Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae)

Nicolás García ^{a,b,*}, Ryan A. Folk ^b, Alan W. Meerow ^d, Srikar Chamala ^{c,e}, Matthew A. Gitzendanner ^{b,c}, Renata Souza de Oliveira ^f, Douglas E. Soltis ^{b,c,g}, and Pamela S. Soltis ^{b,c,g}

^a Facultad de Ciencias Forestales y de la Conservación de la Naturaleza, Universidad de Chile, Av. Santa Rosa 11315, La Pintana, Santiago, Chile

^b Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA

^c Department of Biology, University of Florida, Gainesville, FL 32611, USA

^d USDA-ARS-SHRS-National Germplasm Repository, 13601 Old Cutler Road, Miami, FL 33158, USA

^e Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL 32610, USA

^f GaTE Laboratory, Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, IBUSP, Rua do Matão 277, CEP: 05508-090 São Paulo, SP, Brazil

^g Genetics Institute, University of Florida, Gainesville, FL 32610, USA

* Corresponding author.

E-mail addresses: nicogarciab@gmail.com (N. García), Alan.Meerow@ars.usda.gov (A.W. Meerow), ryanfolk@ufl.edu (R.A. Folk), srikarchamala@gmail.com (S. Chamala), magitiz@ufl.edu (M.A. Gitzendanner), rsolive1@gmail.com (R.S. de Oliveira), dsoltis@ufl.edu (D.E. Soltis), psoltis@flmnh.ufl.edu (P.S. Soltis)

Abstract

Hybridization is a frequent and important force in plant evolution. Next-generation sequencing (NGS) methods offer new possibilities for clade resolution and ambitious sampling of gene genealogies, yet difficulty remains in detecting deep reticulation events using currently available methods. We reconstructed the phylogeny of diploid representatives of Amaryllidaceae tribe Hippeastreae to test the hypothesis of ancient hybridizations preceding the radiation of its major subclade, Hippeastrinae. Through hybrid enrichment of DNA libraries and NGS, we obtained data for 18 nuclear loci through a curated assembly approach and nearly complete plastid genomes for 35 ingroup taxa plus 5 outgroups. Additionally, we obtained alignments for 39 loci through an automated assembly algorithm. These data were analyzed with diverse phylogenetic methods, including concatenation, coalescence-based species tree estimation, Bayesian concordance analysis, and network reconstructions, to provide insights into the evolutionary relationships of Hippeastreae. Causes for gene tree heterogeneity and cytonuclear discordance were examined through a Bayesian posterior predictive approach (JML) and coalescent simulations. Two major clades were found, Hippeastrinae and Traubiinae, as previously reported. Our results suggest the presence of two major nuclear lineages in Hippeastrinae characterized by different chromosome numbers: 1) *Tocantinia* and *Hippeastrum* with $2n = 22$, and 2) *Eithea*, *Habranthus*, *Rhodophiala*, and *Zephyranthes* mostly with $2n = 12$, 14, and 18. Strong cytonuclear discordance was confirmed in Hippeastrinae, and a network scenario with at least six hybridization events is proposed to reconcile nuclear and plastid signals, along a

backbone that may also have been affected by incomplete lineage sorting at the base of each major subclade.

Keywords: hybridization; coalescence; species tree; target-enrichment; cytonuclear discordance

1. Introduction

Hybridization is increasingly acknowledged as a major force in evolution, operating through a diversity of processes such as hybrid speciation (Mallet, 2007; Soltis and Soltis, 2009) and introgression (Anderson, 1949; Arnold et al., 2004), which can occur with whole-genome duplication (allopolyploidy) or without (homoploid hybridization). However, the detection of hybridization, especially homoploid hybridization, has generally been methodologically difficult (Soltis and Soltis, 2009). Tests to distinguish incomplete lineage sorting (ILS) from hybridization using gene tree discordance usually require data from multiple nuclear genes (e.g., Buckley et al., 2006; Holland et al., 2008; Joly et al., 2009; Kubatko, 2009; Meng and Kubatko, 2009; Gerard et al., 2011) to provide gene sample sizes large enough to use coalescent methods effectively, yet suitably large data sets have only recently been attainable for most taxa. From a theoretical perspective, nuclear loci have important advantages over chloroplast (cpDNA) and nuclear ribosomal (nrDNA) DNA for exploring reticulate evolution (Linder and Rieseberg, 2004), in that they allow the collection of statistically robust gene samples of nearly unlimited size. In addition, they do not experience unusual modes of gene evolution such as uniparental inheritance and concerted evolution (Álvarez and Wendel, 2003; Small et al., 2004, respectively) that could complicate interpretations of

gene discordance. Yet, owing to their distinctive modes of inheritance, organellar genomes could remain valuable for detecting instances of hybridization that manifest as cytonuclear discordance (Soltis and Kuzoff, 1995; Seehausen, 2004). For reasons that remain unclear, some taxa appear to have experienced introgression of entire organellar genomes while leaving little or no evidence in the nuclear genome (Good et al., 2015; Folk et al., unpublished).

Next-generation sequencing (NGS) technologies have revolutionized phylogenetics and systematics (e.g., Buggs et al., 2012; Godden et al., 2012; Harrison and Kidner, 2011; Moriarty Lemmon and Lemmon, 2013) due to the possibility of multiplexing several samples and/or loci (e.g., Bybee et al., 2011; Griffin et al., 2011; Straub et al., 2012; McCormack et al., 2013). In-solution hybrid capture, a method of target enrichment, coupled with NGS (Gnirke et al., 2009; Turner et al., 2009; Mamanova et al., 2010) is an effective cost- and time-effective approach to obtain sequence data for many independent nuclear loci (e.g., Folk et al., 2016; Mandel et al., 2014; Stephens et al., 2015; Weitemier et al., 2014). Additionally, bioinformatic processing of target-enriched NGS data usually allows the recovery of nearly complete plastid data sets to perform more robust tests of cytonuclear concordance/discordance (e.g., Weitemier et al., 2014; Folk et al., 2015). Hence, hybrid enrichment data sets place highly resolved phylogenetic trees from all three plant genomes within the reach of researchers – a critical resource, given that the detection of hybridization requires a robust sampling of loci and should consider evidence from both organellar and nuclear genomes.

Resolved phylogenetic trees have proved elusive for Amaryllidaceae tribe Hippeastreae, a Neotropical group of asparagoid lilies (i.e., Asparagales; APG IV, 2016) (García et al., 2014; Meerow et al., 2000; Meerow, 2010). Two potential confounding processes that explain the elusiveness of a robust phylogeny for Hippeastrinae are ILS and ancient hybridization (García et al., 2014; Meerow, 2010); this group has recently been posited as a dramatic case of ancient hybridization based on evidence for plastid capture (García et al., 2014). Despite the considerable systematic attention that Hippeastreae has received because of its horticultural importance, its generic classification has been historically debated (Traub, 1963; Meerow and Snijman, 1998; Ravenna, 2003), mostly due to a lack of both unequivocal diagnostic characters and a robust phylogenetic framework. Chromosome numbers and karyotypes may be important for diagnosing certain lineages (e.g., Naranjo and Andrada, 1975; Greizerstein and Naranjo, 1987; Naranjo and Poggio, 2000; Baeza et al., 2009; Cisternas et al., 2010; Felix et al., 2011), although this group contains clades with dysploid variation ($n = 6 - 30$) and extensive polyploidy, especially in *Zephyranthes* and *Habranthus* (Flory, 1977; Meerow, 2010). Phylogenetic analyses of Hippeastreae strongly support the presence of two subclades, recognized as subtribes Traubiinae and Hippeastrinae (García et al., 2014). Although relationships in Traubiinae are fairly well resolved (García et al., 2014), those in Hippeastrinae are not.

In this paper, we use NGS methods to test the hypothesis that ancient diploid hybridization preceded the main radiation of Hippeastrinae (García et al., 2014) and evaluate the possible roles of hybridization and ILS in generating cytonuclear

discordance throughout the clade. We ultimately seek to determine the most likely evolutionary pattern for Hippeastreae, while recognizing that both hybridization and ILS may have occurred.

2. Materials and methods

2.1. Materials

We selected 35 ingroup diploid taxa and five outgroups (one species of Lycorideae, one of Narcisseae, and three members of Griffinieae; Table 1) for hybrid enrichment experiments. In addition to these 40 species, a transcriptome of *Narcissus viridiflorus* (Table S1) was also used in phylogenetic analyses of a set of curated nuclear loci (but not in plastid analyses) and as a reference to discern coding and non-coding regions in the alignments. The ingroup sampling represents all main diploid lineages (i.e., uncontroversial, monophyletic groups of extant species) identified in García et al. (2014). Given our emphasis on patterns of cladogenesis and homoploid hybridization, taxon sampling has been constrained to diploid taxa. Future studies will address the parentage of polyploid species of *Habranthus*, *Sprekelia*, and *Zephyranthes*.

We also included critical taxa that had not been included in previous phylogenetic studies, such as *Habranthus gracilifolius* (i.e., the type species of *Habranthus*), *Tocantinia dutilhiana* (Büneker et al., 2016), and two undescribed species: 1) *Eithea* “Piracicaba”, a second species from this currently monotypic genus collected at the locality of Piracicaba, Sao Paulo, Brazil; and 2) *Hippeastrum* “CardosoMoreira”, an

early-diverging species of *Hippeastrum* (Oliveira, 2012) collected at the locality of Cardoso Moreira, Rio de Janeiro, Brazil.

2.2. Identification and selection of nuclear loci

Ten transcriptomic data sets from Amaryllidaceae s. l. (Table S1) were obtained from the 1KP project (onekp.com), four of which are from Hippeastreae (*Traubia modesta*, *Phycella cyrtanthoides*, *Rhodophiala splendens*, and *Zephyranthes treatiae*). We included an additional transcriptomic data set for *Hippeastrum brasilianum* (generated at Beijing Genomics Institute, Shenzhen, China; Table S1). All data sets were assembled as scaffolds using SOAPdenovo-trans (Xie et al., 2014). Orthologous groups of transcriptomic scaffolds were identified using a prototypic version of MarkerMiner (Chamala et al., 2015) using *Arabidopsis thaliana* as a reference genome. The single-copy nuclear (SCN) locus filtering step used the Duarte et al. (2010) list of 959 *Arabidopsis*, *Populus*, *Vitis*, and *Oryza* (APVO) genes (Table S2). In an additional run, the filtering step was skipped to explore orthogroups identified as *Arabidopsis* genes that are not necessarily single-copy.

The MarkerMiner algorithm (Chamala et al., 2015) found 7,485 orthogroups, of which 577 corresponded to APVO genes. Using information contained in The *Arabidopsis* Information Resource (TAIR) website (<http://www.arabidopsis.org>), we selected 48 genes (from now on referred to as “reference genes”) that each had a CDS length > 1500 bp and was represented by more than six Amaryllidaceae s. l. transcripts; these genes span a range of different functional groups and structural features, such as

number of introns (Table S2). The corresponding transcripts were aligned against *Arabidopsis thaliana* CDS and genomic sequences using the MAFFT E-INS-i iterative refinement method (Katoh et al., 2009; Katoh et al., 2002). After checking the overall similarity of the Amaryllidaceae transcriptomic scaffolds with the reference, *Arabidopsis* sequences were removed, and the transcripts were re-aligned using the MAFFT G-INS-i algorithm.

2.3. Probe design, library construction, sequence capture, and sequencing

Alignments of the 48 selected reference genes were used to design custom biotinylated RNA bait/probe libraries using the MYbaits hybrid enrichment system (MYcroarray, Ann Arbor, Michigan). The 5,045 baits had a length of 98 to 120 bp and were 2x tiled following the manufacturer's recommendations.

Genomic DNA was extracted from silica-dried leaves following a modified 2x CTAB method (Cullings, 1992; Doyle, 1987; García et al., 2014). Two or three extractions were performed for each sample and combined after resuspension to obtain at least 5 µg of starting DNA to build the libraries. The combined DNA extractions were quantified using 2 µL in a Qubit® 2.0 Fluorometer (Life Technologies, Grand Island, New York) and the dsDNA Broad Range Assay.

Sequencing libraries were constructed using half reactions of the NEBNext® DNA Library Prep Master Mix Set for Illumina® kit (New England Biolabs, Ipswich, Massachusetts) following the manufacturer's protocol. Sequence capture was

performed following the manufacturer's protocol (MYcroarray) and the first additional step suggested by Mandel et al. (2014) (i.e., the amount of Hyb #1 used in Step III of the MYbaits manual v.1.3.7 was reduced from 20 μ L to 15 μ L per sample). Sequencing of the target-enriched libraries was conducted at the Biotechnology Center, University of Wisconsin-Madison (UW-Madison), using the Illumina MiSeq platform with either 150- or 250-bp paired-end reads. Reads were demultiplexed, and barcodes were trimmed from the sequences at UW-Madison prior to delivery of the data. Details on library construction, sequence capture, and sequencing are reported in Supplementary Material.

2.4. Assemblies, extension of contigs, and alignments

Prior to analyses, adaptor sequences were trimmed from reads using Cutadapt (Martin, 2011), and sequences were then filtered using Sickle (Joshi and Fass, 2011) to trim reads when the average quality in a sliding window (0.1x read length) fell below 20. Quality-filtered reads were subject to a BLAT (Kent, 2002) analysis using the longest transcriptomic sequence per gene (among those used to design the baits) as a reference. On-target reads (i.e., reads that had a BLAT hit to the transcriptomic reference) were pulled out from the quality-filtered reads file using the BLAT output files and a custom Python script. To compare sequencing results among species, coverage was roughly estimated, considering the number of on-target reads, a quality-filtered average read length (i.e., 135 and 225 for 150- and 250-bp reads, respectively), and the total CDS length of the 48 targeted genes as estimated in *Arabidopsis thaliana* (i.e., ~102,000 bp).

Curated assembly approach—Velvet (Zerbino and Birney, 2008; Zerbino, 2010) was used to perform *de novo* assemblies of the on-target reads. Estimated insert lengths of 300 and 450 bp were used for samples sequenced with 150- and 250-bp reads, respectively. Kmer values between 29 and 125 were evaluated every 4 units. The best two assemblies per species were chosen based on the n50 and max values and were exported into Geneious v. 8.1.7 (Drummond et al., 2011).

Given that some of the 48 targeted loci were unlikely to have been captured successfully due to short exons and long introns (<http://www.arabidopsis.org>) and given a low proportion of on-target reads, 11 genes that had at least one exon longer than 300 bp and short introns (i.e., <100 bp) or only coding sequences based on the *Arabidopsis thaliana* genome were selected to be explored in detail. Assembly scaffolds were mapped to the Amaryllidaceae transcriptomic reference with Geneious, using high sensitivity and “Move many gaps” fine-tuning parameters. Scaffolds with coverage of 1.0 were deleted from alignments. In cases where it seemed that there was more than a single sequence per locus (i.e., reads mapping to the same region but differing by multiple mismatches), the longest scaffolds per variant were retained for downstream analyses (i.e., short putative paralogs or chimeric sequences were discarded).

After scaffolds were extracted for several species for a particular locus, they were aligned against the reference transcriptomic sequence with the MAFFT E-INS-I algorithm (Katoh et al., 2002) in Geneious. Alignments were checked visually, and

overlapping scaffolds without mismatches were combined for each species. Sequence ends that did not align well were trimmed, and the alignments were edited manually in Geneious, especially in regions with gaps and introns. Non-overlapping scaffolds were united by NNNs using the taxa with the longest contigs as the reference.

Once more than 20 species were aligned and edited, a 50% consensus sequence of the alignment was extracted to use as a reference for mapping scaffolds from the remaining species. We deleted sequences that represented less than 20% of the alignment length. Putative pseudogenes were identified by the presence of premature stop codons and unusually high substitution rates (Gojobori et al., 1982; Balakirev and Ayala, 2003) in amino acid alignments (excluding introns when present) in Geneious. Paralogous loci (and allelic variants) within a reference gene were determined through phylogenetic analysis as defined in section 2.5.

Automated assembly approach—As an alternative to the above approach, we used aTRAM (Allen et al., 2015) to assemble single orthologs from each sample under a primarily automated pipeline. This method used an iterative target-restricted *de novo* approach, restricting the assembly problem only to reads that BLAST to target orthologs, and integrating multiple assembly algorithms. Velvet was used as the assembler, with the same lowest kmer settings as those used above; the reference sequences were also kept the same. Five iterations were used per ortholog. For the contigs recovered, we performed exonerate-based exon extraction (Slater and Birney, 2005) and contig stitching using the pipeline described in Allen et al. (2017), available at

https://github.com/juliema/phylogenomic_pipeline. Alignments were performed in MAFFT; areas of misalignment and sequences with large amounts of missing data were curated as above. Alignment regions with less than 30% of mean pairwise identity (as estimated by Geneious Pro) were trimmed. Following the latter step, only alignments longer than 200 bp were considered for downstream analyses. The transcriptomic reference sequences were removed from the final aTRAM alignments.

Gene tree, ASTRAL, and coalescent simulation analyses were exactly replicated on both the aTRAM-based and curated assemblies to test for sensitivity of the results to the assembly method.

2.5. Nuclear loci analyses

Maximum likelihood (ML) analyses were conducted for each curated and aTRAM alignment using RAxML ver. 8.0.25 (Stamatakis, 2014). Whenever a curated locus included non-coding regions, a partition for exons and a second for introns were used to increase model fit by accommodating within-locus variation. ML searches were conducted using the rapid hill-climbing algorithm (Stamatakis, 2006) with 100 independent searches starting from randomized parsimony trees with the GTRGAMMA model and four discrete rate categories. Likelihood bootstrap (BS) analyses (Felsenstein, 1985) were conducted with 1,000 pseudoreplicates for each data set. Bootstrap values greater than 85% were considered strong support, 70-85% were designated moderate support, and less than 70% as weak.

Whenever the phylogenetic analysis of a given locus generated a tree with two or more subclades containing a duplicate sequence for at least one subtribe, the subclades were considered to represent paralogous loci (e.g., locus 39, OXP1, is represented by two paralogs or curated nuclear loci in the hippeastroids: OXP1A and OXP1B), and each locus was analyzed separately following the methods described above. If a given species was represented by multiple sequence variants within an individual locus alignment, these variants were considered to represent alleles and were included in all analyses (except concatenation and neighbor-net). The *Lycoris* and *Narcissus* sequences of a given locus were repeated in each curated alignment because these were usually single-copy or had clade-specific duplications (i.e., not shared with New World taxa). The trees were rooted in Dendroscope v. 3.2.10 (Huson and Scornavacca, 2012) using the branch leading to the *Lycoris/Narcissus* clade or alternatively *Lycoris radiata*, when the analysis did not recover the Eurasian clade (Meerow et al., 2006).

Coalescent-based species trees were estimated using *BEAST v. 1.8.1 (Heled and Drummond, 2010) with 18 curated nuclear alignments, including multiple alleles where present. Two *BEAST analyses were conducted: 1) including introns, and 2) only coding sequences. Whenever a species was missing from a given partition, it was represented by a string of Ns with equal length to the alignment because *BEAST cannot handle partitions with missing species. We took a conservative approach to the interpretation of posterior probabilities (PP). Thus, nodes with $PP \geq 0.95$ were considered strongly supported, nodes with $0.85 \leq PP < 0.95$ were considered moderately supported, and $PP < 0.85$ was considered weak support.

Each analysis consisted of three runs of 500 million Markov Chain Monte Carlo (MCMC) generations sampled every 50,000 iterations, with random starting trees generated under a constant population size coalescent model for each partition, uncorrelated lognormal relaxed clock models, and a pure birth (Yule) tree prior. Rate variation across branches was modeled as uncorrelated among partitions and lognormally distributed. Appropriate substitution models were selected based on the Bayesian Information Criterion (BIC) score as implemented in jModelTest 2.1.6 (Darriba et al., 2012) without considering invariant-site models (Stamatakis et al., 2008; Yang, 2006) and optimizing trees for likelihood calculations using the Best option. Models were approximated according to the options given by *BEAST, with empirical base frequencies and four gamma categories (Table 2). Rates were estimated relative to SUS2A. Log files were analyzed in Tracer v. 1.6 (Drummond and Rambaut, 2007) to ascertain convergence and adequate effective sample sizes ($ESS > 200$) for the combined runs with a 10% burn-in and to select the two runs that gave the best combination of ESS values.

We also used an alternative species tree method to explore Hippeastrinae in particular, Bayesian Concordance Analysis (BCA; Ané et al., 2007) with BUCKy v.1.4.4 (Larget et al., 2010). A concordance factor (CF) represents the proportion of the genome sampled that supports a given node in a species tree (Ané et al., 2006; Baum, 2007). For the BUCKy analyses, only coding regions from 12 curated loci with complete Hippeastrinae sampling (Table 2) were selected for the generation of tree files using MrBayes v. 3.2.4 (Ronquist et al., 2012) for BCA in BUCKy. Considering that BUCKy can only analyze

sequence markers with identical taxon sampling, BCA was performed on a taxonomically reduced data set comprising 25 taxa (24 Hippeastrinae plus *Lycoris radiata* as outgroup), because some of these loci were putatively lost or pseudogenized in Traubiinae. KAKUSAN 4 (Tanabe, 2011) was used to evaluate nucleotide substitution models for each locus. Two replicate analyses were run on each partition, with eight million iterations and four chains each, sampling every 1,000 iterations, and a burn-in of 10,000 saved samples for summarizing posterior probabilities of both parameter values and trees.

The resulting tree files from the two runs for each partition contained a total of 16,000 trees, which were transformed to BUCKy infiles using the mbsum module of BUCKy, with a further burn-in of 1,000 trees. BUCKy was run four times for 10 million MCMC updates, each with four chains. Analyses were conducted for two different alpha levels (0 and infinity; this parameter is a prior on the expected degree of incongruence, with 0 being complete congruence, and infinity representing completely independent topologies).

As an alternative way to explore the data and identify areas of the phylogeny potentially affected by ILS (Kubatko and Degnan, 2007), we performed a concatenated analysis for 40 species (i.e., excluding *Narcissus viridiflorus*) with three curated loci excluded due to putative losses or pseudogenizations (*OXF1A*, *VPS11B*, and *sf3BS1A*; Table 2). Other details followed those used in the RAxML analyses above. A single sequence per species was selected from each curated locus based on completeness, coverage, and

pseudogenization of sequence variants. Two analyses were performed using a matrix of 15 concatenated curated nuclear loci: 1) including exons and introns as two separate partitions per gene (25,077 bp; 23 partitions), and 2) consisting of only coding sequences (22,853 bp; 15 partitions). An additional concatenated analysis was performed with 39 loci obtained through automated assembly (32,574 bp; 39 partitions).

2.6. Assembly and analyses of plastid genomes

We built a genomic paired-end library of *Hippeastrum cipoanum* (voucher: Lovo *et al.* 455, SPF; Diamantina-Brazil) with the TruSeq DNA PCR-free sample preparation kit (Illumina, San Diego, California), targeting fragments approximately 500 bp in length. The library was sequenced on an Illumina MiSeq instrument using MiSeq Reagent Kit v3 (Illumina) and 600 cycles at the *Centro de Facilidades de Apoio a Pesquisa* (ICB – USP, Brazil). We filtered the reads with Bowtie2 (Langmead and Salzberg, 2012) incorporated on CodonCode Aligner (CodonCode Corporation, Centerville, Massachusetts), using the complete *Allium cepa* (accession number: NC_024813) plastid genome as a reference. *De novo* assembly was conducted on CodonCode Aligner with default parameters. The largest resulting contig, with an average coverage depth of 93x, was checked with bl2seq to find the boundaries between single and repeated regions, and the missing Inverted Repeat (IR) was completed manually from the successfully assembled copy.

The plastid genome of *Hippeastrum cipoanum* was annotated using DOGMA (Wyman *et al.*, 2004) and used as a reference to pull out plastid reads from the quality-filtered

Illumina data with BLAT (as for on-target nuclear reads). Plastid reads were imported into Geneious where duplicate reads were removed. The remaining reads were mapped onto the *H. cipoanum* reference with the Map to Reference tool in Geneious, using medium sensitivity and “Move many gaps” fine-tuning. The resulting assemblies were checked visually to remove any spurious ambiguous base pairs due to mismatching read ends and local low coverage. Regions with no reads were filled with N’s. After visual inspection, the 50% consensus sequence was extracted from the assemblies.

The resultant 40 plastome sequences were aligned in Geneious using Mauve progressive, assuming collinear genomes (which is enforced by the read mapping method) but otherwise using default options. The resulting alignment consisted of extremely long gaps (>10 kb) due to missing data and was therefore edited manually in Geneious, by dragging long stretches of Ns to the correct position and deleting artificial gaps. The alignment was further edited to exclude IR_A, autapomorphic insertions, and several regions that were difficult to align (i.e., trimmed alignment).

The trimmed alignment was analyzed in RAxML ver. 8.0.25 (Stamatakis, 2014) as for each nuclear locus and the concatenated data set. Four analyses were conducted following different partition strategies: 1) whole alignment, unpartitioned; 2) whole alignment with partitions corresponding to coding Large Single-Copy (LSC), coding Small Single-Copy (SSC), coding IR, non-coding LSC, non-coding SSC, and non-coding IR partitions (6 partitions); 3) only coding regions corresponding to LSC, SSC,

and IR partitions (3 partitions); and 4) only non-coding regions corresponding to LSC, SSC, and IR partitions (3 partitions).

2.7. Nuclear-plastid concatenated and network analyses

The data set of 15 curated nuclear loci (excluding introns) used for concatenated analysis (Table 2) was appended to plastid sequences and analyzed in RAxML as for the nuclear gene trees. Two data sets were analyzed to explore the effect of varying amounts of plastid data over the tree inference: 1) nuclear loci plus the 15 longest plastid protein-coding genes (i.e., 3'*ycf1*, *rpoC2*, *rpoB*, *psaA*, *psaB*, *ndhF*, *rpoC1*, *matK*, *psbB*, *atpA*, *ndhD*, *atpB*, *accD*, *rbcL*, *psbC*) in the SSC and LSC regions (30 partitions), and 2) nuclear loci plus complete coding regions of the SSC and LSC (17 partitions).

Both concatenated data sets were subjected to phylogenetic network analyses to explore conflicting signals within and between partitions (Morrison, 2010). Networks were computed with the Neighbor-Net algorithm (Bryant and Moulton, 2004) in SplitsTree v.4.14.2 (Huson and Bryant, 2006) using uncorrected *p*-distances. The highly divergent Eurasian taxa and base positions with over 50% missing data were excluded from these analyses. In addition, the data set of 15 concatenated nuclear loci including only coding regions was analyzed in the same way to explore the effect of excluding plastid data on the network inference. A 95% confidence network (Huson and Bryant, 2006) was computed from 1,000 bootstrap pseudoreplicates for each of these three data sets.

Additionally, we illustrated the observed cytonuclear discordance through a Hybridization Network (García et al., 2014; Huson et al., 2010), which is an approach that explicitly represents reticulate evolution and that assumes that most topological conflict is due to hybridization (Huson et al., 2010; Huson and Scornavacca, 2011). A Hybridization Network was computed in SplitsTree using the *BEAST species tree derived from coding regions and collapsed at nodes < 0.90 PP and the 80% BS plastid tree (full alignment, six partitions).

2.8. Distinguishing hybridization from ILS

We employed the software JML (Joly, 2012) to attempt to determine whether gene tree discordance in *Hippeastreae* is caused by hybridization or ILS. The program implements the method proposed by Joly et al. (2009) to identify introgressed sequences under an ILS hypothesis. This method uses the posterior distribution of species trees (thereby accounting for uncertainty in species tree inference) and population sizes and branch lengths to simulate replicate sequence data sets under a coalescent model with no migration. Tests are performed pairwise for extant tip taxa, using gene tree topologies and branch lengths. The minimum pairwise distance between sequences of two species is evaluated on simulated data sets and compared to that estimated from the empirical data. Departure from the bifurcating species tree model could be interpreted as evidence of hybridization, assuming that the test quantity is a good predictor of such events (Joly, 2012; Joly et al., 2009).

JML analyses were performed for the coding alignments of the 12 longest curated nuclear loci (Table 2), including all sequences and the complete coding regions of the plastid genome. Only the longest curated loci were analyzed in JML because, according to Joly (2012), alignment length may have a profound effect over the performance of the algorithm to detect introgressed sequences. We used the collection of species trees (with branch lengths and population sizes) estimated by *BEAST for the analysis that includes nuclear coding sequences after combining two runs with a burn-in of 10% each (18,002 trees total; see above). These trees were used as input for JML, with the thinning parameter set to 2 so that only 9,001 trees were sampled for the 10,000 simulations. For each locus, sequences of the same length as the original were simulated according to the best substitution model and parameter values as determined by the BIC in jModelTest. A heredity scalar of $\frac{1}{2}$ was used for simulating plastid sequences, while 1 was used for nuclear loci. The relative mutation rate used in the simulations for each gene was set to the median posterior value obtained from a separate *BEAST run (500 million generations, sampled every 50,000, 25% burn-in; rates relative to SUS2A) that included the 13 tested coding partitions (12 curated nuclear loci plus the plastid genome; Table S5); the log file of this analysis was read in Tracer to obtain the mutation rate values used in JML. Settings and evolution models were as previously described for *BEAST analyses, except that GTR + Gamma and mitochondrial ploidy type were set for the plastid partition. The overall fit of the model (bifurcating species tree without gene flow) was investigated, and all observed distances that had a PP < 0.1 of being generated by the posterior distribution are reported (Joly, 2012).

As an additional way of examining ILS vs. hybridization, we simulated plastid gene trees to obtain direct estimates of expected levels of ILS throughout the tree. We used ASTRAL-II (Mirarab and Warnow, 2015) with an allele table comparable to the *BEAST analysis for the curated loci to obtain a species tree with annotated coalescent branch lengths. It is important to account for mode of inheritance of different loci; haploidy and uniparental inheritance are expected under an organellar gene tree model, whereas these processes do not occur in genes informing our species tree estimates. For strictly hermaphroditic taxa, generally a branch length scaling factor of 2 is used (cf., previous section, representing half the effective population size N_e since uniparental inheritance is not expected to affect N_e when all individuals can pass on the cytoplasm, where coalescent branch lengths estimated from diploid loci are defined as $T/2N_e$ with T = number of generations). However, the ratio of N_e among loci is known to be sensitive to factors such as sex bias (Storz et al., 2001) and sexual selection (Charlesworth, 2009). Therefore, we also assessed sensitivity of the results to scaling factors of 1, 2, and 4. Two ASTRAL analyses and subsequent series of coalescent simulations were performed to evaluate sensitivity to the assembly approach: 1) 12 alignments (including introns) obtained via curated assembly (Table 2) and 2) 39 alignments obtained via automated assembly (Table S2). One thousand gene trees were simulated on the scaled trees under the coalescent using a python-based simulation script using dendropy, based on Mirarab et al. (2014). We summarized this gene tree distribution in two ways: by plotting the overall simulated gene tree distribution with DensiTree 2 (Bouckaert and Heled, 2014), and by querying the gene tree distribution for key clades

by plotting simulated clade frequencies on the ASTRAL species trees and plastid tree (using RAxML). Under a scenario of ILS alone, any relationships in the empirical plastid tree should be present in the simulated plastid trees; under a hybridization scenario, relationships unique to the plastid genome should be low (or 0) in frequency (i.e., unexpected under ILS alone), and in the simulated trees should be contradicted by clades consistent with the estimated species tree that are much more frequent (i.e., clades that are expected under ILS).

2.9. Data archiving

Scripts, probe sequences, alignments, partition files, tree and network files, JML results, and simulated gene trees are archived in Dryad, [http/XXXXX](http://XXXXX).

3. Results

3.1. Sequencing and nuclear loci assemblies

An average yield of 223 Mb and a total of ~1,140,000 reads were obtained per sample. Around 24,000 reads (~2% of total) were lost after the quality-filtering step. Only 9,700 reads on average (1.02% of total) were pulled out as being on-target based on the BLAT analyses against the 48 transcriptomic reference sequences. In spite of the low number of reads, median exon coverage of 15.1X was estimated.

Lower kmer values (kmer 1) ranging between 37 – 85 and upper values (kmer 2) between 57 – 121 were selected to explore nuclear gene *de novo* assemblies (Table S4). On average, we obtained per species a total of 1,023 nodes with an N50 length of

244 bp for kmer 1 values, while 470 nodes with an N50 of 277 bp were retrieved for kmer 2 (Table S4).

3.2. Nuclear loci analyses

Gene trees inferred from total curated alignments of 11 loci and for each paralog (giving a total of 18 curated nuclear loci) are shown in Supplementary Figures S1 through S26. Summary statistics for curated nuclear alignments, including length, number of introns, variable and parsimony-informative sites, model of sequence evolution, phylogenetic coverage, and a check-list of analyses performed with each locus, are shown in Table 2. Most loci (except *RPB2*, *EGY1*, and *DDB2*) had two paralogous copies in all Hippeastreae or in one of the two subtribes. Putative pseudogenes were detected for Hippeastreae sequences in *DDB2*, *CER9A*, *OXF1B*, *Rhod7A*, *Rhod7B*, *RPB2*, *sf3BS1A*, *TOC75-3A*, and *VPS11B*. Among the 18 curated nuclear loci alignments, missing data ranged between 15.5 and 41.9% with a mean value of 26.9% (Table 2). Using automated assembly with aTRAM, we obtained alignments for 39 loci; number of taxa, alignment length, and percentage of missing data per locus are available in Table S2. All 39 gene trees are available in Dryad and in general are less resolved than gene trees obtained from curated loci (e.g., only five gene trees retrieve either a monophyletic Hippeastreae or monophyletic subtribes).

Coalescent-based species tree analyses of curated nuclear loci in *BEAST recovered clades corresponding to tribe Hippeastreae and both subtribes Hippeastrinae and Traubiinae with strong support (Fig. 1). Also, all analyses agreed on the general

framework of Traubiinae phylogeny: a grade is formed by *Rhodolirium laetum*, *Traubia modesta*, and *Rhodolirium montanum*, and a strongly supported clade with short internal branches and low resolution is formed by *Phycella*, *Placea*, *Rhodolirium speciosum*, and *Famatina maulensis* (i.e., *Phycella* group). ASTRAL results were largely consistent under both assembly methodologies (Fig. 2).

Both *BEAST analyses recovered the same topology with two major clades in Hippeastrinae: A) *Tocantinia dutilhiana* sister to *Hippeastrum* (i.e., $2n = 22$ clade), and B) *Eithea*, *Rhodophiala bifida*, *Habranthus*, *Zephyranthes*, and core *Rhodophiala* (i.e., $2n = 12-18$ clade). The $2n = 22$ clade and *Hippeastrum* are strongly supported (PP = 0.97-1); in contrast, the $2n = 12-18$ clade is well supported only when introns are excluded from the analysis (PP = 1; Fig. 1) and is weakly supported when introns are included (PP = 0.83). The $2n = 12-18$ clade consists of two subclades: 1) *Zephyranthes* spp. sister to *Habranthus martinezii*, *H. robustus*, and *H. brachyandrus* (i.e., *Zephyranthes*-*H. robustus* group; PP = 0.96-0.98); and 2) *Rhodophiala bifida*, *Habranthus pedunculatus*, and *Eithea* forming a grade along with a strongly supported clade (PP= 1.0) that includes *Habranthus gracilifolius* and *H. jamesonii* (i.e., *H. gracilifolius* group) as sister to the core *Rhodophiala* clade. Both ASTRAL analyses recovered the same two major clades within Hippeastrinae; however, these are weakly supported (Fig. 2). The ASTRAL tree inferred from curated loci is topologically consistent with the *BEAST trees (Fig. 2A). In contrast, the ASTRAL tree inferred from automated assemblies shows different relationships within both Hippeastrinae

subclades, albeit with shorter coalescent times and lower resolution and statistical support (Fig. 2B).

In the BUCKy analyses, there were no differences in CF scores or tree topologies between the two different alpha parameterizations. Concordance factor scores were low at most internal nodes of the BUCKy primary concordance factor tree (Fig. S27A). Highest scores ($CF > 0.8$) are for *Eithea* (0.908), *Zephyranthes* (0.831), and core *Rhodophiala* (0.846). However, the primary concordance tree does not resolve the $2n = 22$ and $2n = 12-18$ subclades. The population tree (Fig. S27B) does resolve these two subclades and is in agreement with the *BEAST species tree (Fig. 1), with the exception of *Rhodophiala bifida* and *H. pedunculatus* as unresolved sisters to the clade formed by *Zephyranthes* and the *H. robustus* group.

Both concatenated analyses of 15 curated loci recovered slightly different topologies from the *BEAST tree (Fig. S28; Table 3). Within Traubiinae, there is strong support for the sister relationship between *Rhodolirium montanum* and the *Phycella* group, and the latter shows a subclade that excludes *Placea*, which is strongly supported in the coding-only analysis (Fig. S28). Within Hippeastrinae, there are several differences as well. The $2n = 22$ clade is not present, because *Tocantinia dutilhiana* is included within a weakly supported $2n = 12-18$ clade. Additionally, within the expanded $2n = 12-18$ clade, *Eithea*, the *H. gracilifolius* group, and core *Rhodophiala* form a monophyletic group (BS = 84-96%) that is weakly supported in *BEAST analyses. In contrast, the concatenated analysis of aTRAM assemblies retrieved a very different topology within both subtribal

clades, in general less resolved and supported than in analyses of curated loci (Fig. S29). Most notable differences are within Hippeastrinae, where this analysis fails to recover clades that are found in most previous analyses of nuclear loci, such as core *Rhodophiala* and *Hippeastrum*.

3.3. Plastid tree

On average, ~15,900 (1.86%) reads per sample were identified as plastid (Table S3). Mean plastid assembly coverage ranged between 7.2 – 40.3X with a median value among taxa of 14.7X. The total plastid alignment had a length of 162,360 bp. After exclusion of problematic regions and IR_A, the trimmed alignment was 130,380 bp with 2.3% missing data or gaps (Table 3).

All plastid analyses recovered the same topology, but with differences in BS support (Fig. 3). Results for analyses of only coding and only non-coding regions are shown in Supplementary Figure S30. The topology obtained here is very similar to plastid trees obtained with just a few loci (García et al., 2014) despite the differences in taxon sampling and amount of data. However, unlike in García et al. (2014), the backbone of the present plastid tree is strongly supported, especially when non-coding regions are included in the analysis (Fig. S30).

Several clear differences between the plastid tree and the nuclear *BEAST tree are notable, including the following plastid relationships: 1) polyphyly of the $2n = 22$ clade and *Hippeastrum* (*H. brasilianum* is sister to the remaining Hippeastrinae), 2) the *H.*

gracilifolius group sister to *Zephyranthes*, 3) *Habranthus pedunculatus* forms a clade with the *Habranthus robustus* group, and this clade is sister to *Rhodophiala bifida*, and 4) *Eithea* spp. are recovered in a clade with *Hippeastrum reticulatum* and *H.* “CardosoMoreira”.

3.4. Nuclear-plastid concatenated and network analyses

The main effect of the plastid signal on Hippeastrinae phylogenetic inference involves the non-monophyly of the *Zephyranthes*-*H. robustus* clade in *BEAST and concatenated analyses of nuclear loci (Fig. S31). In contrast, when plastid data are included, two separate clades are inferred: 1) *Zephyranthes* spp. and 2) *Rhodophiala bifida*, *Habranthus pedunculatus*, and the *H. robustus* group. The support for the clade that includes *Rhodophiala bifida*, *H. pedunculatus*, and *H. robustus*, and for backbone relationships in Hippeastrinae increases when complete SSC and LSC regions are considered (Fig. S32). Including more plastid data in the analysis also has the effect of recovering a weakly supported $2n = 22$ clade (BS = 43-44; Fig. S32).

The 95% confidence network for 15 concatenated curated nuclear loci plus the 15 longest plastid genes is shown in Figure 4A. NeighborNet networks (Fig. 4A, S33, S34, S35, S36, S37) suggest a mostly tree-like pattern of evolution in Hippeastreae and unresolved relationships at the base of Hippeastrinae. The topology for Traubiinae is similar among analyses and consists of three early-diverging lineages of reticulate or unresolved relationships, *Rhodolirium laetum*, *Traubia modesta*, and *Rhodolirium*

montanum. The *Phycella* group is consistently recovered and has very short terminal branches.

Three main lineages are inferred within Hippeastrinae when a moderate amount of plastid data are included in the matrix (Fig. 4A): 1) the $2n = 22$ clade with *Tocantinia dutilhiana* sister to *Hippeastrum*; 2) *Eithea* sister to the *H. gracilifolius* group and core *Rhodophiala*; and 3) *Rhodophiala bifida*, *Habranthus pedunculatus* plus *Zephyranthes* - *Habranthus robustus* group. The positions of *Rhodophiala bifida* and *Habranthus pedunculatus* (Fig. S33, S35) are unresolved with only nuclear data. However, the plastid signal seems to suggest a strong relationship of *Rhodophiala bifida* and *Habranthus pedunculatus* with the *H. robustus* group, although with prevailing conflict as expressed by the reticulate nodes shown in the corresponding 95% confidence network (Fig. 4A). The main difference noted when more plastid data are added concerns relationships within the $2n = 22$ clade, with *Hippeastrum* and *Tocantinia dutilhiana* in a polytomy at the base of Hippeastrinae (Fig. S37).

The hybridization network shows widespread cytonuclear discordance along the backbone of Hippeastrinae (Fig. 4B). In contrast, a tree-like pattern of evolution is inferred for Traubiinae. Within Hippeastrinae, five clusters can be identified, which also reflect different chromosome numbers: (1) *Hippeastrum* + *Tocantinia* ($2n = 22$), (2) core *Rhodophiala* ($2n = 18$), (3) *Rhodophiala bifida* ($2n = 16, 18$), (4) *Eithea* ($2n = 18$), and (5) *Habranthus* + *Zephyranthes* ($2n = 12, 14$).

3.5. Distinguishing hybridization from ILS

JML found no significant results at the $\alpha = 0.01$ level and few distances between sequences that were smaller than the 1st quantile in the posterior predictive distributions of 12 curated nuclear loci tested and complete plastid coding sequences (Table S3). Under the model of hybridization used by JML (which assumes introgression among pairs of extant species), a bifurcating species tree model without gene flow cannot be rejected.

However, simulating a plastid gene tree distribution directly from the ASTRAL species trees shows that several observed plastid discordances are almost never expected to occur under the coalescent alone (frequency ~ 0 ; Table 4, Fig. 6). For instance, a strong majority of gene trees contained a monophyletic *Hippeastrum*, and all nodes in the observed plastid tree that contradict this clade have 0 frequency among simulated plastid gene trees for all branch length scalars, demonstrating insensitivity to estimates of N_e . By contrast, the corresponding clades in the nucleus had clade frequencies $\gg 0$; this difference was greatest when using an X4 branch length scalar for curated assemblies, but was also true of the other branch length scalars and the aTRAM assembly. Similar observations were found for all clades with relatively long stem lineage lengths (uncorrected $\tau > 1$ for curated assemblies; $\tau > 0.5$ for aTRAM assemblies), but for early-branching lineages of Hippeastreae, where branches were considerably shorter, ILS was predicted to be frequent, given the estimated species tree (observed plastid clade frequencies > 0). The difference in estimated coalescent branch lengths between curated and aTRAM assemblies likely reflects differences in gene tree

estimation error between both resultant alignment sets; higher error is expected to result in branch length underestimates (Zhong et al., 2014).

4. Discussion

4.1. Gene tree and cytonuclear discordance

Assuming that gene trees are known without error, three main phenomena can cause gene tree heterogeneity and consequent gene tree-species tree discordance: gene duplication/extinction, horizontal gene transfer/hybridization, and ILS (Maddison, 1997; Maddison and Knowles, 2006; Edwards, 2009; Knowles, 2009). For the curated assembly approach, we strove to eliminate potential issues of gene duplications via the thorough examination of sequence variants in our data. Nonetheless, the impact of putative gene losses/pseudogenizations is difficult to evaluate at this point. Yet we stress that paralogy issues are unlikely to affect the topology of entire plastid genomes, given that this would require lineages to maintain heteroplasmy for thousands to millions of years.

Methods that explicitly model hybridization are desirable for testing hypotheses of reticulation. Nevertheless, currently available methods suffer from critical limitations for the simple fact that it is currently too difficult to enumerate and search through the space of all possible hybridization scenarios. Explicit tests of hybridization generally evaluate a constrained set of models that may not include all reasonable scenarios for the group of interest, effectively assuming that other hybridization scenarios are impossible. The clear identification of these assumptions is critical to interpreting results

correctly. JML conceptualizes hybridization as a process that results in the introgression of a sequence from one extant sampled lineage to another. Therefore, while we found no significant results, our statistical tests could not incorporate a major potential process for this clade (García et al., 2014) – a hybridization event ancestral to a clade, involving an ancestral lineage that is no longer extant. Similar limitations are present in other methods such as STEM-hy (Kubatko, 2009).

By contrast, by not explicitly modeling the process of hybridization, a direct simulation of organellar gene trees has the advantage of being applicable to a wide set of scenarios with few assumptions, including one that is relevant to the case of Hippeastrinae: the level of ILS expected for stem branches of clades. Given the species tree we estimated, ILS is expected among early-diverging lineages of Hippeastrinae. Gene tree simulation (Table 4, Fig. 6) demonstrates that the pattern observed in the plastid tree is largely unexpected given the coalescent alone. The most dramatically unusual relationships observed in the plastid tree, such as the polyphyly of *Hippeastrum*, are at 0 frequency in the simulated plastid trees; this was insensitive to assembly method and branch length scalar. This is unsurprising given that coalescent branch lengths were relatively long for the stem lineage leading to these clades (unscaled $\tau \sim 1$ for curated assemblies; $\tau > 0.5$ for aTRAM assemblies). This indicates that ILS is unlikely for these relationships, yet paralogy (which would imply long periods of heteroplasmy across multiple lineages) is likewise an unlikely explanation for aberrant organellar topologies. Hence, notwithstanding the effect of ILS, coalescent simulations suggest that key instances of cytonuclear discordance (e.g., non-monophyly of *Hippeastrum*) cannot be explained by

ILS alone, and therefore that ancient hybridization and consequent plastid capture need to be invoked to explain such conflicting signals.

Hence, the disagreement observed between JML and the distribution of simulated organellar gene trees mainly lies in the conceptualization of hybridization, and critically in whether gene tree expectations for ancestral lineages (clade frequencies in the simulated approach, or sequence distances in JML) are directly accounted for.

Notwithstanding these methodological issues, coalescent simulations lead to the conclusion that multiple hybridization events were involved in the early diversification of Hippeastrinae, while not ruling out a role for ILS, particularly among the most early-diverging lineages. A logical consequence of invoking ancient hybridization (e.g., Sun et al., 2015) is that a network model (Figs. 4B, 6) is more appropriate to represent deeper relationships within this group (Sneath, 1975; Linder and Rieseberg, 2004; Huson and Bryant, 2006; Huson et al., 2010; Nakhleh, 2011, 2013); thus, any available tree inference method will have problems inferring a bifurcating history for Hippeastrinae given its underlying reticulate pattern of evolution.

4.2. Hippeastreae phylogeny

4.2.1. Hippeastrinae

We propose a network scenario that conservatively identifies six well-supported instances of plastid transfers. This hypothesis does not exclude ILS for Hippeastrinae; the putative horizontal transfers occurred among relatively derived lineages, while discordance among earlier-diverging lineages in the backbone of the species tree is

attributed to ILS (Fig. 6). In view of this evidence, the hybridization network (Fig. 4B) and reticulogram (Fig. 6) are more effective than tree-like models of phylogeny in showing relationships among Hippeastrinae.

Within the $2n = 22$ clade, *Hippeastrum* may have at least three different origins of its plastid and, therefore, at least three maternal contributors of genetic material: 1) *H. brasilianum* from an early-diverging, perhaps extinct or unsampled, plastid lineage; 2) *H. reticulatum* and *H. "CardosoMoreira"* from an ancestral *Zephyranthes* lineage (see below); and 3) the core *Hippeastrum* clade, perhaps representing the "original" *Hippeastrum* plastid lineage. Inferences of chromosome number evolution (not shown) and other literature (e.g., Flory, 1977; Naranjo and Poggio, 2000) have suggested that $2n = 22$ is likely the ancestral number for Amaryllidaceae and for Hippeastreae in particular. Despite its multiple plastid origins, *Hippeastrum* is a very robust lineage in terms of its nuclear genealogy and chromosomal configuration (Poggio et al., 2014).

Regarding relationships within the $2n = 12-18$ clade, species tree analyses of curated nuclear loci (Fig. 1) suggest a polytomy of five lineages: 1) *Habranthus pedunculatus*, 2) *Rhodophiala bifida*, 3) *Eithea*, 4) the *H. gracilifolius* group sister to core *Rhodophiala*, and 5) *Zephyranthes* sister to the *H. robustus* group. The lack of resolution in the backbone could be the result of ancient rapid radiation (Whitfield and Lockhart, 2007; Oliver, 2013), likewise consistent with short coalescent branch lengths. The close relationship of *H. pedunculatus* to the *Habranthus robustus* group is strongly suggested by the plastid signal, as well as by ITS data (García et al., 2014). The affinity of *H.*

pedunculatus with other *Habranthus* and *Zephyranthes* species makes sense from a cytological standpoint, considering that it has a karyotype of $2n = 14$ (Flory and Flagg, 1958; Naranjo, 1974), but also $2n = 12$ (S. Arroyo-Leuenberger, pers. comm.), which is otherwise found in *Zephyranthes flavissima* and *Habranthus jamesonii* (Greizerstein and Naranjo, 1987; Naranjo, 1974).

Plastid data also strongly suggest *Rhodophiala bifida* as sister to a clade that includes *H. pedunculatus* and the *H. robustus* group (Fig. 3). Conflict remains strong in this part of the phylogeny, as evidenced by reticulate relationships in 95% confidence networks (Fig. 4A), probably representing plastid transfer from an ancestral $2n = 12-18$ lineage to *Zephyranthes* (i.e., the *Zephyranthes* plastid lineage; Fig. 6). The *Zephyranthes* plastid lineage is crucial to understanding reticulate relationships within Hippeastrinae because it was involved as the maternal contributor in at least five introgression events (Fig. 6): 1) the primary transfer from an unsampled or extinct lineage to *Zephyranthes* (i.e., excluding *Habranthus robustus* and allies); 2) from *Zephyranthes* to the most recent common ancestor (MRCA) of the *Habranthus gracilifolius* group; 3) from ancestral *Zephyranthes* to *Hippeastrum reticulatum*; 4) from *H. reticulatum* to *H. "CardosoMoreira"*; and 5) from ancestral *H. "CardosoMoreira"* to *Eithea*. This scenario invokes a "stepping stone" hypothesis for the transfer of a plastid lineage across multiple taxa (cf. Rieseberg and Soltis, 1991; Soltis and Kuzoff, 1995).

4.2.2. Traubiinae

Relationships within subtribe Traubiinae are more straightforward than for Hippeastrinae, because different methods and data sets in general agree regarding the topology and hybridization seems to be absent from this clade. It is clear that this subtribe comprises a grade of *Traubia*, *Rhodolirium laetum*, *Rhodolirium montanum* (i.e., *Rhodolirium s. s.*; García et al. (2014)), and a clade of *Placea* spp., *Phycella* spp., *Famatina maulensis*, and *Rhodolirium speciosum* (i.e., the *Phycella* group) (Figs. 4, 6). Concatenated analyses of curated nuclear loci provide moderate to strong support for a sister relationship of *Rhodolirium s. s.* and the *Phycella* group.

The *Phycella* group is robust in all analyses (except concatenation of aTRAM assemblies), and our results do not support the reciprocal monophyly of *Placea* and *Phycella*, even if *Famatina maulensis* and *Rhodolirium speciosum* are transferred to the latter genus. However, coalescent simulations and relatively short terminal branches suggest that a rapid radiation was involved in the diversification of this clade. Therefore, we cannot rule out that both genera could be reciprocally monophyletic if more data were analyzed. *Placea* species are characterized by an infundibuliform perigone and a showy, tubular, reddish corona that can be entire to divided, whereas *Phycella* typically have a tubular, red perigone and usually lack a corona (e.g., *Phycella australis*, *Phycella scarlatina*) (Ravenna, 2003). However, some *Phycella* species (e.g., *P. cyrtanthoides*) do have a corona formed by paired subulate appendages at the base of each filament that are showy in *Rhodolirium speciosum*, while *Placea davidii* does not form a tube but also shows a set of paired bifid appendages (Ravenna, 1981). Corona

morphology seems to be highly labile in the *Phycella* group, which may be associated with pollinator-mediated adaptations (i.e., shift(s) to hummingbird pollination).

4.3. Enrichment, sequencing, and *de novo* assemblies of nuclear loci

We demonstrate that nuclear phylogenetic markers can be successfully developed for non-model organisms based on mining of transcriptomic data and bait-capture/NGS approaches. This study is also a proof of concept for MarkerMiner to develop nuclear sequence markers from transcriptomic data (Chamala et al., 2015). Our results indicate that loci identified by MarkerMiner (i.e., putatively angiosperm-wide SCN genes) can be useful for phylogenetic inference. An increasing number of studies has tested putatively shared SCN genes with varying results regarding their phylogenetic informativeness (e.g., Kiefer and Koch, 2012; Naumann et al., 2013; Ripma et al., 2014; Salas-Leiva et al., 2014). We concur with Folk et al. (2015) that with careful marker design, these loci can be applied across a variety of taxonomic levels, including the species level, where anonymous population-genetic markers such as those generated by RADseq have often been applied (e.g., Hipp et al., 2014; Takahashi and Moreno, 2015; DaCosta and Sorenson, 2016); our success in recovering resolved, yet discordant, gene trees favors the method over anonymous markers for analyzing systems with reticulate evolution.

Other hybrid enrichment studies in angiosperms have not explored the presence of sequence variants (i.e., paralogs or alleles) and have assumed for downstream analysis that a single sequence is present for the gene and species in question (Mandel et al., 2014; Weitemier et al., 2014). This study calls for caution in assuming the single-copy

status of a gene for any particular group of study when using lists of shared single-copy genes (e.g., de Smet et al., 2013; Duarte et al., 2010). This is especially relevant for plant phylogenetics given the chance of small-scale duplications, such as segmental, tandem, and retro-duplications (Hudson et al., 2011), and widespread occurrence of whole-genome duplications throughout angiosperm phylogeny in general (e.g., Jiao et al., 2011; Soltis et al., 2009). The occurrence of at least two major paralogs in over 80% of the loci analyzed through the curated assembly approach could point towards polyploidy or gene duplication events that predated the MRCA of Griffinieae and Hippeastreae. The former explanation seems unlikely considering that the putative ancestral chromosome number for Amaryllidaceae s. s. ($2n = 22$), or fewer ($2n = 20$ in *Griffinia*; Meerow et al., 2002), characterize both tribes (García et al., 2014). However, diploidization of the chromosome number following a whole-genome duplication event is an alternative explanation that has been invoked in plants (e.g., Brassicaceae: Lysak et al., 2009) and animals (e.g., salmonids: Lien et al., 2016).

A noteworthy issue is the low coverage obtained for on-target sequences.

Representation of on-target reads ranged mostly between 1-2% of all reads per sample (Table S3). Individual Velvet *de novo* scaffolds had a coverage that ranged mostly between 2x and 10x. While Illumina sequencing error is among the lowest when compared to other NGS platforms ($\leq 0.1\%$; Glenn, 2011), the coverage values obtained here for on-target sequences contrast with other studies that have reported general enrichment efficiencies fluctuating around 20-60% per sample for ingroup taxa (Weitemier et al., 2014; Folk et al., 2015). This large difference is probably due to

several factors. First, these other studies targeted hundreds of genes in contrast to the 48 loci captured here, which collectively, based on the *A. thaliana* genome (<http://www.arabidopsis.org>), made up to ~100 Kb of coding regions and ~170 Kb of genomic DNA length; a rarer genomic target can be expected to be more difficult to enrich. Second and probably more importantly, Hippeastreae have significantly larger genomes than the taxa previously analyzed. For example, the species of Asteraceae and *Asclepias* studied by Mandel et al. (2014) and Weitemier et al. (2014), respectively, have genomes that are at least 10-20 times smaller than some of those studied here (e.g., Poggio et al., 2014). Hybrid enrichment is more challenging when target loci are very rare in the raw DNA pool. The genes targeted in the present study represent at most 0.001% of the genomes (contrasted with 0.06%, 60-fold greater, empirically estimated in Folk et al., 2015); the fact that we obtained 1-2% on-target reads indicates that the enrichment experiments indeed worked. Notwithstanding the issues mentioned above, the present study is among the first to explore hybrid capture approaches in large genomes (e.g., Fu et al., 2010; Neves et al., 2013; Winfield et al., 2012), and our results are encouraging by showing that low-coverage Illumina sequencing of nuclear genes can still provide useful data for phylogenetic studies. Future studies should address the optimization of hybrid enrichment on large genomes (McCartney-Melstad et al., 2015), perhaps by performing the capture twice on a library pool or experimenting with reaction conditions such as hybridization temperature (Paijmans et al., 2016).

5. Conclusions

Despite recent calls to replace tree-like models of evolution with reticulate models (e.g., Huson and Scornavacca, 2011; Nakhleh, 2011, 2013), methodological difficulties persist in this field. We therefore advocate a pluralistic approach to analyzing potential instances of reticulation. Phylogenetic analyses of multiple nuclear loci and nearly complete plastome sequences, obtained through a combination of in-solution hybrid capture and long-read Illumina sequencing, have provided new insights into relationships and have identified conflict in the phylogeny of Hippeastreae. The major radiation of Hippeastrinae probably involved both stochastic sorting of alleles and diploid hybridization, being therefore more appropriately represented as a phylogenetic network. Strong cytonuclear discordance in this clade was corroborated in the present study. Through phylogenetic analyses and coalescent simulations, we identified at least six instances of conflict that may correspond to ancient hybridization events. The phylogenetic framework reconstructed here will certainly be useful for reevaluating the generic taxonomy and morphological evolutionary framework of Hippeastreae, notwithstanding the challenge posed by a complex network-like pattern of evolution.

Acknowledgments

We thank Silvia Arroyo-Leuenberger, Julie Dutilh, Mauro Peixoto, Antonio Campos Rocha, Harri Lorenzi, Marcelo Baeza, and Patricio Novoa for field assistance and/or providing samples; Julie Allen consulted on assembly strategies. This research was supported by NSF Doctoral Dissertation Improvement grant DEB-1310839 to PSS and NG. Collection of material and extraction of some DNA samples were partially supported by National Science Foundation Grants DEB-968787 and 0129179 to AWM.

We also thank Simon Joly and an anonymous reviewer for their helpful comments to improve this manuscript.

References

- Allen, J. M., Boyd, B., Nguyen, N.-P., Vachaspati, P., Warnow, T., Huang, D. I., Grady, P. G., Bell, K. C., Cronk, Q. C., Mugisha, L., Pittendrigh, B. R., Leonardi, M. S., Reed, D. L., Johnson, K. P., 2017. Phylogenomics from whole genome sequences using aTRAM. *Systematic Biology*, syw105. doi: 10.1093/sysbio/syw105
- Allen, J. M., Huang, D. I., Cronk, Q. C., Johnson, K. P., 2015. aTRAM - automated target restricted assembly method: a fast method for assembling loci across divergent taxa from next-generation sequencing data. *BMC Bioinformatics* 16, 98. doi: 10.1186/s12859-015-0515-2
- Anderson, E., 1949. Introgressive hybridization. Wiley, New York.
- Ané, C., Larget, B., Baum, D.A., Smith, S.D., Rokas, A., 2006. Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution* 24, 412–426. doi:10.1093/molbev/msl170
- APG IV, 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* 181, 1–21.
- Arnold, M.L., Bouck, A.C., Scott Cornman, R., 2004. Verne Grant and Louisiana Irises: Is there anything new under the sun? *New Phytologist* 161, 143–149. doi:10.1046/j.1469-8137.2003.00856.x
- Álvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic

- inference. *Molecular Phylogenetics and Evolution* 29, 417–434.
- Baeza, C.M., Mariangel, C., Ruiz, E., Negritto, M., 2009. El cariotipo fundamental en *Rhodolirium speciosum* (Herb.) Ravenna y *R. andicola* (Poepp.) Ravenna (Amaryllidaceae). *Gayana Botánica* 66, 99–102.
- Balakirev, E.S., Ayala, F.J., 2003. Pseudogenes: Are they “junk” or functional DNA? *Annual Review of Genetics* 37, 123–151.
- Baum, D.A., 2007. Concordance trees, concordance factors, and the exploration of reticulate genealogy. *Taxon* 56, 416–427.
- Bouckaert, R., Heled, J., 2014. DensiTree 2: Seeing trees through the forest. *bioRxiv* 012401. doi:10.1101/012401
- Bryant, D., Moulton, V., 2004. Neighbor-net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* 21, 255–265.
- Buckley, T.R., Cordeiro, M., Marshall, D.C., Simon, C., 2006. Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (*Maoricicada* Dugdale). *Systematic Biology* 55, 411–425.
doi:10.1080/10635150600697283
- Buggs, R.J.A., Renny-Byfield, S., Chester, M., Jordon-Thaden, I.E., Viccini, L.F., Chamala, S., Leitch, A.R., Schnable, P.S., Barbazuk, W.B., Soltis, P.S., 2012. Next-generation sequencing and genome evolution in allopolyploids. *American Journal of Botany* 99, 372–382. doi:10.3732/ajb.1100395
- Büneker, H.M., Bastian, R.E., Soares, K.P., Costa, C.M., 2016. The genus *Tocantinia* (Amaryllidaceae, Amaryllidoideae) and two new species from Brazil. *Balduinia* 53,

1–14. doi: 10.5902/2358198022860

- Bybee, S.M., Bracken-Grissom, H., Haynes, B.D., Hermansen, R.A., Byers, R.L., Clement, M.J., Udall, J.A., Wilcox, E.R., Crandall, K.A., 2011. Targeted Amplicon Sequencing (TAS): A scalable next-gen approach to multilocus, multitaxa phylogenetics. *Genome Biology and Evolution* 3, 1312–1323.
doi:10.1093/gbe/evr106
- Chamala, S., García, N., Godden, G.T., Krishnakumar, V., Jordon-Thaden, I.E., Smet, R.D., Barbazuk, W.B., Soltis, D.E., Soltis, P.S., 2015. MarkerMiner 1.0: A new application for phylogenetic marker development using angiosperm transcriptomes. *Applications in Plant Sciences* 3, 1400115. doi:10.3732/apps.1400115
- Charlesworth, B., 2009. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* 10, 195–205. doi: 10.1038/nrg2526
- Cisternas, M., Araneda, L., García, N., Baeza, C.M., 2010. Karyotypic studies in the Chilean genus *Placea* (Amaryllidaceae). *Gayana. Botánica* 67, 198–205.
- Cullings, K.W., 1992. Design and testing of a plant-specific PCR primer for ecological and evolutionary studies. *Molecular Ecology* 1, 233–240.
- DaCosta, J.M., Sorenson, M.D., 2016. ddRAD-seq phylogenetics based on nucleotide, indel, and presence-absence polymorphisms: Analyses of two avian genera with contrasting histories. *Molecular Phylogenetics and Evolution* 94, 122–135.
doi:10.1016/j.ympev.2015.07.026
- Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods* 9, 772–772.
doi:10.1038/nmeth.2109

- de Smet, R., Adams, K.L., Vandepoele, K., Van Montagu, M.C., Maere, S., Van de Peer, Y., 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* 110, 2898–2903.
doi:10.1073/pnas.1300127110/-/DCSupplemental/sapp.pdf
- Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19, 11–15.
- Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S., 2011. Geneious v5. 4.
- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7, 1. doi:10.1186/1471-2148-7-214
- Duarte, J.M., Wall, P.K., Edger, P.P., Landherr, L.L., Ma, H., Pires, J.C., Leebens-Mack, J., dePamphilis, C.W., 2010. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology* 10, 61. doi:10.1186/1471-2148-10-61
- Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? *Evolution* 63, 1–19. doi:10.1111/j.1558-5646.2008.00549.x
- Felix, W.J.P., Felix, L.P., Melo, N.F., Oliveira, M.B.M., Dutilh, J.H.A., Carvalho, R., 2011. Karyotype variability in species of the genus *Zephyranthes* Herb. (Amaryllidaceae–Hippeastreae). *Plant Syst Evol* 294, 263–271.
doi:10.1007/s00606-011-0467-6
- Felsenstein, J., 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 783–791.

- Flory, W.S., 1977. Overview of chromosome evolution in the Amaryllidaceae. *Nucleus* 20, 70–88.
- Flory, W.S., Flagg, R.O., 1958. A cytological study of the genus *Habranthus*. *The Nucleus* 1, 267–280.
- Folk, R.A., Mandel, J.R., Freudenstein, J.V., 2016. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. *Systematic Biology* syw083. doi: 10.1093/sysbio/syw083
- Folk, R.A., Mandel, J.R., Freudenstein, J.V., 2015. A protocol for targeted enrichment of intron-containing sequence markers for recent radiations: A phylogenomic example from *Heuchera* (Saxifragaceae). *Applications in Plant Sciences* 3, 1500039. doi:10.3732/apps.1500039
- Fu, Y., Springer, N.M., Gerhardt, D.J., Ying, K., Yeh, C.-T., Wu, W., Swanson-Wagner, R., D'Ascenzo, M., Millard, T., Freeberg, L., 2010. Repeat subtraction-mediated sequence capture from a complex genome. *The Plant Journal* 62, 898–909.
- García, N., Meerow, A.W., Soltis, D.E., Soltis, P.S., 2014. Testing deep reticulate evolution in Amaryllidaceae tribe Hippeastreae (Asparagales) with ITS and chloroplast sequence data. *Systematic Botany* 39, 75–89. doi:10.1600/036364414X678099
- Gerard, D., Gibbs, H.L., Kubatko, L., 2011. Estimating hybridization in the presence of coalescence using phylogenetic intraspecific sampling. *BMC Evolutionary Biology* 11, 291. doi:10.1186/1471-2148-11-291
- Glenn, T.C., 2011. Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* 11, 759–769. doi:10.1111/j.1755-0998.2011.03024.x

- Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D.B., Lander, E.S., Nusbaum, C., 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* 27, 182–189. doi:10.1038/nbt.1523
- Godden, G.T., Jordon-Thaden, I.E., Chamala, S., Crowl, A.A., García, N., Germain-Aubrey, C.C., Heaney, J.M., Latvis, M., Qi, X., Gitzendanner, M.A., 2012. Making next-generation sequencing work for you: Approaches and practical considerations for marker development and phylogenetics. *Plant Ecology & Diversity* 5, 427–450. doi:10.1080/17550874.2012.745909
- Gojobori, T., Li, W.-H., Graur, D., 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. *J Mol Evol* 18, 360–369.
- Good, J.M., Vanderpool, D., Keeble, S., Bi, K., 2015. Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. *Evolution* 69, 1961–1972. doi:10.1111/evo.12712
- Greizerstein, E.J., Naranjo, C.A., 1987. Estudios cromosómicos en especies de *Zephyranthes* (Amaryllidaceae). *Darwiniana* 169–186.
- Griffin, P.C., Robin, C., Hoffmann, A.A., 2011. A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to *Poa* grasses. *BMC Biology* 9, 19.
- Harrison, N., Kidner, C.A., 2011. Next-generation sequencing and systematics: What can a billion base pairs of DNA sequence data do for you? *Taxon* 60, 1552–1566.
- Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus

- data. *Molecular Biology and Evolution* 27, 570–580. doi:10.1093/molbev/msp274
- Hipp, A.L., Eaton, D.A.R., Cavender-Bares, J., Fitzek, E., Nipper, R., Manos, P.S., 2014. A framework phylogeny of the American oak clade based on sequenced RAD data. *PLoS ONE* 9, e93975. doi:10.1371/journal.pone.0093975
- Holland, B.R., Benthin, S., Lockhart, P.J., Moulton, V., Huber, K.T., 2008. Using supernetworks to distinguish hybridization from lineage-sorting. *BMC Evolutionary Biology* 8, 202. doi:10.1186/1471-2148-8-202
- Hudson, C.M., Puckett, E.E., Bekaert, M., Pires, J.C., Conant, G.C., 2011. Selection for higher gene copy number after different types of plant gene duplications. *Genome Biology and Evolution* 3, 1369–1380. doi:10.1093/gbe/evr115
- Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23, 254–267.
- Huson, D.H., Rupp, R., Scornavacca, C., 2010. *Phylogenetic networks: Concepts, algorithms and applications*. Cambridge University Press, Cambridge.
- Huson, D.H., Scornavacca, C., 2011. A survey of combinatorial methods for phylogenetic networks. *Genome Biology and Evolution* 3, 23–35. doi:10.1093/gbe/evq077
- Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100.
- Joly, S., 2012. JML: Testing hybridization from species trees. *Molecular Ecology Resources* 12, 179–184. doi:10.1111/j.1755-0998.2011.03065.x
- Joly, S., McLenachan, P.A., Lockhart, P.J., 2009. A statistical approach for

- distinguishing hybridization and incomplete lineage sorting. *The American Naturalist* 174, E54–E70. doi:10.1086/600082
- Joshi, N.A., Fass, J.N., 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. Available from: github.com/najoshi/sickle.
- Katoh, K., Asimenos, G., Toh, H., 2009. Multiple alignment of DNA sequences with MAFFT, in: Posada, D. (Ed.), *Bioinformatics for DNA Sequence Analysis*. Humana Press, pp. 39–64. doi:10.1007/978-1-59745-251-9_3
- Katoh, K., Misawa, K., Kuma, K.I., Miyata, T., 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30, 3059–3066.
- Kent, W.J., 2002. BLAT—the BLAST-like alignment tool. *Genome Research* 12, 656–664.
- Kiefer, C., Koch, M.A., 2012. A continental-wide perspective: The genepool of nuclear encoded ribosomal DNA and single-copy gene sequences in North American *Boechera* (Brassicaceae). *PLoS ONE* 7, e36491. doi:10.1371/journal.pone.0036491
- Knowles, L.L., 2009. Estimating species trees: Methods of phylogenetic analysis when there is incongruence across genes. *Systematic Biology* 58, 463–467. doi:10.1093/sysbio/syp061
- Kubatko, L.S., 2009. Identifying hybridization events in the presence of coalescence via model selection. *Systematic Biology* 58, 478–488.
- Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology* 56, 17–24. doi:10.1080/10635150601146041

- Lien, S., Koop, B.F., Sandve, S.R., Miller, J.R., Kent, M.P., Nome, T., Hvidsten, T.R., Leong, J.S., Minkley, D.R., Zimin, A., Grammes, F., Grove, H., Gjuvsland, A., Walenz, B., Hermansen, R.A., Schalburg, von, K., Rondeau, E.B., Di Genova, A., Samy, J.K.A., Olav Vik, J., Vigeland, M.D., Caler, L., Grimholt, U., Jentoft, S., Inge Våge, D., de Jong, P., Moen, T., Baranski, M., Palti, Y., Smith, D.R., Yorke, J.A., Nederbragt, A.J., Tooming-Klunderud, A., Jakobsen, K.S., Jiang, X., Fan, D., Hu, Y., Liberles, D.A., Vidal, R., Iturra, P., Jones, S.J.M., Jonassen, I., Maass, A., Omholt, S.W., Davidson, W.S., 2016. The Atlantic salmon genome provides insights into rediploidization. *Nature* 533, 200–205. doi:10.1038/nature17164
- Linder, C.R., Rieseberg, L.H., 2004. Reconstructing patterns of reticulate evolution in plants. *American Journal of Botany* 91, 1700–1708.
- Lysak, M.A., Koch, M.A., Beaulieu, J.M., Meister, A., Leitch, I.J., 2008. The dynamic ups and downs of genome size evolution in Brassicaceae. *Molecular Biology and Evolution* 26, 85–98. doi:10.1093/molbev/msn223
- Maddison, W., Knowles, L., 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* 55, 21–30. doi:10.1080/10635150500354928
- Maddison, W.P., 1997. Gene trees in species trees. *Systematic Biology* 46, 523–536.
- Mallet, J., 2007. Hybrid speciation. *Nature* 446, 279–283. doi:10.1038/nature05706
- Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J., Turner, D.J., 2010. Target-enrichment strategies for next-generation sequencing. *Nature Methods* 7, 111–118. doi:10.1038/nmeth.1419
- Mandel, J.R., Dikow, R.B., Funk, V.A., Masalia, R.R., Staton, S.E., Kozik, A., Michelmore, R.W., Rieseberg, L.H., Burke, J.M., 2014. A target enrichment method

- for gathering phylogenetic information from hundreds of loci: An example from the Compositae. *Applications in Plant Sciences* 2, 1300085.
doi:10.3732/apps.1300085.s1
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* 17, 10–12.
- McCartney-Melstad, E., Mount, G.G., Shaffer, H.B., 2015. Exon capture optimization in large-genome amphibians. *bioRxiv* 021253. doi:10.1101/021253
- McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T., 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* 66, 526–538.
- Meerow, A.W., 2010. Convergence or reticulation? Mosaic evolution in the canalized American Amaryllidaceae, in: Seberg, O., Petersen, G., Barfod, A.S., Davis, J.I. (Eds.), *Diversity, Phylogeny and Evolution in the Monocotyledons*. Aarhus University Press, Aarhus, pp. 145–168.
- Meerow, A.W., Francisco-Ortega, J., Kuhn, D.N., Schnell, R.J., 2006. Phylogenetic relationships and biogeography within the Eurasian clade of Amaryllidaceae based on plastid *ndhF* and nrDNA ITS sequences: Lineage sorting in a reticulate area? *Systematic Botany* 31, 42–60.
- Meerow, A.W., Guy, C.L., Li, Q.B., Yang, S.L., 2000. Phylogeny of the American Amaryllidaceae based on nrDNA ITS sequences. *Systematic Botany* 25, 708–726.
- Meerow, A.W., Preuss, K.D., Tombolato, F.C., 2002. *Griffinia* (Amaryllidaceae), a critically endangered Brazilian geophyte with horticultural potential, in: Littlejohn, G. (Ed.). *Acta Horticulturae (ISHS)*, 57–64.

- Meerow, A.W., Snijman, D.A., 1998. Amaryllidaceae, in: Kubitzki, K. (Ed.), Families and Genera of Vascular Plants. Springer-Verlag, Berlin, pp. 83–110.
- Meng, C., Kubatko, L.S., 2009. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: A model. *Theoretical Population Biology* 75, 35–45. doi:10.1016/j.tpb.2008.10.004
- Mirarab, S., Bayzid, M.S., Boussau, B., Warnow, T., 2014. Statistical binning enables an accurate coalescent-based estimation of the avian tree. *Science* 346, 1250463–1250463. doi:10.1126/science.1250463
- Mirarab, S., Warnow, T., 2015. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31, i44–i52. doi:10.1093/bioinformatics/btv234
- Moriarty Lemmon, E., Lemmon, A.R., 2013. High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 44, 99–121. doi:10.1146/annurev-ecolsys-110512-135822
- Morrison, D.A., 2010. Using data-display networks for exploratory data analysis in phylogenetic studies. *Molecular Biology and Evolution* 27, 1044–1057. doi:10.1093/molbev/msp309
- Nakhleh, L., 2013. Computational approaches to species phylogeny inference and gene tree reconciliation. *Trends in Ecology & Evolution* 28, 719–728. doi:10.1016/j.tree.2013.09.004
- Nakhleh, L., 2011. Evolutionary phylogenetic networks: Models and issues, in: Heath, L.S., Ramakrishnan, N. (Eds.), *Problem Solving Handbook in Computational Biology and Bioinformatics*. Springer Science+Business Media, pp. 125–158.

doi:10.1007/978-0-387-09760-2_7

- Naranjo, C.A., 1974. Karyotypes of 4 Argentine species of *Habranthus* and *Zephyranthes* (Amaryllidaceae). *Phyton-International Journal of Experimental Botany* 32, 61–71.
- Naranjo, C.A., Andrada, A.B., 1975. El cariotipo fundamental en el género *Hippeastrum* Herb. (Amaryllidaceae). *Darwiniana* 19, 566–582.
- Naranjo, C.A., Poggio, L., 2000. Karyotypes of five *Rhodophiala* species (Amaryllidaceae). *Boletín de la Sociedad Argentina de Botánica* 35, 335–343.
- Naumann, J., Salomo, K., Der, J.P., Wafula, E.K., Bolin, J.F., Maass, E., Frenzke, L., Samain, M.-S., Neinhuis, C., dePamphilis, C.W., Wanke, S., 2013. Single-copy nuclear genes place haustorial Hydnoraceae within Piperales and reveal a Cretaceous origin of multiple parasitic angiosperm lineages. *PLoS ONE* 8, e79204. doi:10.1371/journal.pone.0079204.s007
- Neves, L.G., Davis, J.M., Barbazuk, W.B., Kirst, M., 2013. Whole-exome targeted sequencing of the uncharacterized pine genome. *The Plant Journal* 75, 146–156. doi:10.1111/tpj.12193
- Oliveira, R.S. de, 2012. O gênero *Hippeastrum* Herb. (Amaryllidaceae) no Brasil: Evidência de evolução reticulada e análise de caracteres florais. Universidade Estadual de Campinas, Campinas.
- Oliver, J.C., 2013. Microevolutionary processes generate phylogenomic discordance at ancient divergences. *Evolution* 67, 1823–1830. doi:10.1111/evo.12047
- Paijmans, J.L., Fickel, J., Courtiol, A., Hofreiter, M., Förster, D.W., 2016. Impact of enrichment conditions on cross-species capture of fresh and degraded DNA.

- Molecular Ecology Resources 16, 42–55.
- Poggio, L., Realini, M.F., Fourastie, M.F., Garcia, A.M., Gonzalez, G.E., 2014. Genome downsizing and karyotype constancy in diploid and polyploid congeners: a model of genome size variation. *AoB PLANTS* 6, plu029–plu029. doi:10.1093/aobpla/plu029
- Ravenna, P., 2003. Elucidation and systematics of the Chilean genera of Amaryllidaceae. *Botanica Australis* 2, 1–21.
- Ravenna, P., 1981. Contributions to South American Amaryllidaceae: 7. *Plant Life* 37, 57–83.
- Rieseberg, L.H., Soltis, D.E., 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* 5, 65–84.
- Ripma, L.A., Simpson, M.G., Hasenstab-Lehman, K., 2014. Geneious! Simplified genome skimming methods for phylogenetic systematic studies: a case study in *Oreocarya* (Boraginaceae). *Applications in Plant Sciences* 2, 1400062. doi:10.3732/apps.1400062
- 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. *Systematic Biology* 61, 539–542.
- Salas-Leiva, D.E., Meerow, A.W., Francisco-Ortega, J., Calonje, M., Griffith, M.P., Stevenson, D.W., Nakamura, K., 2014. Conserved genetic regions across angiosperms as tools to develop single-copy nuclear markers in gymnosperms: An example using cycads. *Molecular Ecology Resources* 14, 831–845. doi:10.1111/1755-0998.12228

- Seehausen, O., 2004. Hybridization and adaptive radiation. *Trends in Ecology & Evolution* 19, 198–207. doi:10.1016/j.tree.2004.01.003
- Slater, G. S. C., Birney, E., 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6, 31. doi: 10.1186/1471-2105-6-31
- Small, R.L., Cronn, R.C., Wendel, J.F., 2004. LAS Johnson Review No. 2. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* 17, 145–170.
- Sneath, P.H.A., 1975. Cladistic representation of reticulate evolution. *Systematic Zoology* 24, 360–368.
- Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., dePamphilis, C.W., Wall, P.K., Soltis, P.S., 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96, 336–348. doi:10.3732/ajb.0800079
- Soltis, D.E., Kuzoff, R.K., 1995. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution* 49, 727–742.
- Soltis, P.S., Soltis, D.E., 2009. The role of hybridization in plant speciation. *Annual Review of Plant Biology* 60, 561–588. doi:10.1146/annurev.arplant.043008.092039
- 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
- Stamatakis, A., 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. doi:10.1093/bioinformatics/btl446

- Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* 57, 758–771.
doi:10.1080/10635150802429642
- Stephens, J.D., Rogers, W.L., Heyduk, K., Cruse-Sanders, J.M., Determann, R.O., Glenn, T.C., Malmberg, R.L., 2015. Resolving phylogenetic relationships of the recently radiated carnivorous plant genus *Sarracenia* using target enrichment. *Molecular Phylogenetics and Evolution* 85, 76–87. doi:10.1016/j.ympev.2015.01.015
- Storz, J. F., Ramakrishnan, U., Alberts, S. C., 2001. Determinants of effective population size for loci with different modes of inheritance. *Journal of Heredity* 92, 497–502. doi: 10.1093/jhered/92.6.497
- 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. *American Journal of Botany* 99, 349–364. doi:10.3732/ajb.1100335
- Sun, M., Soltis, D.E., Soltis, P.S., Zhu, X., Burleigh, J.G., 2015. Deep phylogenetic incongruence in the angiosperm clade Rosidae. *Molecular Phylogenetics and Evolution* 83, 156–166. doi:10.1016/j.ympev.2014.11.003
- Takahashi, T., Moreno, E., 2015. A RAD-based phylogenetics for *Orestias* fishes from Lake Titicaca. *Molecular Phylogenetics and Evolution* 93, 307–317.
doi:10.1016/j.ympev.2015.08.012
- Tanabe, A.S., 2011. Kakusan4 and Aminosan: Two programs for comparing nonpartitioned, proportional and separate models for combined molecular phylogenetic analyses of multilocus sequence data. *Molecular Ecology Resources* 11, 914–921. doi:10.1111/j.1755-0998.2011.03021.x

- Traub, H.P., 1963. The genera of the Amaryllidaceae. American Plant Life Society, La Jolla.
- Turner, E.H., Ng, S.B., Nickerson, D.A., Shendure, J., 2009. Methods for genomic partitioning. *Annual Review of Genomics and Human Genetics* 10, 263–284.
doi:10.1146/annurev-genom-082908-150112
- 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. *Applications in Plant Sciences* 2, 1400042.
doi:10.3732/apps.1400042.s1
- Whitfield, J.B., Lockhart, P.J., 2007. Deciphering ancient rapid radiations. *Trends in Ecology & Evolution* 22, 258–265. doi:10.1016/j.tree.2007.01.012
- Winfield, M.O., Wilkinson, P.A., Allen, A.M., Barker, G.L., Coghill, J.A., Burridge, A., Hall, A., Brenchley, R.C., D'Amore, R., Hall, N., 2012. Targeted re-sequencing of the allohexaploid wheat exome. *Plant Biotechnology Journal* 10, 733–742.
- Wyman, S.K., Jansen, R.K., Boore, J.L., 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20, 3252–3255.
- Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., 2014. SOAPdenovo-Trans: *De novo* transcriptome assembly with short RNA-Seq reads. *Bioinformatics* 30, 1660–1666.
- Yang, Z., 2006. Computational molecular evolution. Oxford University Press, Oxford.
- Zerbino, D.R., 2010. Using the Velvet *de novo* assembler for short-read sequencing technologies. *Current Protocols in Bioinformatics* 11.5. 1–11.5. 12.
- Zerbino, D.R., Birney, E., 2008. Velvet: Algorithms for *de novo* short read assembly

using de Bruijn graphs. *Genome Research* 18, 821–829.

Zhong, B., Liu, L., Penny, D., 2014. The multispecies coalescent model and land plant origins: a reply to Springer and Gatesy. *Trends in Plant Science* 19, 270–272.

Table 1. Samples used in capture enrichment and MiSeq Illumina sequencing experiments.

Table 2. Alignment and analysis information for curated nuclear loci.

Table 3. Alignment information for plastid and concatenated data sets.

Table 4. Summary of results of plastid gene tree simulations based on the nuclear species tree. "Expected clade frequencies" refer to the observed frequency of a relationship among simulated gene trees. In cases where we posited hybridization as an explanation of the data, plastid relationships were unexpected based on the simulated plastid gene tree distribution, whereas contradicting nuclear relationships were very frequent among simulated trees (indicating a low level of expected ILS). Observed clade support columns summarize maximum likelihood bootstrap (BS) values from the 6-partition plastid tree and posterior probabilities (PP) from the *BEAST analysis including introns.

Figure 1. Maximum clade credibility tree inferred in *BEAST using 18 curated nuclear loci and only coding sequences. Posterior probabilities (PP) $\geq 50\%$ are shown above branches for the total (including introns) and only coding analyses, respectively. “*” = 100% PP; “-” = clade absent. Inset shows tree based on only coding sequences, with branch lengths, excluding the Eurasian clade; scale not shown.

Figure 2. Species trees inferred in ASTRAL-II using (A) 12 nuclear loci (including introns) from curated assembly and (B) 39 nuclear loci from automated assembly. Values above branches correspond to bootstrap (BS)/local posterior probabilities (LPP); only branches with either BS $> 70\%$ and/or LPP $> 85\%$ are labelled. “*” = 100% BS or LPP; “-” = clade absent from bootstrap tree. Scale in coalescent units.

Figure 3. Maximum likelihood tree of Hippeastreae inferred using RAxML and whole plastid genomes (130,380 bp). Bootstrap (BS) values = 100%, except where shown above branches for both the single-partition and six-partition analyses, respectively. Inset shows six-partition tree with branch lengths, excluding the Eurasian clade; scale not shown.

Figure 4. Phylogenetic network analyses of Hippeastreae. A. Confidence (95%) network of 15 curated nuclear loci plus 15 plastid genes inferred from 1,000 NeighborNet bootstrap pseudoreplicates in SplitsTree. B. Hybridization network

of 0.95 PP *BEAST tree (18 curated nuclear loci; only coding sequences) and 80% bootstrap plastid tree (130,380 bp; 6 partitions) computed in SplitsTree. Color shading indicates chromosome numbers.

Figure 5. Comparison of the observed plastid topology (red) with a distribution of organellar gene trees (blue, drawn with DensiTree) simulated with a branch length scaling factor of 4 from ASTRAL nuclear species tree inferred from 12 curated nuclear loci. While gene tree heterogeneity is high among basal lineages, for others (e.g., *Hippeastrum*) predicted ILS is quite low, and observed relationships are essentially absent from the gene tree distribution.

Figure 6. Reticulogram summarizing a hypothesized reconciliation of the observed plastid phylogeny on the species tree estimated from nuclear data. Instances of incongruence hypothesized to result from ILS are marked as such on species tree branches; for hypothesized instances of plastid capture, the plastid genome source and hypothesized directions of transfer are indicated by dashed lines and arrows. Question marks indicate unobserved (extinct or unsampled) lineages required to reconcile the position of a taxon.

Figure 1 *BEAST

Figure 2 ASTRAL

A

B

Figure 3 plastid

$2n = 12, 14$
$2n = 16$
$2n = 18$
$2n = 20$
$2n = 22$

B.

Figure 5 coal simulation vs plastid

Figure 6 reticulogram

Table 1.

Species binomial	Voucher (herbarium)	SRA code
Outgroup taxa:		
<i>Lycoris radiata</i>	Meerow 2606 (FTG)	SRR3183975
<i>Narcissus papyraceus</i>	N. García 4397 (FLAS)	SRR3185382
<i>Griffinia hyacinthina</i>	Meerow 2106 (FLAS, FTG)	SRR3183890
<i>Griffinia nocturna</i>	J. Dutilh s.n. (UEC)	SRR3185235
<i>Griffinia rochae</i>	J. Dutilh s.n. (UEC)	SRR3189192
Ingroup taxa:		
<u>Traubiiinae</u>		
<i>Famatina maulensis</i>	N. García 4384 (CONC, FLAS)	SRR3185225
<i>Phycella australis</i>	M. Baeza 4258a (CONC)	SRR3189429
<i>Phycella cyrtanthoides</i>	N. García 4404 (FLAS)	SRR3184070
<i>Phycella scarlatina</i>	N. García 862 (CONC)	SRR3189430
<i>Placea arzae</i>	N. García 3025 (CONC)	SRR3184080
<i>Placea davidii</i>	N. García 3031 (CONC, FLAS)	SRR3189431
<i>Placea ornata</i>	O. Fernández 68 (JBN-Chile)	SRR3185390
<i>Rhodolirium laetum</i>	N. García 1022 (CONC, FLAS)	SRR3185392
<i>Rhodolirium montanum</i>	N. García 4379 (FLAS)	SRR3184083
<i>Rhodolirium speciosum</i>	I. Lizama s.n. (CONC)	SRR3185393
<i>Traubia modesta</i>	O. Fernández 151 (JBN-Chile)	SRR3184145
<u>Hippeastrinae</u>		
<i>Eithea blumenavia</i>	J. Dutilh s.n. (UEC)	SRR3183496
<i>Eithea</i> "Piracicaba"	J. Dutilh s.n. (UEC)	SRR3189190
<i>Famatina cisandina</i>	N. García 2811 (SGO)	SRR3183838
<i>Famatina herbertiana</i>	Leuenberger 3738 (B)	SRR3185213
<i>Habranthus brachyandrus</i>	Meerow 2400 (FTG)	SRR3183891
<i>Habranthus gracilifolius</i>	Leuenberger 4804g (B)	SRR3189437
<i>Habranthus jamesonii</i>	Leuenberger 3960 (B)	SRR3189443
<i>Habranthus martinezii</i>	Meerow 2437 (FTG)	SRR3189195
<i>Habranthus pedunculatus</i>	Leuenberger 4496 (B)	SRR3183893
<i>Habranthus robustus</i>	N. García 4387 (FLAS)	SRR3183894
<i>Hippeastrum brasilianum</i>	Meerow 2405 (FTG)	SRR3183895
<i>Hippeastrum</i> "CardosoMoreira"	J. Dutilh s.n. (UEC)	SRR3189293
<i>Hippeastrum cipoanum</i>	J. Dutilh s.n. (UEC)	SRR3189312
<i>Hippeastrum evansiae</i>	Meerow 3118 (NA)	SRR3189316
<i>Hippeastrum papilio</i>	Meerow 2406 (FTG)	SRR3183904
		SRR3189365
<i>Hippeastrum parodii</i>	Meerow 2434 (FTG)	SRR3185282
<i>Hippeastrum reticulatum</i>	Meerow s.n. (NA)	SRR3185380
<i>Rhodophiala advena</i>	N. García 2964 (FLAS, CONC)	SRR3184124
<i>Rhodophiala ananuca</i>	M. Rosas 7042 (CONC)	SRR3189436
<i>Rhodophiala araucana</i>	N. García 4345 (FLAS, CONC)	SRR3185395
<i>Rhodophiala bifida</i>	Meerow 3102 (FTG)	SRR3184143
<i>Tocantinia dutilhiana</i>	N. García 4400 (FLAS)	SRR3189454
<i>Zephyranthes flavissima</i>	N. García 4388 (FLAS)	SRR3185396
<i>Zephyranthes mesochloa</i>	Leuenberger 4494 (B)	SRR3185397

Table 2

Table 2.

Reference gene (ATID)	Name	Set	Alignment length (bp)	Intron n°/bp	Variable/ parsimony- informative sites (%)	Missing data (%) ¹	Model (JModelTest ² /*BEAST)	Taxonomic coverage	Analyses				
									*BEAST	ASTRAL	BUCKy	Concat	JML
AT1G80070.1	SUS2	total	2543	5/451	37.8/23.1	26.8		Complete, duplicated					
		copy A	2528	5/436	24.7/13.9	23.5	TIM3 + G/GTR + G	<i>Griffinia</i> , Hippeastreae	x	x	x	x	x
		copy B	2555	5/463	26.2/15.3	23.1	TPM1uf + G/GTR + G	Hippeastreae	x	x	x	x	x
AT2G05170.1	VPS11	total	2866	2/194	34.6/18.2	30.3		Complete, duplicated					
		copy A	2866	2/194	26.9/12.9	27.9	TrN + G/id	<i>Griffinia</i> , Hippeastreae	x	x	x	x	x
		copy B	2866	2/194	21.1/9.1	33.7	TrN + G/id	<i>Griffinia</i> , Traubiinae, <i>Hippeastrum</i> , <i>Tocantinia</i>	x				
AT2G40760.1	Rhod7	total	375	0	65.3/40.0	23		Complete, duplicated					
		copy A	375	0	60.0/33.1	22.4	HKY + G/id	<i>Griffinia</i> , Hippeastreae	x		x	x	
AT3G46740.1	TOC75-3	total	820	0	40.4/23.4	25		Complete, duplicated					
		copy A	820	0	33.3/18.4	22.5	TIM1ef + G/GTRef + G	<i>Griffinia</i> , Hippeastreae	x	x	x	x	x
		copy B	820	0	27.8/16.0	27	TIM1ef + G/GTRef + G	<i>Griffinia</i> , Hippeastreae	x	x	x	x	x
AT4G21710.1	RPB2	total	1061	4/364	33.1/15.8	44.4	TrN + G/id	Complete, single copy	x			x	x
AT4G33030.1	SQD1	total	1020	0	39.1/23.4	16.7		Complete, duplicated					
		copy A	1020	0	29.9/17.8	15.5	K80 + G/HKYef + G	<i>Griffinia</i> , Hippeastreae	x	x	x	x	x
		copy B	1020	0	24.6/12.8	17.7	TPM2uf + G/GTR + G	Hippeastrinae	x	x	x	x	x
AT4G34100.1	CER9	total	1538	2/189	58.5/36.1	42.5		Complete, ~3 copies					
		copy A	1538	2/189	40.2/17.4	41.9	TIM2 + G/GTR + G	<i>Griffinia</i> , Hippeastreae	x		x	x	x

Table 3.

Dataset	Alignment length (bp)	Variable/parsimony- informative sites (%)	Missing data (%) ¹
cpDNA/plastome (trimmed)	130,380	5.8/2.4	2.3
Coding	68,441	4.9/2.0	1.3
Non-coding	61,939	6.8/2.8	3.4
15 LCNGs with introns	25,077	24.9/12.6	28.4
15 LCNGs only coding	22,853	23.2/11.9	26.6
15 LCNGs + 15 cpDNA, coding	55,292	12.7/5.8	12.3
15 LCNGs + SSC + LSC, coding	79,834	10.4/4.6	8.9

¹Includes gaps.

Table 4

Table 4.

Observed plastid relationship	Observed plastid clade support (BS)	Expected plastid clade frequency given the species tree		Observed nuclear relationship	Observed nuclear clade support (PP)
		Curated assembly; 1X/2X/4X scalar	aTRAM assembly; 1X/2X/4X scalar		
Polyphyletic <i>Hippeastrum</i>	95,97,99,100,100,79	0/0/0	0/0/0	Monophyletic <i>Hippeastrum</i> ‡	1
<i>H. brasilianum</i> sister to remaining <i>Hippeastrineae</i>	95	0/0/0	0/0/0	<i>H. brasilianum</i> sister to <i>H. "CardosoMoreira"</i> ‡	1
<i>Eithea</i> sister to <i>Hippeastrum</i> "CardosoMoreira"	79	0/0/0	0/0/0	<i>Eithea</i> sister to core <i>Rhodophiala</i> and	0.59
<i>Hippeastrum reticulatum</i> sister to (<i>Eithea</i> + <i>H. "CardosoMoreira"</i>)	100	0/0/0	0/0/0	<i>H. reticulatum</i> within <i>Hippeastrum</i> *	1
<i>Zephyranthes</i> sister to <i>Habranthus gracilifolius</i> group	100	0/0/0	0/0/0	<i>Zephyranthes</i> sister to <i>Habranthus robustus</i> group	0.97

* The position of *H. reticulatum* was not supported in *BEAST, but concatenated analyses placed it as sister to core *Hippeastrum* (BS 100).

‡ The position of *H. brasilianum* was not supported in ASTRAL for aTRAM assemblies (but was decisive for curated assemblies); this taxon was not considere

** The position of *Eithea* was supported in ASTRAL (BS 87) and concatenated (BS 84-96) analyses of curated assemblies.