



Plastome Rearrangements in the “*Adenocalymma-Neojobertia*” Clade (Bignoniaceae, Bignoniaceae) and Its Phylogenetic Implications

Luiz H. M. Fonseca* and Lúcia G. Lohmann*

Laboratório de Sistemática Vegetal, Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

OPEN ACCESS

Edited by:

Daniel Pinero,
Universidad Nacional Autónoma de
México, Mexico

Reviewed by:

Marcelo Reginato,
New York Botanical Garden,
United States
Ming Kang,
South China Institute of Botany (CAS),
China

*Correspondence:

Luiz H. M. Fonseca
luizhmf@gmail.com
Lúcia G. Lohmann
llohmann@usp.br

Specialty section:

This article was submitted to
Plant Genetics and Genomics,
a section of the journal
Frontiers in Plant Science

Received: 17 April 2017

Accepted: 16 October 2017

Published: 01 November 2017

Citation:

Fonseca LHM and Lohmann LG
(2017) Plastome Rearrangements in
the “*Adenocalymma-Neojobertia*”
Clade (Bignoniaceae, Bignoniaceae) and
Its Phylogenetic Implications.
Front. Plant Sci. 8:1875.
doi: 10.3389/fpls.2017.01875

The chloroplast is one of the most important organelles of plants. This organelle has a circular DNA with approximately 130 genes. The use of plastid genomic data in phylogenetic and evolutionary studies became possible with high-throughput sequencing methods, which allowed us to rapidly obtain complete genomes at a reasonable cost. Here, we use high-throughput sequencing to study the “*Adenocalymma-Neojobertia*” clade (Bignoniaceae, Bignoniaceae). More specifically, we use Hi-Seq Illumina technology to sequence 10 complete plastid genomes. Plastomes were assembled using selected plastid reads and *de novo* approach with SPAdes. The 10 assembled genomes were analyzed in a phylogenetic context using five different partition schemes: (1) 91 protein-coding genes (“coding”); (2) 76 introns and spacers with alignment manually edited (“non-coding edited”); (3) 76 non-coding regions with poorly aligned regions removed using T-Coffee (“non-coding filtered”); (4) 91 coding regions plus 76 non-coding regions edited (“coding + non-coding edited”); and, (5) 91 protein-coding regions plus the 76 filtered non-coding regions (“coding + non-coding filtered”). Fragmented regions were aligned using Mafft. Phylogenetic analyses were conducted using Maximum Likelihood (ML) and Bayesian Criteria (BC). The analyses of the individual plastomes consistently recovered an expansion of the Inverted Repeated (IRs) regions and a compression of the Small Single Copy (SSC) region. Major genomic translocations were observed at the Large Single Copy (LSC) and IRs. ML phylogenetic analyses of the individual datasets led to the same topology, with the exception of the analysis of the “non-coding filtered” dataset. Overall, relationships were strongly supported, with the highest support values obtained through the analysis of the “coding + non-coding edited” dataset. Four regions at the LSC, SSC, and IR were selected for primer development. The “*Adenocalymma-Neojobertia*” clade shows an unusual pattern of plastid structure variation, including four major genomic translocations. These rearrangements challenge the current view of conserved plastid genome architecture in terms of gene order. It also complicates both genomic assemblies using reference genomes and sequence alignments using whole plastomes. Therefore, strategies that employ *de novo* assemblies and manual evaluation of sequence alignments are required to prevent assembly and alignment errors.

Keywords: cp genome, genomic rearrangements, DNA sequence alignment, phylogenomics, plastid primers

INTRODUCTION

The plastome is the portion of the plant genome that contains all the genetic information included in the chloroplast (Bock, 2007). The chloroplast is an organelle of prokaryotic origin with a crucial role in photosynthesis and cell storage (Wise, 2006). It contains the biochemical machinery necessary to replicate its own genome, transcribe genes, and translate those genes into proteins (Wise, 2006). Plastomes have a circular genome of double-stranded DNA that ranges from 72 to 217 kb in flowering plants (Chumley et al., 2006), with approximately 130 genes (Sugiura, 1992, 1995). Genes found in the plastomes encode the core proteins of photosynthetic complexes, including Photosystem I and II, Cytochrome b_6/f , NADH dehydrogenase, ATP synthase and RUBISCO (Grenn, 2011). Chloroplast genomes typically include a quadripartite structure that consists of a small single copy region (SSC) with approximately 16–27 kb, a large single copy region (LSC) with approximately 80–90 kb, and a pair of inverted repeats (IRs) with approximately 20 to 28 kb each. Expansions and contractions of the IRs, as well as gene and intron losses have been documented in Angiosperms (Jansen et al., 2011; Liu et al., 2016). However, the overall chloroplast structure, gene content, and organization are thought to be highly conserved among flowering plants (Odintsova and Yurina, 2003; Wicke et al., 2011; Smith and Keeling, 2015; Reginato et al., 2016).

The conserved structure of the chloroplast genome facilitates PCR primer design and sequencing within Angiosperms (Small et al., 1998; Shaw et al., 2005, 2007). Efforts to resolve Angiosperm phylogenetic relationships at different taxonomic levels have traditionally used plastome coding and non-coding regions as sources of evidence (e.g., Soltis et al., 1999; Shaw et al., 2007). While these regions are very informative at higher taxonomic levels, they often lack sufficient variation to resolve relationships at the species or population levels, even when rapidly evolving non-coding DNA regions are considered (Small et al., 1998; Shaw et al., 2005, 2007). More recently, high-throughput sequencing methods have allowed researchers to rapidly obtain complete genomes at a reasonable cost (Cronn et al., 2008; Parks et al., 2009). These genomes have been used as basis for phylogenomic studies, leading to highly resolved and strongly supported phylogenies of several plant groups (Moore et al., 2010; Harrison et al., 2015; Wysocki et al., 2015).

The “*Adenocalymma-Neojobertia*” clade (Bignoniaceae) is a lineage of lianas, shrubs and treelets that includes approximately 75 species. The genus exhibits substantial diversity in ecology, with species distributed from deciduous forests (e.g., Brazilian cerrados and caatingas) to tropical rain forests (e.g., Amazonia and Atlantic forest) (Lohmann and Taylor, 2014). A phylogenetic study of the whole tribe Bignoniaceae based on sequences of the nuclear intron *pepC* and the plastid gene *ndhF* was the only study to sample species of *Adenocalymma* Mart. ex Meisn. and *Neojobertia* Baill. to date (Lohmann, 2006). This study sampled 12 of the 75 species currently recognized and recovered a monophyletic “*Adenocalymma-Neojobertia*” clade (Lohmann, 2006). While generic-level clades were strongly supported in this study, resolution was weak within the “*Adenocalymma-Neojobertia*”

clade (Lohmann, 2006). Full plastomes generally include a high number of phylogenetic informative characters and can improve estimates of phylogenetic relationship at various taxonomic levels (e.g., Ma et al., 2014; Reginato et al., 2016). However, only the plastid genomes of *Tanaecium tetragonolobum* (Jacq.) L.G. Lohmann (NC_027955.1; Nazareno et al., 2015) and *Crescentia cujete* L. (KT182634.2; Moreira et al., 2016) are currently available for members of the plant family Bignoniaceae.

In this study, we used high-throughput sequencing technology to sequence ten complete plastomes of members of the “*Adenocalymma-Neojobertia*” clade in order to: (i) characterize the gene content, levels of sequence variation, and structure of plastomes within this clade; (ii) compare the plastomes of members of the “*Adenocalymma-Neojobertia*” clade with those available for other Bignoniaceae; (iii) explore the potential of genomic data for phylogenomic studies within the “*Adenocalymma-Neojobertia*” clade and the Bignoniaceae as a whole; and, (iv) identify informative markers for future species level phylogenetic studies.

MATERIALS AND METHODS

Taxon Sampling and Genome Sequencing

We sampled 10 accessions of members of the “*Adenocalymma-Neojobertia*” clade, representing nine species of *Adenocalymma* plus one species of *Neojobertia* (NCBI accession numbers at **Table 1**). These species were selected in order to represent the breadth of morphological diversity and geographical distribution within the clade. Total genomic DNA was extracted from silica-dried leaflets or herbarium specimens using the Invisorb® Spin Plant Mini Kit (Invitek, Berlin, Germany). Approximately 60 ng of leaf tissue were pulverized with TissueLyzer® (Qiagen, Duesseldorf, Germany) for 3 min at 60 Hz. Five micrograms of total DNA were fragmented using a Covaris S-series sonicator, generating DNA fragments of approximately 300 bp. Libraries were constructed using the NEBNext DNA Library Prep Master Mix Set and the NEBNext Multiplex oligos for Illumina (New England BioLabs Inc., Ipswich, MA) following the manufacturer's protocol. DNA library concentration was determined using the Kapa Library Quantification Kit (Kapa Biosystems Inc., Wilmington, MA) on an Applied Biosystems 7500 Real-Time PCR System. The final libraries were diluted to a concentration of 10 nM and put together in pools of 20 samples. Each pool of species was sequenced in a lane using pair-end (2 × 100) on an Illumina HiSeq 2000 system (Illumina Inc., San Diego, CA).

Plastome Assembly and Annotation

Plastomes were assembled using the Fast-Plast pipeline (<https://github.com/mrmckain/Fast-Plast>; McKain and Wilson, unpublished). For each species, adaptors were removed and low quality sequences trimmed using Trimmomatic 0.35 (Bolger et al., 2014) with the SLIDINGWINDOW:10:20 and MINLEN:40 parameters. Trimmed reads were mapped against a database that included *C. cujete* L., *Erythranthe lutea* (L.) G.L. Nesom (NC_030212.1; Vallejo-Marín et al., 2016), *Olea europaea* L. (NC_013707.2; Messina, unpublished), *Sesamum indicum* L. (NC_016433.2; Yi and Kim, 2012), *Salvia miltiorhiza* Bunge (NC_020431.1; Qian et al., 2013), and *T. tetragonolobum* (Jacq.)

TABLE 1 | Taxa, vouchers, collection sites, and accession numbers of *Adenocalymma* and *Neojobertia* specimens sampled.

Taxa	Vouchers	Collection Sites	Accession
<i>A. allamandiflorum</i>	B.M. Gomes 671 (SPF)	Brazil; Pará; Santarém	MG008307
<i>A. aurantiacum</i>	L.G. Lohmann 658 (SPF)	Brazil; Espírito Santo; Linhares	MG008308
<i>A. biternatum</i>	M.R. Pace 521 (SPF)	Peru; Loreto; Iquitos	MG008309
<i>A. bracteatum</i>	L.H.M. Fonseca 100 (SPF)	Brazil; Paraná; Jundiá do Sul	MG008310
<i>A. cristicalyx</i>	L.G. Lohmann 705 (SPF)	Brazil; Paraíba; Alagoa Grande	MG008311
<i>A. nervosum</i>	L.H.M. Fonseca 262 (SPF)	Brazil; Minas Gerais; Catuji	MG008312
<i>A. pedunculatum</i>	L.H.M. Fonseca 267 (SPF)	Brazil; Minas Gerais; Diamantina	MG008313
<i>A. peregrinum</i>	L.H.M. Fonseca 444 (SPF)	Brazil; Goiás; São Jorge da Chapada	MG008314
<i>A. subspicatum</i>	R.G. Udulutsch 2758 (SPF)	Brazil; Ceará; Tianguá	MG008315
<i>N. candolleana</i>	L.G. Lohmann 363 (SPF)	Brazil; Bahia; Mucugê	MG008316

L.G. Lohmann using Bowtie2 2.1.0 (Langmead and Salzberg, 2012) with the default parameters. Mapped reads were assembled into contigs using SPAdes 3.1.0 (Bankevich et al., 2012) with k-mer sizes of 55 and 87, using the “only-assembler” option. Resulting contigs were assembled with the software afn (<https://bitbucket.org/afn/afn>) and the default parameters -l 50, -f 0.1, -d 100, -x 100, and -i 2. For species for which it was harder to obtain comprehensive contigs, we tested different values for maximum percentage of mismatches (-g), and minimum overlap of contig (-p) parameters. For some species, the *de novo* assembly returned a large contig that contained the complete plastome. These contigs were checked and finalized with Geneious 9.0.2 (Kearse et al., 2012). The plastome assembly was verified through a coverage analysis conducted in Jellyfish 2.1.3 (Marçais and Kingsford, 2011). The estimate of 25-mer abundance was used to map a 25-bp sliding window of coverage across the plastome of each species.

Plastome annotation was initially conducted in DOGMA (Wyman et al., 2004). These annotations were checked in Geneious 9.0.2 using *O. europaea* and *Solanum lycopersicum* L. (NC_007898.3; Daniell et al., 2006) as references. Promising open reading frames at non-coding regions were verified with BLAST (Altschul et al., 1990) available at NCBI (<https://www.ncbi.nlm.nih.gov/>). Maps of the annotated plastomes were created using OGDRAW (Lohse et al., 2007). We characterized the overall plastome structure, gene content, and general gene information of the 10 species sampled and compared our results with the information available for two other Bignoniaceae (i.e., *C. cujete* and *T. tetragonolobum*), and one Oleaceae (i.e., *O. europaea*). Points of potential rearrangements and junctions between the IRs, the LSC, and SSC were tested iteratively using afn (<https://bitbucket.org/afn/afn>), and checked with PCR amplifications and electrophoresis. Coverage values for these regions were also assessed.

Phylogenetic Analyses

We used the LSC, SSC and one IR to infer the phylogenetic tree of the “*Adenocalymma-Neojobertia*” clade. We excluded one IR to avoid duplication of data. We used three chloroplast genomes of members of the Lamiales (*C. cujete*, *O. europaea*, and *T. tetragonolobum*) as outgroups. Pseudogenes and its orthologous

were treated as non-coding regions. Genes with overlapping portions were treated as neighbors to avoid character duplication.

For the phylogenetic analyses, annotated plastomes were fragmented into coding and non-coding regions, excluding regions smaller than 50 bp. The retained regions were grouped by sequence similarity (with a threshold of 65% of global similarity and default alignment costs) using the annotated plastome of *Adenocalymma biternatum* (A. Samp.) L.G. Lohmann as reference and considering the pool of regions for all species. Plastome partitioning and sequence grouping was conducted using the R package Biostrings (R Development Core Team, 2017; Pagès et al., unpublished). Coding regions were aligned with MAFFT 7 (Katoh and Standley, 2013) using the G-INS-i 1,000 strategy, while non-coding regions were aligned using the E-INS-i 1,000 strategy. We removed poorly aligned regions of the coding and non-coding alignments using GBLOCKS (Castresana, 2000) default settings in order to circumvent homology assessment problems due to random similarity of sequences or indels. Alignments of non-coding regions with rearrangements were edited manually or misaligned sequences were removed using the outlier search option implemented in T-Coffee (Notredame et al., 2000). This was necessary since GBLOCKS is not able to recognize rare outlier sequences (Castresana, 2000). Three different partition schemes were built as follows: (1) 91 coding regions (“coding”); (2) 76 introns and spacers with alignment edited by hand (“non-coding edited”); and (3) 76 non-coding regions with poorly aligned sequences removed with T-Coffee (“non-coding filtered”). Combined datasets were also analyzed as follows: (4) 91 coding regions plus 76 non-coding regions (“coding + non-coding edited”); and (5) 91 coding regions plus the 76 filtered non-coding regions (“coding + non-coding filtered”). The five datasets were compared based on tree topology and node support.

All phylogenetic analyses were performed with Maximum Likelihood (ML) using RAxML 8.2.9 (Stamatakis, 2014), and Bayesian Criteria (BC) using MrBayes 3.2 (Ronquist et al., 2011). ML node support was estimated through a rapid bootstrap analysis with 1,000 replicates. BC were run using uniform priors and two independent runs of 10 million generations with four chains per run, sampling trees every 1,000 generations. BC support was estimated using posterior probabilities. For BC,

chain convergence and stationarity were assessed using the R package Coda (R Development Core Team, 2017; Plummer et al., unpublished) by visually examining plots of parameter values and log-likelihood against the number of generations. For Bayesian analysis we employed the reversible jump strategy (Ronquist et al., 2011), which does not require the establishment of evolutionary models or partition schemes *a priori*. For ML the GTRCAT evolutionary model was used (Stamatakis, 2014), avoiding pre-defined partitions.

Identification of Markers for Species Level Phylogenetic Studies

Among the 76 introns and spacers recovered, we retained the 31 regions that were recombination free and with suitable length for PCR amplification (amplicons with size between 500 and 1,100 bp). These partitions were analyzed to identify highly informative regions that may serve as useful markers for future species level phylogenetic analyses. ML trees were inferred for each of the 31 partitions using RAxML 8.2.9 and the GTRCAT evolutionary model. For each partition, alignment length, variable sites, topological distance, and branch length distance (Kendall and Colijn, 2016) were estimated. Metrics were computed using the R packages (R Development Core Team, 2017) Ape (Paradis et al., 2004) and Treescap (Jombart et al., unpublished). Partitions were ranked using standardized values of the number of informative characters, as well as the topological and branch length distances between the tree derived from the analysis of each partition and the best tree estimated in this study (i.e., the tree derived from the analysis of the “coding + non-coding edited” dataset; see results). All metrics were computed for the “*Adenocalymma-Neojobertia*” clade exclusively. Four non-coding regions were selected with Geneious 9.0.2 for primer design.

RESULTS

Plastome Assembly

We sequenced the complete plastomes of 10 species of the “*Adenocalymma-Neojobertia*” clade using an Illumina HiSeq

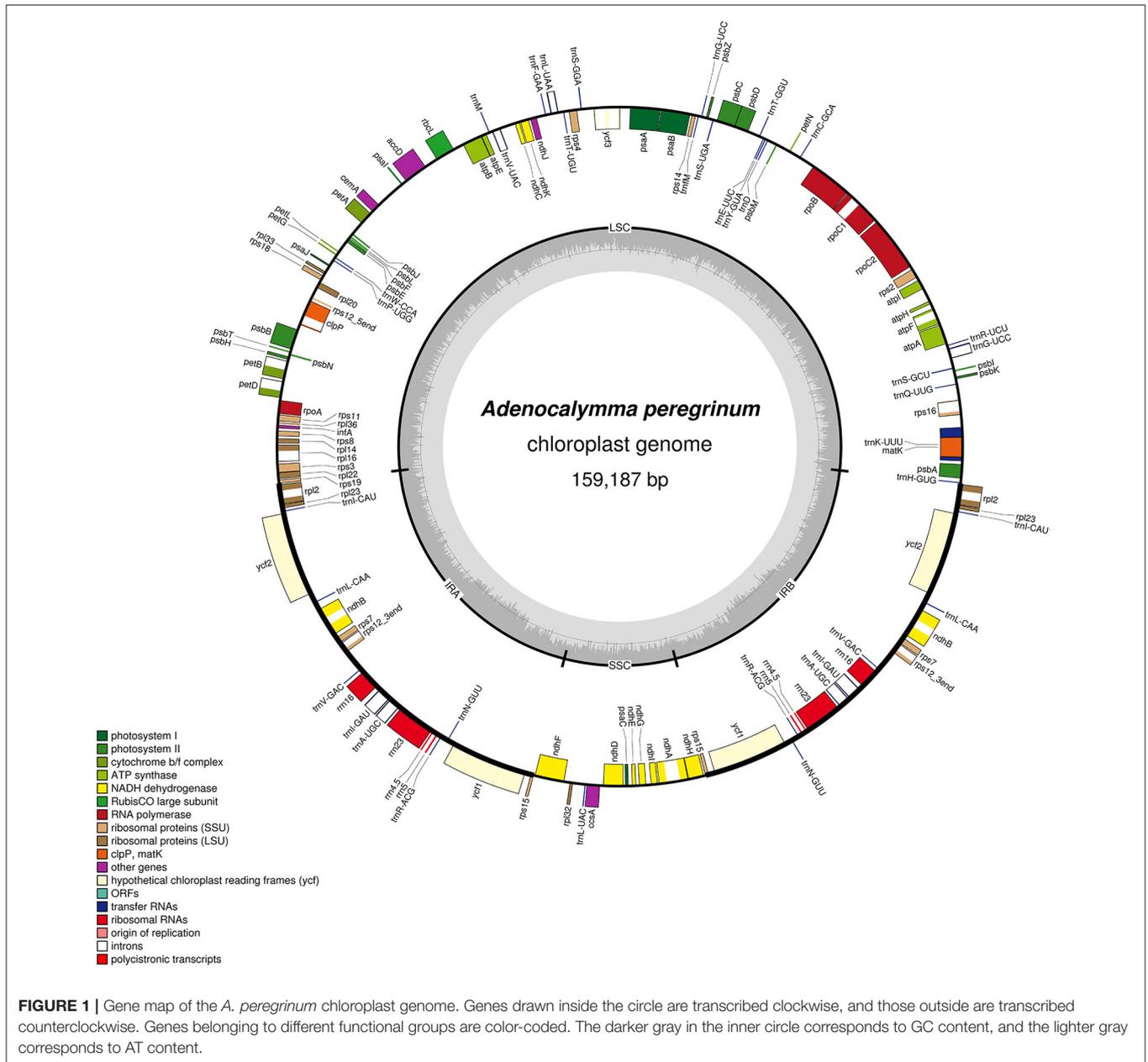
2000 system, namely: *A. allamandiflorum* (Bureau ex K. Schum.) L.G. Lohmann, *A. aurantiacum* Udulutsch and Assis, *A. biternatum* (A. Samp.) L.G. Lohmann, *A. bracteatum* (Cham.) DC., *A. cristicalyx* (A.H. Gentry) L.G. Lohmann, *A. nervosum* Bureau and K. Schum., *A. pedunculatum* (Vell.) L.G. Lohmann, *A. peregrinum* (Miers) L.G. Lohmann, *A. subspicatum* A.H. Gentry, and *N. candolleana* (Mart. ex DC.) Bureau and K. Schum. (Table 2). A minimum of 8,532,329, and a maximum of 30,862,472 paired end raw reads (with an average length of 101 bp) were generated for *N. candolleana* and *A. biternatum*, respectively. After mapping reads against the reference genomes of *C. cujete*, *T. tetragonolobum*, and *O. europaea*, a minimum of 239,286 reads and a maximum of 762,288 reads were retained for *A. subspicatum* and *A. bracteatum*, respectively. Plastome coverage ranged from $307.7 \times$ to $964 \times$ for *A. subspicatum* and *A. bracteatum*, respectively (Table 2). Junctions of the quadripartite structure and the regions with potential rearrangements were tested interactively and recovered in all combinations of parameters used. A high mean coverage value was obtained for all species, providing additional support for the plastome assemblies (Table 2). High coverage values were also observed at junctions of the quadripartite structure and regions with rearrangements. PCR and electrophoresis recovered the amplicons expected for each junction of the quadripartite structure and regions with potential rearrangements. No regions with low coverage (<20x) were recovered. The finished, high quality organelle genome sequences were used for downstream analyses.

Plastome Features

The plastomes of the 10 species of the “*Adenocalymma-Neojobertia*” clade ranged in size from 157,025 (*A. biternatum*) to 159,725 bp (*A. bracteatum*). All cp genomes have the typical quadripartite structure of Angiosperms, which consists of a pair of IR regions (30,084–30,954 bp) separated by a LSC region (84,059–85,665 bp), and a SSC region (12,585–12,804 bp). A circular map of *A. peregrinum* is shown (Figure 1), and those of all other species sampled are available as Supplemental Material (Supplementary Figures 1–9). The average GC content was ~38% for all species, with the exception of *A. nervosum* (41.6%). The GC

TABLE 2 | Summary of sequenced plastomes of *Adenocalymma* and *Neojobertia*.

Species	No. of raw reads	No. of mapped reads	Plastome coverage (x)	Plastome length (bp)	LSC length (bp)	SSC length (bp)	IR length (bp)	GC content (%)	CDS	tRNA	rRNA
<i>A. allamandiflorum</i>	14,546,510	284,977	364.4	157,952	84,668	12,804	30,240	38.2	85	37	8
<i>A. aurantiacum</i>	9,142,072	713,911	904.6	159,407	84,934	12,585	30,954	37.4	85	37	8
<i>A. biternatum</i>	30,862,472	501,539	645.2	157,025	84,059	12,723	30,097	38.2	84	37	8
<i>A. bracteatum</i>	17,633,356	762,288	964	159,725	85,665	12,632	30,835	38.3	85	37	8
<i>A. cristicalyx</i>	10,369,744	399,680	507.7	159,010	85,654	12,677	30,310	38	85	37	8
<i>A. nervosum</i>	19,780,568	432,099	549.7	158,786	84,999	12,616	30,561	41.6	85	37	8
<i>A. pedunculatum</i>	13,881,445	305,321	390	158,103	85,043	12,730	30,147	37.6	85	37	8
<i>A. peregrinum</i>	18,872,525	506,711	643	159,187	85,106	12,804	30,614	38.3	84	37	8
<i>A. subspicatum</i>	11,761,741	239,286	307.7	157,089	84,219	12,765	30,084	37.7	85	37	8
<i>N. candolleana</i>	8,532,329	738,770	942	158,409	85,192	12,737	30,211	38.1	85	37	8



content values suggest an AT-rich plastome organization, which is similar to that found in the other Bignoniaceae plastomes available to date (Nazareno et al., 2015; Moreira et al., 2016). All plastomes studied include 131–132 genes, with 86–87 coding regions, 37 tRNA, and 8 rRNA (Tables 2, 3). The difference in number of coding genes among species is due to the complete loss of the *ycf4* gene in *A. biternatum* and *A. peregrinum*. A copy of the duplicated gene *rps15* is a pseudogene, with only an incomplete sequence found at the border of the IRA and the SSC (Figure 2). Thirteen genes have one intron, while three genes (i.e., *clpP*, *rps12* and *ycf3*) have two introns (Table 3). In *rps12* a trans-splicing event was observed with the 5' end located in the LSC region and the duplicated 3' end in the IRs. The *trnQ-UUG* gene is

duplicated in the LSC of all species, with one intron found in one of the copies (Table 3). Among protein-coding genes, 84–85 genes start with the standard initiator codon AUG; however, the *rps19* starts with GUG, while the *ndhD* starts with ACG. The stop codon UAA was the most common, followed by UAG and UGA.

The gene structure of the IR/SSC boundary regions were well conserved within the “*Adenocalymma-Neojobertia*” clade, only differing slightly in a few base pairs (Figure 2). However, major structural differences were observed when those plastomes were compared with those of *C. cujete*, *O. europaea*, and *T. tetragonolobum*. Species of the “*Adenocalymma-Neojobertia* clade” included the gene *ycf1* and part of the gene *rps15* at

TABLE 3 | Genes recovered within the “*Adenocalymma-Neojobertia*” clade.

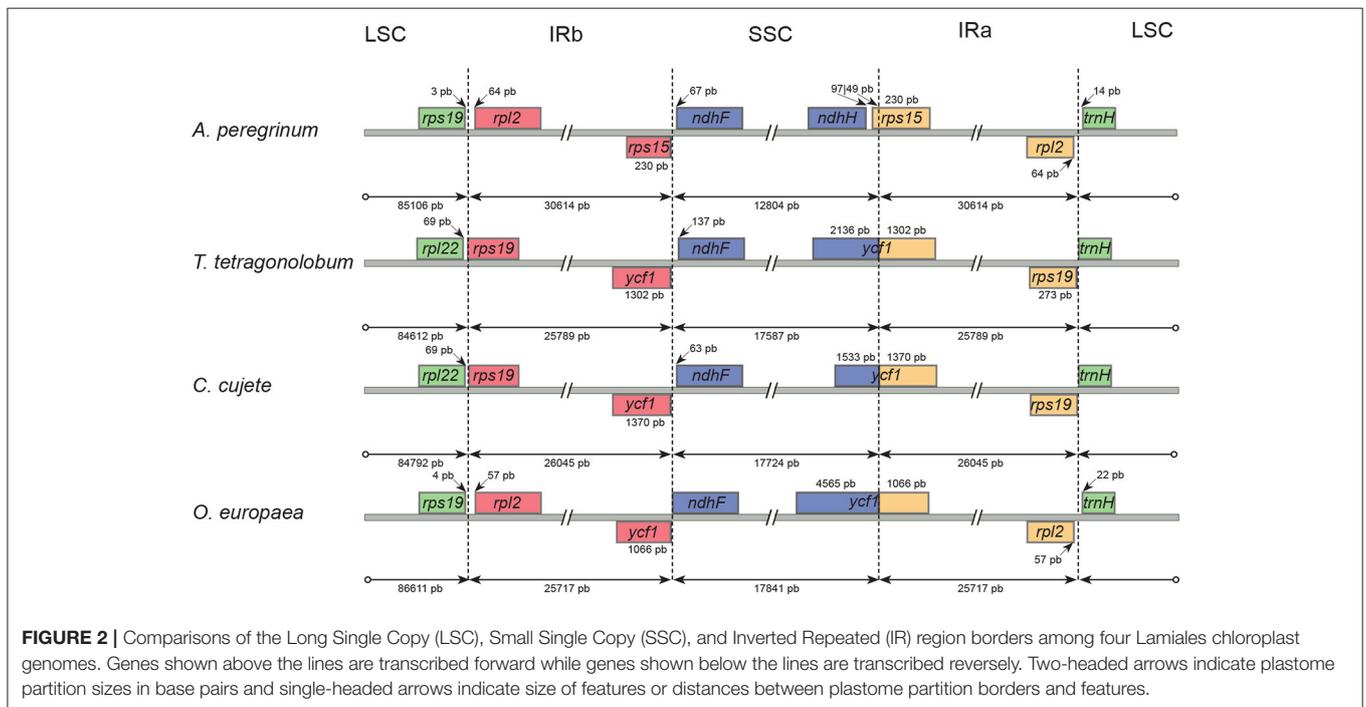
Gene function	Gene type	Gene
Self-replication	<ul style="list-style-type: none"> rRNA genes tRNA genes Small ribosomal subunit Large ribosomal subunit DNA dependent RNA 	<i>rrn4.5^a, rrn5^a, rrn16^a, rrn23^a</i> <i>trnA-UGC^a, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnM-CAU, trnG-UCC, trnG-UCC*, trnH-GUG, trnI-CAU^a, trnI-GAU^a, trnK-UUU*, trnL-CAA^a, trnL-UAA*, trnL-UAG, trnM-CAU, trnN-GUU^a, trnP-UGG, trnQ-UUG, trnR-ACG^a trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC^a, trnV-UAC*, trnW-CCA, trnY-GUA</i> <i>rps2, rps3, rps4, rps7^a, rps8, rps11, rps12^a, rps14, rps15^a, rps16[*], rps18, rps19^b</i> <i>rpl2^a, rpl14, rpl16*, rpl20, rpl22, rpl23^a, rpl32, rpl33, rpl36</i> <i>rpoA, rpoB, rpoC1*, rpoC2</i>
Photosynthesis	<ul style="list-style-type: none"> Photosystem I Photosystem II NADH-dehydrogenase Cytochrome b6/f complex ATP synthase Rubisco 	<i>psaA, psaB, psaC, psal, psaj, ycf3**</i> <i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i> <i>ndhA*, ndhB^a, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i> <i>petA, petB*, petD*, petG, petL, petN</i> <i>atpA, atpB, atpE, atpF*, atpH, atpI</i> <i>rbcL</i>
Other genes	<ul style="list-style-type: none"> Translational initiator Maturase Protease Envelope membrane protein Acetyl-CoA-carboxylase c-type cytochrome synthesis 	<i>infA</i> <i>matK</i> <i>clpP**</i> <i>cemA</i> <i>accD</i> <i>ccsA</i>
Unknown function	Conserved open read frames	<i>ycf1^a, ycf2^a, ycf4^b</i>

^{*}Gene with one intron.

^{**}Gene with two introns.

^aGene with two copies.

^bPseudogene in some species.



the borders of the IR and SSC regions; this led to a smaller SSC (12,585–12,804 bp) (Table 2) when compared to the SSC of *C. kujete* (17,724 bp), *O. europaea* (17,841 bp), and *T. tetragonolobum* (17,587 bp) (Figure 2). The IR regions and LSC borders found in members of the “*Adenocalymma-Neojobertia*” clade also differed from those of *C. kujete*, and *T. tetragonolobum*, with the *rps19* gene lacking from the IR regions of all species of the “*Adenocalymma-Neojobertia*” clade sampled (Figure 2). These rearrangements at the IR regions led to larger plastomes for all taxa analyzed (Table 2), when compared to those of *C. kujete* (154,662 bp), *O. europaea* (155,889 bp) and *T. tetragonolobum* (153,776 bp).

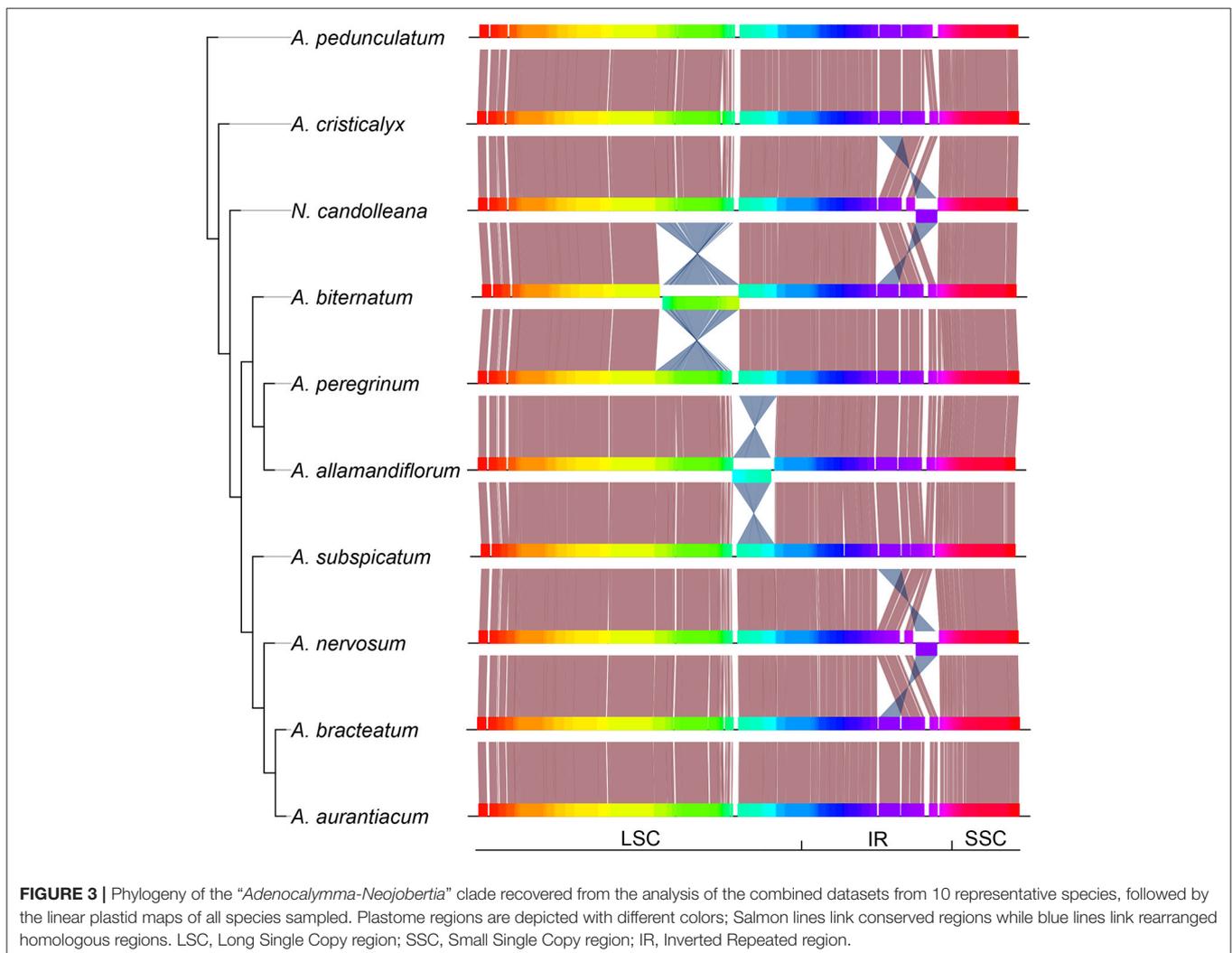
At least four major inversions were detected in some species of the “*Adenocalymma-Neojobertia*” clade. Two of those inversions were found at the LSC and two at the IRs (Figure 3). Rearrangements at the LSC occurred at different positions and were associated with different gene blocks (Figure 3). On the other hand, the rearrangements at the IRs involved the same gene blocks, except from the rearrangement found at *trnV-GAC*, indicating a parallel event. All structural changes involved

non-coding regions. Furthermore, no genes were shut down by the inclusion of major genomic parts.

Phylogenetic Analyses

We conducted phylogenetic analyses of five different datasets derived from plastome data of 10 species belonging to the “*Adenocalymma-Neojobertia*” clade, plus three outgroups (i.e., *C. kujete*, *T. tetragonolobum*, and *O. europaea*) using ML and BC. Among all datasets, the “non-coding edited” and “non-coding filtered” datasets contained the highest number of variable sites (39.3%), followed by the “coding + non-coding edited” and “coding + non-coding filtered” datasets (32.5%), and the “coding” dataset (27.9%) (Table 4). The analyses of all datasets led to the same topology (Figure 4, Supplementary Figure 10), except from the topology reconstructed based on the “non-coding filtered” dataset, which led to a slightly different tree (Figure 4).

All topologies derived from the BC and ML analyses recovered *A. pedunculatum* as sister to all other species of the “*Adenocalymma-Neojobertia* clade” (node A). For the majority



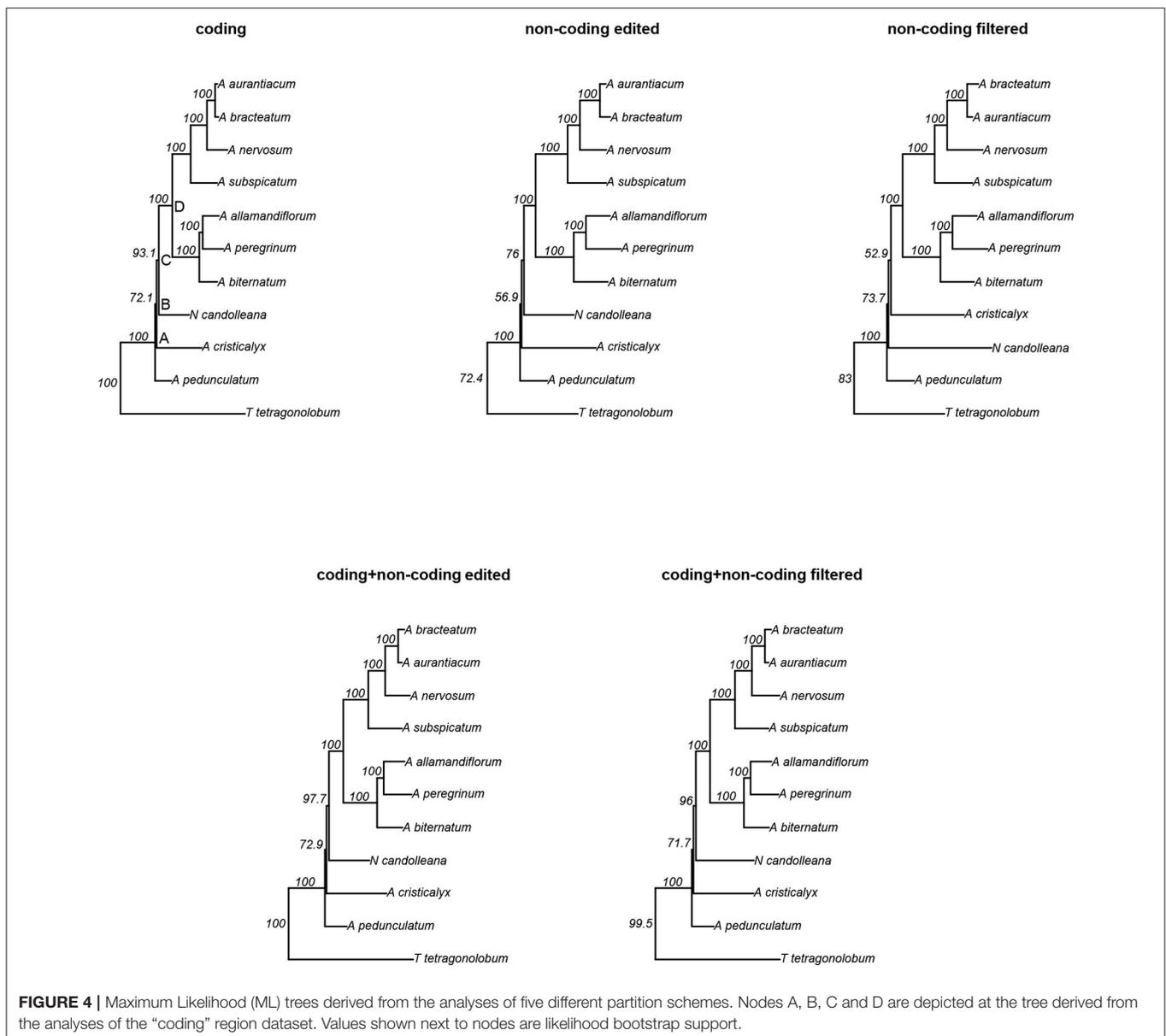
of the topologies, node A is followed by the divergence of *A. cristicalyx* (node B), which is followed by the divergence of *N. candolleana* (node C). The remaining species are included in a clade (node D) that is divided into two sub-clades, one including

(*A. biternatum*, *A. allamandiflorum*, *A. peregrinum*) and the other including (*A. subspicatum*, *A. nervosum*, *A. aurantiacum*, *A. bracteatum*) (Figure 4, Supplementary Figure 10). Node D and all clades included herein were recovered from the analyses of all datasets. However, nodes B and C were not recovered in the tree that resulted from the analyses of the “non-coding filtered” dataset for both BC and ML; instead, the analyses of the “non-coding filtered” dataset recovered *N. candolleana* as the second diverging lineage (right after node A) within the “*Adenocalymma-Neojobertia*” clade. This node is followed by the divergence of *A. cristicalyx* (Figure 4, Supplementary Figure 10).

Bootstrap mean values of trees derived from the analyses of the datasets are: 96.3% for the “coding” dataset, 92.54% for the “non-coding edited” dataset, 91.85% for the “non-coding filtered” dataset, 96.73% for the “coding + non-coding edited” dataset, 96.73% for the “coding + non-coding filtered” dataset,

TABLE 4 | Summary of partition schemes.

Partition	Alignment length (bp)	Variable sites	% of variation
Coding	71,395	19,912	27.9
Non-coding edited	48,469	19,052	39.3
Non-coding filtered	48,319	19,005	39.3
Coding + non-coding edited	119,864	38,964	32.5
Coding + non-coding filtered	119,714	38,917	32.5



and 96.41% “coding + non-coding filtered” dataset. Most nodes have maximum support in all trees, except from nodes B and C, where all differences in support are found. Among the combined datasets, the tree derived from the analysis of the “coding + non-coding edited” dataset has bootstrap values of 72.9 and 97.7 for nodes B and C, respectively and the “coding + non-coding filtered” dataset has bootstrap values of 71.7 and 96 for nodes B and C, respectively (Figure 4).

Identification of Markers for Species Level Phylogenetic Studies

The five (out of 31) regions with highest potential for species level phylogenetic studies based on the percentage of sequence variation, topological and branch length distances were: *clpP* intron 1, *ndhA* intron, *petN-psbM* spacer, *rpl32-trnL* spacer, and *trnG* intron (Table 5, Supplementary Table 2). Three out of the five regions selected are part of the LSC; with *ndhA* intron and *rpl32-trnL* spacer included in the SSC. The *ndhA* intron was the region with the greatest percentage of sequence variation, followed by the *petN-psbM* spacer, *trnG* intron, *clpP* intron 1, and *rpl32-trnL* spacer (Table 5). The topologies obtained from the analysis of the *petN-psbM*, and *trnG* intron identical to the best plastome tree (i.e., the tree derived from the analysis of the “coding + non-coding edited” dataset). Among the five regions selected, the trees derived from the analyses of the *clpP* intron 1, and *trnG* intron spacer datasets were the most similar to the best plastome tree in terms of branch lengths (Table 5). Primers for PCR amplification were designed for four regions selected (Table 6).

DISCUSSION

In this study, we sequenced, assembled and annotated the plastomes of nine species of *Adenocalymma* and the plastome of *Neojobertia candolleana*. The assembled plastomes were compared with those from *C. cujete*, *O. europaea*, and *T. tetragonolobum*. Phylogenetic studies using five data partition schemes were conducted and compared in terms of topology and bootstrap support. Overall, the “coding + non-coding edited” dataset led to the best estimate of phylogenetic relationships

within the “*Adenocalymma-Neojobertia*” clade, representing the best dataset for phylogenetic studies. A search for variable regions for phylogenetic studies identified the five markers with the highest potential for species level phylogenetic studies. Primers were designed for four regions and are now available for future phylogenetic studies within the “*Adenocalymma-Neojobertia*” clade and the Bignoniaceae as a whole. These results establish a foundation for future studies on the evolution of plastome structure and phylogenomics within the Bignoniaceae.

Plastome Features

Seed plant plastomes typically encode up to 80 protein coding genes, 30 tRNAs and eight rRNAs (Wu and Chaw, 2015; Asaf et al., 2016; Reginato et al., 2016). Differences in plastome size are usually a result of IR expansions or contractions (Kim and Lee, 2004). Plastome architecture is highly conserved in Seed Plants (Odintsova and Yurina, 2003; Wicke et al., 2011; Smith and Keeling, 2015; Wu and Chaw, 2015; Reginato et al., 2016), with only a few examples of plastic genome architecture available for Angiosperms (e.g., Guisinger et al., 2011) and Gymnosperms (e.g., Wu and Chaw, 2016). The plastomes of selected members of the “*Adenocalymma-Neojobertia*” clade include similar numbers of genes than previously sequenced plastomes (Hu et al., 2015). More specifically, the plastomes of members of the “*Adenocalymma-Neojobertia*” clade include 86–87 protein coding genes, 37 tRNAs and eight rRNAs (Table 2). However, when the newly sequenced plastomes are compared with those from other Bignoniaceae (i.e., *C. cujete* and *T. tetragonolobum*), a pronounced expansion of the IRs and a contraction of the SSC were encountered, with the complete inclusion of the gene *ycf1* and part of the *rps15* in the IRs (Figure 2). Although unusual, the expansion of the IRs toward the SSC has also been reported in *Pelargonium* L'Hér. (Chumley et al., 2006) and members of Apiales (Downie and Jansen, 2015). Furthermore, a pseudogene was found in the plastomes of all species of the “*Adenocalymma-Neojobertia*” clade, with the partial loss of *rps15* from the IR of all species sampled, and complete loss of *ycf4* from the LSC in *A. biternatum* and *A. peregrinum* (Table 3). Pseudogenization events (gene duplication followed by loss of function) have been reported in several plant lineages. A notable example is the transfer of the *accD* gene

TABLE 5 | Summary statistics of the five most useful introns and intergenic spacers for phylogeny reconstruction.

Region	Alignment length (bp)	Min. seq. length (bp)	Max. seq. length (bp)	% of variation*	Topology dist.**	Branch lengths dist.**
<i>ndhA</i> intron	1173	958	1109	1	0.85	0.89
<i>petN-psbM</i>	1233	986	994	0.39	1	0.79
<i>trnG</i> intron	706	670	699	0.2	1	0.91
<i>clpP</i> intron 1	852	732	750	0.21	0.89	0.9
<i>rpl32-trnL</i>	961	863	889	0.64	0.59	0.75

Regions were selected based on a standardized mean of three variables:

(1) Percentage of variable sites;

(2) Phylogenetic tree topology distance and

(3) Phylogenetic tree branch lengths distance.

*Values standardized.

**Distances computed using Kendall and Colijn (2016) method.

TABLE 6 | Sequences of primer pairs designed in this study for selected regions.

Region	Forward primer	Reverse primer	Tm (°C)
<i>ndhA</i> intron	TCAATATCTCTACGTGCGA TTCG	CCGTCGCTATTACAGAA CCGT	59.5
<i>petN-psbM</i>	TGGGGAAGAAGTGGGC TCTA	AGTTCCTACCGCTTTTCT ACTT	59.9
<i>tmG</i> intron	TGTAGCGGGTATAGTTTAG TGGT	CGCATCGTTAGCTTGG AAGG	59
<i>clpP</i> intron 1	AGGACCGTACATGCAC CTTT	CCTCTGTTTCGCCAA GAAA	58.9

from the plastid to the nucleus of *Primula sinensis* Sabine ex Lindley (Liu et al., 2016). Pseudogenes are also common at the IRa/IRb and LSC junction regions, with loss of function due to the accumulation of premature stop codons or gene loss, which is particularly common for *ycf1* and *rps19* (Nazareno et al., 2015; Moreira et al., 2016).

The structure of the whole plastome was also found to be quite variable, with rearrangements in the LSC and IRs regions (Figure 3). This plastic architecture has also been reported for the Geraniaceae (Guisinger et al., 2011) and Mimosoid Legumes (Dugas et al., 2015). Genic regions are usually conserved, with rearrangements occurring predominantly at intergenic regions (Dugas et al., 2015). Furthermore, several genes are transcribed in operons due to the endosymbiotic origin of plastomes (Sugita and Sugiura, 1996; Sugiura et al., 1998). These gene clusters are stretches of the plastome consisting of several genes (Sugita and Sugiura, 1996; Sugiura et al., 1998), explaining the relative conserved pattern of gene groups and the frequent rearrangements that are found in spacers between gene clusters (Dugas et al., 2015).

Phylogenetic Analyses

Plastome sequences have been successfully used to address phylogenetic questions at different taxonomic scales using both protein coding and non-coding sequences (e.g., Soltis et al., 1999; Shaw et al., 2007). Here, we used plastome sequences of ten species of the “*Adenocalymma-Neojobertia*” clade and three outgroups (i.e., *C. cujete*, *O. europaea* and *T. tetragonolobum*) to reconstruct major phylogenetic affinities within this clade. We also compared five different data partition schemes in order to determine the best dataset for phylogenetic studies. The most variable regions were the introns and spacers (39.3%), with protein coding regions showing a much lower number of informative sites (27.9%) (Table 4). Higher rates of molecular evolution in intronic and intergenic regions have also been reported for several other plant groups (e.g., *Begonia* L., Harrison et al., 2015; *Epimedium*, Zhang et al., 2016; Melastomataceae, Reginato et al., 2016). There is growing evidence that organellar genomes, including plastomes, are not a direct product of natural selection, but may have been shaped by adaptive and non-adaptive processes (Lynch et al., 2006; Lynch, 2007). As a result, non-coding regions may be more prone to indel events and a

higher number of DNA substitutions when compared to coding regions.

Phylogenies were estimated using five data partitions independently. The topologies recovered using ML and BC are highly concordant, regardless of the dataset used (Figure 4, Supplementary Figure 10). The “non-coding filtered” dataset was the only data partition that led to a different topology when compared with other datasets in both criteria (Figure 4, Supplementary Figure 10). For this dataset poorly aligned regions and indels were removed using Gblocks and outlier sequences were removed using T-Coffee. However, even after a pure mechanistic approach non-homologous portions derived from rearrangements remained aligned, leading to the difference in topologies observed. Indeed, rearrangements can lead to a loss of homology correspondence in particular genomic regions which, when aligned, increase the number of gaps and “saturated” regions in sequence alignments (Castresana, 2000; Xia et al., 2003; Jeffroy et al., 2006; Misof et al., 2014). Indels and saturated regions are putatively eliminated with Gblocks (Castresana, 2000), but with some limitations to deal with rare misaligned sequences. T-Coffee was used to remove the sequences (Notredame et al., 2000), however even using different thresholds of sequence similarity some outliers remained, leading to a different topology when compared to “coding” and “non-coding edited” datasets (Figure 4, Supplementary Figure 10).

The analyses of all combined datasets (i.e., “coding + non-coding edited,” and “coding + non-coding filtered”) recovered identical topologies and similar branch lengths in all BC and ML searches (Figure 4, Supplementary Figure 10), thus revealing the importance of the phylogenetic signal of the coding regions (Figure 4, Supplementary Figure 10). However, a small increase in bootstrap support at nodes B and C is observed in the tree that resulted from the analysis of the “coding + non-coding edited” dataset (Figure 4), suggesting a decrease of phylogenetic noise in the dataset with non-homologous sequences derived from rearrangements removed by hand (Figure 4, Supplementary Figure 10) when compared with the dataset computationally edited. Overall, our results suggest that the “coding + non-coding edited” dataset is the most reliable data partition for phylogenetic estimation within the “*Adenocalymma-Neojobertia*” clade due to the greater node support (Jeffroy et al., 2006; Misof et al., 2014). In the case of inclusion of non-coding regions, alignment visual inspection is necessary to prevent non-homologous regions prevented from rearrangements being included after constructing the datasets by sequence similarity.

Identification of Markers for Species Level Phylogenetic Studies

The genomic data obtained in this study allowed us to identify the four most promising plastome regions for phylogeny reconstruction within the “*Adenocalymma-Neojobertia*” clade. Despite the limited sampling (approximately 15% of the known species), the sampled taxa cover the breadth of morphological diversity found within the “*Adenocalymma-Neojobertia*” clade and are broadly distributed through the phylogeny of this clade (Fonseca and Lohmann, in prep.). Therefore, the regions selected

likely represent good markers for phylogeny reconstruction within the whole clade. Among the regions selected, the *ndhA* intron also showed a high potential for phylogeny reconstruction in the Melastomataceae (Reginato et al., 2016), *rpl32-trnL* is an intergenic region widely used among angiosperms (Shaw et al., 2007). The *rpl32-trnL* marker has been successfully used in phylogenetic studies within the Bignoniaceae (Fonseca and Lohmann, 2015; Medeiros and Lohmann, 2015). While high-throughput sequencing methods allow the generation of an enormous amount of data, budget and computational limitations can reduce the taxonomic coverage of studies of this nature. To ease some of these limitations, a hybrid NGS and Sanger sequencing approach is recommended and has been successfully used to reconstruct the phylogeny of a variety of plant lineages, including the Malpighiales (Xi et al., 2012), Arundinarieae-Poaceae (Ma et al., 2014), and Goodeniaceae (Gardner et al., 2016). Indeed, a combination of NGS and Sanger data may represent the most cost-efficient approach to estimate species-level phylogenies.

CONCLUSIONS AND FUTURE DIRECTIONS

Ten full plastomes of species from the “*Adenocalymma-Neojobertia*” clade led to a strongly supported phylogeny for this lineage. The plastomes assembled also allowed the identification of four suitable molecular markers for future phylogenetic studies. The plastic nature of the genomic architecture of members of this clade has direct implications for plastome assembly. More specifically, the recurrent rearrangements indicate the importance of *de novo* strategies for plastome assembly. Given that rearrangements occur even between closely related species, reference based approaches are not advisable. This variable architecture also has implications for phylogenomics as the lack of correspondence between gene junctions leads to problematic sequence alignments and errors in sequence homology assessment. The resulting bias can be reduced by the exclusion of poorly aligned regions.

The results derived from this study also serve as basis for future phylogenetic work within the “*Adenocalymma-Neojobertia*” clade. Ongoing studies, based on a broader sampling of taxa (approximately 90% of the known species of the “*Adenocalymma-Neojobertia*” clade) and a combination of Sanger and NGS sequencing data, aim to reconstruct a comprehensive phylogeny for the whole clade (Fonseca and Lohmann, in prep.). A robust phylogeny of this taxonomically complicated group, based a broad sample of taxa and markers,

is critical to evaluate the monophyly of taxa, identify potential morphological synapomorphies for lineages, and subsidize taxonomic studies in this group (Fonseca and Lohmann, in prep.).

Our results also have major implications for broader phylogenetic studies within the whole Bignoniaceae. More specifically, the four molecular markers identified as suitable for phylogenetic studies within the “*Adenocalymma-Neojobertia*” clade, could also be used to reconstruct phylogenetic relationships within the whole family. A broad phylogeny is already available for the Bignoniaceae (Olmstead et al., 2009), however support of deeper relationships could be substantially improved by an increase in sampling of taxa and markers. A robust phylogeny of the whole Bignoniaceae is critical for an improved understanding of the biogeographic and evolutionary history of this ecologically diverse clade of Neotropical trees, shrubs and lianas (Gentry, 1980).

AUTHOR CONTRIBUTIONS

LF and LL conceived and designed the experiment, collected the materials, and wrote the paper. LF performed the experiments, assembled sequences, and analyzed the data.

ACKNOWLEDGMENTS

The authors thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), ASPT (American Society of Plant Taxonomists), BSA (Botanical Society of America), and IAPT (International Association of Plant Taxonomists) for graduate fellowships or research grants to LF as well as CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for a Pq-1C grant to LL (307781/2013-5), and FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) for a regular research grant to LL (2011/50859-2), and a collaborative FAPESP-NSF-NASA grant to LL (2012/50260-6). We also thank Monica Carlsen for assistance with library preparation, Michael McKain for allowing us to use the unpublished Fast-Plast pipeline, and the Core Facility for Scientific Research from the Universidade de São Paulo (CEFAP-USP/GENIAL) for allowing us to use the Covaris S2 sonicator, Qubit and SEAL server.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2017.01875/full#supplementary-material>

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Asaf, S., Khan, A. L., Khan, A. R., Waqas, M., Kang, S.-M., Khan, M. A., et al. (2016). Complete chloroplast genome of *Nicotiana otophora* and its comparison with related species. *Front. Plant Sci.* 7:843. doi: 10.3389/fpls.2016.00843
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Bock, R. (2007). Structure, function, and inheritance of plastid genomes. *Top. Curr. Genet.* 19, 29–63. doi: 10.1007/4735_2007_0223

- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 5, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334
- Chumley, T., Palmer, J., Mower, J., Fourcade, H., Calie, P., Boore, J., et al. (2006). The complete chloroplast genome sequence of *Pelargonium x hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol. Biol. Evol.* 23, 2175–2190. doi: 10.1093/molbev/msl089
- Cronn, R., Liston, A., Parks, M., Gernandt, D.S., Shen, R., and Mockler, T. (2008). Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Res.* 36:e122. doi: 10.1093/nar/gkn502
- Daniell, H., Lee, S., Grevich, J., Sasaki, C., Quesada-Vargas, T., Guda, C., et al. (2006). Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. *Theor. Appl. Genet.* 112, 1503–1518. doi: 10.1007/s00122-006-0254-x
- Downie, S. R., and Jansen, R. K. (2015). A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent non-coding regions. *Syst. Bot.* 40, 336–351. doi: 10.1600/036364415X686620
- Dugas, D. V., Hernandez, D., Koenen, E. J. M., Schwartz, E., Straub, S., Hughes, C. E., et al. (2015). Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in *clpP*. *Sci. Rep.* 5:16958. doi: 10.1038/srep16958
- Fonseca, L. H. M., and Lohmann, L. G. (2015). Biogeography and evolution of *Dolichandra* (Bignoniaceae, Bignoniaceae). *Bot. J. Linnean. Soc.* 179, 403–420. doi: 10.1111/boj.12338
- Gardner, A., Sessa, E. B., Michener, P., Johnson, E., Shepherd, K. A., Howarth, D. G., et al. (2016). Utilizing next-generation sequencing to resolve the backbone of the core Goodeniaceae and inform future taxonomic and floral form studies. *Mol. Phylogenet. Evol.* 94, 605–617. doi: 10.1016/j.ympev.2015.10.003
- Gentry, A. H. (1980). *Bignoniaceae – Part I. Flora Neotropica Monographs* 25. New York, NY: New York Botanical Garden.
- Grenn, B. R. (2011). Chloroplast genomes of photosynthetic eukaryotes. *Plant J.* 66, 34–44. doi: 10.1111/j.1365-313X.2011.04541.x
- Guisinger, M. M., Kuehl, J. V., Boore, J. L., and Jansen, R. K. (2011). Extreme reconfiguration of plastid genomes in the Angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Mol. Biol. Evol.* 28, 583–600. doi: 10.1093/molbev/msq229
- Harrison, N., Harrison, R. J., and Kidner, C. A. (2015). Comparative analysis of *Begonia* plastid genomes and their utility for species-level phylogenetics. *PLoS ONE* 11:e0153248. doi: 10.1371/journal.pone.0153248
- Hu, S., Sablok, G., Wang, B., Qu, D., Barbaro, E., and Viola, R. (2015). Plastome organization and evolution of chloroplast genes in *Cardamine* species adapted to contrasting habitats. *BMC Genom.* 16:1. doi: 10.1186/s12864-015-1498-0
- Jansen, R. K., Sasaki, C., Lee, S., Hansen, A. K., and Daniell, H. (2011). Complete plastid genome sequences of three Rosids (*Castanea*, *Prunus*, *Theobroma*): evidence for at least two independent transfers of *rpl22* to the nucleus. *Mol. Biol. Evol.* 28, 835–847. doi: 10.1093/molbev/msq261
- Jeffroy, O., Brinkmann, H., Delsuc, F., and Philippe, H. (2006). Phylogenomics: the beginning of incongruence? *Trends Genet.* 22, 225–231. doi: 10.1016/j.tig.2006.02.003
- 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
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199
- Kendall, M., and Colijn, C. (2016). Mapping phylogenetic trees to reveal distinct patterns of evolution. *Mol. Biol. Evol.* 33, 2735–2743. doi: 10.1093/molbev/msw124
- Kim, K. J., and Lee, H. L. (2004). Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* 11, 247–261. doi: 10.1093/dnares/11.4.247
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Liu, T. J., Zhang, C. Y., Yan, H. F., Zhang, L., Ge, X. J., and Hao, G. (2016). Complete plastid genome of *Primula sinensis* (Primulaceae): structure, comparison, sequence variation and evidence for *accD* transfer to nucleous. *PeerJ.* 4:e2101. doi: 10.7717/peerj.2101
- Lohmann, L.G. (2006). Untangling the phylogeny of Neotropical lianas (Bignoniaceae, Bignoniaceae). *Am. J. Bot.* 93, 304–318. doi: 10.3732/ajb.93.2.304
- Lohmann, L. G., and Taylor, C. (2014). A new generic classification of tribe Bignoniaceae (Bignoniaceae). *Ann. Missouri Bot. Gard.* 99, 348–489. doi: 10.3417/2003187
- Lohse, M., Drechsel, O., and Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* 52, 267–274. doi: 10.1007/s00294-007-0161-y
- Lynch, M. (2007). *The Origins of Genome Architecture*. Sunderland: Sinauer.
- Lynch, M., Koskella, B., and Schaack, S. (2006). Mutation pressure and the evolution of organelle genomic architecture. *Science* 311, 1727–1730. doi: 10.1126/science.1118884
- Ma, P. F., Zhang, Y. X., Zeng, C. X., Guo, Z. H., and Li, D. Z. (2014). Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable Bamboo tribe Arundinarieae (Poaceae). *Syst. Biol.* 63, 933–950. doi: 10.1093/sysbio/syu054
- Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* 27, 764–770. doi: 10.1093/bioinformatics/btr011
- Medeiros, M. C. M. P., and Lohmann, L. G. (2015). Phylogeny and biogeography of *Tynanthus* Miers (Bignoniaceae, Bignoniaceae). *Mol. Phylogenet. Evol.* 85, 32–40. doi: 10.1016/j.ympev.2015.01.010
- Misof, B., Meusemann, K., von Reumont, B. M., Kück, P., Prohaska, S. J., and Stadler, P. F. (2014). A priori assessment of data quality in molecular phylogenetics. *Algorithms Mol. Biol.* 9, 22. doi: 10.1186/s13015-014-0022-4
- Moore, M. J., Soltis, P. S., Bell, C. D., Burleigh, J. G., and Soltis, D. E. (2010). Phylogenetic analysis of 83 plastid genes further resolves the early diversification of Eudicots. *Proc. Natl Acad. Sci. U.S.A.* 107, 4623–4628. doi: 10.1073/pnas.0907801107
- Moreira, P. A., Mariac, C., Scarcelli, N., Couderc, M., Rodrigues, D. P., Clement, C. R., et al. (2016). Chloroplast sequence of tregourd (*Crescentia cujete*, Bignoniaceae) to study phylogeography and domestication. *Appl. Plant Sci.* 4:1600048. doi: 10.3732/apps.1600048
- Nazareno, A. G., Carlsen, M., and Lohmann, L. G. (2015). Complete chloroplast genome of *Tanaecium tetragonolobum*: the first Bignoniaceae plastome. *PLoS ONE* 10:e0129930. doi: 10.1371/journal.pone.0129930
- Notredame, C., Higgins, D. G., and Heringa, J. (2000). T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205–217. doi: 10.1006/jmbi.2000.4042
- Odintsova, M. S., and Yurina, N. P. (2003). Plastid genomes of higher plants and algae: structure and function. *Mol. Biol.* 37, 649–662. doi: 10.1023/A:1026020623631
- Olmstead, R. G., Zjhra, M. L., Lohmann, L. G., Grose, S. O., and Eckert, A. J. (2009). A molecular phylogeny and classification of Bignoniaceae. *Am. J. Bot.* 96, 1731–1743. doi: 10.3732/ajb.0900004
- 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 sequencing of chloroplast genomes. *BMC Biol.* 7:84. doi: 10.1186/1741-7007-7-84
- Qian, J., Song, J., Gao, H., Zhu, Y., Xu, J., Pang, X., et al. (2013). The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS ONE* 8:e57607. doi: 10.1371/journal.pone.0057607
- R Development Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: The R Foundation for Statistical Computing. Available online at: <http://www.R-project.org/>.
- Reginato, M., Neubig, K. M., Majure, L. C., and Michelangeli, F. (2016). The first complete plastid genomes of Melastomataceae are highly structurally conserved. *Peer J.* 4:e2715. doi: 10.7717/peerj.2715

- Ronquist, F., Teslenko, M., Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2011). MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Shaw, J., Lickey, E. B., Beck, J., Farmer, S. B., Liu, W., Miller, J., et al. (2005). The tortoise and the hare II: relative utility of 21 non-coding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* 92, 142–166. doi: 10.3732/ajb.92.1.142
- Shaw, J., Lickey, E. B., Schilling, E. E., and Small, R. L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am. J. Bot.* 94, 275–288. doi: 10.3732/ajb.94.3.275
- Small, R. L., Ryburn, J. A., Cronn, R. C., Seelanan, T., and Wendel, J. F. (1998). The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. *Am. J. Bot.* 85, 1301–1315. doi: 10.2307/2446640
- Smith, D. R., and Keeling, P. J. (2015). Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10177–10184. doi: 10.1073/pnas.1422049112
- Soltis, P. S., Soltis, D. E., and Chase, M. W. (1999). Angiosperms phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402, 402–404. doi: 10.1038/46528
- 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
- Sugita, M., and Sugiura, M. (1996). Regulation of gene expression in chloroplast of higher plants. *Plant Mol. Biol.* 32, 315–326. doi: 10.1007/BF00039388
- Sugiura, M. (1992). The chloroplast genome. *Plant Mol. Biol.* 19, 149–168. doi: 10.1007/BF00015612
- Sugiura, M. (1995). The chloroplast genome. *Essays Biochem.* 30, 49–57.
- Sugiura, M., Hirose, T., and Sugita, M. (1998). Evolution and mechanism of translation in chloroplast. *Annul. Rev. Genet.* 32, 437–459. doi: 10.1146/annurev.genet.32.1.437
- Vallejo-Marín, M., Cooley, A. M., Lee, M. Y., Folmer, M., McKain, M. R., and Puzey, J. R. (2016). Strongly asymmetric hybridization barriers shape the origin of a new polyploid species and its hybrid ancestor. *Am. J. Bot.* 103, 1272–1288. doi: 10.3732/ajb.1500471
- 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
- Wise, R. (2006). “The diversity of plastid form and function,” in *The Structure and Function of Plastids*, eds R. Wise, and J. K. Hooper (New York, NY: Springer press), 2–25.
- Wu, C.-S., and Chaw, S.-M. (2015). Evolutionary stasis in *Cycad* plastomes and the first case of plastome GC-biased gene conversion. *Genome. Biol. Evol.* 7, 2000–2009. doi: 10.1093/gbe/evv125
- Wu, C.-S., and Chaw, S.-M. (2016). Large-Scale comparative analysis reveals the mechanisms driving plastomic compaction, reduction, and inversions in Conifers II (Cupressophytes). *Genome. Biol. Evol.* 8, 3740–3750. doi: 10.1093/gbe/evw278
- Wyman, S. K., Jansen, R. K., and Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20, 3252–3255. doi: 10.1093/bioinformatics/bth352
- Wysocki, W. P., Clark, L. G., Attigala, L., Ruiz-Sanchez, E., and Duvall, M.R. (2015). Evolution of the bamboos (Bambusoideae, Poaceae): a full plastome phylogenetic analysis. *BMC Evol. Biol.* 15:50. doi: 10.1186/s12862-015-0321-5
- Xi, Z., Ruhfel, B. R., Schaefer, H., Amorim, A. M., Sugumaran, M., Wurdack, K. J., et al. (2012). Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17519–17524. doi: 10.1073/pnas.1205818109
- Xia, X., Xie, Z., Salemi, M., Chen, L., and Wang, Y. (2003). An index of substitution saturation and its application. *Mol. Phylogenet. Evol.* 26, 1–7. doi: 10.1016/S1055-7903(02)00326-3
- Yi, D. K., and Kim, K. J. (2012). Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. *PLoS ONE* 7:e35872. doi: 10.1371/journal.pone.0035872
- Zhang, Y., Du, L., Liu, A., Chen, J., Wu, L., Hu, W., et al. (2016). The complete chloroplast genome sequences of five *Epimedium* species: lights into phylogenetic and taxonomic analyses. *Front. Plant Sci.* 7:306. doi: 10.3389/fpls.2016.00306

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Fonseca and Lohmann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.