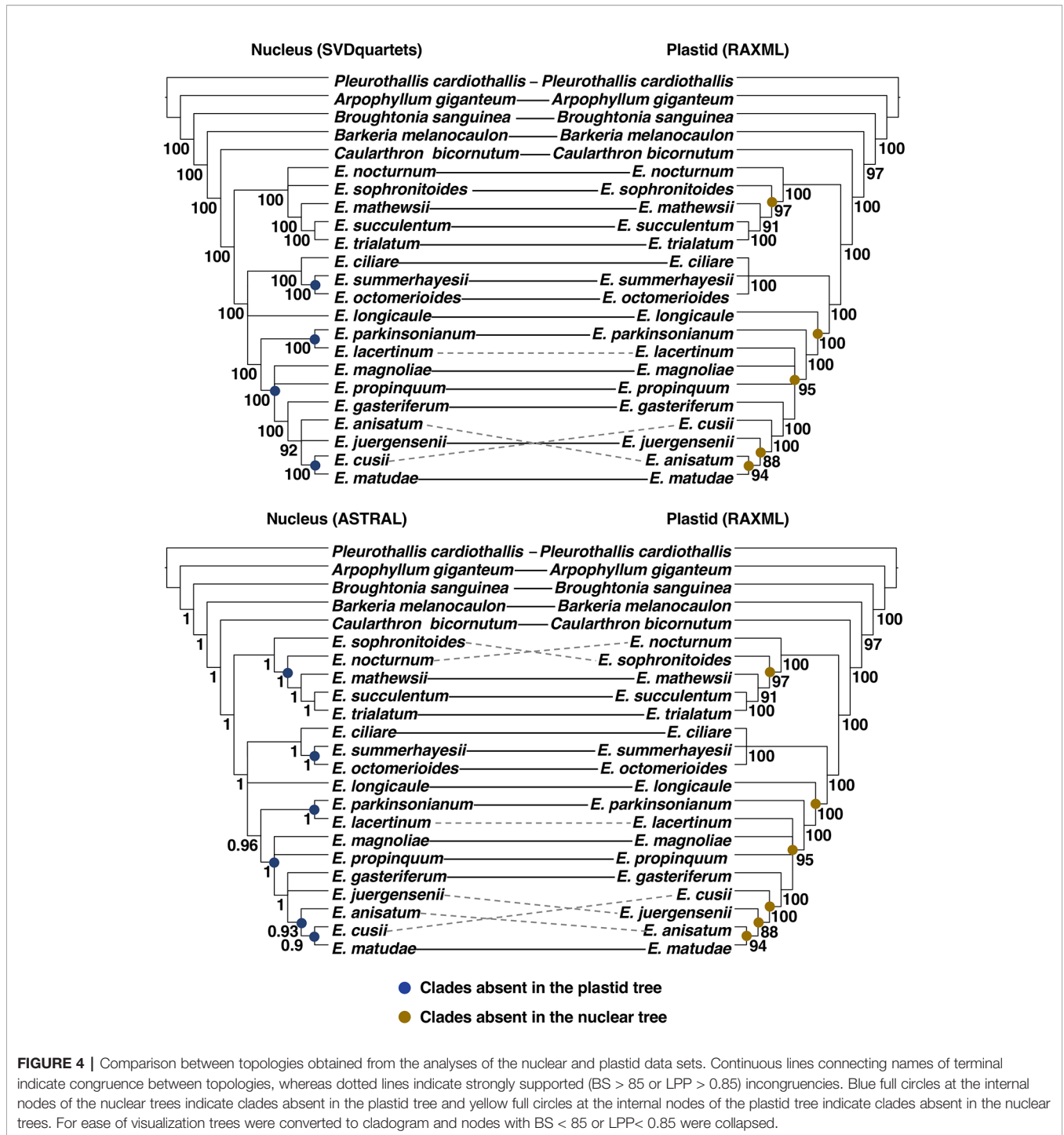

FIGURE 2 | Attributes of retrieved plastid loci. **(A, B)** Histograms showing length and number of species in the alignments, respectively. **(C)** Missing data, including bases called as N plus missing flanking regions of loci, in terms of base pairs (bp, yellow points and lines) and number of loci (blue points and lines) per species. **(D)** % of reads on target as function of % of missing data, in terms of bp.

FIGURE 3 | (A) Topology obtained in the nuclear SVDQuartets analysis with branch lengths optimized in RAxML and posteriorly converted to ultrametric (see *Materials and Methods* section). Nodes denoted by an asterisk (*) received BS < 85%. **(B)** Net phylogenetic informativeness profiles of nuclear target (light to dark blue), nuclear non-target (light to dark green), and plastid (light to dark yellow) partitions. Yellow and black dashed curves correspond to the *ycf1* and *matK* genes, respectively, discussed in the main text. Distribution of loci maximum net phylogenetic informativeness values and time at which these values were reached is shown with quantiles 2 and 3 to the right and below the informativeness profiles, respectively. Whiskers denote maximum and minimum values. Time scale of the informativeness profiles match that of the ultrametric tree in **(A)**. Vertical dotted lines denote the divergence times at which target (blue) and non-target (green) nuclear and plastid (yellow) partitions were more informative.

FIGURE 4 | Comparison between topologies obtained from the analyses of the nuclear and plastid data sets. Continuous lines connecting names of terminal indicate congruence between topologies, whereas dotted lines indicate strongly supported (BS > 85 or LPP > 0.85) incongruencies. Blue full circles at the internal nodes of the nuclear trees indicate clades absent in the plastid tree and yellow full circles at the internal nodes of the plastid tree indicate clades absent in the nuclear trees. For ease of visualization trees were converted to cladogram and nodes with BS < 85 or LPP< 0.85 were collapsed.

(clade A; **Figure 3A**). The other clade consists of a sister pair in which one lineage contains *Epidendrum ciliare* as sister to the clade of *E. summerhayesii* and *E. octomerioides*, and the other lineage contains *E. longicaule* (BS = 49) as sister to a clade where [*E. lacertinum*–*E. parkinsonianum*] are sister to a grade of *Epidendrum propinquum, E. magnoliae* (BS = 52), *Epidendrum gasteriferum, E. anisatum, E. juergensenii* (BS = 51), and *Epidendrum matudae* sister to *E. cusii* (clade B; **Figure 3A**).

Phylogenetic Informativeness of Nuclear and Plastid Loci

Net phylogenetic informativeness widely varied across loci (**Figure 3**). Nuclear target and flanking non-target regions showed profiles with steep increases corresponding to the shallowest divergences and gradual decreases towards the root, as well as profiles with rather flat curves. With few exceptions, plastid loci showed lower and mostly flat curves, having a gradual

TABLE 3 | Previous studies applying the plant anchored hybrid enrichment (AHE) method (Buddenhagen et al., 2016) sorted by divergence times between the focal group and the closest set of reference species used in the kit design.

Study	Focal group (family, order)	Closest set of reference species (family, order)	Divergence time (Ma)	# retrieved loci*
Fragoso-Martínez et al. (2017)	*Salvia* subg. Calosphace (Lamiaceae, Lamiales)	*Mimulus guttatus* (Phrymaceae, Lamiales)	40.3[1]	448
Kriebel et al. (2019)	*Salvia* (Lamiaceae, Lamiales)	*Mimulus guttatus* (Phrymaceae, Lamiales)	40.3[1]	498
Shahi Shavvon et al. (2017)	*Oxytropis* (Fabaceae, Fabales)	*Medicago truncatula* and *Glycine max* (Fabaceae, Fabales)	68[2]	527
Léveillé-Bourret et al. (2018)	Cariceae–Dulichieae–Scirpeae clade (Cyperaceae, Poales)	*Carex lurida, C. tenax,* and *Rhynchospora chalarocephala* (Cyperaceae, Poales)	88[2]	462
Crowl et al. (2019)	*Quercus* (Fagaceae, Fagales)	*Cucumis sativus* (Cucurbitaceae, Cucurbitales)	109.1[1]	493
This study	*Epidendrum* (Orchidaceae, Asparagales)	*Phoenix dactylifera* (Arecaceae, Arecales)	114.6[1]	335
Mitchell et al. (2017)	*Protea* (Proteaceae, Proteales)	*Nelumbo nucifera* (Nelumbonaceae, Proteales)	117.4[1]	498
Cardillo et al. (2017)	*Hakea* (Proteaceae, Proteales)	*Nelumbo nucifera* (Nelumbonaceae, Proteales)	117.4[1]	450
Wanke et al. (2017)	*Aristolochia* subsec. *Pentandrae* (Aristolochiaceae, Piperales)	*Amborella trichopoda* (Amborellaceae, Amborellales)	139.4[1]	233

[1]*Family divergence times taken from Magallón et al. (2015).*
[2]*Oldest crown-family divergence time reported in http://www.mobot.org/MOBOT/research/APweb.*
* *including paralogs.*

increase and an attenuated decrease towards the root. Maximum net phylogenetic informativeness values were in general higher for nuclear partitions, both target and non-target regions (median 14.71 and 18.56, respectively), compared to plastid loci (median 7.98). However, the *ycf1* gene strongly deviated from this general pattern showing 399.45 of maximum net phylogenetic informativeness. When the length of the markers is considered (per site phylogenetic informativeness), the *ycf1* gene is not only the plastid marker with the highest per site phylogenetic informativeness, but it also surpassed the informativeness of 40 nuclear partitions (data not shown). Most nuclear partitions reached their maximum net phylogenetic informativeness closer to the present, with median value of 0.44 for flanking non-coding regions and 0.53 for target regions, compared to plastid markers that had median value of 0.79 (**Figure 3B**).

DISCUSSION

Performance of Nuclear Hybrid Enrichment

Our study is the first one to apply the Angiosperm v.1 probe kit to the orchid genus *Epidendrum*. A substantial proportion (ca. 63%) of the original targeted nuclear loci could be captured and used for phylogenetic inference. To our knowledge, eight previous studies have applied this capture kit to angiosperm lineages (**Table 3**). A general trend is observed in these studies, where larger evolutionary distance of studied taxa to the closest reference lineage used for kit design results in reduced loci recovery (including paralogs; **Supplementary Material S9**). *Epidendrum* diverged ~115 mya from the closest reference species, *Phoenix dactylifera* (Arecaceae, Arecales; Magallón et al., 2015). Two previous studies in the family Proteaceae (Cardillo et al., 2017; Mitchell et al., 2017) showed a slightly larger evolutionary distance between their focal groups and their respective reference species (~117 mya) than *Epidendrum*. However, those studies recovered a larger number of loci (498

and 450, respectively). This deviation from the general trend could be attributed to several factors, including potential loci conservation within the order Proteales since a reference species within this order was available, differences in genomic DNA (gDNA) isolates quality, and/or sequencing depth. Another possible explanation is the retrieval of a larger number of paralogs that sum up to the total number of retrieved loci. However, information about how many multicopy loci were recovered has not been consistently reported in previous studies. In *Epidendrum*, only eight loci were recovered as multicopy and the mean copies recovered per species was less than 0.88, suggesting that duplication in this set of loci is not a common process in this lineage. Regardless, the existence of a relatively large evolutionary distance (~115 mya) between *Epidendrum* and its closest monocot reference species supports the claims of Buddenhagen et al. (2016) and Wanke et al. (2017) that the Angiosperm v.1 kit is universally applicable among angiosperms.

Efficiency in terms of number of captured loci was relatively homogeneous across targeted species (335–314), except for *E. magnoliae* from which only 285 loci were recovered. Failure of capturing some loci for this species could be explained by the low percentage (the lowest among all the species) of reads on target, also reflected in its high percentage of missing data (see below; **Figure 1D**).

Regarding sequence quality, all species from subtribe Laeliinae, except *E. magnoliae* (9.5%), showed a relatively low percentage of missing data (6.8-1%), whereas the outgroup species *P. cardiothallis* from subtribe Pleurothallidinae showed a substantially higher proportion of missing data (12.5%). There is no evidence that genome sizes in Pleurothallidinae are larger than in Laeliinae (Leitch et al., 2019) and polyploidy has not been reported in the former subtribe (Felix and Guerra, 2010). Therefore, larger genome size or higher ploidy level seem unlikely to explain the lower enrichment efficiency for *P. cardiothallis*. Although genomic DNA of *P. cardiothallis* met similar quality standards as other species analyzed and was subjected to the same wet lab and bioinformatic processing, there may be lineage-specific variation in capture efficiency.

Plastid Loci Mining Success

All four rRNAs and the 68 selected plastid protein encoding genes were successfully mined from off-target reads. Regarding the *ndh* gene family, our results agree with previous studies (e.g., Kim et al., 2015; Lin et al., 2015; Kim et al., 2017; Niu et al., 2017; Zhitao et al., 2017; Dong et al., 2018) which have shown that this set of genes are commonly pseudogenized, lost, or translocated in orchids, since most *ndh* genes for most of the target species were recovered most likely as pseudogenes and very rarely as functional genes. The *ndhA*, *ndhG*, *ndhH*, and *ndhI* genes were missing in one or several species; however, this cannot be assumed with confidence to be the consequence of gene loss, because there is a possibility that they were simply not recovered among the off-target reads. A comparative analysis of full *Epidendrum* plastomes could help to elucidate this in the future.

Our mining results are promising considering that the entire libraries were enriched with the target AHE nuclear markers and that a non-*Epidendrum* reference species (*M. coccinea* of subtribe Pleurothallidinae) was employed for extracting the plastid orthologs. We aimed at extracting plastid exons only, which are generally better conserved than non-coding regions such as introns or intergenic spacers (Shaw et al., 2005). Although not performed here, orchid studies aiming to extract the more variable non-coding plastid regions might be more challenging due to the expected higher sequence divergence between target and reference species. A way to overcome this would be using closer relatives as reference species. However, for *Epidendrum* this will only be possible when complete plastomes become available. To date, plastomes of 163 orchid species representing 46 genera have been sequenced (NCBI database; accessed June, 2019), therefore new orchid studies that apply our plastome mining approach will have a wealth of potential publicly available reference species. Another approach might be to perform further scaffolding to extract flanking non-coding regions, a strategy that we will follow in an upcoming publication.

It is remarkable that extremely long genes, such as *ycf2* (6,990 aligned bp), could be assembled and used for phylogenetic inference. Furthermore, most individual loci had a full taxon representation and the number of loci recovered per species was rather high (>70), except for *E. parkinsonianum* from which 68 loci were recovered. In general, percentage of missing data increased with the percentage of reads on target, or in other words, missing data increased if fewer non-target reads were available for assembling plastome regions.

Phylogenetic Utility of Nuclear and Plastid Loci

Relationships here obtained among the included subtribe Laeliinae genera are mostly in agreement with those recovered by Hágsater and Soto-Arenas (2005) and van den Berg et al. (2009), but with stronger statistical support. Previous studies, all of them based on a few Sanger-sequenced loci, often failed to provide strong statistical support for intergeneric relationships, as well as for many internal relationships of *Epidendrum*. Our study overcame

this limitation, albeit on a limited taxon sampling, by analyzing the largest set of nuclear and plastid loci generated to date for subtribe Laeliinae, with few nodes within *Epidendrum* receiving weak statistical support (BS < 85, **Figure 3**). *Caularthron* is sister to *Epidendrum* with strong support, in agreement with the results of van den Berg et al. (2009). Likewise, the two major clades recovered within *Epidendrum* in our analyses (marked as A and B in **Figure 3**) more or less correspond to the two main clades found by Hágsater and Soto-Arenas (2005), except for the placement in our trees of *E. nocturnum* in clade A instead of clade B. It is also worth noting that our data recover a strongly-supported clade of mostly Mexican species (*E. parkinsonianum-E. lacertinum* to *E. matudae*), which did not group but formed a grade at the base of clade B in the analysis of Hágsater and Soto-Arenas (2005). The five species of the *E. anisatum* group included here form a strongly-supported clade, in agreement with the various morphological and eco-geographical features they share (Quiroga-González, 2017), although the internal relationships in this clade include one incongruence between our plastid and nuclear trees, as discussed below (**Figure 4**).

Most of the resulting phylogenetic relationships were consistent across different analyses (summary and site-based) or data sources (plastid *vs.* nuclear). However, some strongly supported incongruences were detected when comparing the nuclear and plastid phylogenies. The higher number of strongly supported incongruences between the nuclear ASTRAL and the plastid tree (six incongruences), compared to those found between the nuclear SVDQuartets and the plastid tree (three incongruences), could be attributed to potential noise introduced by gene tree estimation error associated to summary methods (Roch and Warnow, 2015). Measures known to increase accuracy of ASTRAL analyses were herein applied, such as using input trees where nodes with BS < 33 are collapsed. However, nuclear loci varied widely in their aligned length, and differences in their number of potentially informative sites and their proportion of missing data could lead to estimation error in some short or highly incomplete alignments.

Because site-based methods are known to avoid difficulties associated to gene tree estimation, incongruences between the obtained nuclear and plastid hypotheses will be further discussed based on the SVDQuartets and RAxML trees. In these trees, incongruences involved recent divergences within *Epidendrum* corresponding to the alternative positions of *E. lacertinum*, *E. anisatum,* and *E. cusii*. When focused, for instance, on the alternative relationship of *E. matudae* as sister to *E. anisatum* in the plastid tree or as sister to *E. cusii* in the nuclear trees, we found that 46 (out of 335) nuclear gene trees recovered the alternative sister relationship between *E. matudae* and *E. anisatum* of the plastid tree. Although only three of these 46 nuclear gene trees recovered this relationship with high support (BS > 85), this indicates that the phylogenetic signal of the plastid hypothesis is shared with a small proportion of nuclear gene trees.

At these shallow evolutionary levels, processes such as incomplete lineage sorting may play a role if lineages are associated to deeper rapid radiations (Degnan and Rosenberg,

2009). A rapid radiation has not been formally tested in *Epidendrum*, but short internodes are characteristic of its phylogenetic tree (e.g., Hágsater and Soto-Arenas, 2005). Alternatively, branch length heterogeneity (in coalescent units) within *Epidendrum* observed in the nuclear ASTRAL tree may suggest potential changes of the effective population size across the evolutionary history of this genus, with short branches indicating increased gene tree discordance (Degnan and Rosenberg, 2009). A further potential source of conflict between biparentally and uniparentally inherited DNA data is hybridization and introgression. As revisited by Pinheiro and Cozzolino (2013), hybridization likely is a key process shaping the diversification of some groups of *Epidendrum*. Similar patterns of strongly-supported incongruence between nuclear and plastid partitions at shallow phylogenetic levels have been observed within most of the genera of Laeliinae that have been investigated using Sanger sequencing, e.g., *Cattleya* (van den Berg et al., 2009; van den Berg, 2014), *Encyclia* (Leopardi-Verde et al., 2017), and *Laelia* (Peraza-Flores et al., 2016). This seems to be caused by the combination of few genetic incompatibility barriers and very low variation in chromosome numbers across the subtribe, including *Epidendrum* (Felix and Guerra, 2010; De Assis et al., 2013), and confirmed by the large number of natural hybrids (including intergenerics) reported in the subtribe (Adams and Anderson, 1958; van den Berg, 2014).

Phylogenetic informativeness profiles allowed us to confirm a general trend that has been widely documented in vascular plant evolutionary studies, i.e., that nuclear data generally provides higher informativeness than plastid data, and that nuclear data better informs more recent evolutionary events than plastid data (e.g., Clegg et al., 1994; Sang, 2002; Duarte et al., 2010; Zimmer and Wen, 2012; Rothfels et al., 2013; Salas-Leiva et al., 2013; Soltis et al., 2013; Lu et al., 2014). Nuclear target *vs.* non-target regions contributed with similar levels of phylogenetic utility; however, nuclear non-target regions informed a wider range of more recent divergences compared to nuclear target regions.

Phylogenetic Utility of the *ycf1* Gene at Low Taxonomic Levels

One remarkable exception to the general trend observed between plastid and nuclear markers is the *ycf1* gene, which by far surpassed the net informativeness of all other loci, either plastid or nuclear. Although until recently the possible function of *ycf1* was unknown (its gene abbreviation stands for hypothetical chloroplast open reading frame 1), mounting evidence indicates that it is part of an inner membrane envelope translocon, or TIC, i.e., a complex of proteins associated with the translocation of polypeptides across the inner membrane of the chloroplast (de Vries et al., 2015; Nakai, 2015). The *ycf1* gene alignment was the second longest of all the analyzed loci, only surpassed by the *ycf2* gene (see also Wicke et al., 2011), however, phylogenetic informativeness of *ycf2* is substantially lower than that of *ycf1*. To account for the differences in alignment length we calculated the per site phylogenetic informativeness (data not shown), finding that informativeness provided by the *ycf1* gene surpasses that of all plastid and several (40) nuclear partitions.

The unusually high phylogenetic utility of the plastid *ycf1* gene for shallow taxonomic levels in orchids was first highlighted by Neubig et al. (2009), who found two portions close to the 5' and 3' ends of this gene to be more variable than other genes commonly used for phylogenetic inference, such as *matK* and *rbcL*. Subsequently, this marker was used to inform phylogenetic relationships at low taxonomic levels in the orchid subtribes Oncidiinae (Chase et al., 2009; Neubig et al., 2012) and Maxillariinae (Arévalo et al., 2015), and partial (Gernandt et al., 2009; Drew and Sytsma, 2011; Drew and Sytsma, 2012; Drew and Sytsma, 2013; Shi et al., 2013) or entire (Parks et al., 2009) exon sequences of this gene have been used successfully to estimate phylogenetic relationships at intermediate to low taxonomic levels in other plant lineages. More recently, Dong et al. (2015) identified two segments of this gene as promising DNA barcodes for plants. Roma et al. (2018), in their study of the orchid genus *Ophrys*, attributed the unusually high sequence divergence of *ycf1* relative to other genes to its location in the junction of the inverted repeat and the small single copy regions, which additionally cause a high sequence length variation and potential pseudogenization (see also Jheng et al., 2012). Although our mining strategy does not provide information about gene location, it allowed us to compare the informativeness of complete sequences of most known plastid genes and confirm that the *ycf1* gene not only provides greater phylogenetic resolution than the commonly used *matK* and *rbcL* genes, as pointed out by Neubig et al. (2009) for orchids in general, but it is also more informative than any other plastid gene and several nuclear loci in *Epidendrum*. None of the species analyzed here showed signs of pseudogenization or loss of the *ycf1* gene. After its proposed origin before the diversification of green plants (Wicke et al., 2011), the *ycf1* gene is known to be absent in the Poaceae (Dong et al., 2015) and a few holoparasitic (*Orobanche purpurea*, Orobanchaceae) and photosynthetic eudicots (*Vaccinium macrocarpon*, Ericaceae, and *Erodium* spp., Geraniaceae; see de Vries et al., 2015), confirming it as an excellent candidate for phylogenetic inference at low taxonomic levels not only in *Epidendrum*, but in many other angiosperm lineages.

Integrating information from the chloroplast and nuclear genomes increased the range of evolutionary depths that could be estimated and contrasted in our study. It is worth noticing that, except for a few outliers, none of the analyzed partitions (target or non-target nuclear regions or plastid loci) had its maximum informativeness within the diversification time of subtribe Laeliinae. This is to be expected, because the inclusion of non-coding regions in our analyses was moderate. As explained in the *Materials and Methods* section, we did not aim to mine non-coding plastid regions, hence the only sources of non-coding data in our analyses were the adjacent non-target regions of the nuclear loci. A future study will focus on resolving shallower divergences among an expanded taxon sample of *Epidendrum*, by mining plastid non-coding regions to increase the resolution power at fine evolutionary levels using a newly sequenced *Epidendrum* plastome as reference.

CONCLUSION

Our study demonstrated for the first time the technical implementation of the Angiosperm v.1 probe kit (Buddenhagen et al., 2016) to the orchid genus *Epidendrum*, supporting the universal applicability of this kit across angiosperms. Moreover, we confirmed the feasibility of mining plastome loci from off-target reads when using this kit, generating complementary sequence data of uniparental inheritance at no extra sequencing cost. Our analyses are in general congruent across methods and data sources. The few strongly supported incongruences detected suggest the possibility of incomplete lineage sorting or potential hybridization and introgression events among closely related species. Our ample survey of the phylogenetic utility of coding nuclear and plastid loci in *Epidendrum* allowed us to identify the *ycf1* gene as a strongly useful locus for resolving relationships at low taxonomic levels, surpassing the net informativeness of every other plastid and nuclear loci analyzed. Hyb-seq approaches (Weitemier et al., 2014) thus appear as a promising option for generating informative data sets derived from different genome compartments. Although our taxonomic sample is too small to attempt to draw conclusions about organismal and evolutionary aspects of the genus as a whole, our results provide a foundation for a much more inclusive sampling strategy aimed at covering the structural diversity of the genus throughout the Neotropics in a forthcoming phase of our research program.

DATA AVAILABILITY STATEMENT

Raw data associated to this article can be found under the NCBI Sequence Read Archive BioProject PRJNA589279.

AUTHOR CONTRIBUTIONS

GS, EH, CM, and SW conceived and designed the study. GS, EH, and CB designed the taxon sampling and collected or provided the samples. CM, EL, and AL performed the laboratory work. AL and MJ performed the bioinformatic process of nuclear and plastid data, respectively. CM performed the phylogenetic and informativeness analyses. CM drafted the manuscript and MJ, EH, SM, SW, CB, EL, AL, and GS proof read and approved the final manuscript.

REFERENCES

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019. 01761/full#supplementary-material

SUPPLEMENTARY MATERIAL S1 | A .zip file containing the code, scripts and a description file (README.txt) used to generate the nuclear alignments from raw data.

SUPPLEMENTARY MATERIAL S2 | Inferred nuclear gene trees.

SUPPLEMENTARY MATERIAL S3 | A concatenated matrix of all nuclear loci partitioned by locus.

SUPPLEMENTARY MATERIAL S4 | A concatenated matrix of all plastid loci partitioned by locus.

SUPPLEMENTARY MATERIAL S5 | Table with number of raw reads and percentage of reads on target for each sampled species.

SUPPLEMENTARY MATERIAL S6 | Resulting nuclear ASTRAL tree.

SUPPLEMENTARY MATERIAL S7 | Resulting plastid RAXML tree.

SUPPLEMENTARY MATERIAL S8 | Resulting nuclear SVDQuartets tree.

SUPPLEMENTARY MATERIAL S9 | A figure comparing this and previous plant Anchored Hybrid Enrichment studies.

Adams, H., and Anderson, E. (1958). A conspectus of hybridization in the Orchidaceae. *Evolution* 12, 512–518. doi: 10.2307/2405962

Arévalo, R., Carnevali Fernández-Concha, G., and M. Cameron, K. (2015). Three new species of mormolyca (Orchidaceae: Maxillariinae) with an updated molecular phylogenetic analysis. *Syst. Bot.* 40, 692–705

Bogarín, D., Pérez-Escobar, O. A., Groenenberg, D., Holland, S. D., Karremans, A. P., Lemmon, E. M., et al. (2018). Anchored hybrid enrichment generated nuclear, plastid and mitochondrial markers resolve the *Lepanthes horrida* (Orchidaceae: Pleurothallidinae) species complex. *Mol. Phylogenet. Evol.* 129, 27–47. doi: 10.1016/j.ympev.2018.07.014

Buddenhagen, C., Lemmon, A. R., Lemmon, E. M., Bruhl, J., Cappa, J., Clement, W. L., et al. (2016). Anchored phylogenomics of angiosperms I: assessing the robustness of phylogenetic estimates. *bioRxiv*, 086298. doi: 10.1101/086298

Cardillo, M., Weston, P. H., Reynolds, Z. K. M., Olde, P. M., Mast, A. R., Lemmon, E. M., et al. (2017). The phylogeny and biogeography of *Hakea* (Proteaceae) reveals the role of biome shifts in a continental plant radiation. *Evolution* 71, 1928–1943. doi: 10.1111/evo.13276

Chase, M. W., Cameron, K. M., Freudenstein, J. V., Pridgeon, A. M., Salazar, G. A., van den Berg, C., et al. (2015). An updated classification of Orchidaceae. *Bot. J. Linn. Soc* 177, 151–174. doi: 10.1111/boj.12234

Chase, M. W., Williams, N. H., de Faria, A. D., Neubig, K. M., Amaral, M., do, C. E., and Whitten, W. M. (2009). Floral convergence in Oncidiinae (Cymbidieae; Orchidaceae): an expanded concept of Gomesa and a new genus Nohawilliamsia. *Ann. Bot.* 104, 387–402. doi: 10.1093/aob/mcp067

Chifman, J., and Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30, 3317–3324. doi: 10.1093/bioinformatics/btu530

Clegg, M. T., Gaut, B. S., Learn, G. H., and Morton, B. R. (1994). Rates and patterns of chloroplast DNA evolution. *Proc. Natl. Acad. Sci. U.S.A.* 91, 6795–6801. doi: 10.1073/pnas.91.15.6795

Crowl, A. A., Manos, P. S., McVay, J. D., Lemmon, A. R., Lemmon, E. M., and Hipp, A. L. (2019). Uncovering the genomic signature of ancient introgression between white oak lineages (*Quercus*). *New Phytol.* doi: 10.1111/nph.15842

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Meth.* 9, 772–772. doi: 10.1038/nmeth.2109

De Assis, F. N. M., Souza, B. C. Q., Medeiros-Neto, E., Pinheiro, F., Silva, A. E. B., and Felix, L. P. (2013). Karyology of the genus *Epidendrum* (Orchidaceae: Laeliinae) with emphasis on subgenus *Amphiglottium* and chromosome number variability in *Epidendrum secundum*. *Bot. J. Linn. Soc* 172, 329–344. doi: 10.1111/boj.12045

de Sousa, F., Bertrand, Y. J. K., Nylinder, S., Oxelman, B., Eriksson, J. S., and Pfeil, B. E. (2014). Phylogenetic properties of 50 nuclear loci in *Medicago* (Leguminosae) generated using multiplexed sequence capture and next-generation sequencing. *PloS One* 9, e109704. doi: 10.1371/journal.pone.0109704

de Vries, J., Sousa, F. L., Bölter, B., Soll, J., and Gould, S. B. (2015). *YCF1*: A green TIC? *Plant Cell* 27, 1827. doi: 10.1105/tpc.114.135541

Degnan, J. H., and Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* 24, 332–340. doi: 10.1016/j.tree.2009.01.009

Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S., et al. (2015). *YCF1*, the most promising plastid DNA barcode of land plants. *Sci. Rep.* 5, 8348. doi: 10.1038/srep08348

Dong, W.-L., Wang, R.-N., Zhang, N.-Y., Fan, W.-B., Fang, M.-F., and Li, Z.-H. (2018). Molecular evolution of chloroplast genomes of orchid species: Insights into phylogenetic relationship and adaptive evolution. *Int. J. Mol. Sci.* 19, 716. doi: 10.3390/ijms19030716

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.

Drew, B. T., and Sytsma, K. J. (2011). Testing the monophyly and placement of *Lepechinia* in the tribe Mentheae (Lamiaceae). *Syst. Bot.* 36, 1038–1049. doi: 10.1600/036364411X605047

Drew, B. T., and Sytsma, K. J. (2012). Phylogenetics, biogeography, and staminal evolution in the tribe Mentheae (Lamiaceae). *Am. J. Bot.* 99, 933–953. doi: 10.3732/ajb.1100549

Drew, B. T., and Sytsma, K. J. (2013). The South American radiation of *Lepechinia* (Lamiaceae): phylogenetics, divergence times and evolution of dioecy. *Bot. J. Linn. Soc* 171, 171–190. doi: 10.1111/j.1095-8339.2012.01325.x

Duarte, J., Wall, P. K., Edger, P., Landherr, L., Ma, H., Pires, J. C., et al. (2010). Identification of shared single-copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evol. Biol.* 10, 61. doi: 10.1186/1471-2148-10-61

Felix, L. P., and Guerra, M. (2010). Variation in chromosome number and the basic number of subfamily Epidendroideae (Orchidaceae). *Bot. J. Linn. Soc* 163, 234–278. doi: 10.1111/j.1095-8339.2010.01059.x

Fragoso-Martínez, I., Salazar, G. A., Martínez-Gordillo, M., Magallón, S., Sánchez-Reyes, L., Moriarty Lemmon, E., et al. (2017). A pilot study applying the plant Anchored Hybrid Enrichment method to New World sages (*Salvia* subgenus *Calosphace*; Lamiaceae). 25th Anniv. Issue. *Mol. Phylogenet. Evol.* 117, 124–134. doi: 10.1016/j.ympev.2017.02.006

Gernandt, D. S., Hernández-León, S., Salgado-Hernández, E., and Pérez de la Rosa, P. (2009). Phylogenetic relationships of *Pinus* subsection *Ponderosae* inferred from rapidly evolving cpDNA regions. *Syst. Bot.* 34, 481–491. 10.1600/036364409789271290

Grover, C. E., Gallagher, J. P., Jareczek, J. J., Page, J. T., Udall, J. A., Gore, M. A., et al. (2015). Re-evaluating the phylogeny of allopolyploid *Gossypium* L. *Mol. Phylogenet. Evol.* 92, 45–52. doi: 10.1016/j.ympev.2015.05.023

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520

Hágsater, E., and Soto-Arenas, M. Á. (2005). "*Epidendrum* L," in *Genera Orchidacearum*. Eds. A. M. Pridgeon, P. J. Cribb, M. W. Chase and F. N. Rasmussen (Oxford: Oxford University Press), 236–251.

Hamilton, C. A., Lemmon, A. R., Lemmon, E. M., and Bond, J. E. (2016). Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. *BMC Evol. Biol.* 16, 212. doi: 10.1186/s12862-016-0769-y

Heyduk, K., Trapnell, D. W., Barrett, C. F., and Leebens-Mack, J. (2016). Phylogenomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. *Biol. J. Linn. Soc* 117, 106–120. doi: 10.1111/bij.12551

Jheng, C.-F., Chen, T.-C., Lin, J.-Y., Chen, T.-C., Wu, W.-L., and Chang, C.-C. (2012). The comparative chloroplast genomic analysis of photosynthetic orchids and developing DNA markers to distinguish *Phalaenopsis* orchids. *Plant Sci.* 190, 62–73. doi: 10.1016/j.plantsci.2012.04.001

Johnson, M. G., Pokorny, L., Dodsworth, S., Botigué, L. R., Cowan, R. S., Devault, A., et al. (2019). A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Syst. Biol.* 68, 594–606. doi: 10.1093/sysbio/syy086

Katoh, K., and Standley, D. M. (2013). MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kim, H. T., Kim, J. S., Moore, M. J., Neubig, K. M., Williams, N. H., Whitten, W. M., et al. (2015). Seven new complete plastome sequences reveal rampant independent loss of the *ndh* gene family across orchids and associated instability of the inverted repeat/small single-copy region boundaries. *PloS One* 10, e0142215. doi: 10.1371/journal.pone.0142215

Kim, Y.-K., Kwak, M. H., Hong, J.-R., Kim, H.-W., Jo, S., Sohn, J.-Y., et al. (2017). The complete plastome sequence of the endangered orchid *Kuhlhasseltia nakaiana* (Orchidaceae). *Mitochondrial DNA Part B* 2, 701–703. doi: 10.1080/23802359.2017.1390408

Kriebel, R., Drew, B. T., Drummond, C. P., González-Gallegos, J. G., Celep, F., Mahdjoub, M. M., et al. (2019). Tracking temporal shifts in area, biomes, and pollinators in the radiation of *Salvia* (sages) across continents: leveraging anchored hybrid enrichment and targeted sequence data. *Am. J. Bot.* 106, 573–597. doi: 10.1002/ajb2.1268

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2017). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34, 772–773. doi: 10.1093/molbev/msw260

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30, 3276–3278. doi: 10.1093/bioinformatics/btu531

Leitch, I. J., Johnston, E., Pellicer, J., Hidalgo, O., and Bennett, M. D. (2019). Angiosperm DNA C-values database. Available at: https://cvalues.science.kew.org/[Accessed July 27, 2019].

Lemmon, E. M., and Lemmon, A. R. (2013). High-throughput genomic data in systematics and phylogenetics. *Annu. Rev. Ecol. Evol. Syst.* 44 44, 99–9+. doi: 10.1146/annurev-ecolsys-110512-135822

Leopardi-Verde, C. L., Carnevali, G., and Romero-González, G. A. (2017). A phylogeny of the genus *Encyclia* (Orchidaceae: Laeliinae), with emphasis on the species of the Northern Hemisphere. *J. Syst. Evol.* 55, 110–123. doi: 10.1111/jse.12225

Léveillé-Bourret, É., Starr, J. R., Ford, B. A., Moriarty Lemmon, E., and Lemmon, A. R. (2018). Resolving rapid radiations within angiosperm families using anchored phylogenomics. *Syst. Biol.* 67, 94–112. doi: 10.1093/sysbio/syx050

Li, Y.-X., Li, Z.-H., Schuiteman, A., Chase, M. W., Li, J.-W., Huang, W.-C., et al. (2019). Phylogenomics of Orchidaceae based on plastid and mitochondrial genomes. *Mol. Phylogenet. Evol.* 139, 106540. doi: 10.1016/j.ympev.2019.106540

Lin, C.-S., Chen, J. J. W., Huang, Y.-T., Chan, M.-T., Daniell, H., Chang, W.-J., et al. (2015). The location and translocation of *ndh* genes of chloroplast origin in the Orchidaceae family. *Sci. Rep.* 5, 9040. doi: 10.1038/srep09040

Liu, L., Yu, L., Kubatko, L., Pearl, D. K., and Edwards, S. V. (2009). Coalescent methods for estimating phylogenetic trees. *Mol. Phylogenet. Evol.* 53, 320–328. doi: 10.1016/j.ympev.2009.05.033

López-Giráldez, F., and Townsend, J. (2011). PhyDesign: an online application for profiling phylogenetic informativeness. *BMC Evol. Biol.* 11, 152. doi: 10.1186/1471-2148-11-152

Lu, Y., Ran, J.-H., Guo, D.-M., Yang, Z.-Y., and Wang, X.-Q. (2014). Phylogeny and divergence times of Gymnosperms inferred from single-copy nuclear genes. *PloS One* 9, e107679. doi: 10.1371/journal.pone.0107679

Maddison, W. P. (1997). Gene trees in species trees. *Syst. Biol.* 46, 523–536. doi: 10.1093/sysbio/46.3.523

Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., and Hernández-Hernández, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* 207, 437–453. doi: 10.1111/nph.13264

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., et al. (2010). Target-enrichment strategies for next-generation sequencing. *Nat. Meth.* 7, 111–118. doi: 10.1038/nmeth.1419

Mandel, J. R., Dikow, R. B., Funk, V. A., Masalia, R. R., Staton, S. E., Kozik, A., et al. (2014). A target enrichment method for gathering phylogenetic information from hundreds of loci: an example from the Compositae. *Appl. Plant Sci.* 2, 1300085. doi: 10.3732/apps.1300085

Mitchell, N., Lewis, P. O., Lemmon, E. M., Lemmon, A. R., and Holsinger, K. E. (2017). Anchored phylogenomics improves the resolution of evolutionary relationships in the rapid radiation of *Protea* L. *Am. J. Bot.* 104, 102–115. doi: 10.3732/ajb.1600227

Molloy, E. K., and Warnow, T. (2017). To include or not to include: the impact of gene filtering on species tree estimation methods. *Syst. Biol.* 67, 285–303. doi: 10.1093/sysbio/syx077

Nakai, M. (2015). YCF1: a green TIC: response to the de Vries et al. commentary. *Plant Cell* 27, 1834–1838. doi: 10.1105/tpc.15.00363

Neubig, K. M., Whitten, W. M., Carlsward, B. S., Blanco, M. A., Endara, L., Williams, N. H., et al. (2009). Phylogenetic utility of *ycf1* in orchids: a plastid gene more variable than *matK*. *Plant Syst. Evol.* 277, 75–84. doi: 10.1007/s00606-008-0105-0

Neubig, K. M., Whitten, W. M., Williams, N. H., Blanco, M. A., Endara, L., BURLEIGH, J. G., et al. (2012). Generic recircumscriptions of Oncidiinae (Orchidaceae: Cymbidieae) based on maximum likelihood analysis of combined DNA datasets. *Bot. J. Linn. Soc.* 168, 117–146. doi: 10.1111/j.1095-8339.2011.01194.x

Nikolov, L. A., Shushkov, P., Nevado, B., Gan, X., Al-Shehbaz, I. A., Filatov, D., et al. (2019). Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. *New Phytol.* 222, 1638–1651. doi: 10.1111/nph.15732

Niu, Z., Xue, Q., Zhu, S., Sun, J., Liu, W., and Ding, X. (2017). The complete plastome sequences of four orchid species: insights into the evolution of the Orchidaceae and the utility of plastomic mutational hotspots. *Front. Plant Sci.* 8, 715–715. doi: 10.3389/fpls.2017.00715

Pagnussat Klein, V., Pessoa, E. M., Oreste Demarchi, L., Sader, M., and Fernandez Piedade, M. T. (2019). *Encyclia, Epidendrum,* or *Prosthechea*? clarifying the phylogenetic position of a rare Amazonian orchid (Laeliinae-Epidendroideae-Orchidaceae). *Syst. Bot.* 44, 297–309. doi: 10.1600/036364419X15562054132983

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412

Parks, M., Cronn, R., and Liston, A. (2009). Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* 7, 84. doi: 10.1186/1741-7007-7-84

Peraza-Flores, L. N., Carnevali, G., and van den Berg, C. (2016). A molecular phylogeny of the *Laelia* alliance (Orchidaceae) and a reassessment of *Laelia* and *Schomburgkia*. *Taxon* 65, 1249–1262. doi: 10.12705/656.3

Pérez García, E. (1993). Arquitectura vegetal en *Epidendrum* L. (Orchidaceae).

Pinheiro, F., and Cozzolino, S. (2013). *Epidendrum* (Orchidaceae) as a model system for ecological and evolutionary studies in the Neotropics. *Taxon* 62, 77–88. doi: 10.1002/tax.621007

Pond, S. L. K., Frost, S. D. W., and Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679. doi: 10.1093/bioinformatics/bti079

Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M., et al. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* 526, 569–573. doi: 10.1038/nature15697

Quiroga-González, S. (2017). *Sistemática molecular del grupo de Epidendrum anisatum (Orchidaceae)*. B.Sc. thesis, Universidad Nacional Autónoma de México. Available at http://132.248.9.195/ptd2017/junio/0760012/Index.html.

Roch, S., and Warnow, T. (2015). On the robustness to gene tree estimation error (or lack thereof) of coalescent-based species tree methods. *Syst. Biol.* 64, 663–676. doi: 10.1093/sysbio/syv016

Rokyta, D. R., Lemmon, A. R., Margres, M. J., and Aronow, K. (2012). The venom-gland transcriptome of the eastern diamondback rattlesnake (Crotalus adamanteus). *BMC Genomics* 13, 1–23. doi: 10.1186/1471-2164-13-312

Roma, L., Cozzolino, S., Schlüter, P. M., Scopece, G., and Cafasso, D. (2018). The complete plastid genomes of *Ophrys iricolor* and *O. sphegodes* (Orchidaceae) and comparative analyses with other orchids. *PloS One* 13, e0204174–e0204174. doi: 10.1371/journal.pone.0204174

Rothfels, C. J., Larsson, A., Li, F.-W., Sigel, E. M., Huiet, L., Burge, D. O., et al. (2013). Transcriptome-mining for single-copy nuclear markers in ferns. *PloS One* 8, UNSP e76957. doi: 10.1371/journal.pone.0076957

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. doi: 10.1093/oxfordjournals.molbev.a040454

Salas-Leiva, D. E., Meerow, A. W., Calonje, M., Griffith, M. P., Francisco-Ortega, J., Nakamura, K., et al. (2013). Phylogeny of the cycads based on multiple single-copy nuclear genes: congruence of concatenated parsimony, likelihood and species tree inference methods. *Ann. Bot.* 112, 1263–1278. doi: 10.1093/aob/mct192

Sang, T. (2002). Utility of low-copy nuclear gene sequences in plant phylogenetics. *Crit. Rev. Biochem. Mol. Biol.* 37, 121–147. doi: 10.1080/10409230290771474

Sass, C., Iles, W. J., Barrett, C. F., Smith, S. Y., and Specht, C. D. (2016). Revisiting the Zingiberales: using multiplexed exon capture to resolve ancient and recent phylogenetic splits in a charismatic plant lineage. *PeerJ* 4, e1584. doi: 10.7717/peerj.1584

Sayyari, E., and Mirarab, S. (2016). Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* 33, 1654–1668. doi: 10.1093/molbev/msw079

Shahi Shavvon, R., Kazempour Osaloo, S., Maassoumii, A. A., Moharrek, F., Karaman Erkul, S., Lemmon, A. R., et al. (2017). Increasing phylogenetic support for explosively radiating taxa: the promise of high-throughput sequencing for *Oxytropis* (Fabaceae). *J. Syst. Evol.* 55, 385–404. doi: 10.1111/jse.12269

Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W. S., Miller, J., et al. (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* 92, 142–166. doi: 10.3732/ajb.92.1.142

Shi, S., Li, J., Sun, J., Yu, J., and Zhou, S. (2013). Phylogeny and classification of *Prunus* sensu lato (Rosaceae). *J. Integr. Plant Biol.* 55, 1069–1079. doi: 10.1111/jipb.12095

Soltis, D. E., Gitzendanner, M. A., Stull, G., Chester, M., Chanderbali, A., Chamala, S., et al. (2013). The potential of genomics in plant systematics. *Taxon* 62, 886–898. doi: 10.12705/625.13

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. doi: 10.1093/bioinformatics/btl446

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Stephens, J. D., Rogers, W. L., Mason, C. M., Donovan, L. A., and Malmberg, R. L. (2015). Species tree estimation of diploid *Helianthus* (Asteraceae) using target enrichment. *Am. J. Bot.* 102, 910–920. doi: 10.3732/ajb.1500031

Stull, G. W., Moore, M. J., Mandala, V. S., Douglas, N. A., Kates, H.-R., Qi, X., et al. (2013). A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes. *Appl. Plant Sci.* 1, 1200497. doi: 10.3732/apps.1200497

Summerer, D. (2009). Enabling technologies of genomic-scale sequence enrichment for targeted high-throughput sequencing. *Genomics* 94, 363–368. doi: 10.1016/j.ygeno.2009.08.012

Swofford, D. L. (2003). *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.* (Sunderland, Massachusetts: Sinauer Associates).

Townsend, J. P. (2007). Profiling phylogenetic informativeness. *Syst. Biol.* 56, 222–231. doi: 10.1080/10635150701311362

van den Berg, C., Higgins, W. E., Dressler, R. L., Whitten, W. M., Soto Arenas, M. A., Culham, A., et al. (2000). A phylogenetic analysis of Laeliinae (Orchidaceae) based on sequence data from internal transcribed spacers (ITS) of nuclear ribosomal DNA. *Lindleyana* 15, 96–114.

van den Berg, C., Higgins, W. E., Dressler, R. L., Whitten, W. M., Soto-Arenas, M. A., and Chase, M. W. (2009). A phylogenetic study of Laeliinae (Orchidaceae) based on combined nuclear and plastid DNA sequences. *Ann. Bot.* 104, 417–430. doi: 10.1093/aob/mcp101

van den Berg, C. (2000). Molecular phylogenetics of tribe Epidendreae with emphasis on subtribe Laeliinae (Orchidaceae).

van den Berg, C. (2014). The importance of hybridization for the evolution of *Cattleya*. *Renziana* 4, 74–79.

Wanke, S., Granados Mendoza, C., Müller, S., Paizanni Guillén, A., Neinhuis, C., Lemmon, A. R., et al. (2017). Recalcitrant deep and shallow nodes in *Aristolochia* (Aristolochiaceae) illuminated using anchored hybrid enrichment. 25th Anniv. Issue. *Mol. Phylogenet. Evol.* 117, 111–123. doi: 10.1016/j.ympev.2017.05.014

Weitemier, K., Straub, S. C. K., Cronn, R. C., Fishbein, M., Schmickl, R., McDonnell, A., et al. (2014). Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Appl. Plant Sci.* 2, 1400042. doi: 10.3732/apps.1400042

Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Müller, K. F., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol. Biol.* 76, 273–297. doi: 10.1007/s11103-011-9762-4

Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinf.* 19, 153. doi: 10.1186/s12859-018-2129-y

Zhitao, N., Shuying, Z., Jiajia, P., Ludan, L., Jing, S., and Xiaoyu, D. (2017). Comparative analysis of *Dendrobium* plastomes and utility of plastomic mutational hotspots. *Sci. Rep.* 7, 2073. doi: 10.1038/s41598-017-02252-8

Zimmer, E. A., and Wen, J. (2012). Using nuclear gene data for plant phylogenetics: progress and prospects. *Mol. Phylogenet. Evol.* 65, 774–785. doi: 10.1016/j.ympev.2012.07.015