Molecular Phylogenetics and Evolution 107 (2017) 551–563

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogenetic relationships of *Burmeistera* (Campanulaceae: Lobelioideae): Combining whole plastome with targeted loci data in a recent radiation

Simon Uribe-Convers [a,*], Monica M. Carlsen [a,1], Laura P. Lagomarsino [a,b], Nathan Muchhala [a]

[a] University of Missouri-St. Louis, Biology Department, One University Blvd., Research Building, St. Louis, MO 63121, USA
[b] University of Gothenburg, Department of Biological and Environmental Sciences, Carl Skottsbergs gata 22B, P.O. Box 461, SE 405 30 Göteborg, Sweden

ARTICLE INFO

Article history:
Received 23 September 2016
Revised 6 December 2016
Accepted 9 December 2016
Available online 21 December 2016

Keywords:
Andes
Bellflowers
Character evolution
Genome skimming
Plastome
Phylogenomics

ABSTRACT

The field of molecular systematics has benefited greatly with the advent of high-throughput sequencing (HTS), making large genomic datasets commonplace. However, a large number of targeted Sanger sequences produced by many studies over the last two decades are publicly available and should not be overlooked. In this study, we elucidate the phylogenetic relationships of the plant genus *Burmeistera* (Campanulaceae: Lobelioideae), while investigating how to best combine targeted Sanger loci with HTS data. We sequence, annotate, and analyze complete to nearly complete plastomes for a subset of the genus. We then combine these data with a much larger taxonomic dataset for which only Sanger sequences are available, making this the most comprehensively sampled study in the genus to date. We show that using a phylogeny inferred from the species with plastome data as a topological constraint for the larger dataset increases the resolution of our data and produces a more robust evolutionary hypothesis for the group. We then use the resulting phylogeny to study the evolution of morphological traits thought to be important in *Burmeistera*, and assess their usefulness in the current taxonomic classification of the genus. The main morphological character used to delimit subgeneric sections, the presence or absence of hairs on the apex of the two ventral anthers, shows a complex evolutionary history with many changes in the tree, suggesting that this character should not be used for taxonomic classification. Although it is too soon to propose a new subgeneric classification for *Burmeistera*, our results highlight some morphological traits shared by whole clades that could potentially be used in future taxonomic work.

1. Introduction

The advent of high-throughput sequencing (HTS) has rendered the generation of large-scale molecular datasets significantly easier, and greatly increased the use of multilocus nuclear datasets (e.g., Schmickl et al., 2016; Uribe-Convers et al., 2016; Weitemier et al., 2014) as well as complete chloroplast genomes (e.g., Knox, 2014; Nazareno et al., 2015; Uribe-Convers et al., 2014). However, in the decades prior to HTS systematists amassed large quantities of molecular sequence data for many taxa via targeted Sanger sequencing (e.g., Chase et al., 1993; Graham and Olmstead, 2000; Shaw et al., 2005; Soltis et al., 2011); data which is typically publically available and can still be useful to systematic studies to expand information per species and/or taxonomic coverage of clades. In this study, we take advantage of the benefits of both HTS and the large pool of molecular data available for *Burmeistera* Triana (Campanulaceae: Lobelioideae) to infer phylogenetic relationships in the genus using both plastomes and newly generated and publicly available Sanger sequences.

Complete chloroplast genomes, also known as plastomes, have become particularly dominant in plant molecular systematics for several reasons. Although plastomes have slower rates of molecular evolution than the nuclear genome (Wolfe et al., 1987; Wolfe et al., 1989), plastid sequences have been the workhorse of plant molecular systematics for more than three decades (e.g., Chase et al., 1993), and have been informative at all taxonomic levels (Crowl et al., 2016; Downie and Palmer, 1992; Marx et al., 2010; Moore et al., 2007; Nürk et al., 2015; Parks et al., 2009;

* Corresponding author.
E-mail addresses: uribe.convers@gmail.com (S. Uribe-Convers), monica.carlsen@mobot.org (M.M. Carlsen), lagomarsino.l@gmail.com (L.P. Lagomarsino), muchhalan@umsl.edu (N. Muchhala).
[1] Present address: Smithsonian Institution, Botany Department, National Museum of Natural History, Washington, D.C. 20560, USA.

http://dx.doi.org/10.1016/j.ympev.2016.12.011
1055-7903/

Uribe-Convers and Tank, 2015; Uribe-Convers and Tank, 2016). Furthermore, the haploid nature and uniparental inheritance of the plastome provide relatively straightforward phylogenetic inference, particularly in the face of gene tree-species tree discordance due to coalescent stochasticity and/or hybridization. Moreover, their high copy numbers per cell have made sequencing plastid regions, and more recently full plastomes, easy and even trivial. They are often obtained using a genome skimming approach (Straub et al., 2011; Straub et al., 2012) in which HTS reads belonging to the plastome are 'skimmed' out of a low coverage genome (i.e., a shotgun library) by mapping them to a plastid reference. Genome skimming recovers high quality plastomes at relatively low sequencing depth, allowing many samples to be multiplexed in a single sequencing run (Straub et al., 2012).

Burmeistera is distributed from Guatemala to Peru, with its highest species richness in the northern Andes of Colombia and Ecuador, where approximately half of the species occur (Garzón Venegas et al., 2014). The genus comprises ~116 herbaceous to suffruticose species, 77 of which were recognized in the earliest monograph (Wimmer, 1943). They are mostly found in the understory of cloud forests, where they can be terrestrial, scandent, or hemi-epiphytic. Their often green flowers are primarily pollinated by nectar bats (Phyllostomidae), and the pollination biology of many species has been well-documented (Muchhala, 2003, 2006, 2008; Muchhala and Potts, 2007). Members of the genus have been included in previous phylogenetic studies focused on the Lobelioideae subfamily of Campanulaceae (Antonelli, 2008; Knox et al., 2008; Lagomarsino et al., 2014). The monophyly of *Burmeistera* has been recovered with high support in these studies, and is further bolstered by a series of morphological synapomorphies—most notably, their dilated anther orifice (Knox et al., 2008; Lagomarsino et al., 2014). The genus is closely related to green-flowered species of *Centropogon* C. Presl in the burmeisterid subclade of the centropogonids (Lagomarsino et al., 2014). Additionally, species-level taxonomy within the genus has received abundant attention in recent years (Garzón Venegas and González, 2012; Garzón Venegas et al., 2012; Garzón Venegas et al., 2014; Garzón Venegas et al., 2013; González and Garzón Venegas, 2015; Lagomarsino et al., 2015; Muchhala and Lammers, 2005; Muchhala and Pérez, 2015). *Burmeistera* was originally divided into two taxonomic sections based on the pubescence of the anthers (Wimmer, 1943): section *Barbatae* E. Wimm., characterized by tufted hairs on the apex of the two ventral anthers, and section *'Imberbes'* nom. invalid, in which all five anthers are either glabrous or soft-pubescent at the apex. The latter name is invalid because it includes the type species of the genus (*B. ibaguensis* Triana & H. Karst.), and under Article 22.1 of the International Code of Botanical Nomenclature (McNeill et al., 2012), an infrageneric taxon that includes the type of the genus must be an autonym of the genus name (i.e., *Burmeistera* section *Burmeistera*). Regardless, previous work suggests that these two sections do not in fact form clades (Knox et al., 2008; Lagomarsino et al., 2014) and, thus, are not a good representation of the evolutionary history within *Burmeistera*.

In this study, we were able to include all molecular data currently available for *Burmeistera* by incorporating newly sequenced and previously published plastid regions generated via Sanger sequencing with 19 complete plastomes. This approach maximizes the taxonomic representation of the genus. Using these data, we generated three different molecular matrices with varying degrees of missing data and analyzed them to assess the impact of missing data on our phylogenetic inferences. We then reconstructed the ancestral states of various characters along the phylogeny, identifying possible morphological traits that will be useful for future taxonomic revision. Although *Burmeistera* has been included in previous studies, this is the first to include HTS-scale data and, further, has the densest taxon sampling to date.

2. Materials and methods

2.1. Taxon sampling

The primary goal of this study is to increase the phylogenetic resolution in *Burmeistera* by maximizing both the number of taxa sampled and the amount of molecular data. Towards this end, we assembled data for two outgroup species (*Centropogon nigricans* Zahlbr. and *Siphocampylus krauseanus* E.Wimm.) and 45 species of *Burmeistera* that span the taxonomic, geographic, and morphological range of the genus. The final dataset includes 19 newly-generated complete or nearly complete plastomes and seven individual plastid loci with known phylogenetic utility (two newly-generated, and five publically available on GenBank). Voucher and source information for each sample can be found in Table 1. Field-collected tissue samples were stored in airtight bags filled with silica gel desiccant for DNA extractions.

2.2. Molecular methods

2.2.1. Targeted Sanger sequencing

Total genomic DNA was extracted from 0.02 g of silica gel-dried tissue using a modified 2× CTAB method (Doyle and Doyle, 1987; Uribe-Convers and Tank, 2015). Two plastid regions that harbor rapidly evolving genes (Knox, 2014), *ndhC–psaI* and *trnQ–rps16*, were amplified via polymerase chain reaction (PCR) using the primers 234F and 235R for the former and 310F and 311R for the latter (Supplementary Table 1). The PCR program used for amplification was 95 °C for 4 min, then 35 cycles of 95 °C for 1 min, 53 °C for 1 min, 72 °C for 1.5 min, followed by a final extension of 72 °C for 5 min. Sequencing of the PCR products was done on a ABI 3130xl capillary DNA sequencer (Applied Biosystems, Foster City, California, USA) using ABI BigDye v3.1 cycle sequencing chemistry. We sequenced both strands of the PCR products to ensure accuracy, using the same PCR primers and/or various internal primers where needed (see Supplementary Table 1). Raw sequence data were assembled and edited using Geneious R7 v.7.1.9 (http://www.geneious.com, Kearse et al., 2012) or Sequencher v.5.3.2 (Gene Codes Corp., Ann Arbor, Michigan, USA), and consensus sequences were generated and submitted to GenBank (Table 1). Additional Sanger sequences used in this study were obtained from GenBank and belong to five plastid loci: the *atpB* and *rbcL* genes and their intergenic spacer, and the intergenic spacers between *rps16-trnK*, *trnS (GCU)-trnG (UCC)*, *rpl32-trnL (UAG)*, and *rpl32-ndhF*.

2.2.2. High-throughput plastome sequencing

We sequenced and assembled 19 plastomes, including 17 species of *Burmeistera* and two outgroup species (*Centropogon nigricans* and *Siphocampylus krauseanus*). The species used for high-throughput sequencing were chosen to include as much taxonomic, morphological, and geographical variation from the genus. The taxa used for HTS were not included in the Sanger sequencing experiments. Total genomic DNA was fragmented using a sonicator and 300 bp insert libraries were constructed using NEBNext Ultra DNA Library Prep Kit and Multiplex oligos for Illumina (New England BioLabs Inc., Ipswich, MA, USA) following the manufacturer's recommendations. DNA library concentration and expected size were verified using an Agilent High Sensitivity (HS) DNA Kit run on an Agilent 2100 BioAnalyzer (Agilent Technologies Inc., Santa Clara, CA, USA), as well as a Qubit dsDNA HS Assay Kit run on a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA). The 19 libraries were diluted to a concentration of 10 nM, pooled together in a single lane, and sequenced (100 bp single-end) on an Illumina HiSeq 2000 system (Illumina Inc., San Diego, CA,

Table 1
Taxa and voucher information for plant material from which DNA was extracted. Sequences produced in this study are denoted by an asterisk. GenBank accessions for sequences not generated in this study are also given.

Genus	Species	Author	Collection Voucher	Origin	GenBank accessions							
					Complete Plastome	*atpB-rbcL*	*rps16-trnK*	*trnQ–rps16*	*trnS-trnG-trnG*	*ndhC–psaI*	*rpl32-trnL*	*rpl32-ndhF*
Burmeistera	*aff bullatifolia*	J. Garzón & F. Gonzalez	Bacon 264A (GB)	Colombia	–	–	KP014501	–	–	–	KP014167	KP014861
Burmeistera	*almedae*	Wilbur	Lagomarsino 86 (GH)	Costa Rica	–	–	KP014511	KX396048*	–	KX396039*	KP014195	KP014848
Burmeistera	*auriculata*	Muchhala & Lammers	Muchhala 120 (QCA)	Ecuador	KY045495	–	–	–	–	–	–	–
Burmeistera	*borjensis*	Jeppesen	Muchhala 154 (QCA)	Ecuador	KY045494	–	–	–	–	–	–	–
Burmeistera	*ceratocarpa*	Zahlbr.	Muchhala 177 (QCA)	Ecuador	KY045493	–	–	–	–	–	–	–
Burmeistera	*chiriquiensis*	Wilbur	Santamaria 8996 (INB)	Costa Rica	–	–	KP014521	–	KP014754	–	KP014192	KP014858
Burmeistera	*crassifolia*	E. Wimm.	Muchhala 121 (QCA)	Ecuador	–	EF174659.1	–	–	–	–	–	–
Burmeistera	*crebra*	McVaugh	Lagomarsino 70 (GH)	Costa Rica	–	–	KP014507	KX396044*	KP014758	KX396035*	KP014234	KP014870
Burmeistera	*crispiloba*	Zahlbr.	Muchhala 123 (QCA)	Ecuador	KY045492	–	–	–	–	–	–	–
Burmeistera	*cyclostigmata*	Donn. Sm.	Muchhala 145 (QCA)	Ecuador	KY045491	–	–	–	–	–	–	–
Burmeistera	*cylindrocarpa*	Zahlbr.	Muchhala 202 (QCA)	Ecuador	KY045490	–	–	–	–	–	–	–
Burmeistera	*dendrophila*	E. Wimm.	Lagomarsino 255 (GH)	Panama	–	–	KP014524	–	KP014750	–	KP014163	KP014855
Burmeistera	*domingensis*	Jeppesen	Muchhala 102 (QCA)	Ecuador	KY045489	–	–	–	–	–	–	–
Burmeistera	*dukei*	Zahlbr.	Clark 12662 (MO)	Panama	–	–	–	KX396041*	–	KX396031*	–	–
Burmeistera	*fuchsioides*	Garzón & F. González	Bacon 258 (GB)	Colombia	–	–	KP014503	–	–	–	KP014186	KP014865
Burmeistera	*fuscoapicata*	E. Wimm.	Cl 12955	Colombia	KY045488	–	–	–	–	–	–	–
Burmeistera	*glabrata*	(Kunth) Benth. & Hook. f. ex B.D. Jacks.	Muchhala 165 (QCA)	Ecuador	–	EF174656.1	–	–	–	–	–	–
Burmeistera	*holm-nielsenii*	Jeppesen	Muchhala 134 (QCA)	Ecuador	–	EF174647.1	–	–	–	–	–	–
Burmeistera	*loejtnantii*	Jeppesen	Luteyn 8415 (F)	Ecuador	KY045487	–	–	–	–	–	–	–
Burmeistera	*lutosa*	E. Wimm.	Muchhala 133 (QCA)	Ecuador	KY045486	–	–	–	–	–	–	–
Burmeistera	*mcvaughii*	Wilbur	Lagomarsino 257 (GH)	Panama	–	–	KP014526	–	KP014768	–	KP014166	KP014853
Burmeistera	*microphylla*	Donn. Sm.	Lagomarsino 34 (GH)	Costa Rica	–	–	KP014350	KX396045*	KP014720	KX396033*	KP014187	KP014850
Burmeistera	*minutiflora*	Garzón & F. González	Bacon 262 (GB)	Colombia	–	–	KP014508	–	KP014712	–	KP014235	KP014869
Burmeistera	*morii*	Wilbur	Lagomarsino 273 (GH)	Panama	–	–	KP014525	–	KP014771	–	KP014165	KP014864
Burmeistera	*multiflora*	Zahlbr.	Muchhala 114 (QCA)	Ecuador	–	–	KP014530	–	KP014583	–	KP014160	KP014879
Burmeistera	*obtusifolia*	E. Wimm.	Lagomarsino 61 (GH)	Costa Rica	–	–	KP014510	KX396047*	–	–	KP014197	KP014969
Burmeistera	*oyacachensis*	Jeppesen	Muchhala 171 (QCA)	Ecuador	KY045485	–	–	–	–	–	–	–
Burmeistera	*panamensis*	Wilbur	Muchhala 494 (MO)	Panama	–	–	–	KX396042*	–	KX396034*	–	–
Burmeistera	*parviflora*	E. Wimm. ex Standl.	Lagomarsino 53 (GH)	Costa Rica	KY045484	–	–	–	–	–	–	–
Burmeistera	*pirrensis*	Wilbur	Muchhala 491 (HUA)	Colombia	KY045483	–	–	–	–	–	–	–
Burmeistera	*racemiflora*	Lammers	Madison 7145 (AAU)	Ecuador	–	–	KP014504	–	KP014760	–	KP014170	KP014845
Burmeistera	*refracta*	E. Wimm.	Muchhala 110 (QCA)	Ecuador	–	EF174653.1	KP014494	–	KP014711	–	KP014176	KP014880
Burmeistera	*resupinata*	Zahlbr.	Muchhala 223 (QCA)	Ecuador	KY045482	–	–	–	–	–	–	–
Burmeistera	*rubrosepala*	(E. Wimm.) E. Wimm	Muchhala 220 (QCA)	Ecuador	KY045481	–	–	–	–	–	–	–
Burmeistera	*smaragdi*	Lammers	Muchhala 205 (QCA)	Ecuador	KY045480	–	–	–	–	–	–	–
Burmeistera	*sodiroana*	Zahlbr.	Muchhala 115 (QCA)	Ecuador	KY045479	–	–	–	–	–	–	–
Burmeistera	*succulenta*	H. Karst. & Triana	Bacon 297 (GB)	Colombia	–	–	KP014529	–	–	–	KP014240	KP014866
Burmeistera	*tenuiflora*	Donn. Sm.	Lagomarsino 276 (GH)	Panama	–	–	KP014509	–	KP014721	–	KP014184	KP014881
Burmeistera	*toroensis*	Wilbur	Lagomarsino 92 (GH)	Panama	–	–	KP014497	–	KP014709	–	KP014241	KP014868
Burmeistera	*truncata*	Zahlbr.	Muchhala 201 (QCA)	Ecuador	–	EF174649.1	–	–	–	–	–	–
Burmeistera	*utleyi*	Wilbur	Lagomarsino 253 (GH)	Panama	–	–	KP014523	–	KP014751	–	KP014164	KP014856
Burmeistera	*variabilis*	(Gleason) E. Wimm.	Bacon 298 (GB)	Colombia	–	KU670751	KP014505	–	KP014759	–	KP014169	KP014846

(*continued on next page*)

Table 1 (*continued*)

Genus	Species	Author	Collection Voucher	Origin	GenBank accessions							
					Complete Plastome	*atpB-rbcL*	*rps16-trnK*	*trnQ–rps16*	*trnS-trnG-trnG*	*ndhC–psaI*	*rpl32-trnL*	*rpl32-ndhF*
Burmeistera	*vulgaris*	E. Wimm.	Santamaria S-980 (GB)	Costa Rica	–	–	KP014495	–	KP014714	–	KP014174	KP014872
Burmeistera	*zamorensis*	Muchhala & A. J. Pérez	Muchhala 467 (QCA)	Ecuador	–	–	–	–	–	KX396036*	–	–
Burmeistera	*zurquiensis*	Wilbur	Lagomarsino 26 (GH)	Costa Rica	–	KU670805	KP014512	KX396049*	KP014702	KX396040*	KP014193	KP014849
Centropogon	*nigricans*	Zahlbr.	Muchhala 118 (QCA)	Ecuador	KY045478	–	–	–	–	–	–	–
Siphocampylus	*krauseanus*	E. Wimm.	Smith and Leiva 509 (MO)	Peru	KY045477	–	–	–	–	–	–	–

* denotes that this sequence was generated in this study.

USA) at the University of Missouri – Columbia DNA Core Facility. The raw sequencing reads were deposited in NCBI's Sequence Read Archive (SRA) (Accession numbers: SRR4253254–SRR4253273).

2.2.3. Plastome assembly and annotation

Following a combination of assembly techniques (Nazareno et al., 2015; Welker et al., 2016), single-end Illumina raw reads were excised from adaptors and barcodes using cutadapt (Martin, 2011) and then quality-filtered with custom Perl scripts (available at https://github.com/mrmckain/Fast-Plast) that trimmed reads from the ends until there were three consecutive bases with a Phred quality score >20, and then removed all reads that were shorter than 40 bp, had a median quality score $\leqslant$21, or had more than three uncalled bases. Resulting reads were assembled into contigs *de novo* using Velvet v2.3 (Zerbino and Birney, 2008) with a K-mer length of 71 and with SPAdes v3.6.1 (Bankevich et al., 2012) with K-mer lengths of 55, 77, and 89. To exclude mitochondrial and nuclear DNA, we used BLAST (Altschul et al., 1990) to remove contigs that did not align to a plastome reference (*Brighamia insignis* A. Gray, GenBank accession KT372780). The remaining contigs were assembled into larger contiguous sequences in the program Sequencher, based on a minimum overlap of 20 bp and 98% similarity. We tried to fill gaps between contigs that did not assemble in the initial set of analyses using an iterative plastome-walking procedure. This was done by searching the trimmed reads for sequences with identical similarity to the ends of contigs (using the "grep" command in UNIX) and assembling those reads to the contigs within Sequencher. The quality of the assembly was assessed based on consistency of sequencing depth across the entire plastome. Using Jellyfish (Marçais and Kingsford, 2011), we created a 20-kmer count look-up table and visualized significant changes in sequencing depth with a custom script (available at https://github.com/mrmckain/Chloroplast-Genome-Assembly). Based on abrupt changes in sequencing depth, we identified the boundaries between the inverted repeats (IR) and the single copy regions, i.e., the Large Single Copy (LSC) and Small Single Copy (SSC) regions. Final plastomes were annotated in Geneious using *Brighamia insignis* and two other closely-related species (*Centropogon granulosus* C.Presl and *Lobelia polyphylla* Hook. & Arn., Knox E, unpublished) as references, and a circular representation was visualized in GenomeVx (Conant and Wolfe, 2008). Finally, gene content, order, and variability were analyzed in Geneious and R (R Core Team, 2016).

2.3. Phylogenetic inference

2.3.1. Complete plastomes

We began phylogenetic analyses by inferring the evolutionary relationships among the 19 taxa with fully-sequenced plastomes. These plastomes were trimmed to include only one copy of the IR and aligned using MAFFT v.7.272 (Katoh and Standley, 2013) in its default settings. The resulting alignment was visually inspected in Geneious. Based on gene annotations, we created a partition scheme consisting of 182 distinct partitions—of which 79 were coding regions and 103 were non-coding. However, to reduce the number of parameters in downstream phylogenetic analyses, we used PartitionFinder (Lanfear et al., 2012; Lanfear et al., 2014) under the Bayesian information criterion (BIC) to optimize our partitioning scheme while simultaneously identifying the best-fit model of sequence evolution for each partition. To analyze the partitioned dataset in a Bayesian framework, we used MrBayes v.3.2.5 (Ronquist et al., 2012) with individual parameters unlinked across the data partitions. We ran two independent runs for 100 million generations with eight Markov chains each, using default priors and heating values. The runs were started from a randomly generated tree and were sampled every 1000

generations. Convergence of the chains was determined by analyzing the plots of all parameters and the −lnL values using the program Tracer v1.6 (Rambaut et al., 2014). Stationarity of the chains was verified when all parameter values and the −lnL had stabilized; the likelihoods of independent runs were considered indistinguishable when the average standard deviation of split frequencies was <0.001. A consensus tree was obtained using the 'sumt' command in MrBayes, after the first 25% of the trees had been discarded as burn-in. To analyze the partitioned dataset under maximum likelihood (ML), we used RAxML v.8.1.21 (Stamatakis, 2014) under the GTRCAT model with 1000 replicates of nonparametric bootstrapping using the rapid bootstrap algorithm (Stamatakis et al., 2008). Every fifth bootstrap tree generated by the rapid bootstrap analyses was used as a starting tree for full ML searches, and the trees with the highest ML scores were chosen.

2.3.2. Integration of whole plastome and targeted Sanger sequences

To improve taxonomic coverage of *Burmeistera*, we inferred a more inclusive phylogeny that combined data from the complete plastomes described in Section 2.2.2 with the seven plastid loci described in Section 2.2.1. We extracted these seven loci from the complete plastomes, and aligned them to each corresponding Sanger locus independently with Muscle v.3.8.31 (Edgar, 2004). These new alignments were then cleaned with Phyutility v.2.2.4 (Smith and Dunn, 2008) at a 50 percent similarity threshold to minimize missing data due to ambiguously aligned sites, visually inspected in Geneious, and concatenated using Phyutility. PartitionFinder was used to infer the best partitioning scheme for the concatenated dataset and the appropriate model of sequence evolution for each partition. Extracting the seven loci from the plastomes was preferable to aligning the loci to the complete plastomes because they represent less than 7.5% of the plastome, and such large amounts of missing data can adversely impact phylogenetic inference (Lemmon et al., 2009). We analyzed this partitioned, concatenated dataset using both Bayesian inference (BI) and ML, in the programs MrBayes and RAxML, respectively. We followed the same steps described in Section 2.3.1 with the exception that each of the two MrBayes runs had four chains and was run for 20 million generations. We will refer to these resulting trees as the extracted-unconstrained phylogenies hereafter.

We conducted a second analysis to incorporate all the information contained in the whole plastome alignment, but without introducing the possible problems of large amounts of missing data. For this approach, we used the complete plastome phylogeny (Section 2.3.1) as a constraint tree for the extracted dataset, and analyzed it with RAxML using the 'constraint' option ('-r' flag). This analysis followed the same parameters and commands previously described for the unconstrained analyses, and the resulting tree will be referred to as the extracted-constrained phylogeny hereafter.

We performed a third analysis using the extracted plastid regions but this time using data from only the 19 taxa for which complete plastomes are available, which we refer to as the plastome-extracted phylogeny hereafter. The rationale behind this analysis is to assess the amount of information contained in those seven plastid regions and to compare the resulting topology to the complete plastome one. Furthermore, it allows us to study the effect of adding more data to the same set of taxa. We ran phylogenetic analyses on this dataset using both MrBayes and RaxML as described in the previous approaches (extracted-unconstrained phylogenies).

Lastly, the extracted dataset was used to infer an ultrametric phylogeny with the program BEAST v.2.3.3 (Bouckaert et al., 2014; Drummond and Rambaut, 2007; Drummond et al., 2012) in order to be able to perform ancestral state reconstructions (see Section 2.4). Because the only known macrofossil of Campanulaceae (i.e., *Campanula paleopyradimalis*; Lancucka-Srodoniowa, 1977) is quite distant from our study group and the use of secondary calibrations has been shown to produce unreliable results (Schenk, 2016), we opted to analyze this dataset to produce an ultrametric tree with relative, as opposed to absolute, ages. We used the extracted-constrained phylogeny as a topological constraint to infer relative divergence times in an analysis consisting of a single run of 20 million generations sampled every 1000 trees. The models of nucleotide substitution were unlinked for each partition, we used an unlinked relaxed lognormal clock model for each partition, and the linked tree priors were kept as default under the Yule model. Convergence of the parameters was assessed in Tracer v.1.6, and a maximum clade credibility tree was generated using TreeAnnotator v.2.3.3 (Bouckaert et al., 2014), after 25% of the trees had been discarded as burn-in. Since we did not use any calibration points in this analysis, the resulting node heights (ages) are given in relative time. Every tree produced in this study was visualized using the program FigTree v.1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

2.4. Ancestral state reconstruction

We used ancestral state reconstruction to explore the evolution of a handful of ecologically and taxonomically important morphological characters in an effort to better understand the evolution of *Burmeistera* and to identify potential synapomorphies of subclades that may aid in a future infrageneric classification of the genus. We coded character states for ten traits: anther pubescence (presence/absence of tufted hairs on the ventral two anthers), habit (terrestrial, scandent, or hemi-epiphytic); recurved petals (presence/absence); bibracteate pedicels (presence/absence); flower color (red, pink, or green); hypanthium base (rounded or tapered); hypanthium shape (globose, conical, cylindric, campanulate, hemispherical, or ovoid), inflated fruits with thin exocarp (presence/absence); and fruit color at a coarse scale (green, white, or reddish), and fruit color at a fine scale (green, white, pink, blue, yellow, or purple). Traits were coded based on field notes and photos, observations of herbarium specimens at the Missouri Botanical Garden (MO) and on the JSTOR Global Plants database (http://plants.jstor.org/), and descriptions from the taxonomic literature (Garzón Venegas and González, 2012; Garzón Venegas et al., 2013; Jeppesen, 1981; Wilbur, 1975, 1976, 1981; Wimmer, 1943). Additionally, because the shape of the hypanthium cannot be determined with certainty from herbarium specimens, hypanthium shape character states followed the already established concepts of Wimmer (1943) and Jeppesen (1981). Ancestral states were reconstructed using the ultrametric tree obtained in the BEAST analysis of the extracted dataset (Section 2.3.2), with the distantly related *Siphocampylus krauseanus* removed. We excluded *S. krauseanus* because this taxon is extremely divergent (~5 Myr) from the other species in this study, making it an inappropriate taxon with which to polarize characters within *Burmeistera*. This problem is compounded by the propensity towards homoplasy/convergent evolution within this larger group (Lagomarsino et al., 2014; Lagomarsino et al., 2016). The traits were analyzed with the 'ace' function in the R package APE (Paradis et al., 2004) under the equal rates (ER), symmetric (SYM), and all-rates-different (ARD; for traits with 3+ states) models, and a likelihood ratio test was subsequently done to determine the best-fit model.

3. Results

3.1. Taxon sampling

The 47 accessions used in this study include 45 *Burmeistera* species and two outgroup species. Sampling in *Burmeistera* has been

limited in the past (Knox et al., 2008; Lagomarsino et al., 2014) but here we have included almost 40% of the ~116 recognized species, making this the most comprehensively sampled study to date.

3.2. Plastome assembly and annotation

3.2.1. Plastome structure and variability

Burmeistera plastomes vary in size from 163,961 to 166,128 bp, with a typical quadripartite structure in which the large single-copy (LSC; 83,436–84,065 bp) and small single-copy (SSC; 7347–7469 bp) regions are separated by the inverted repeat (IR; 36,528–37,434 bp) regions, and the two outgroup species fall within the same range (Table 2). The linearized plastomes start at the *trnH (GUG)* end of the LSC, with the LSC/IR boundary located in the 5′ end of *rps19*, the IR/SSC boundary located in the middle of *ndhE*, the SSC ending downstream of *ndhF*, followed by the second copy of the IR (Fig. 1). These plastomes are collinear with *Lobelia thuliniana* E. B. Knox and many other lobelia species in a clade with two large inversions in the LSC (Knox, 2014; Knox et al., 1993; Knox and Palmer, 1999). The total GC content of these plastomes was very similar across the taxa, with a mean GC content of 39.2% (39.1–39.2%) that is slightly higher in the IR (40.7% [40.6–40.8%]) and lower in the SSC (32.7% [32.6–32.8%]).

Of the unique, linear plastome sequence (i.e., the LSC-IR-SSC, without the other copy of the IR), 53.5% encodes proteins, 1.7% encodes transfer RNAs (tRNA), and 5.5% encodes ribosomal RNAs (rRNA). The remaining 39.3% is noncoding. The coding regions contain 112 canonical plastid genes, three foreign ORFs, one nearly intact pseudogene (*rpl23*), and the pseudogene remnants of *infA*.

Sequence variability analyses between *Burmeistera* indicated that the plastomes are relatively conserved (~97.9% similarity), with some regions showing more variation than other. As seen in other flowering plants, coding regions are more conserved than their noncoding counterparts. The greatest variation for coding regions was found in the *rps16*, *rpl16*, and *rps18* genes, while the greatest variation for noncoding regions was found in the *rps16-trnQ (UUG)*, *trnQ (UUG)-psbK*, and *rps8-rpl36* intergenic spacers (Fig. 2).

3.3. Phylogenetic inference

3.3.1. Complete plastomes

The alignment of the unique plastome sequence (only one copy of the IR) for the 19 taxa was 133,478 bp in length, with 4265 (3.2%) variable characters, of which 1804 (1.4%) were potentially parsimony informative (PPI). PartitionFinder divided this matrix into 11 partitions, of which seven shared the GTR+Γ model and four shared GTR+I+Γ. As expected (Salichos and Rokas, 2013), support for both trees is extremely high, with most of the branches within *Burmeistera* supported by 100 bootstrap support and 1.0 posterior probability (BS/PP hereafter; Fig. 3). The Bayesian and ML analyses of the complete plastome alignment produced identical topologies, thus we only use the Bayesian phylogeny in further analyses.

3.3.2. Integration of whole plastome and targeted Sanger sequences

The extracted dataset was 9308 bp in length with 1097 variable sites (11.8%) and 557 PPI sites (6.0%) and was divided into three partitions: two sharing the GTR+I+Γ model and one with GTR+Γ. Both the unconstrained and constrained analyses of the extracted dataset recovered *Burmeistera* as a monophyletic group (100/1.0) that is sister to *C. nigricans* (100/1.0). However, despite the monophyly at the generic level, the two infrageneric sections described by Wimmer (1943) were not recovered as natural groups, which is consistent with previous studies (Knox et al., 2008; Lagomarsino et al., 2014). Within *Burmeistera*, the unconstrained and constrained analyses resulted in very similar topologies with two main exceptions: (1) the placement of a clade formed by two species (Clade A, Fig. 4), and (2) the collapse of a moderately supported clade within clade D (see the star in Fig. 4b). Overall, the main relationships of *Burmeistera* can be split into four broad clades + *B. panamensis* (Clades A–D; Fig. 4), but the relationships among them were recovered as a polytomy. The first two groups are small clades with two and three species, Clade A (100 BS; 100 BS/1.0 PP [BS constrained; BS/PP unconstrained hereafter) and Clade B (98; 99/1.0), respectively; these species are all distributed in the Darien region of Panama (*B. dukei* Wilbur) and in the northern Andes of Colombia and Ecuador (the remaining four species). The four taxa composing Clade C (90; 870.69) occur in Colombia or in the region of Ecuador abutting the Colombian border, and are robust herbs to subshrubs (terrestrial, not scandent or hemiepiphytic), with two bracts on the pedicels (bibracteate). Lastly, although marginally supported (76; 74/0.92), Clade D includes the remaining 35 species sampled in this study, and encompasses most of the morphological diversity and geographic range within *Burmeistera*, from southern Ecuador through Costa Rica. Because of the weak support for this group, we focus our description on the smaller, well-supported clades within Clade D. One of these

Table 2
Burmeistera and allied genera plastome descriptions. LSC = Large Single Copy, SSC = Small Single Copy, IR = Inverted Repeat. Length of regions is given in number of base pairs (bp).

Species	Total Length	Total GC%	Length LSC	LSC GC%	Length IR	IR GC%	Length SSC	SSC GC%	Copies of trnQ
Burmeistera auriculata	165,772	39.2	83,839	38.4	37,256	40.7	7421	32.8	2
Burmeistera borjensis	163,961	39.1	83,436	38.4	36,528	40.7	7469	32.7	1
Burmeistera ceratocarpa	165,820	39.1	83,915	38.3	37,241	40.6	7423	32.6	1
Burmeistera crispiloba	165,349	39.2	83,708	38.4	37,112	40.8	7417	32.7	1
Burmeistera cyclostigmata	165,032	39.2	83,552	38.4	37,033	40.8	7414	32.7	2
Burmeistera cylindrocarpa	164,866	39.2	83,441	38.4	37,039	40.8	7347	32.8	1
Burmeistera dominguensis	165,547	39.2	83,957	38.4	37,096	40.7	7398	32.7	1
Burmeistera fuscoapicata	165,222	39.1	83,771	38.3	37,017	40.7	7417	32.6	1
Burmeistera loejtnantii	165,888	39.1	84,025	38.3	37,207	40.7	7449	32.7	2
Burmeistera lutosa	165,366	39.1	83,836	38.3	37,052	40.7	7426	32.8	3
Burmeistera oyacachensis	164,048	39.2	83,492	38.4	36,560	40.7	7436	32.7	1
Burmeistera parviflora	164,989	39.2	83,642	38.4	36,965	40.8	7417	32.7	2
Burmeistera pirrensis	165,353	39.2	83,825	38.4	37,056	40.8	7416	32.8	2
Burmeistera resupinata	166,128	39.2	83,845	38.4	37,434	40.7	7415	32.8	2
Burmeistera rubrosepala	165,466	39.2	83,797	38.4	37,122	40.7	7425	32.7	3
Burmeistera smaragdi	165,288	39.2	83,514	38.3	37,185	40.7	7404	32.6	1
Burmeistera sodiroana	165,511	39.2	83,747	38.4	37,173	40.8	7418	32.7	2
Centropogon nigricans	165,515	39.2	84,065	38.4	37,017	40.7	7416	32.8	1
Siphocampylus krauseanus	164,871	39.2	83,677	38.3	36,887	40.7	7420	32.8	1

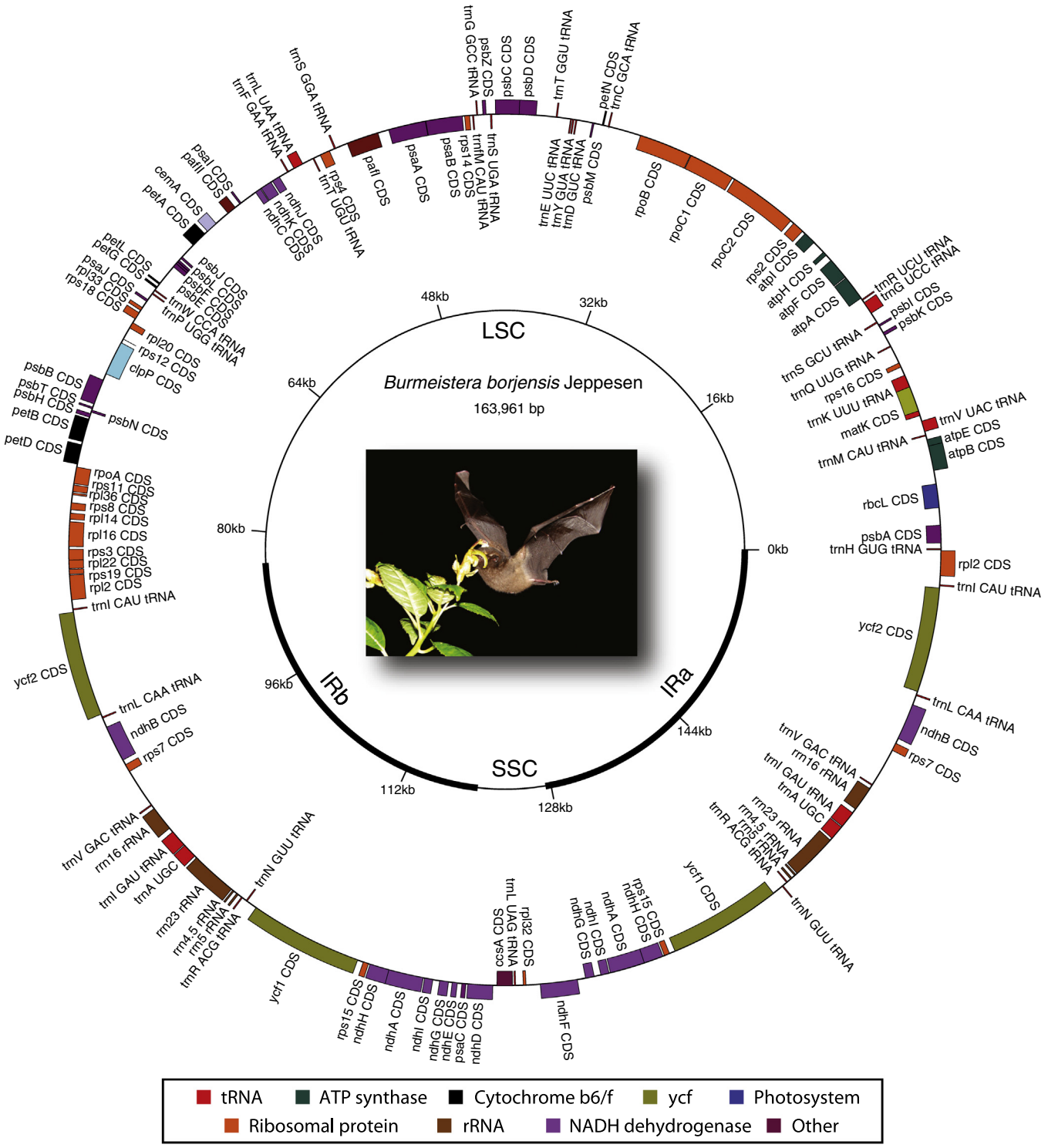

Fig. 1. The annotated plastome of *Burmeistera borjensis* Jeppesen. The inner circle shows the large single copy region (LSC), small single copy region (SSC), and two inverted repeats (IR). Gene functional groups are denoted by their colors. Genes drawn within the circle are transcribed clockwise, while genes drawn outside are transcribed counterclockwise. Plastome drawn using GenomeVx. An image of a *Burmeistera borjensis* Jeppesen flower being pollinated by *Anoura caudifer* Geoffroy is shown in the middle of the plastome.

is Clade D1 (97; 96/0.99), a clade containing primarily Central American species that are relatively short (<1 m) terrestrial plants, with leaves reduced in size towards the flowering apex, round fruits, and relatively long, narrow corolla tubes. Species in Clade D2 (85; 86/1.0) are generally obligate hemi-epiphytes with long, tapered, turbinate hypanthia, whereas taxa composing Clade D3 (100; 100/1.0) have recurved petals and leaves typically with brochidodromous venation, i.e., the secondary veins do not terminate at the margin but rather join together in a series of upward arches. Finally, Clade D4 (99; 99/1.0) and D5 (98; 97/1.0) are both composed of taxa with greatly inflated fruits, though this character is also present in the distantly related *B.* aff. *bullatifolia* in Clade C.

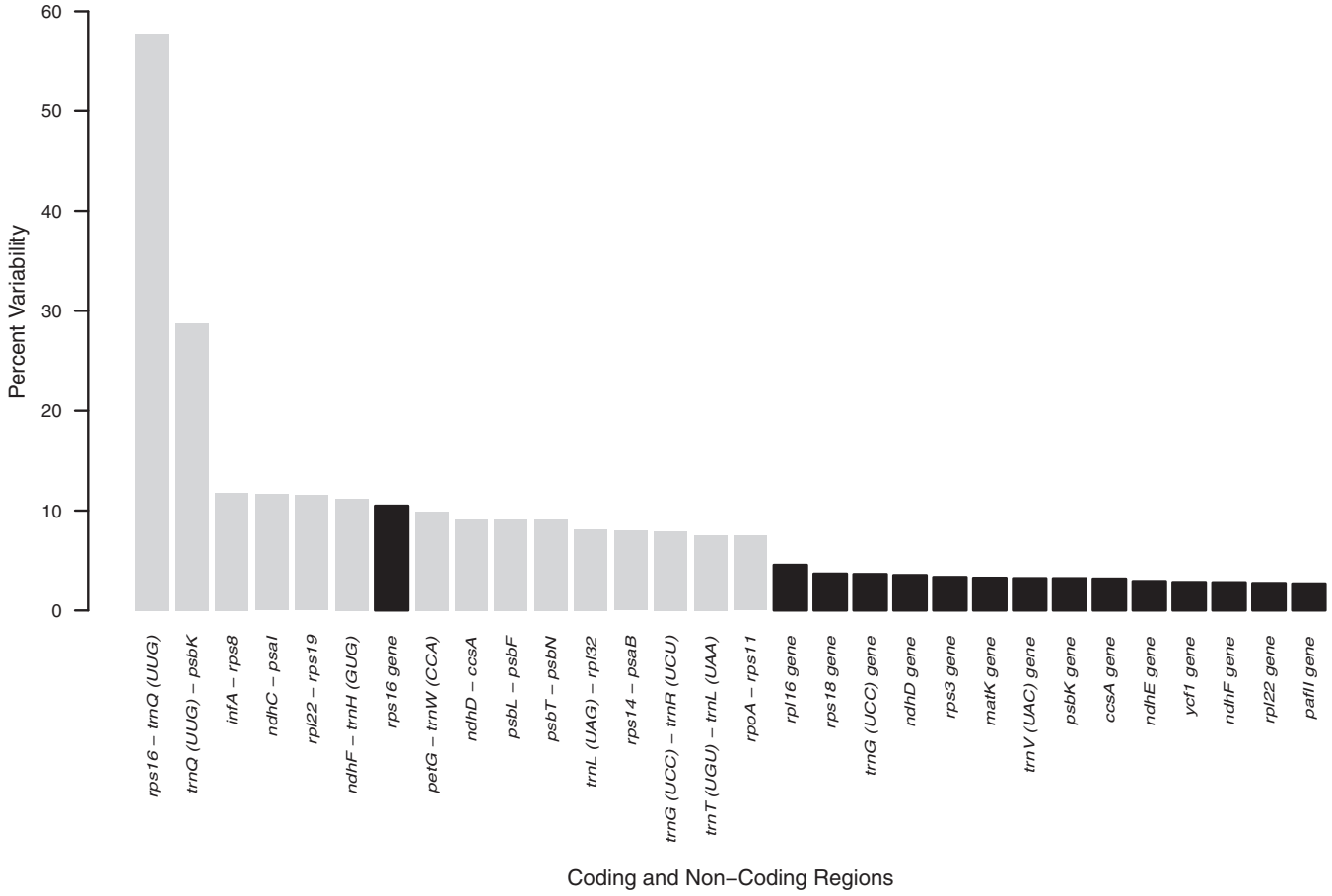

Fig. 2. The 15 most variable non-coding (gray) and 15 coding (black) regions from a comparison of the 17 *Burmeistera* plastomes.

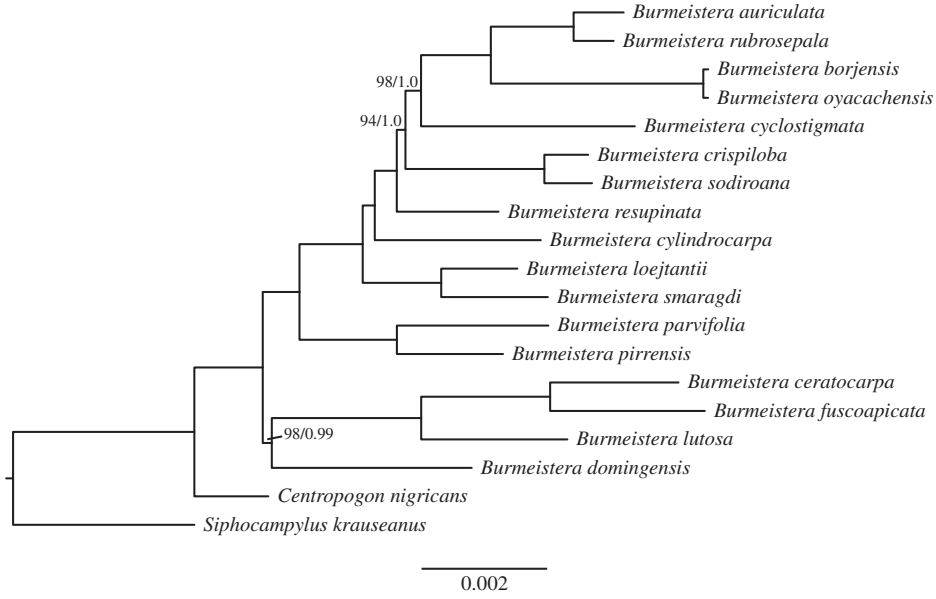

Fig. 3. Majority rule consensus tree (excluding burn-in trees) with mean branch lengths from the partitioned Bayesian analysis of the complete plastome alignment. Branch lengths are proportional to the number of substitutions per site as measured by the scale bar. Every branch has 100 bootstrap support and 1.0 Bayesian posterior probabilities (BS/PP, respectively) except for the ones where values are shown above the branches.

Inflated fruits may be a synapomorphy for a larger clade composed of clade D4 and D5, as they were recovered as sister, albeit with low support (29; 32/0.69; results not shown). Additional evidence for the close relationship between Clades D4 and D5 is provided by the complete plastome analysis; though we lack the complete taxon sampling that the combined analyses affords, all four mem-

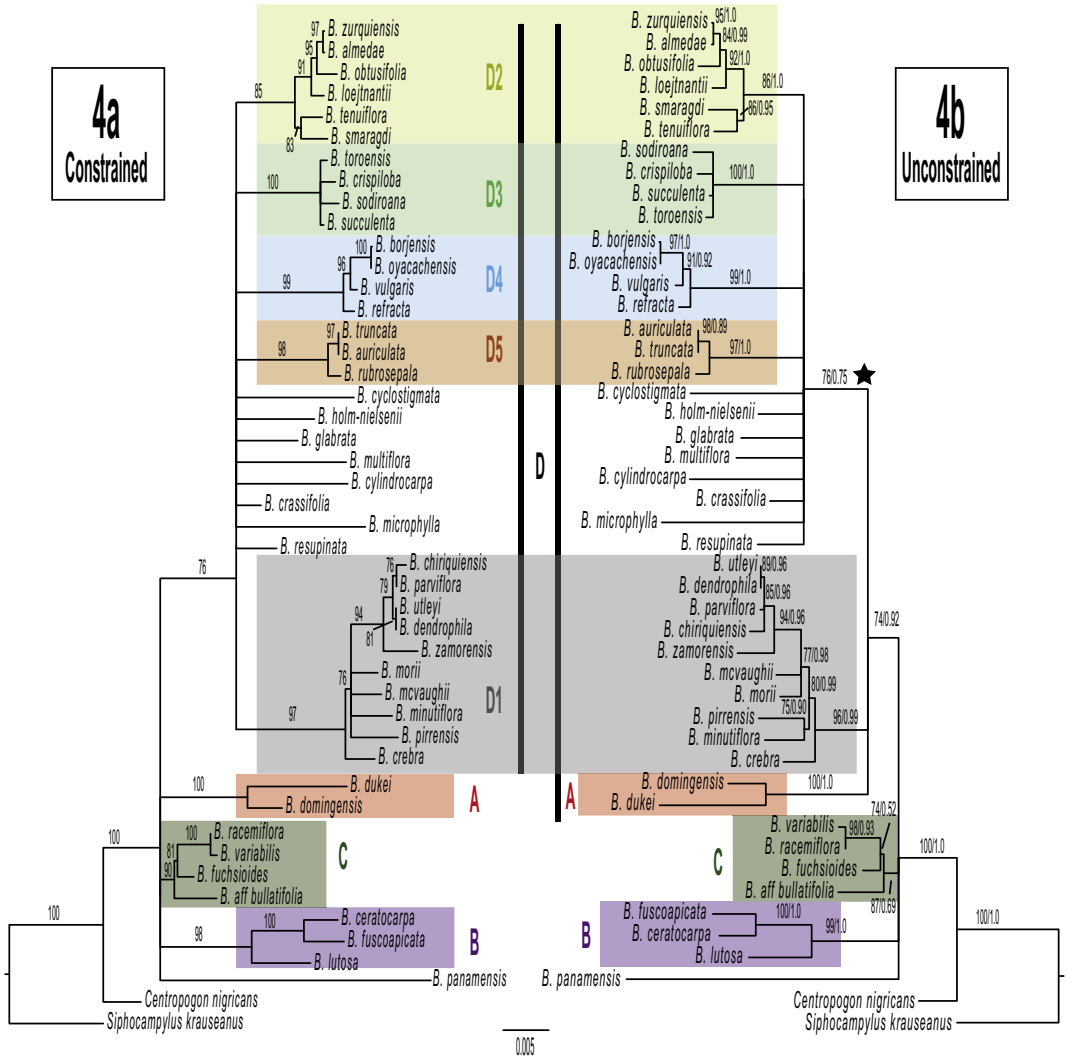

Fig. 4. Phylogenies obtained from our maximum likelihood (ML) analyses of the extracted dataset using a complete plastome topological constraint (4a) and without the constraint (4b). Colored boxes depict clades shared between the two topologies and are labeled A through D1-5. The star in 4b depicts the clade that collapses in the constrained analysis. Branch lengths are proportional to the number of substitutions per site as measured by the scale bar. Numbers on the branches represent ML bootstrap support and Bayesian posterior probabilities, respectively. Relationships with bootstrap support lower than 70 were collapsed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

bers of Clades D4 and D5 form a monophyletic unit (100/1.0; Fig 3), suggesting that the lack of support in the analyses with full taxon sampling may be a result of lower information content in the seven individual plastid loci.

The main differences between these two analyses is that Clade A is either in a polytomy with the other three main clades in the constrained analysis (Fig. 4a) or is included in a trichotomy with the members of Clade D (Fig. 4b) in the unconstrained one. The second main difference, which is a consequence of the first, is that we cannot confidently say that Clade D is monophyletic in the unconstrained analysis. Instead its members fall into a trichotomy with Clade A, Clade D1, and a monophyletic polytomy including all other members of Clade D (i.e., Clade D2–5 and unplaced members of Clade D).

The third analysis on the plastome-extracted dataset—including data for the 19 taxa for which complete plastomes are available—resulted in a tree with similar relationships to the extracted-unconstrained phylogeny, in that *B. domingensis* Jeppesen (Clade A) is placed within Clade D (results not shown). The alternate placement of these species in the complete plastome phylogeny—and thus, in the constrained analysis—suggests that the phylogenetic signal in the seven extracted loci is not representative of the signal found in the whole plastome. While we have equivocal support for the monophyly of Clade D in the unconstrained (and plastome-extracted analyses), its monophyly is marginally supported in the constrained analysis (Fig. 4a) and strongly supported in the complete plastome analysis (Fig. 3).

3.4. Ancestral state reconstruction

The majority of characters we coded showed complex evolutionary histories. In fact, only two of the traits that we analyzed are clade specific: bibracteate pedicels for Clade C and recurved petals for Clade D3 (Fig. 5). The former trait is likely plesiomorphic for the genus, as all species in the other two genera comprising the centropogonid clade (i.e., *Centropogon* and *Siphocampylus*) share this character state, thus its presence in Clade C suggests either a reversal to the plesiomorphic state or two independent losses in the remaining clades of *Burmeistera*. A third trait, that of greatly inflated fruits with thin walls, arose 2–3 times. While we have strong support for at least two independent origins (i.e., *B. aff. bullatifolia* and at least once in Clade D), we lack the phylogenetic

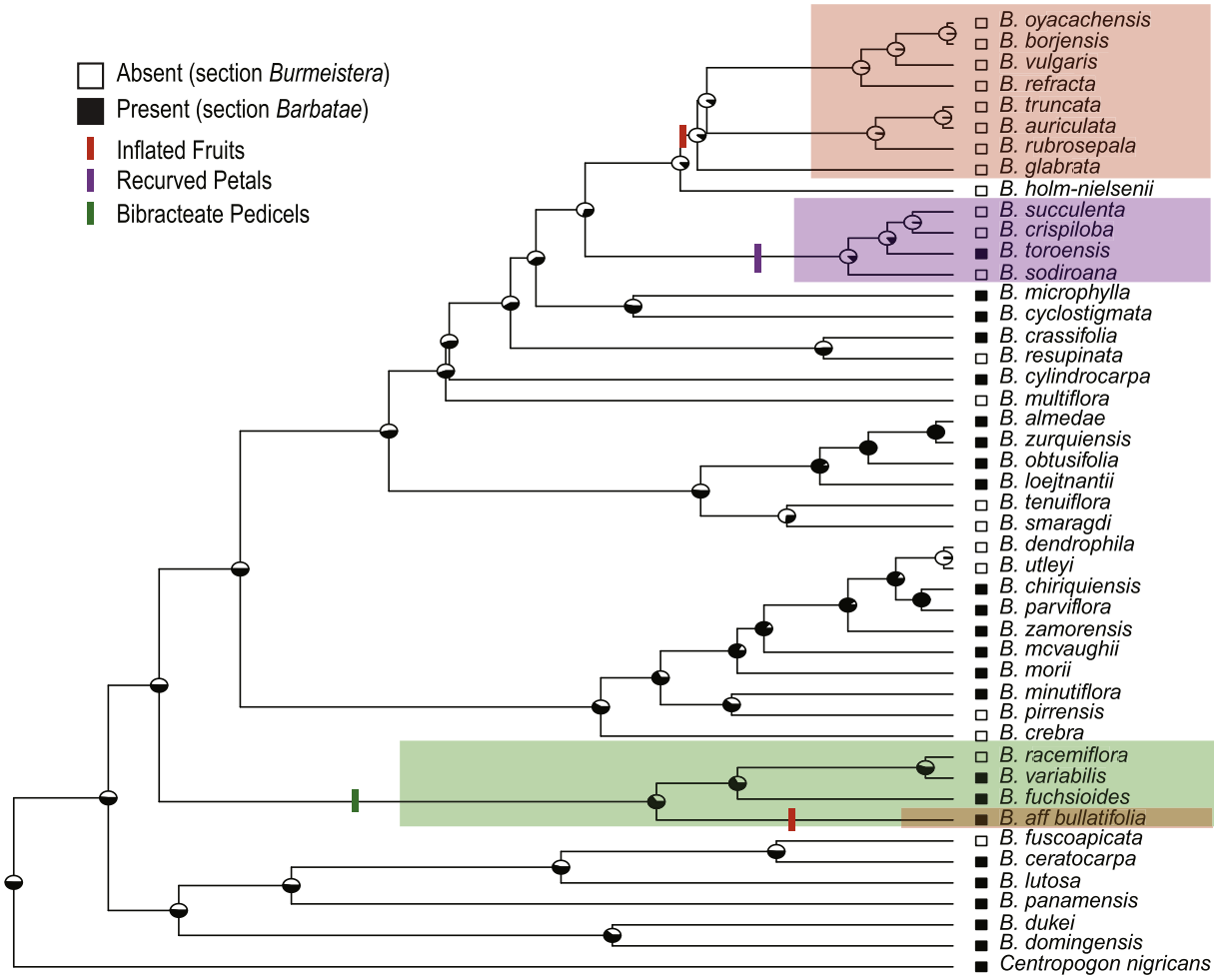

Fig. 5. Ultrametric phylogeny obtained from the extracted-constrained dataset depicting the relationships of *Burmeistera* and *Centropogon nigricans*, with the most distantly related outgroup (*Siphocampylus krauseanus*) removed. Ancestral state reconstruction of anther pubescence is represented by the pie charts on the nodes: black = presence, white = absence. The evolution of three additional traits is illustrated by the colored rectangles on the branches and the colored boxes on the clades: red = inflated fruit, purple = recurved petals, green = bibracteate pedicels. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

resolution within Clade D to infer whether this fruit type evolved once or twice within this clade. Conversely, anther pubescence, the morphological character used to delimit sections in the genus, shows a highly dynamic history and has evolved and been lost many times. The remaining traits in this analysis also show highly labile evolution, as presented in Supplementary Fig. 1. While phylogenetic uncertainty may exacerbate the degree of lability inferred in these analyses, all characters we explored (except for bibracteate pedicels and recurved petals) displayed at least some degree of well-supported homoplasy.

4. Discussion

We assembled and annotated 19 plastomes, which provided a large dataset for phylogenetic analyses, and allowed us to infer a robust and well-supported phylogenetic backbone for the genus *Burmeistera* (Fig. 3). Additional phylogenetic analyses combining these data with Sanger-sequenced plastid regions for another 28 species provides the most densely-sampled phylogenetic estimate for the genus to date (Fig. 4), covering nearly half of the known species diversity. Results are congruent with previous studies of allied genera (Antonelli, 2008, 2009; Knox et al., 2008; Lagomarsino et al., 2014): they give strong support to the monophyly of *Burmeistera*, while at the same time demonstrating that the current infrageneric classification does not adequately represent the evolutionary relationships within the genus. Ancestral state reconstruction demonstrates high levels of homoplasy in many morphological characters, with only a handful corresponding to synapomorphies for infrageneric clades. Below, we discuss results for plastome structure, phylogenetic analyses, and character evolution in greater detail.

4.1. Plastome assembly, structure, and variation

The plastomes of *Burmeistera* and allied genera included in this study show a high level of similarity, with up to 97.9% pairwise identity. Such lack of variability among plastomes is not uncommon (Wolfe et al., 1987), and great efforts have been made to identify more variable regions that might be 'universally' informative for plant phylogenetic studies (Shaw et al., 2005; Shaw et al., 2007; Shaw et al., 2014). Our study generally finds the highest variation in regions already known to be the most variable (Dong et al., 2013; Liu et al., 2013; Nazareno et al., 2015; Wu et al., 2010), particularly the *rps16*, *rpl16*, and *rps18* genes for coding parts of the plastome, and the *rps16-trnQ (UUG)*, *trnQ (UUG)-psbK*, and *infA-rps8* regions in the non-coding areas. The *rps16-trnQ (UUG)* region, the most variable region in the *Burmeistera* plastomes, is one of the regions in which insertion of foreign open reading frames (ORFs) have been reported in the family (Knox, 2014). Interestingly, additional copies (up to three) of the *trnQ (UUG)* gene were found in

most species (Table 2), although these copies are non-functional. The molecular evolution of Campanulaceae plastomes has previously been shown to include inversions, gene rearrangements, and trafficking of foreign elements (Knox, 2014), and *Burmeistera* shows many of the same changes.

4.2. Phylogenetic inference

One of our main goals was to include as much taxonomic sampling and sequence data as possible, combining newly generated Sanger and HTS data with available Sanger sequences from GenBank. In a recent study of the plant genus *Acacia* Mill., Williams et al. (2016) approached the issue of combining HTS with Sanger data by creating multiple datasets with different amounts of missing data. They demonstrated that extracting loci from complete plastomes and merging them with their Sanger counterparts resulted in trees with higher support than trees inferred from the Sanger loci aligned to the complete plastomes, likely because the latter alignments had such high amounts of missing data (~97%; (Williams et al., 2016). They also showed that using a complete plastome tree to constrain their analyses improved the placement of rogue taxa, i.e., samples that have different phylogenetic relationships in different analyses. Following Williams et al. (2016), we analyzed the available Sanger data for *Burmeistera* combined with the seven extracted loci from the complete plastome alignment, both with and without a topological constraint.

The different datasets and analyses in this study yielded similar results in terms of relationships and support. With 45 of the 116 species of *Burmeistera*, our study represents the most comprehensive sampling of the genus, and includes all the sequence data available to date. Nevertheless, we were not able to resolve the relationships among the earliest diverging clades in the genus (Clade A–D + *B. panamensis*), and eight species within Clade D are not placed in well supported clades (Fig. 4). We did see, however, that the use of a complete plastome topological constraint in the extracted analyses had implications in the placement of Clade A. By comparing the topologies obtained from the plastome-extracted and the complete plastome datasets, it is evident that the placement of Clade A and the creation of a trichotomy in Clade D (Fig. 4b) are a direct result of reducing the amount of sequence data, and suggests that the phylogenetic signal contained in the seven extracted loci is weaker compared to the signal in the complete plastome. Based on this result and the fact that the plastome is a uniparentally inherited, non-recombining locus expected to have a single evolutionary history, it is reasonable to use the more powerful signal present in the complete plastome as a topological constraint for further analyses.

Even though the complete plastome phylogeny has very high support at almost every branch, the inclusion of more taxa—and with them, the increase in missing data and/or the reduction in sequence length—decreases branch support drastically. By only using data from the seven Sanger-sequenced loci, we lowered the phylogenetic signal available in the data. Additionally, the plastome's slow mutation rate, which can be up to three times slower than nuclear genes (Wolfe et al., 1987; Wolfe et al., 1989), may not be sufficient to entirely elucidate relationships at the species level in this genus. *Burmeistera* is a relatively young radiation estimated to have diverged approximately 2.6 Myr (2.4–3.5 Myr) (Lagomarsino et al., 2016). This young age, coupled with the plastome's slow rate of molecular evolution, highlights the necessity to include several nuclear loci to conduct a multilocus analysis on the genus. Moreover, hybridization, introgression, and/or chloroplast capture are processes that may hinder phylogenetic inference in the group. To overcome these problems, we are currently sequencing over 800 nuclear genes using a Hyb-Seq approach (Schmickl et al., 2016; Weitemier et al., 2014), which will result in a robust dataset to better elucidate phylogenetic relationships in *Burmeistera*.

4.3. Evolutionary trends and implications for future reclassification in Burmeistera

Character evolution in *Burmeistera* has generally been quite dynamic, which is not unexpected given the known taxonomic difficulty of the group (Wilbur, 1976) and the frequent convergent evolution among *Burmeistera* and its closest relatives (Lagomarsino et al., 2014; Lagomarsino et al., 2016). The evolution of some of the traits have obvious underlying ecological hypotheses; for example, the multiple shifts to brightly colored flowers from ancestrally green, dull-colored flowers (Fig. S1f) are likely all associated with pollination shifts from bats to hummingbirds (Muchhala, 2006). Similarly, the evolution of hemi-epiphytism and a scandent habit (Fig. S1a) may have opened up new niche spaces for the diversification of *Burmeistera*, allowing these species to be better competitors for light against other terrestrial plants, or to be more easily located by their bat or hummingbird pollinators. The evolutionary and ecological pressures that could have resulted in the dynamic evolution of other traits, such as shape of the ovary and the repeated loss of tufted hairs on the ventral anthers, are less apparent. A potential explanation is that the phylogeny of *Burmeistera* is not adequately represented by a fully bifurcating tree, and a network of introgression may be a more appropriate representation, although further investigation is needed to shed light into this hypothesis.

The labile nature of character evolution has made taxonomic efforts focused on *Burmeistera* difficult. In fact, the single trait that currently delimits sections—the degree of anther apex pubescence—is particularly labile, changing at least 11 times along the phylogeny (Fig. 5). As first suggested by Knox et al. (2008), the repeated loss of tufted hairs on the ventral anthers may be consequence of the dilated anther orifice which characterizes *Burmeistera*: these hairs, which normally function as a "trigger" to release pollen in Lobelioideae with closed anthers, are rendered obsolete by the dilated anther opening, and thus, represent an evolutionary vestige and are lost over time within the genus. Supporting this hypothesis, we see that the vast majority of transitions in anther pubescence are losses, though a handful of gains are inferred (e.g., *B. cyclostigmata*); this may reflect phylogenetic uncertainty. In contrast, all other members of the centropogonid clade (i.e., *Centropogon* and *Siphocampylus*), including *C. nigricans* sampled here, have closed anthers with ventral anther pubescence. Though this pattern of trait evolution with *Burmeistera* is interesting, anther pubescence will clearly not be a useful trait in a subgeneric classification that reflects the evolutionary history of the genus, which highlights the need to revisit the subgeneric taxonomy of *Burmeistera*. Fortunately, several traits investigated in this study were relatively stable. These include greatly inflated fruits (Figs. 5 and S1g), bibracteate pedicels (Figs. 5 and S1c), and recurved petals (Figs. 5 and S1b); each of these diagnoses the clades in which they occur, though greatly inflated fruits evolved at least twice independently. These characters may serve as the basis for reclassification efforts in the future, when a phylogeny with nearly complete taxon sampling and more robust branch support is available. Biogeographic occurrence may also serve to inform classification; multiples clades are geographically restricted, including the primarily Central American Clade D1. That Clades A–C are restricted to the northern Andes and the adjacent Darien region of Colombia and Panama suggests a potential origin of the genus in northern South America, although a proper biogeographic analysis is needed to test this hypothesis.

The clades demarcated in Fig. 4 are generally well-defined units based on their gross morphology. For example, Clade D1 generally

comprises low-lying terrestrial herbs with finely divided juvenile leaves, adult leaves with distinctly toothed leaf margins, rounded ovaries, and long, narrow corolla tubes; all are endemic to Central America, except for Ecuadorian *B. zamorensis* Muchhala & A.J. Pérez. Further, Clade D3 comprises species whose flowers have recurved petals, reduced calyx, and brochidodromous venation, while Clade D2 are generally obligate hemi-epiphytes taxa with long-tapered, turbinate hypanthia. Finally, Clades D4 and D5, which are likely sister clades (30/0.69), both produce greatly inflated fruits, a particularly striking potential synapomorphy. Within these groups, species in Clade D5 (*B. rubrosepala* (E.Wimm.) E.Wimm., *B. truncata* Zahlbr., *B. auriculata* Muchhala & Lammers) are distributed on the western slope of the Andes and have leaves that are distichous and ovate to lanceolate in shape, while species in Clade D4 (*B. oyacachensis* Jeppesen, *B. borjensis* Jeppesen, *B. refracta* E. Wimm., *B. vulgaris* E. Wimm.) are found on the eastern Andean slopes (with the exception of widespread *B. vulgaris*) and have elliptical leaves that are spirally arranged.

Although it is still premature for a formal subgeneric reclassification of *Burmeistera*, efforts are underway to incorporate multilocus nuclear sequences using HTS data and phylogeny reconstruction methods that accommodate gene tree-species tree incongruence caused by coalescent stochasticity, hybridization, and/or introgression. Furthermore, many of the traits relevant to the recircumscription may be most accurately captured as continuous traits, requiring further study. The in-depth exploration of such traits across species and geography (sensu Zapata and Jimenez, 2012) will be useful, and perhaps even necessary, for subdividing *Burmeistera*—long acknowledged to be taxonomically difficult—into constituent monophyletic units that are also well defined by morphology and geography.

Acknowledgements

We would like to thank Eric Knox and two anonymous reviewers who provided valuable comments on earlier version of the manuscript, and Michael McKain for help in scripting and assistance in plastome assembly. The Missouri Botanical Garden (MO) provided important access to their herbarium collection. We thank Stacey D. Smith for providing resources to generate Sanger sequences in her lab at the University of Nebraska, and Tom Lammers for nomenclatural advice. Funding for this study was provided by a University of Missouri Research Board Grant. LPL was funded by an NSF Postdoctoral Research Fellowship in Biology under Grant No. 1523880.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2016.12.011.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Antonelli, A., 2008. Higher level phylogeny and evolutionary trends in *Campanulaceae subfam.* Lobelioideae: molecular signal overshadows morphology. Mol. Phylogenet. Evol. 46, 1–18.
Antonelli, A., 2009. Have giant lobelias evolved several times independently? Life form shifts and historical biogeography of the cosmopolitan and highly diverse subfamily Lobelioideae (Campanulaceae). BMC Biol. 7, 82.
Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477.
Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comp. Biol., e1003537
Chase, M.W., Soltis, D.E., Olmstead, R.G., Morgan, D., Les, D.H., Mishler, B.D., Duvall, M.R., Price, R.A., Hills, H.G., Qiu, Y.-L., 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. Ann. Mo. Bot. Gard., 528–580
Conant, G.C., Wolfe, K.H., 2008. GenomeVx: simple web-based creation of editable circular chromosome maps. Bioinformatics 24, 861–862.
Crowl, A.A., Miles, N.W., Visger, C.J., Hansen, K., Ayers, T., Haberle, R., Cellinese, N., 2016. A global perspective on Campanulaceae: biogeographic, genomic, and floral evolution. Am. J. Bot., 233–245
Dong, W., Xu, C., Cheng, T., Zhou, S., 2013. Complete chloroplast genome of *Sedum sarmentosum* and chloroplast genome evolution in Saxifragales. PLoS ONE, e77965.
Downie, S.R., Palmer, J.D., 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In: Soltis, P.S., Soltis, D.E., Doyle, J.J. (Eds.). Chapman and Hall, New York NY., pp. 14–35.
Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19, 11–15.
Drummond, A., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.
Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol., 1969–1973
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. 32, 1792–1797.
Garzón Venegas, J., González, F., 2012. Five new species and three new records of *Burmeistera* (Campanulaceae-Lobelioideae) from Colombia. Caldasia, 309–324.
Garzón Venegas, J., González, F., Velez Puerta, J.M., 2012. *Burmeistera minutiflora* (Campanulaceae-Lobelioideae), a new species from the high Andes of Antioquia (Colombia) with the smallest flowers in the genus. Anales Del Jardin Botanico de Madrid, 243–246.
Garzón Venegas, J., Luteyn, J.L., González, F., 2014. A new species of *Burmeistera* (Campanulaceae, Lobelioideae) from the Western Cordillera of Colombia. Novon: J. Botan. Nomencl., 165–170
Garzón Venegas, J., Velez Puerta, J.M., González, F., 2013. Three new species of *Burmeistera* (Campanulaceae-Lobelioideae) from Colombia. Brittonia, 119–127.
González, F., Garzón Venegas, J., 2015. On the typification of *Burmeistera* (Campanulaceae-Lobelioideae) and the identity of *B. ceratocarpa* var. *dentata*, *B. ibaguensis* and *B. rivina*. Caldasia, 251–260.
Graham, S.W., Olmstead, R.G., 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. Am. J. Bot. 87, 1712–1730.
Jeppesen, S., 1981. Lobeliaceae. In: Harling, G., Sparre, B. (Eds.), Flora of Ecuador 14: 9–170. Swedish Natural Science Research Council, Stockholm.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.
Knox, E.B., 2014. The dynamic history of plastid genomes in the Campanulaceae sensu lato is unique among angiosperms. Proc. Natl. Acad. Sci. 111, 11097–11102.
Knox, E.B., Downie, S.R., Palmer, J.D., 1993. Chloroplast genome rearrangements and the evolution of giant lobelias from herbaceous ancestors. Mol. Biol. Evol., 414–430
Knox, E.B., Muasya, A.M., Muchhala, N., 2008. The predominantly South American clade of Lobeliaceae. Syst. Bot. 33, 462–468.
Knox, E.B., Palmer, J.D., 1999. The chloroplast genome arrangement of *Lobelia thuliniana* (Lobeliaceae): expansion of the inverted repeat in an ancestor of the Campanulales. Plant Syst. Evol., 49–64
Lagomarsino, L.P., Aguilar, D.S., Muchhala, N., 2015. Two new species of *Burmeistera* (Campanulaceae: Lobelioideae) from the Cordillera de Talamanca of Costa Rica and Panama, with a key to the Central American species. Syst. Bot. 40, 914–921.
Lagomarsino, L.P., Antonelli, A., Muchhala, N., Timmermann, A., Mathews, S., Davis, C.C., 2014. Phylogeny, classification, and fruit evolution of the species-rich Neotropical bellflowers (Campanulaceae: Lobelioideae). Am. J. Bot. 101, 2097–2112.
Lagomarsino, L.P., Condamine, F.L., Antonelli, A., Mulch, A., Davis, C.C., 2016. The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). New Phytol., pp n/a-n/a.
Lancucka-Srodoniowa, M., 1977. New herbs described from the Tertiary of Poland. Acta Palaeobatanica 18, 37–44.
Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701.
Lanfear, R., Calcott, B., Kainer, D., Mayer, C., Stamatakis, A., 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol. Biol. 14, 82.
Lemmon, A.R., Brown, J.M., Stanger-Hall, K., Lemmon, E.M., 2009. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. Syst. Biol. 58, 130–145.
Liu, Y., Huo, N., Dong, L., Wang, Y., Zhang, S., Young, H.A., Feng, X., Gu, Y.Q., 2013. Complete chloroplast genome sequences of Mongolia medicine *Artemisia frigida* and phylogenetic relationships with other plants. PLoS ONE, e57533.

Marçais, G., Kingsford, C., 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27, 764–770.
Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17, 10–12.
Marx, H.E., O'Leary, N., Yuan, Y.-W., Lu-Irving, P., Tank, D.C., Múlgura, M.E., Olmstead, R.G., 2010. A molecular phylogeny and classification of Verbenaceae. Am. J. Bot. 97, 1647–1663.
McNeill, J., Barrie, F.R., Buck, W.R., Demoulin, V., Greuter, W., Hawksworth, D.L., Herendeen, P.S., Knapp, S., Marhold, K., Prado, J., Prud'homme van Reine, W.F., Smith, G.F., Wiersema, J.H., Turland, N.J., 2012. International Code of Nomenclature for algae, fungi, and plants (Melbourne Code) - Regnum Vegetabile 154. Koeltz Scientific Books, Oberreifenberg, Germany.
Moore, M.J., Bell, C.D., Soltis, P.S., Soltis, D.E., 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc. Natl. Acad. Sci. 104, 19363–19368.
Muchhala, N., 2003. Exploring the boundary between pollination syndromes: bats and hummingbirds as pollinators of *Burmeistera cyclostigmata* and *B. tenuiflora* (Campanulaceae). Oecologia, 373–380.
Muchhala, N., 2006. The pollination biology of *Burmeistera* (Campanulaceae): specialization and syndromes. Am. J. Bot., 1081–1089
Muchhala, N., 2008. Functional significance of interspecific variation in *Burmeistera* Flower morphology: evidence from nectar bat captures in Ecuador. Biotropica, 332–337.
Muchhala, N., Lammers, T.G., 2005. A new species of *Burmeistera* (Campanulaceae: Lobelioideae) from Ecuador. Novon 15, 176–179.
Muchhala, N., Pérez, Á.J., 2015. *Burmeistera zamorensis* (Campanulaceae, Lobelioideae), a New Species from Southern Ecuador. Novon: J. Botan. Nomencl., 36–38
Muchhala, N., Potts, M.D., 2007. Character displacement among bat-pollinated flowers of the genus *Burmeistera*: analysis of mechanism, process and pattern. In: Proc. Roy. Soc. B: Biol. Sci., pp. 2731–2737.
Nazareno, A.G., Carlsen, M., Lohmann, L.G., 2015. Complete chloroplast genome of *Tanaecium tetragonolobum*: the first Bignoniaceae plastome. PLoS ONE 10, e0129930.
Nürk, N.M., Uribe-Convers, S., Gehrke, B., Tank, D.C., Blattner, F., 2015. Oligocene niche shift, Miocene diversification - cold tolerance and accelerated speciation rates in the St. John's Worts (*Hypericum*, Hypericaceae). BMC Evol. Biol. 15, 80.
Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics, 289–290.
Parks, M., Cronn, R., Liston, A., 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol. 7, 84.
R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria <http://www.r-project.org/>. ISBN 3-900051-07-0.
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.
Salichos, L., Rokas, A., 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature, 327–331 (Nature Publishing Group).
Schenk, J.J., 2016. Consequences of secondary calibrations on divergence time estimates. PLoS ONE 11, e0148228.
Schmickl, R., Liston, A., Zeisek, V., Oberlander, K., Weitemier, K., Straub, S.C.K., Cronn, R.C., Dreyer, L.L., Suda, J., 2016. Phylogenetic marker development for target enrichment from transcriptome and genome skim data: the pipeline and its application in southern African *Oxalis* (Oxalidaceae). Mol. Ecol. Resour. 16, 1124–1135.
Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W., Miller, J., Siripun, K.C., Winder, C. T., Schilling, E.E., Small, R.L., 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am. J. Bot. 92, 142–166.
Shaw, J., Lickey, E.B., Schilling, E.E., 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.
Shaw, J., Shafer, H.L., Leonard, O.R., Kovach, M.J., Schorr, M., Morris, A.B., 2014. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. Am. J. Bot. 101, 1987–2004.
Smith, S., Dunn, C., 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. Bioinformatics 24, 715.
Soltis, D.E., Smith, S.A., Cellinese, N., Wurdack, K.J., Tank, D.C., Brockington, S.F., Refulio-Rodriguez, N.F., Walker, J.B., Moore, M.J., Carlsward, B.S., Bell, C.D., Latvis, M., Crawley, S., Blanco, M.A., Diouf, D., Xi, Z., Rushworth, C.A., Gitzendanner, M.A., Sytsma, K.J., Qiu, Y.L., Hilu, K.W., Davis, C.C., Sanderson, M.J., Beaman, R.S., Olmstead, R.G., Judd, W.S., Donoghue, M.J., Soltis, P.S., 2011. Angiosperm phylogeny: 17 genes, 640 taxa. Am. J. Bot. 98, 704–730.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57, 758–771.
Straub, S.C.K., Fishbein, M., Livshultz, T., et al., 2011. Building a model: developing genomic resources for common milkweed (*Asclepias syriaca*) with low coverage genome sequencing. BMC Genom. 12.
Straub, S.C.K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C., Liston, A., 2012. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. Am. J. Bot. 99, 349–364.
Uribe-Convers, S., Duke, J.R., Moore, M.J., Tank, D.C., 2014. A long PCR-based approach for DNA enrichment prior to next-generation sequencing for systematic studies. Appl. Plant Sci. 2, 1300063.
Uribe-Convers, S., Settles, M.L., Tank, D.C., 2016. A phylogenomic approach based on PCR target enrichment and high throughput sequencing: resolving the diversity within the South American species of *Bartsia* L. (Orobanchaceae). PLoS ONE 11, e0148203.
Uribe-Convers, S., Tank, D.C., 2015. Shifts in diversification rates linked to biogeographic movement into new areas: an example of a recent radiation in the Andes. Am. J. Bot. 102, 1854–1869.
Uribe-Convers, S., Tank, D.C., 2016. Phylogenetic revision of the genus *Bartsia* (Orobanchaceae): disjunct distributions correlate to independent lineages. Syst. Bot. 41, 672–684.
Weitemier, K., Straub, S.C.K., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A., Liston, A., 2014. Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. Appl. Plant Sci. 2, 1400042.
Welker, C.A.D., Souza-Chies, T.T., Longhi-Wagner, H.M., Peichoto, M.C., McKain, M. R., Kellogg, E.A., 2016. Multilocus phylogeny and phylogenomics of *Eriochrysis* P. Beauv. (Poaceae-Andropogoneae): taxonomic implications and evidence of interspecific hybridization. Mol. Phylogenet. Evol. 99, 155–167.
Wilbur, R.L., 1975. A synopsis of the Costa Rican species of *Burmeistera* (Campanulaceae: Lobelioideae). Bull. Torrey Bot. Club, 225–231.
Wilbur, R.L., 1976. Flora of Panama. Part IX. Family 183. Campanulaceae. Ann. Mo. Bot. Gard., 593–655
Wilbur, R.L., 1981. Additional Panamanian species of *Burmeistera* (Campanulaceae: Lobelioideae). Ann. Mo. Bot. Gard., 167–171
Williams, A.V., Miller, J.T., Small, I., Nevill, P.G., Boykin, L.M., 2016. Integration of complete chloroplast genome sequences with small amplicon datasets improves phylogenetic resolution in *Acacia*. Mol. Phylogenet. Evol. 96, 1–8.
Wimmer, F.E., 1943. Campanulaceae-Lobelioideae. I. Teil. In: Mansfeld R. (Ed.) Das Pflanzenreich IV.276b. Wilhem Engelmann, Leipzig, Germany, pp. 1–260.
Wolfe, K.H., Li, W.H., Sharp, P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA 84, 9054–9058.
Wolfe, K.H., Sharp, P.M., Li, W.-H., 1989. Rates of synonymous substitution in plant nuclear genes. J. Mol. Evol. 29, 208–211.
Wu, F.-H., Chan, M.-T., Liao, D.-C., Hsu, C.-T., Lee, Y.-W., Daniell, H., Duvall, M.R., Lin, C.-S., 2010. Complete chloroplast genome of *Oncidium* Gower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiinae. BMC Plant Biol., 68
Zapata, F., Jimenez, I., 2012. Species delimitation: inferring gaps in morphology across geography. Syst. Biol. 61, 179–194.
Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.