Received Date : 15-Jan-2017

Revised Date : 18-Feb-2017

Accepted Date : 19-Feb-2017

Article type : MS - Regular Manuscript

Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*)

Oscar M. Vargas[1,2], Edgardo M. Ortiz[2] and Beryl B. Simpson[2]

[1]Integrative Biology and Plant Resources Center, The University of Texas at Austin, Austin, TX 78712, USA; [2]Department of Ecology and Evolutionary Biology, University of Michigan, 830 N. University Ave, Ann Arbor, MI 48109, USA

Author for correspondence:

Oscar M. Vargas

Tel: +1 512 300 9236

Email: oscarvargash@gmail.com

Received: *15 January 2017*

Accepted: *19 February 2017*

Summary

- High-throughput sequencing is helping biologists to overcome the difficulties of inferring the phylogenies of recently diverged taxa. The present study analyzes the phylogenetic signal of genomic regions with different inheritance patterns using genome skimming and ddRAD-seq in a species-rich Andean genus (*Diplostephium*) and its allies.
- We analyzed the complete nuclear ribosomal cistron, the complete chloroplast genome, a partial mitochondrial genome, and a nuclear-ddRAD matrix separately with phylogenetic methods. We applied several approaches to understand the causes of incongruence among datasets, including simulations and the detection of introgression using the D-statistic (ABBA-BABA test).
- We found significant incongruence among the nuclear, chloroplast, and mitochondrial phylogenies. The strong signal of hybridization found by simulations and the D-statistic among genera and inside the main clades of *Diplostephium* indicate reticulate evolution as a main cause for phylogenetic incongruence.
- Our results add evidence for a major role of reticulate evolution in events of rapid diversification. Hybridization and introgression confound chloroplast and mitochondrial phylogenies in relation to the species tree due to the uniparental inheritance of these genomic regions. Practical implications regarding the prevalence of hybridization are discussed in relation to the phylogenetic method.

Key words: ddRAD, *Diplostephium*, genome skimming, hybridization, introgression, phylogenomics, rapid diversification, reticulate evolution.

Introduction

Rapid diversifications usually occur in landscapes like archipelagos and mountain ranges that provide disjunct terranes suitable for speciation via isolation and ecological divergence (Givnish, 1997). The Andes Cordillera is one of these landscapes in which numerous plant groups have experienced high rates of speciation during and after its recent uplift (Madriñán *et al.*, 2013; Luebert & Weigend, 2014; Hughes & Atchinson, 2015). Phylogenies of Andean taxa based on Sanger sequencing often suffer from lack of resolution and support, especially in crown clades where accelerated speciation occurred (e.g. Rauscher, 2002; Emshwiller, 2002; Sánchez-Baracaldo, 2004; Bell & Donoghue, 2005; Hughes & Eastwood, 2006; Zapata, 2013; Nürk *et al.*, 2013). Poorly resolved relationships within Andean taxa confirm the prediction that phylogenies resulting from rapid radiations are difficult to estimate because of low and conflicting signal caused by short internodes (Whitfield & Lockhart, 2007). High-throughput sequencing, a technology capable of producing orders of magnitude more data than obtained by Sanger sequencing, has created new opportunities for overcoming the difficulties of working with recently diversified taxa (Bock *et al.*, 2014; Ma *et al.*, 2014; Mort *et al.*, 2015). Here, we implement genome skimming (Straub *et al.*, 2012) and double-digest Restriction site Associated DNA sequencing (ddRAD, Peterson *et al.*, 2012) to infer the phylogenetic patterns of one of the most species-rich genera of Andean plants and its relatives.

Diplostephium is a main component of the tropical high Andean flora. The genus traditionally comprises 111 species (Vargas, 2011) characterized by a woody habit (from 10 cm decumbent subshrubs to 10 m tall trees) and radiate capitula with white to purple rays (Blake, 1928; Cuatrecasas, 1969; Vargas & Madriñán, 2006). *Diplostephium* inhabits the high elevations of the Talamanca Cordillera (Costa Rica), the Northern Andes (Venezuela-Colombia-Ecuador), and the Central Andes (Peru-Bolivia). Most species of the genus (*c*. 60) inhabit the páramo, a Northern Andean ecosystem known for its high plant diversity (Luteyn, 1999), island-like geography (Simpson, 1974), and a large number of species-rich genera (Madriñán *et al.*, 2013; Luebert & Weigend, 2014). In addition to the páramo, some *Diplostephium* species inhabit the Central Andean puna and the upper limit of the high Andean forest. *Diplostephiun* belongs to Astereae, where it has traditionally been classified as part of the *Chiliotrichum* group, a subset of the subtribe Hinterhuberinae (Nesom, 1994; Nesom & Robinson, 2007). Molecular phylogenies have shown that Astereae subtribes and their subdivisions are polyphyletic (Noyes & Rieseberg,

1999; Sancho & Karaman-Castro, 2008; Brouillet *et al.*, 2009; Karaman-Castro & Urbatsch, 2009; Sancho *et al.*, 2010; Vargas & Madriñán, 2012) and need to be recircumscribed. *Diplostephium* is positioned in the 'South American lineages' grade (Brouillet *et al.*, 2009) but its ambiguous position in the Astereae phylogeny (Noyes & Rieseberg, 1999; Brouillet *et al*., 2009; Karaman-Castro & Urbatsch, 2009; Vargas & Madriñán, 2012) has obscured its place among its closest genera. The low molecular variation found by Vargas & Madriñán (2012) in *Diplostephium* highlighted the problems of inferring the phylogeny of a rapidly diversified taxon with limited sampling using Sanger sequencing.

Genome skimming or shallow shotgun sequencing, is an approach in which whole genomic DNA is sequenced in order to recover high-copy DNA regions suitable for phylogenetic analyses (Straub *et al.*, 2012). The biparentally inherited (Volkov *et al.*, 2007) complete nuclear ribosomal DNA has proved to be a useful marker for inferring species-level phylogenies (Linder *et al.*, 2000; Straub *et al.*, 2012; Bock *et al.*, 2014). The mostly non-recombinant and uniparentally inherited chloroplast DNA (Birky, 1995; Jansen & Ruhlman, 2012) has been used to reconstruct the phylogeny of Asteraceae at the tribal, generic, and species levels (Kim *et al.*, 2005, Panero & Funk, 2008, Bock *et al.*, 2014; Panero *et al.*, 2014). Finally, the mostly non-recombinant uniparentally inherited mitochondrial DNA (Birky, 1995) has been employed to infer the evolutionary history of angiosperms at the family and order level (Qiu *et al.*, 2010; Sun *et al.*, 2015) and more recently to study phylogenetic patterns among species (Bock *et al.*, 2014). Restriction site Associated DNA sequencing (RAD-Seq, e.g. ddRAD, GBS, etc), on the other hand, is a technique that surveys hundreds or thousands of loci mainly from the nuclear genome (Peterson *et al.,* 2012).

The comparison among markers with different inheritance patterns has the potential to elucidate hybridization and introgression events (Rieseberg & Soltis, 1991; Hardig *et al.*, 2000; Bock *et al.*, 2014; Sun *et al.*, 2015) hypothesized to play an important role in cases of evolutionary bursts of speciation (Anderson & Stebbins, 1954; Seehausen, 2004). Because rapid divergence happens during a short window of time, it is expected that the descendants of a diversification burst are prone to interbreed before reproductive barriers develop. This expectation has been hypothesized as beneficial because hybridization among nascent lineages can result in genotypes potentially adapted to unexploited niches (Anderson & Stebbins, 1954; Seehausen, 2004). *Diplostephium* is

thus an excellent model to test such ideas, as the genus appeared to have undergone a recent diversification associated with the uplift of the Andes (Vargas & Madriñán, 2012).

In this study, we focused on uncovering the phylogenetic patterns of *Diplostephium* and its allied genera. The aims of this paper are to: compare the phylogenetic signal among nuclear-ddRAD, nuclear ribosomal, chloroplast, and mitochondrial DNA; and test for infra- and intergeneric introgressive hybridization in *Diplostephium* and its Andean relatives.

Materials and Methods

Assembly of the genome skimming datasets

A total of 91 samples were sequenced. The ingroup contained 74 samples of *Diplostephium* (69 species, *c*. 62% of the total number of species) and 14 samples from 13 allied genera in Astereae (Supporting Information Table S1). We chose ingroup and outgroup genera based on their phylogenetic positions in Astereae inferred by Brouillet *et al.* (2009). The majority of the samples (63) were collected in the field where leaf tissue was dried using silica gel. The remaining samples were taken from herbarium specimens deposited in ANDES, F, FMB, HUA, HUSA, TEX, US, and USM (Table S1).

We performed total genomic DNA extractions with the DNeasy Plant Mini Kit (Qiagen, CA, USA) following the manufacturer's protocol. To increase the yield of DNA from herbarium material, we added 50 µl of proteinase K (Qiagen; activity > 600 mAU ml^{-1}) to the lysis solution and incubated it overnight at 45ºC after the initial 10-min incubation. Standard genomic DNA Illumina paired-end libraries with an average of *c*. 400 bp length fragment size were prepared at the Genomic Sequencing and Analysis Facility (GSAF) at The University of Texas at Austin (UT) and then sequenced by an Illumina HiSeq 2500. The sequencing targeted 10 million paired-end reads (2×100 bp length) per sample. We inspected the quality of the reads with the program FASTQC v.0.10.1 (Andrews, 2010) and filtered the raw data with the program 'process_shortreads' of the software package STACKS v.1.20 (Catchen *et al.*, 2011). The command discarded any reads with uncalled bases and trimmed terminal nucleotides that averaged a Phred score of ≤10 (90% of confidence in base calls) over a sliding window of 10 bp.

We employed a two-step strategy to create three subsets of reads per sample: nuclear ribosomal, chloroplast, and mitochondrial. By including only the necessary subsets of genomic data to perform the assemblies, we aimed to reduce noise and increase the accuracy and computational efficiency in our downstream analyses. Our first step consisted of performing a *de novo* assembly of each sample using all the genomic data. The software RAY v.2.3.1 (Boisvert *et al.*, 2012) performed the assembly using three different k-mer values visually selected from the report graph provided by KMERGENIE (Chikhi & Medvedev, 2014). The contigs resulting from the RAY assembly were parsed with the '-search' function of RAY using *Helianthus annus* L. nuclear ribosomal, chloroplast, and mitochondrial reference sequences (GenBank accessions HM638217.1, NC_007977.1, and KF815390.1, respectively); this step provided a subset of RAY contigs. The second step consisted in separating the genomic reads into nuclear ribosomal, chloroplast, and mitochondrial subsets by mapping all reads against the RAY contig produced after the first step. For example, whole genome reads of *Diplostephium haenkei* were mapped against their own RAY chloroplast contigs to obtain a chloroplast subset of reads. BOWTIE v.2.2.3 (Langmead & Salzberg, 2012) was used to perform the mapping. Then, we analyzed with different pipelines the three subsets of reads obtained per sample to obtain the corresponding sequence alignments (Fig. S1).

Due to the intraindividual polymorphic nature of the nuclear ribosomal DNA, we assembled these sequences in two steps. First, we created a *de novo* draft assembly with the nuclear ribosomal reads subset in each sample using SPAdes v.3.5.0 (Bankevich *et al.*, 2012) with the '-careful' option, and 21, 33, 55, 77, and 89 as k-mer sizes. The nuclear ribosomal contigs obtained by SPAdes were merged in GENEIOUS v.7.1.4 (Kearse *et al.*, 2012) allowing overlapping regions of two contigs to assemble with some mismatches by using the 'medium sensitivity' setting. Taking into account that the nuclear ribosomal DNA is arranged in tandem repeats, the complete nuclear ribosomal contig produced by GENEIOUS was treated as circular aiming to increase the accuracy of the back-mapping. The first nucleotide position of the circular contig was set at the first base right (5' to 3' direction) of the TATAGGGGG promoter found at the end of the non-transcribed spacer (NTS) (Linder *et al*., 2000). Second, we back-mapped the nuclear ribosomal reads to their circular draft assembly in GENEIOUS using a minimum overlap identity of 95%. From the back-mapping obtained for each sample, we calculated three consensus sequences per sample using different threshold percentages (50%, 75%, and 90%, Fig.

S1). By doing this, we were able to account for the different levels of intraindividual polymorphisms found in the nuclear ribosomal tandem repeats and evaluate their effects on phylogenetic estimation. The 50% consensus only called a nucleotide in a polymorphic position if the nucleotide were present in > 50% of the reads containing that position. Therefore, the 50% consensus sequence set had fewer ambiguities than the 75% and 90% consensus sequences. All the 273 nuclear ribosomal sequences (three per sample) were aligned with MAFFT v.7.017 (Katoh *et al.*, 2002) into a master alignment. We corrected the master alignment by hand using GENEIOUS. Finally, we extracted from the master alignment the three matrices corresponding to the three percentage thresholds (nr50, nr75, and nr90). Each matrix was analyzed separately.

A *de novo* assembly for each sample's chloroplast reads subset was performed with SPAdes with the same options used for the nuclear ribosomal assembly. First, we annotated the chloroplast genome of *Diplostephium haenkei* using DOGMA (Wyman *et al.*, 2004) with subsequent manual correction in GENEIOUS using *Guizotia abyssinica* (L.f.) Cass. as a reference (GenBank accession EU549769.1). Then, we used the chloroplast of *D. haenkei* to merge and annotate the chloroplast-SPAdes contigs of the remaining 90 samples using GENEIOUS (Fig. S1). We codified the gaps produced by non-overlapping contigs as missing data. We employed MAFFT to create the chloroplast genome alignment that was later corrected by hand in GENEIOUS. For the phylogenetic analysis, we only included one inverted repeat in our alignment and manually removed three unalignable regions of *c*. 2,300 sites total.

Due to the high degree of rearrangements and the discontinuity found in the mitochondrial assemblies performed with SPAdes, it was not feasible to make a direct alignment of our mitochondrial genomes. Instead, we *de novo* assembled the mitochondrial genome of *Diplostephium hartwegii* and then used this genome as a reference for assembly by mapping the remaining 90 samples. We chose *D. hartwegii* based on the continuity and the high coverage of its mitochondrial contig obtained by RAY. We expected all the mitochondrial assemblies with the exception of *D. hartwegii* to have missing data at the boundaries of DNA blocks where rearrangements occurred relative to *D. hartwegii*. With this strategy, we intended to obtain a gapped mitochondrial genome assembly for each sample in which the order of the mitochondrial blocks matched that of *D. hartwegii* making their alignment feasible. We employed SPAdes (with the same parameters used for the nuclear ribosomal assemblies) and MITOFYv.1.3.1 (Alverson *et al.*, 2010) to assemble and annotate, respectively, the mitochondrial genome of *D.*

hartwegii. We corrected by hand our MITOFY annotation using GENEIOUS based on that of *Cucumis sativus* L. (Alverson *et al*., 2011). Before mapping, we filtered the mitochondrial reads of each sample because we detected a small proportion of chloroplast reads mixed in. The intraindividual presence of chloroplast reads in the mitochondrial subset is explained by the similarity of some tRNAs and the transference of DNA between these two genomes (Alverson *et al.*, 2011). To filter out the chloroplast reads from the mitochondrial reads subsets, we mapped every sample's mitochondrial reads subset against its *de novo* chloroplast genome assembled in this study (e.g. the mitochondrial reads of *D. colombianum* were mapped to the chloroplast genome of *D. colombianum*) using a 97% minimum overlap identity in GENEIOUS. The reads not mapped to the chloroplast genome were mapped to the mitochondrial reference (*D. hartwegii*) to create a consensus sequence per sample (Fig. S1) using the 'medium sensitivity' setting in GENEIOUS. MAFFT was used to perform the alignment of the 91 mitochondrial sequences. We visually inspected the matrix in GENEIOUS, and regions difficult to align, or with significant amounts of missing data, were excluded from the matrix. We also removed the mitochondrial rRNA genes *rrn5*, *rrnL*, and *rrnS* from the alignment because we found bacterial DNA matching some hyperconserved regions of these genes. We suspect that the source of bacterial DNA in our dataset came from bacteria living on the leaf surfaces of our plant samples.

Assembly of the nuclear-ddRAD dataset

The RAD-Seq protocol known as ddRAD (double-digest RAD-Seq, Peterson *et al.*, 2012) was used to prepare a reference library that included a subset of 44 samples (Table S1) to which we could map all the unused nuclear reads that resulted from genome skimming. The restriction enzymes EcoRI and SphI were used to digest total genomic DNA, with a size selection range for the resulting fragments (excluding adapters) of 354–414 bp. Library sequencing, carried out using an Illumina HiSeq 4000 at the UT-GSAF, produced *c.* 83 million of 2×150 bp pair-end reads. The reference library was demultiplexed with the software deML v.1 (Renaud *et al.*, 2014) that allowed us to confidently assign > 99% of the reads to their corresponding samples while tolerating up to 2 mismatches in the barcodes. The sequence quality assessed using FASTQC v.0.11.5 (Andrews, 2010) revealed low quality over the restriction overhang regions and a large portion of the R2 reads. Therefore, we excluded R2 reads from subsequent analyses and trimmed

the restriction overhang of the R1 files, as well as 40 nucleotides from the end of the reads to obtain sequences of 100 bp using the script 'reformat.sh' from the toolkit BBTools v.36.38 (Bushnell, 2016). To guarantee that the assembly was done using only nuclear reads, we used the script 'bbsplit.sh' from BBTools to create sets of reads that did not match either the nuclear ribosomal, chloroplast, or mitochondrial sequences of our ingroup, or the Coliphage phiX174 viral genome (GenBak NC_001422.1) used in Illumina runs to increase nucleotide diversity. The demultiplexed, trimmed, and filtered reads were used as input for the software ipyrad v.0.5.1 (Eaton & Overcast, 2016) in which the assembly was performed with the parameters specified in Methods S1. This reference ddRAD library contained 32,424 loci. A custom script 'loci_sampler.py' (https://github.com/edgardomortiz/radscripts.git) was used to select the most common sequence of each aligned locus, producing a collection of unique sequences each representing a different locus, named hereafter ddRAD reference.

In order to expand the taxon sampling to the full set of species and to increase the coverage of samples per locus we designed a pipeline called 'shotgun2rad' (https://github.com/edgardomortiz/shotgun2rad.git) to map genome skimming reads to our ddRAD reference (or any other type of RAD data) and produce multiple loci alignments across species. The pipeline performs quality filtering and adapter removal with CUTADAPT v.1.12 (Martin, 2011), merges the quality-filtered pairs that overlap using VSEARCH v.2.0.3 (Rognes *et al.*, 2016), maps the merged and unmerged reads to the ddRAD reference using BWA v.0.7.12 (Li & Durbin, 2009), and parses the resulting BWA alignments to match the 'clustS.gz' format produced by step 3 (within-sample clustering) of pyRAD v.3.0.66 (Eaton, 2014). From that point, shotgun2rad runs pyRAD steps 4–7 to produce loci alignments across species and a variety of matrix formats for phylogenetic analyses. The parameters used for shotgun2rad and pyRAD are indicated in Methods S2 and S3. Two datasets were derived from our ddRAD pipeline, a loci dataset (nuclear-ddRAD-loci) and a matrix containing only variable sites (nuclear-ddRAD). All computational processes related to the assembly of matrices were carried out at the Texas Advance Computing Center at UT (http://www.tacc.utexas.edu).

Phylogenetic analyses

We performed independent phylogenetic analyses for each genome skimming dataset using Bayesian Inference (BI) and Maximum Likelihood (ML). We evaluated a non-partitioned (M0) vs. a partitioned model (M1). The M1 of the chloroplast and mitochondrial matrices contained a coding and a non-coding partition. The M1 of the nuclear ribosomal DNA dataset contained three partitions: rRNA, transcribed spacers (the external transcribed spacer (ETS), the ITS), and the non-transcribed spacer (NTS). We compared M0 and M1 for each dataset by calculating the stepping stone marginal likelihood (Xie *et al.*, 2011) of both model schemes with MrBayes v.3.2.2 (Ronquist *et al.*, 2012) on the CIPRES Science Gateway Server (Miller *et al.*, 2010) using 10 million generations, 2 runs, 4 chains per run, and assuming a GTR+ Γ model. The stepping stone marginal likelihood values were compared using Bayes factors (Fan *et al.*, 2011). To calculate the model of evolution of the partitions we employed a mixed strategy. First, we inferred the use + Γ and +I parameters in our partitions using the corrected Akaike information criterion (AICc) (Hurvich & Tsai, 1989) employed in jModelTest v.2.1.7 (Guindon & Gascuel, 2003; Darriba *et al.*, 2012). Then, we calculated the substitution parameters among nucleotides with reversible jump Markov Chain Monte Carlo (rjMCMC) simulations (Huelsenbeck *et al.*, 2004) using MrBayes with 10 million generations, 2 runs, 4 chains per run, and the + Γ and/or +I parameters if suggested by the AICc. A final MrBayes analysis with 10 million generations, 2 runs, and 4 chains per run was performed using the best partition model along with the substitution parameters inferred for each dataset. We performed the ML analyses with RAxML v.8.1.11 (Stamatakis, 2014) in the CIPRES portal using the partitioned schemes favored, 100 rapid bootstrap replicates, and the GTR+ Γ model of evolution (as recommended by RAxML manual Stamatakis, 2015). For all Bayesian analyses we used a burn-in fraction of 0.25. We employed TRACER v.1.6.0 (Rambaut *et al.*, 2014) to confirm the convergence of parallel MCMC runs.

We analyzed the nuclear-ddRAD matrix with maximum likelihood using the RAxML-HPC2 Workflow available in CIPRES (Miller *et al.*, 2010). Two separate analyses were performed, one consisting of 100 independent runs and one consisting of 100 thorough bootstraps, both run with the Ascertainment Bias option enabled. The 100 independent runs were summarized into a Majority Consensus Rule Tree with 'sumtrees.py' v.4.10 (Sukumaran & Holder, 2010a)

available in the package Dendropy v.4.10 (Sukumaran & Holder, 2010b). Support was obtained from the 100 RAxML thorough bootstraps using 'sumtrees.py.' The same matrix was analyzed with SVDQuartets (Chifman & Kubatko, 2014) implemented in PAUP* v.4.0a150 (Swofford, 2002) using all quartets and performing 100 bootstraps. We excluded outgroups from both analyses because of their high levels of missing data. FigTree v.1.4.2 (Rambaut, 2014) allowed us to inspect, compare, and export trees to image editors.

Congruence assessment and visualization

We visually inspected the congruence of the topologies obtained by different matrices with the multidimensional scaling of tree space using the Robinson–Foulds distance implemented in TreeSetViz v3.0 (Amenta & Klingner, 2002) as a package for MESQUITE v.3.04 (Maddison & Maddison, 2015). The first comparison was made among the results obtained by the BI and ML analysis of the three nuclear ribosomal datasets (nr50, nr75, and nr90). The second comparison contrasted the trees from the analyses of the nuclear ribosomal 90%, chloroplast, mitochondrial, and nuclear-ddRAD datasets. From the BI analyses, we sampled the maximum clade credibility tree (MCCT from sumtrees.py) and 50 random Bayesian topologies from each dataset after a 0.25 burn-in. From the ML and SVDq analyses, we sampled the best tree obtained and 50 random bootstrap replicates from each subset. A hierarchical likelihood-ratio congruence test among the nuclear ribosomal 90%, chloroplast, and mitochondrial datasets was performed with CONCATERPILLAR v.1.8a (Leigh *et al.*, 2008) coupled with RAxML v.7.2.8. (Stamatakis, 2006). We identified rogue taxa by using the Procrustean Approach to Cophylogeny (PACo, Balbuena *et al.*, 2013) with the pipeline designed by Pérez-Escobar *et al.* (2016), specifically aimed at identifying incongruent taxa between nuclear and chloroplast phylogenies. The approach of Pérez-Escobar *et al.* (2016) assumes that the chloroplast phylogeny is dependent on the nuclear phylogeny and transforms the topologies into matrices of patristic distances, that are then transformed into Euclidean Principal Coordinate matrices using the method proposed by de Vienne *et al.* (2011). Taxa with significant incongruent phylogenetic positions are identified by comparing the observed distances against those produced by permutations. Because this approach performs best with phylograms (Pérez-Escobar *et al.*, 2016), we used the BI trees

obtained from the nuclear ribosomal 90 % and chloroplast matrices that have low levels of missing data and therefore a more accurate branch estimation.

Hybridization assessment

To distinguish between incomplete lineage sorting (ILS) and hybridization, we employed the method described by Joly *et al.* (2009) implemented in the software JML v.1.3.0 (Joly, 2012). This method uses the posterior distribution of coalescent species trees, estimated population sizes, and branch length from a *BEAST v.2.3.0 (Heled & Drummond, 2010) output to simulate a coalescent scenario with no migration. Hybridization is identified by comparing the minimum pairwise distance between the simulated and the empirical datasets. If an observed distance is significantly smaller than the simulated one, then ILS can be rejected, suggesting hybridization (Joly *et al.*, 2009; Joly, 2012). Joly *et al.*'s (2009) approach assumes that the marker used to calculate genetic distances does not recombine and its power increases with longer sequences (Joly, 2012), this makes our chloroplast matrix an excellent candidate to simulate and compare genetic distances. We calculated a coalescent tree in *BEAST using the nuclear ribosomal and chloroplast matrices. We ran five independent analyses in *BEAST using 200 million generations, sampling every 1000, and with a GTR+ Γ model. For the JML input, we randomly sampled 2000 trees from the five *BEAST runs using LogCombiner v2.3 (Drummond *et al.*, 2012) after a burn-in of 0.25. With JML, we carried out 2000 simulations of chloroplast genome pairwise distance comparisons. To visualize the results, we created a matrix map in R (R Core Team, 2016) indicating significant pairwise comparisons after the Benjamini-Hochberg correction.

To detect introgression, we calculated the Patterson's D-statistic (ABBA-BABA test, Huson *et al.*, 2005; Green *et al.*, 2010; Durand *et al.*, 2011) on selected quartets of taxa with the nuclear-ddRAD-loci dataset employing the software pyRAD. We chose the quartets based on the JML results and the comparison between nuclear-ddRAD and chloroplast topologies. The D-statistic calculates the proportion ‘ABBA’ and ‘BABA’ patterns in a four-taxon phylogeny. This proportion should be equally frequent in a scenario with ILS and no gene flow; if the proportion between the two patterns is significantly different, then introgression between two of the taxa is

inferred. Significant patterns were identified using a Z-score > 3, which correspond to a conservative α=0.01 (Eaton & Ree, 2013) after the Benjamini–Hochberg correction.

Time calibration of the phylogeny

We could not implement a primary calibration for our dataset because of the poor fossil record for Astereae. Therefore, in order to obtain a calibration point for our ingroup, we extracted the ITS region (*c*. 700 bp) of our nuclear ribosomal alignment and combined it with that of Strijk *et al.* (2012), which contains sequences from Heliantheae, Gnaphalieae, and Anthemideae providing external nodes for the time calibration. We aligned the ITS matrix with MAFFT and employed MrBayes to obtain a topology using 30 million generations, 2 runs, 4 chains, a temperature of 0.001, and GTR+ Γ +I model. We calculated the chronogram with BEAST v.2.3.0 (Drummond *et al.*, 2012) using the MrBayes topology (fixed), a lognormal relaxed clock, a Yule model of speciation, 100 million generations, a sampling frequency of 40 thousand generations, and the same calibration piors used by Strijk *et al.* (2012). We produced the control file using BEAUti v.2.3 (Drummond *et al.*, 2012). The sets of trees from the two runs were combined after a burnin of 0.25 using LogCombiner v.2.3 (Drummond *et al.*, 2012) before calculating the chronogram with TreeAnnotator 2.3 (Drummond *et al.*, 2012).

Because the ITS matrix (*c*. 700 bp) did not provide enough phylogenetic information to produce a robust and resolved topology in the part of the tree corresponding to our ingroup, we calibrated our Bayesian ddRAD tree (fixed) using the complete nuclear ribosomal 90% matrix (*c*. 13 Kb) employing secondary calibration points derived from our ITS chronogram and the literature. We aligned to our matrix the nuclear ribosomal cistron of *Helianthus annuus* (GenBank accession KF767534) using MAFFT and employed the following calibration points on the tree: (1) a mean of 32.4 million yr (Myr) based on Kim *et al.* (2005) with a normal distribution and a sigma of 11 (95% CI:14.4–50.6) to the root of the tree representing the crown clade of Asteroideae, and (2) a mean of 11.21 Myr with a normal distribution and a sigma of 3.1 (95% CI:6.11–16.3) to the most recent common ancestor of the 'South American Lineages' (inferred in our ITS chronogram). To calculate the chronogram, we employed BEAST with the same parameters used in the calibration of the ITS dataset and the model and partition scheme calculated for the phylogenetic analysis of the nuclear ribosomal matrix.

Consensus sequences were deposited in GenBank (Table S1). Subsets of reads, aligned matrices, and control files are available at Dryad (http://dx.doi.org/10.5061/dryad.cn74h).

Results

Characteristics of the datasets

The nuclear ribosomal alignments comprise 13,362 characters of which 1,203–1,425 (9.00–10.66%) are parsimony-informative characters (PICs) depending on the consensus threshold (the number of PICs reported here always exclude the outgroup, Table S2). While the nr50 matrix contains only 25 ambiguities, the nr75 and the nr90 have 3,201 and 5,012 ambiguities, respectively. The chloroplast genome of *Diplostephium haenkei* contains 85 genes plus 36 tRNAs and 8 rRNAs in 152,292 bp (Fig. S2). No significant rearrangements or gene losses were found across the sampled species relative to *Guizotia abyssinica* with the exception of the loss of the *rps*19 gene for *Baccharis genistelloides* and *trnT-GGU* for *B. genistelloides*, *B. tricuneata*, and *Llerasia caucana.* Fifty-six of the 91 genomes have missing data in their assemblies due to the low coverage of reads obtained in regions with a high number of repeats. The chloroplast alignment is 135,440 characters long, of which 2,169 (1.69%) are PICs (Table S2). The *de novo* mitochondrial genome of *Diplostephium hartwegii* has a total length of 277,718 bp, containing 56 genes plus 3 chloroplast-like tRNAs (Fig. S3). The final mitochondrial matrix has a length of 209,392 and contains 1,730 (0.83%) PICs. The nuclear-ddRAD-loci dataset used for introgression analyses has a total of 19,987 loci. The matrix containing only variable SNPs, nuclear-ddRAD, employed for the phylogenetic analyses, has a total of 244,255 total sites with 96,421 (39.5%) PICs (Table S2).

Phylogenetic analysis

In all cases, the Bayes factors favor a partitioned matrix analysis over a non-partitioned one (Tables S3, S4). Overall, all the phylogenetic topologies obtained are well resolved with highly supported nodes (Figs 1, S4–S11). The two topologies obtained by BI or SVDq and ML for each of the matrices analyzed (e.g. the BI-chloroplast tree compared to the ML-chloroplast tree) are

consistent excluding clades or taxa with low support (bootstrap support (BS) <80%, Bayesian posterior probability (BPP) <0.9).

Nuclear ribosomal topologies

The BI and ML backbones of the nr50, nr75, and nr90 trees are all highly congruent with the exception being the position of *Diplostephium meyenii*, which has low support on all of the trees (Figs 1b, S5–S9). The visualization of the Robinson-Foulds tree distances of the six nuclear ribosomal tree subsets (BI-nr50, BI-nr75, BI-nr90, ML-nr50, ML-nr75, and ML-nr90) shows four clusters of topologies (Fig. S12). None of the four clouds are formed exclusively by the trees of one subset. This visualization emphasizes that despite the general agreement among the nuclear ribosomal topologies, there are some incongruences located towards the tip of the trees (Figs 1b, S5–S9). For example, *Diplostephium schultzii*, a species represented by two samples, is polyphyletic in the BI-nr50 and ML-nr50 topologies (Figs S5, S7), whereas it is monophyletic in the BI-nr75, BI-nr90, ML-nr75 and ML-nr90 topologies (Figs 1b, S6, S8, S9). Because the nr50, nr75, and nr90 backbones are congruent and the nr90 matrix is the most conservative in a phylogenetic context capturing a considerable amount of intraindividual polymorphisms, (informative to MrBayes and RAxML, Ronquist *et al.*, 2011; Stamatakis, 2015) we selected the nr90 dataset to make comparisons with the chloroplast and mitochondrial datasets.

For the remaining part of the paper we will therefore refer to the nuclear ribosomal 90% dataset simply as ‘nuclear ribosomal.’

Incongruence among the genomic datasets

To avoid confusion in our results and discussion, we only use the BI MCCT topologies from the genome skimming matrices and the ML topology from the nuclear-ddRAD matrix when comparing all datasets topologies since alternative topologies from the same matrices are almost identical.

Incongruence is significant among the topologies obtained from the nuclear-ddRAD, nuclear ribosomal, chloroplast, and mitochondrial genomic regions with the four trees recovering different number of clades of *Diplostephium* taxa and contrasting generic relationships (Fig. 1). The topology in Fig. 1a shows two labeled groups that we refer to as *Diplostephium* A and B.

While in the nuclear-ddRAD and nuclear ribosomal topologies species of group A comprise a monophyletic group that is sister to a clade containing non-*Diplostephium* genera and species of *Diplostephium* group B, in the chloroplast and mitochondrial topologies species of group A (excepting *D. inesianum* in the mitochondrial topology) are nested within a clade containing *Diplostephium* group B species and the rest of Astereae genera sampled. Topological incongruence among our datasets is also obvious in the visualization of the Robinson-Foulds tree distances (Fig. S13) that depicts four independent clouds of topologies corresponding to the nuclear-ddRAD, nuclear ribosomal, chloroplast, and mitochondrial tree subsets. The chloroplast and mitochondrial clouds are positioned more closely together relative to the nuclear-ddRAD and nuclear ribosomal trees. The hierarchical likelihood-ratio rejects the concatenation of the most congruent genome skimming dataset combination, chloroplast plus mitochondrial DNA (P<0.0001). Normalized squared residuals values obtained by PACo identify 35 (38%) of the 91 tips as rogue taxa (Fig. 2).

Hybridization assessment

A strong signal of hybridization in our data is revealed by JML (Fig. 2); 1,647 (40%) out of 4,095 pairwise chloroplast distances are significantly smaller than expected in a scenario with ILS and no migration ($P < 0.05$) after the Benjamini–Hochberg correction. The Patterson's D-statistic (ABBA-BABA) performed in selected quartets of taxa shows a signal of introgression among genera and inside *Diplostephium* groups A and B for many of the quartets evaluated (Tables 1, S5).

Chronograms

The ITS chronogram shows poor support and resolution in nodes positioned in the ‘South American Lineages’ (Fig. S14), nevertheless, it provides a time estimate for the origin of the ‘South American lineages’ (node R, mean = 11.2 Myr 95% CI = 6.2–16.4) employed to calibrate the nuclear-ddRAD topology with the complete nuclear ribosomal matrix. The time calibration of our nuclear topology using the nuclear ribosomal matrix (Fig. 3) provides a rough estimation for the age of *Diplostephium* groups A (mean = 6.5 Myr 95% CI = 1.3–11.4) and B (mean = 7.3

Myr 95% CI = 2.3–12.5), and the divergence of South American genera (mean = 10.17 Myr 95% CI = 3.2–16.2).

Discussion

Phylogenetic incongruence and hybridization

Incongruence among the nuclear-ddRAD, nuclear ribosomal, chloroplast, and mitochondrial datasets is found at the generic and species level, especially between the nuclear vs. organellar trees. The only clade recovered consistently across phylogenies (excluding *D. inesianum* for the mitochondrial topology), is a clade that contains a group of *Diplostephium* species restricted mostly to the Northern Andes, labeled group A in the nuclear-ddRAD topology (Fig. 1). However, despite this Northern Andean clade being recovered in the four genomic phylogenies, its position relative to other genera and the rest of species of *Diplostephium* is incongruent when nuclear vs. organellar topologies are compared. Incongruence among different markers can be produced by phylogenetic uncertainty, incomplete lineage sorting, and/or hybridization (Rieseberg & Soltis, 1991; Maddison, 1997; Huelsenbeck *et al.*, 2000). To rule out phylogenetic uncertainty, we performed a PACo analysis to identify rogue taxa. Because PACo performs simulations incorporating phylogenetic uncertainty, the presence of significant outliers provides evidence for the phylogenies being affected by ILS or hybridization. The result that 38% of our samples are rogue indicates extensive ILS or hybridization in our dataset.

To distinguish between ILS and hybridization, we employed the method proposed by Joly *et al.* (2009) incorporated in the software JML (Joly, 2012). The significantly small pairwise distances of *c*. 40% of all possible comparisons among chloroplast genomes suggest high levels of hybridization. Although JML results can be difficult to interpret because they only show species pairs affected by hybridization and do not incorporate a model to detect ancient hybridization, our visualization of the JML results mapped next to a phylogeny (Fig. 2) provides a new approach to formulate hybridization hypotheses. This visualization shows a strong signal of ancient hybridization because significant comparisons are present among different genera and among almost all comparisons between species of *Diplostephium* groups A and B.

We employed a nuclear-ddRAD-loci dataset to calculate empirically the amount of introgression among selected taxa using the Patterson's D-statistic. Our results (Table 1) show that introgression has occurred among genera and inside *Diplostephium* groups A and B. Even though we did not test for every possible hybridization event in our D-statistic framework, the significant tests reported (Table 1) suggest at least 12 reticulate events. In the specific case of *Diplostephium* sp. nov. CAJ2, *D. meyenii*, and *Parastrephia quadrangularis* the D-statistic identifies a strong signal of hybridization also supported by morphological data. We hypothesize that *Diplostephium* sp. nov. CAJ2 is the product of a cross between *D. meyenii* and *P. quadrangularis* because *Diplostephium* sp. nov. CAJ2 has scale-like leaves similar to *P. quadrangularis* but heterogamous capitula with ray flowers like *D. meyenii* and group B species (Fig. 4a–c). Similarly, a high Z-score suggests that *D. cinereum* (Fig. 4d) has hybridized with *D. meyenii* or *P. quadrangularis* (Table 1), all of which occur in the Peruvian Andes in parapatry, and morphological hybrids have been observed in the field (V. Quipuscoa pers. comm.). The incongruence among genomic phylogenies, the JML results, and the signal of introgression inferred by the D-statistic, suggest a complex pattern of reticulate evolution in our ingroup.

The high number of instances of hybridization seen in our dataset explains to a great extent the incongruence obtained among our genomic phylogenies. Numerous authors have proposed that horizontal gene transfer via introgression could cause deviation of the chloroplast and mitochondrial phylogenies from the species tree (Rieseberg & Soltis, 1991; Moore, 1995; Hardig *et al*., 2000; Sun *et al.*, 2015). Because organellar DNA inheritance is mostly uniparental and there is generally no recombination following fertilization (Birky, 1995; Jansen & Ruhlman, 2012), an event of hybridization could completely replace the original chloroplast and mitochondrial DNA of a lineage by an alien one, confounding their phylogenies relative to the species tree (Rieseberg & Soltis, 1991). Chloroplast and mitochondrial genomes have the potential to be fixed rapidly because their effective population size (Ne) is one-fourth that of a nuclear autosomal gene (Moore, 1995), making organellar genomes less prone to ILS. Additionally, the discrepancies among the phylogenies of chloroplast and mitochondrial DNA suggest a decoupling of inheritance of these two genomes. The evidence presented here makes us conclude that the trees obtained from the chloroplast and mitochondrial DNA unequivocally deviate from the species tree.

The nuclear-ddRAD tree is to our knowledge the best species tree hypothesis taking into account the numerous loci sampled from the nuclear genome and its phylogenetic robustness (Fig. 1a). This topology fits better the morphological classification and biogeography of our ingroup than all the other phylogenies obtained in this study. The nuclear-ddRAD topology indicates that *Diplostephium* (*sensu* Cuatrecasas 1969) is biphyletic (if *Parastrephia* is included in group B species) with both clades diverging approximately in the late Miocene. Because the nomenclatural type of the genus, *D. ericoides*, is a member of group B species (*Diplostephium* s.s.), group A must be circumscribed as a different taxon (O. M. Vargas, unpublished).

From the genome skimming dataset, the nuclear ribosomal DNA is better at capturing signal to infer the species tree than the chloroplast and mitochondrial DNA because the nuclear ribosomal cistron is biparentally inherited (Volkov *et al.*, 2007), it recombines (Hughes & Petersen, 2001; Ambrose & Crease, 2011), and its copies are located on multiple chromosomes (Phillips *et al.*, 1971; Álvares & Wendel, 2003). The nuclear ribosomal tree is also largely congruent with the nuclear-ddRAD tree (Fig. 1a,b).

Extensive introgression in South American Astereae is probably due to lack of reproductive barriers. Even though the pollination biology of our ingroup is still understudied, *Diplostephium* and its allies lack obvious pollinator specialization (Cuatrecasas, 1969) probably because of the low diversity of pollinators in the high Andes (Berry & Calvo, 1989). Additionally, the complex topography of the Andes might promote speciation via geographic isolation while allowing gene flow in early stages of differentiation before reproductive barriers develop. Available reports of chromosome numbers (Chromosome Counts Database, http://ccdb.tau.ac.il, Fig. 2) suggest that the ancient hybridization identified in our data did not result in polyploidy in the majority of cases because the only sample with a chromosome count report deviating from the base chromosome number in Astereae (n=9, Nesom & Robinson, 2007) is *Diplostephium jaramilloi* (n=18). However, more data about karyotypes are needed to confirm the above statements, especially for *Diplostephium* group A species.

Practical considerations

Our results demonstrate that hybridization is prevalent in *Diplostephium* and its allies deviating the chloroplast and mitochondrial phylogenies from the species tree. In conjunction with

documented cases of reticulate evolution in other organisms like heliconiine butterflies (Mallet *et al.*, 2007) and cichlid fishes (Genner & Turner, 2012), our results support the hypothesis that hybridization might be a common process in events of rapid evolution (Anderson & Stebbins, 1954; Seehausen, 2004). Because hybridization might be frequent during evolutionary bursts of speciation, ancient and recent diversifications (and radiations) are especially prone to the phylogenetic biases explained here. We urge systematists to avoid concatenating data with different inheritance patterns (e.g. chloroplast and nuclear markers) without testing for congruence, and to use caution when using chloroplast and mitochondrial phylogenies as the basis for taxonomic classification, time calibrations, historical biogeographic reconstructions, and comparative phylogenetic analyses.

Phylogenetic methods have focused on a bifurcating model of evolution but dichotomous trees cannot illustrate processes like hybridization and horizontal gene transfer (Doolittle, 1999; Huson & Bryant, 2006; Rieppel, 2010). Recent advances in phylogenetic explicit network estimation (Solís-Lemus & Ané, 2016; Wen *et al.*, 2016) promise better models to understand the evolutionary history of taxa that have undergone hybridization and lateral gene transfer, yet, these approaches are still in their infancy and are only efficient in cases of recent hybridization with few reticulate events.

Approaches that identify rogue terminals (e.g. PACo and PARAfit, Legendre *et al.*, 2002; Balbuena *et al.*, 2013; Pérez-Escobar *et al.*, 2016) are often used to remove taxa from phylogenetic analysis and decrease incongruence among datasets (e.g. Philippe *et al.*, 2009; Regier *et al.*, 2013; Salichos & Rokas, 2013; Reginato & Michelangeli, 2016). While PACo identified 38% of our taxa as rogue, our hybridization analysis suggests that *c*. 94% of our sampled taxa are affected by ancient or recent hybridization. Therefore, removing taxa identified as rogue from our analysis would not eliminate topological incongruence produced by hybridization. We encourage biologists to uncover the reason for incongruence (instead of removing rogue taxa) as the answer to this question could elucidate noteworthy biological processes.

Our study demonstrates that ancient hybridization is worthy of attention and should be modeled in phylogenetic inference. Future research should focus on understanding the role of hybridization on diversification and developing methods to quantify ancient introgression and

accurately infer reticulate networks. The evidence presented in our study builds on numerous reports (e.g. Sessa *et al.*, 2012; Sun *et al.*, 2015; Sochor *et al.*, 2015; Wu *et al.*, 2015) that emphasize a central role of reticulation in plant evolution.

Acknowledgements

We thank Fernando Alzate, Felipe Cardona, Michael Dillon, Alvaro Idarraga, Santiago Madriñán, Haydeé Montoya, Victor Quipuscoa, Adriana Prieto, Rusty Russell, Katya Romoleroux, Luis Sánchez, and Tom Wendt for providing access to their herbaria. Hamilton Beltrán provided the pictures of *Diplostephium cinereum* and *D.* sp. nov. CAJ 2. Jordan Bemmels, Joseph Brown, David Cannatella, Amalia Diaz, Robert Jansen, Thomas Juenger, Teofil Nakov, Juan Palacio, Sarah Sussman, Ning Wang, and Mao-Lun Weng provided help and comments during the development of this study. We also want to thank the three anonymous reviewers for their constructive comments. This project was funded by The University of Texas at Austin (Plant Biology Program Award, the C. L. Lundell Chair of Systematic Botany, and The Linda Escobar Award), The Garden Club of America (2012 Award in Tropical Botany) and the National Science Foundation (post-doctoral support FESD 1338694 and DEB 1240869). Research permits: resolution 0145 of 'El Ministerio de Ambiente y Desarrollo Sostenible' and permit 37 of 'Sistema Ambiental Sina' (Colombia); authorization 10-2012-IC-FLO-DPAP-MA of 'Ministerio del Medio Ambiente' (Ecuador); and resolution 0369-2011-AG-DGFFS-DGEFFS of 'Dirección de Gestión Forestal y de Fauna Silvestre' (Peru).

Author contributions

O.M.V. and E.M.O. performed the research, collected the samples, and analyzed the data. B.B.S. and O.M.V. designed the research and interpreted the results. The writing was primarily carried out by O.M.V.

References

Álvarez I, Wendel JF. **2003**. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* **29**: 417–434.

Alverson AJ, Rice DW, Dickinson S, Barry K, Palmer JD. **2011**. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. *The Plant Cell* **23**: 2499–2513.

Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD. **2010**. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Molecular Biology and Evolution* **27**: 1436–1448.

Ambrose CD, Crease TJ. **2011**. Evolution of the nuclear ribosomal DNA intergenic spacer in four species of the *Daphnia pulex* complex. *BMC Genetics* **12**: 13.

Amenta N, Klingner J. 2002. Case study: visualizing sets of evolutionary trees. *8th IEEE Symposium on Information Visualization* **2002:** 71–74.

Anderson E, Stebbins GL. **1954**. Hybridization as an evolutionary stimulus. *Evolution* **8**: 378–388.

Andrews S. 2010. *FastQC: A quality control tool for high throughput sequence data*. [WWW document] URL http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. [accessed 1 January 2015].

Balbuena JA, Míguez-Lozano R, Blasco-Costa I. **2013**. PACo: a novel procrustes application to cophylogenetic analysis. *PLoS ONE* **8**: e61048.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD *et al.* 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**: 455–477.

Bell CD, Donoghue MJ. **2005**. Phylogeny and biogeography of Valerianaceae (Dipsacales) with special reference to the South American valerians. *Organisms Diversity and Evolution* **5**: 147–159.

Berry PE, Calvo RN. **1989**. Wind pollination, self-Incompatibility, and altitudinal shifts in pollination systems in the high Andean genus *Espeletia* (Asteraceae). *American Journal of Botany* **76**: 1602–1614.

Birky CW. **1995**. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences, USA* **92**: 11331–11338.

Blake SF. **1928**. Review of the genus *Diplostephium*. *American Journal of Botany* **15**: 43–64.

Bock DG, Kane NC, Ebert DP, Rieseberg LH. **2014**. Genome skimming reveals the origin of the Jerusalem artichoke tuber crop species: neither from Jerusalem nor an artichoke. *New Phytologist* **201:** 1021–1030.

Boisvert S, Raymond F, Godzaridis E, Laviolette F, Corbeil J. **2012**. Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biology* **13**: R122.

Brouillet L, Lowrey TK, Urbatsch LE, Karaman-Castro V, Sancho G, Wagstaff SJ, Semple JC. 2009. Astereae. In: Funk VA, Stuessy T, Bayer R, eds. *Systematics, evolution and biogeography of Compositae*. Vienna, Austria: International Association of Plant Taxonomists, 589–629.

Bushnell B. **2016**. BBTools: A suite of bioinformatic tools used for dna and rna sequence data analysis. [WWW document] URL http://jgi.doe.gov/data-and-tools/bbtools/. [accessed 1 July 2016].

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. **2011**. Stacks: building and genotyping loci de novo from short-read sequences. *G3 (Bethesda, Md.)* **1**: 171–82.

Chifman J, Kubatko L. **2014**. Quartet inference from SNP data under the coalescent model. *Bioinformatics* **30**: 3317–3324.

Chikhi R, Medvedev P. **2014**. Informed and automated k-mer size selection for genome assembly. *Bioinformatics* **30**: 31–7.

Cuatrecasas J. 1969. Prima Flora Colombiana. 3. Compositae-Astereae. *Webbia* **24:** 1–335.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**: 772–772.

de Vienne DM, Aguileta G, Ollier S. **2011**. Euclidean nature of phylogenetic distance matrices. *Systematic Biology* **60**: 826–832.

Doolittle WF. **1999**. Phylogenetic classification and the universal tree. *Science* **284**: 2124–2129.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29:** 1969–1973.

Durand EY, Patterson N, Reich D, Slatkin M. **2011**. Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution* **28**: 2239–2252.

Eaton DAR. **2014**. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* **30**: 1844–1849.

Eaton DAR, Overcast I. **2016**. ipyrad: Interactive Assembly and Analysis of RADseq data sets. [WWW document] URL http://ipyrad.readthedocs.io/. [accessed 1 August 2016].

Eaton DAR, Ree RH. **2013**. Inferring phylogeny and introgression using RADseq data : An example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology* **62**: 689–706.

Emshwiller E. **2002**. Biogeography of the *Oxalis tuberosa* alliance. *Botanical Review* **68**: 128–152.

Fan Y, Wu R, Chen MH, Kuo L, Lewis PO. **2011**. Choosing among partition models in Bayesian phylogenetics. *Molecular Biology and Evolution* **28**: 523–532.

Genner MJ, Turner GF. **2012**. Ancient hybridization and phenotypic novelty within lake Malawi's cichlid fish radiation. *Molecular Biology and Evolution* **29**: 195–206.

Givnish TJ. 1997. Adaptive radiation and molecular systematics: aims and conceptual issues. In: Givnish TJ, Systma KJ, eds. *Molecular evolution and adaptive radiation*. New York, USA: Cambridge University Press, 1–54.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y *et al.* 2010. A draft sequence of the Neandertal genome. *Science* **328**: 710–22.

Guindon S, Gascuel O. **2003**. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**: 696–704.

Hardig TM, Soltis PS, Soltis DE. **2000**. Diversification of the North American shrub genus *Ceanothus* (Rhamnaceae): conflicting phylogenies from nuclear ribosomal DNA and chloroplast DNA. *American Journal of Botany* **87**: 108–123.

Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27:** 570–580.

Huelsenbeck JP, Larget B, Alfaro ME. **2004**. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Molecular Biology and Evolution* **21**: 1123–1133.

Huelsenbeck JP, Rannala B, Masly JP. **2000**. Accommodating phylogenetic uncertainty in evolutionary studies. *Science* **288**: 2349–2350.

Hughes CE, Atchison GW. **2015**. The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains. *New Phytologist* **207**: 275–282.

Hughes CE, Eastwood R. **2006**. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences* **103**: 10334–10339.

Hughes KW, Petersen RH. **2001**. Apparent recombination or gene conversion in the ribosomal ITS region of a *Flammulina* (Fungi, Agaricales) hybrid. *Molecular Biology and Evolution* **18**: 94–96.

Hurvich CM, Tsai C. **1989**. Regression and time series model selection in small samples. *Biometrika* **76**: 297–307.

Huson DH, Bryant D. **2006**. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**: 254–267.

Huson DH, Klöpper T, Lockhart PJ, Steel MA. 2005. Reconstruction of reticulate networks from gene trees. In: Miyano S, Mesirov J, Kasif S, Istrail S, Pevzner PA, Waterman M, eds. *Research in computational molecular biology*. Berlin, Heidelberg, Germany: Springer, 233–249.

Jansen RK, Ruhlman TA. **2012**. Plastid genomes of seed plants. In: Bock R, Knoop V, eds. *Genomics of chloroplasts and mitochondria, advances in photosynthesis and respiration*. Dordrecht, the Netherlands: Springer, 103–126.

Joly S. **2012**. JML: Testing hybridization from species trees. *Molecular Ecology Resources* **12**: 179–184.

Joly S, McLenachan PA, Lockhart PJ. **2009**. A statistical approach for distinguishing hybridization and incomplete lineage sorting. *American Naturalist* **174**: E54–E70.

Karaman-Castro V, Urbatsch LE. **2009**. Phylogeny of *Hinterhubera* group and related genera (Hinterhuberinae: Astereae) based on the nrDNA ITS and ETS sequences. *Systematic Botany* **34**: 805–817.

Katoh K, Misawa K, Kuma K, Miyata T. **2002**. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**: 3059–3066.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C *et al.* 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.

Kim KJ, Choi KS, Jansen RK. **2005**. Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). *Molecular Biology and Evolution* **22**: 1783–1792.

Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9:** 357–359.

Legendre P, Desdevises Y, Bazin E. **2002**. A statistical test for host-parasite coevolution. *Systematic Biology* **51**: 217–234.

Leigh JW, Susko E, Baumgartner M, Roger AJ. **2008**. Testing congruence in phylogenomic analysis. *Systematic Biology* **57**: 104–115.

Li H, Durbin R. **2009**. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Linder CR, Goertzen LR, Heuvel B V, Francisco-Ortega J, Jansen RK. **2000**. The complete external transcribed spacer of 18S-26S rDNA: amplification and phylogenetic utility at low taxonomic levels in Asteraceae and closely allied families. *Molecular Phylogenetics and Evolution* **14**: 285–303.

Luebert F, Weigend M. **2014**. Phylogenetic insights into Andean plant diversification. *Frontiers in Ecology and Evolution* **2**: 1–17.

Luteyn, J. 1999. *Páramos: a checklist of plant diversity, geographical distribution, and botanical literature*. New York, USA: New York Botanical Garden.

Ma PF, Zhang YX, Zeng CX, Guo ZH, Li DZ. **2014**. Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (Poaceae). *Systematic Biology* **63**: 933–950.

Maddison WP. **1997**. Gene trees in species trees. *Systematic Biology* **46**: 523–536.

Maddison WP, Maddison D.R. 2015. *Mesquite: a modular system for evolutionary analysis.* Version 3.04. [WWW document] URL http://mesquiteproject.org. [accessed 1 January 2016]

Madriñán S, Cortés AJ, Richardson JE. **2013**. Páramo is the world's fastest evolving and coolest biodiversity hotspot. *Frontiers in Genetics* **4**: 192.

Mallet J, Beltrán M, Neukirchen W, Linares M. **2007**. Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC evolutionary biology* **7**: 28.

Martin M. **2011**. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal* **17**: 10–12.

Miller MA, Feiffer WP, Schwartz T. 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop (GCE)* **2010:** 1–8.

Moore WM. **1995**. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution* **49**: 718–726.

Mort ME, Crawford DJ, Kelly JK, Santos-Guerra A, Menezes de Sequeira M, Moura M, Caujape-Castells J. **2015**. Multiplexed-shotgun-genotyping data resolve phylogeny within a very recently derived insular lineage. *American Journal of Botany* **102**: 634–641.

Nesom GL. **1994**. Subtribal classification of the Astereae (Asteraceae). *Phytologia* **76**: 193–274.

Nesom GL, Robinson H. 2007. Tribe Astereae. In: Kadereit JW, Jeffrey C, eds. *The families and genera of vascular plants vol 8*. Berlin, Germany: Springer, 284–342.

Noyes RD, Rieseberg LH. **1999**. ITS sequence data support a single origin for North American Astereae (Asteraceae) and reflect deep geographic divisions in *Aster* s.l. *American Journal of Botany* **86**: 398–412.

Nürk NM, Scheriau C, Madriñán S. **2013**. Explosive radiation in high Andean *Hypericum* – rates of diversification among New World lineages. *Frontiers in Genetics* **4:** 175.

Panero JL, Freire SE, Ariza Espinar L, Crozier BS, Barboza GE, Cantero JJ. **2014**. Resolution of deep nodes yields an improved backbone phylogeny and a new basal lineage to study early evolution of Asteraceae. *Molecular Phylogenetics and Evolution* **80**: 43–53.

Panero JL, Funk VA. 2008. The value of sampling anomalous taxa in phylogenetic studies: Major clades of the Asteraceae revealed. *Molecular Phylogenetics and Evolution* **47**: 757–782.

Pérez-Escobar OA, Balbuena JA, Gottschling M. **2016**. Rumbling orchids: how to assess divergent evolution between chloroplast endosymbionts and the nuclear host. *Systematic Biology* **65**: 51–65.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. **2012**. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**: e37135.

Philippe H, Derelle R, Lopez P, Pick K, Borchiellini C, Boury-Esnault N, Vacelet J, Renard E, Houliston E, Quéinnec E *et al.* 2009. Phylogenomics revives traditional views on deep animal relationships. *Current Biology* **19**: 706–712.

Phillips RL, Kleese RA, Wang SS. **1971**. The nucleolus organizer region of maize (*Zea mays* L.): Chromosomal site of DNA complementary to ribosomal RNA. *Chromosoma* **36**: 79–88.

Qiu YL, Li LB, Wang B, Xue JY, Hendry T a, Li RQ, Brown JW, Liu Y, Hudson GT, Chen ZD. **2010**. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *Journal of Systematics and Evolution* **48**: 391–425.

R Core Team. 2016. *R: a language and environment for statistical computing.* [WWW document] http://www.R-project.org/. [accessed 1 January 2016]

Rambaut A. 2014. *FigTree v1.4.2.* [WWW document] URL http://tree.bio.ed.ac.uk/. [accessed 1 January 2016]

Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014. *Tracer v1.6.* [WWW document] URL http://beast.bio.ed.ac.uk/Tracer. [accessed 1 January 2016]

Rauscher JT. **2002**. Molecular phylogenetics of the *Espeletia* complex (Asteraceae): evidence from nrDNA its sequences on the closest relatives of an Andean adaptive radiation. *American Journal of Botany* **89**: 1074–1084.

Regier JC, Mitter C, Zwick A, Bazinet AL, Cummings MP, Kawahara AY, Sohn JC, Zwickl DJ, Cho S, Davis DR *et al.* 2013. A large-scale, higher-level, molecular phylogenetic study of the insect order Lepidoptera (moths and butterflies). *PLoS ONE* **8**: e58568.

Reginato M, Michelangeli FA. **2016**. Untangling the phylogeny of *Leandra s.str.* (Melastomataceae, Miconieae). *Molecular Phylogenetics and Evolution* **96**: 17–32.

Renaud G, Stenzel U, Maricic T, Wiebe V, Kelso J. **2014**. deML: Robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics* **31**: 770–772.

Rieppel O. **2010**. Species monophyly. *Journal of Zoological Systematics and Evolutionary Research* **48**: 1–8.

Rieseberg LH, Soltis DE. **1991**. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* **5**: 64–84.

Rognes T, Flouri T, Nichols B, Quince C, Mahé F. **2016**. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**: e2584–22.

Ronquist F, Huelsenbeck JP, Telensko M. 2011. *MrBayes version 3.2 manual: tutorials and model summaries.* [WWW document] URL http://mrbayes.sourceforge.net/index.php. [accessed 1 January 2016]

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA., Huelsenbeck JP. **2012**. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**: 539–542.

Salichos L, Rokas A. **2013**. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* **497**: 327–331.

Sánchez-Baracaldo P. **2004**. Phylogenetics and biogeography of the neotropical fern genera *Jamesonia* and *Eriosorus* (Pteridaceae). *American Journal of Botany* **91**: 274–284.

Sancho G, Hind DJN, Pruski JF. **2010**. Systematics of *Podocoma* (Asteraceae : Astereae): a generic reassessment. *Botanical Journal of the Linnean Society* **163**: 486–513.

Sancho G, Karaman-Castro V. **2008**. A phylogenetic study in American Podocominae (Asteraceae: Astereae) based on morphological and molecular data. *Systematic Botany* **33**: 762–775.

Seehausen O. **2004**. Hybridization and adaptive radiation. *Trends in Ecology and Evolution* **19**: 198–207.

Sessa EB, Zimmer EA, Givnish TJ. **2012**. Reticulate evolution on a global scale: A nuclear phylogeny for New World *Dryopteris* (Dryopteridaceae). *Molecular Phylogenetics and Evolution* **64**: 563–581.

Simpson BB. **1974**. Glacial migrations of plants: island biogeographical evidence. *Science* **185**: 698–700.

Sochor M, Vašut RJ, Sharbel TF, Trávníček B. **2015**. How just a few makes a lot: speciation via reticulation and apomixis on example of European brambles (*Rubus* subgen. *Rubus*, Rosaceae). *Molecular Phylogenetics and Evolution* **89**: 13–27.

Solís-Lemus C, Ané C. **2016**. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genetics* **12**: 1–21.

Stamatakis A. **2006**. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.

Stamatakis A. **2014**. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Stamatakis A. 2015. *The RAxML v8.1.X Manual.* [WWW document] URL http://sco.h-its.org/exelixis/web/software/raxml/index.html. [accessed 1 January 2016]

Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. **2012**. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* **99**: 349–364.

Strijk JS, Noyes RD, Strasberg D, Cruaud C, Gavory F, Chase MW, Abbott RJ, Thébaud C. 2012. In and out of Madagascar: dispersal to peripheral islands, insular speciation and diversification of Indian ocean daisy trees (*Psiadia*, Asteraceae). *PLoS ONE* **7**: e42932.

Sukumaran J, Holder MT. **2010a**. SumTrees: Phylogenetic Tree Summarization v.4.10. [WWW document] URL https://github.com/jeetsukumaran/DendroPy. [accessed 1 July 2016]

Sukumaran J, Holder MT. 2010b. DendroPy: a Python library for phylogenetic computing. *Bioinformatics* **26**: 1569–1571.

Sun M, Soltis DE, Soltis PS, Zhu X, Burleigh JG, Chen Z. **2015**. Deep phylogenetic incongruence in the angiosperm Rosidae clade. *Molecular Phylogenetics and Evolution* **83**: 156–166.

Swofford DL. 2002. *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4.* Sunderland, MA, USA: Sinauer Associates.

Vargas OM. **2011**. A nomenclator of *Diplostephium* (Asteraceae: Astereae): a list of species with their synonyms and distribution. *Lundellia* **14**: 32–51.

Vargas OM, Madriñán S. **2006**. Clave para la identificación de las especies del género *Diplostephium* (Asteraceae, Astereae) en Colombia. *Revista de la academia Colombiana de Ciencias Exactas Fisicas y Naturales* **30**: 489–494.

Vargas OM, Madriñán S. **2012**. Preliminary phylogeny of *Diplostephium* (Asteraceae): speciation rate and character evolution. *Lundellia* **15**: 1–15.

Volkov RA, Komarova NY, Hemleben V. **2007**. Ribosomal DNA in plant hybrids: inheritance, rearrangement, expression. *Systematics and Biodiversity* **5**: 261–276.

Wen D, Yu Y, Hahn MW, Nakhleh L. **2016**. Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. *Molecular Ecology* **25**: 2361–2372.

Whitfield JB, Lockhart PJ. **2007**. Deciphering ancient rapid radiations. *Trends in Ecology and Evolution* **22**: 258–265.

Wu J, Nyman T, Wang D-C, Argus GW, Yang Y-P, Chen J-H. **2015**. Phylogeny of *Salix* subgenus *Salix* s.l. (Salicaceae): delimitation, biogeography, and reticulate evolution. *BMC Evolutionary Biology* **15**: 1–13.

Wyman SK, Jansen RK, Boore JL. **2004**. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **20**: 3252–3255.

Xie W, Lewis PO, Fan Y, Kuo L, Chen MH. **2011**. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology* **60**: 150–160.

Zapata F. **2013**. A multilocus phylogenetic analysis of *Escallonia* (Escalloniaceae): diversification in montane South America. *American Journal of Botany* **100**: 526–545.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Diagram representing the sequence assembly pipeline for the genome skimming dataset.

Fig. S2 Circular representation of the chloroplast genome of *Diplostephium haenkei*.

Fig. S3 Circular representation of the mitochondrial genome of *Diplostephium hartwegii*.

Fig. S4 Nuclear ddRAD tree obtained with SVD quartets.

Fig. S5 Nuclear ribosomal maximum clade credibility tree obtained by Bayesian Inference with the 50% consensus matrix.

Fig. S6 Nuclear ribosomal maximum clade credibility tree obtained by Bayesian Inference with the 75% consensus matrix

Fig. S7 Nuclear ribosomal best tree obtained by Maximum Likelihood with the 50% consensus matrix.

Fig. S8 Nuclear ribosomal best tree obtained by Maximum Likelihood with the 75% consensus matrix.

Fig. S9 Nuclear ribosomal best tree obtained by Maximum Likelihood with the 90% consensus matrix.

Fig. S10 Chloroplast best tree obtained by Maximum Likelihood.

Fig. S11 Mitochondrial best tree obtained by Maximum Likelihood.

Fig. S12 Robinson-Foulds tree distances visualization among the nuclear ribosomal 50%, 75% and 90% consensus datasets.

Fig. S13 Robinson-Foulds tree distances visualization among the ddRAD, nuclear ribosomal 90%, chloroplast and mitochondrial (red) datasets.

Fig. S14. Chronogram of the ITS matrix.

Table S1 List of specimens with their voucher and GenBank information

Table S2 Descriptive statistics of the matrices obtained

Table S3 Bayesian factor comparison between the two partition models calculated with MrBayes in each genome skimming dataset

Table S4 Models of evolution inferred by jModelTest and MrBayes for the genome skimming datasets

Table S5 Complete list of Patterson's D-statistic calculations

Methods S1 Parameters file used for assembly of the reference ddRAD library in ipyrad.

Methods S2 Parameters file used for assembly of the nuclear-ddRAD dataset in shotgun2rad.

Methods S3 Parameters file used for assembly of the nuclear-ddRAD dataset in pyRAD.

Fig. 1 Phylogenies obtained from the different datasets: (a) nuclear-ddRAD – maximum likelihood, (b) nuclear ribosomal 90% – Bayesian inference (BI) maximum clade credibility tree (MCCT), (c) chloroplast – BI-MCCT, and (d) mitochondrial – BI-MCCT. Dots indicate highly supported nodes (bootstrap support > 90, Bayesian posterior probability > 0.9). Shaded areas indicate the distribution of *Diplostephium* taxa according to the map at the right (e).

Fig. 2 JML matrix mapped on the nuclear-ddRAD topology showing significant pairwise comparisons (black squares) in which the chloroplast genetic distance was found to be significantly smaller than expected under a scenario with incomplete lineage sorting with no migration suggesting hybridization. Ingroup rogue taxa identified by PACo are indicated by the orange squares. Numbers following taxon names indicate the haploid number of chromosomes, the star indicates that the count was a made in a different species of the same genus.

Fig. 3 Chronogram calculated on the maximum likelihood ddRAD topology using the nuclear ribosomal 90% matrix. Numbers at nodes indicate the average estimated ages while bars show their 95% confidence interval. Stars mark the calibration points used.

Fig. 4 Photos of taxa with high genetic hybridization signal. (a) *Parastrephia quadrangularis,* (b) *Diplostephium meyenii*, (c) *Diplostephium sp. nov.* CAJ2, and (d) *Diplostephium cinereum.* Notice the hybrid morphology of *Diplostephium sp. nov.* CAJ2 in relation to *P. quadrangularis* and *D. meyenii*.

Table 1 Patterson's D-statistic showing the tests for which introgression between P3 and either P1 or P2 was detected using a Z-score threshold of 3 (see Supporting Information Table S1 for full species details) **[Author, please check inserted text '(see Supporting Information Table S1 for full species details)' is ok.]**

P1	P2	P3	O	Z	BABA	ABBA	nloci
Introgression among genera							
D. ochraceum	*D. violaceum*	*B. bartsiifolia*	*D. haenkei*	4.99	8.5	35.25	198
B. genistelloides	*L. sophiifolia*	*D. ericoides*	*D. ochraceum*	4.8	30.75	84	159
B. genistelloides	*B. tricuneata*	*D. espinosae*	*D. frontinense*	5.06	77.25	169.5	291
B. genistelloides	*L. sophiifolia*	*D. espinosae*	*D. frontinense*	4.35	50.38	106.88	204
B. genistelloides	*B. tricuneata*	*D. ochraceum*	*D. haenkei*	4.58	184	89.5	303
B. genistelloides	*B. tricuneata*	*D. violaceum*	*D. haenkei*	3.55	172.5	96.5	309
B. genistelloides	*B. tricuneata*	*H. alienus*	*E. notobellidiastrum*	5.65	121.88	48.88	190

Introgression inside *Diplostephium* group A							
D. mutiscuanum	*D. oblongifolium*	*D. antioquense*	*D. colombianum*	8.15	223.12	542.88	1447
D. revolutum	*D. rosmarinifolium*	*D. colombianum*	*B. tricuneata*	6.03	167.38	56.88	333
D. revolutum	*D. rosmarinifolium*	*D. eriophorum*	*B. tricuneata*	4.87	146.75	64	337
D. revolutum	*D. rosmarinifolium*	*D. juajibioyi*	*D. colombianum*	3.84	232.12	361.88	1321
D. tenuifolium	*D. ochraceum*	*D. rosmarinifolium*	*D. colombianum*	5.05	478.12	294.12	1550
Introgression inside *Diplostephium* group B							
D. callilepis	*D. goodspeedii*	*D. haenkei*	*D. azureum*	13.5	238.38	715.38	1350
D. meyenii	*P. quadrangularis*	*D. cinereum*	*B. tricuneata*	19.44	498	76.75	444
D. cinereum	*D. sp. nov. CAJ2*	*D. meyenii*	*F. hypsophila*	11.6	337	89.25	452
D. foliosissimum	*D. sagasteguii*	*D. meyenii*	*B. tricuneata*	11.14	78.62	309.12	473
D. barclayanum	*D. sagasteguii*	*D. meyenii*	*B. tricuneata*	9.42	81.38	277.62	473
D. haenkei	*D. callilepis*	*D. meyenii*	*F. hypsophila*	9.3	300.25	92.25	490
D. goodspeedii	*D. callilepis*	*D. meyenii*	*F. hypsophila*	5.28	242.38	108.62	466
D. meyenii	*P. quadrangularis*	*D. sp. nov. CAJ2*	*B. tricuneata*	19.44	498	76.75	444
D. sp. nov. JUN3	*D. sp. nov. JUN*	*D. sp. nov. JUN2*	*D. azureum*	3.35	411.5	551.25	1294

(a) Nuclear ddRAD

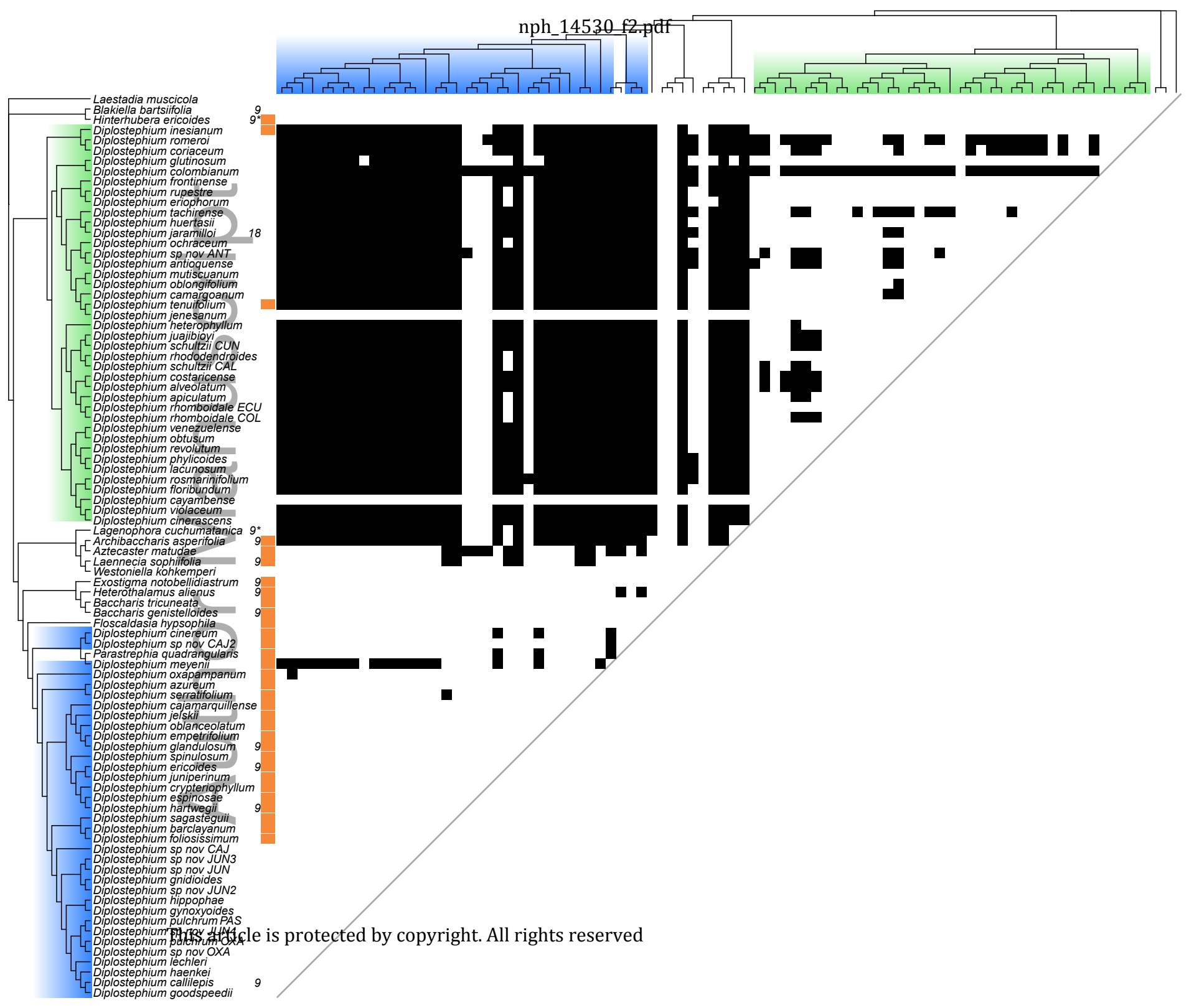
(b) Nuclear ribosomal

(c) Chloroplast

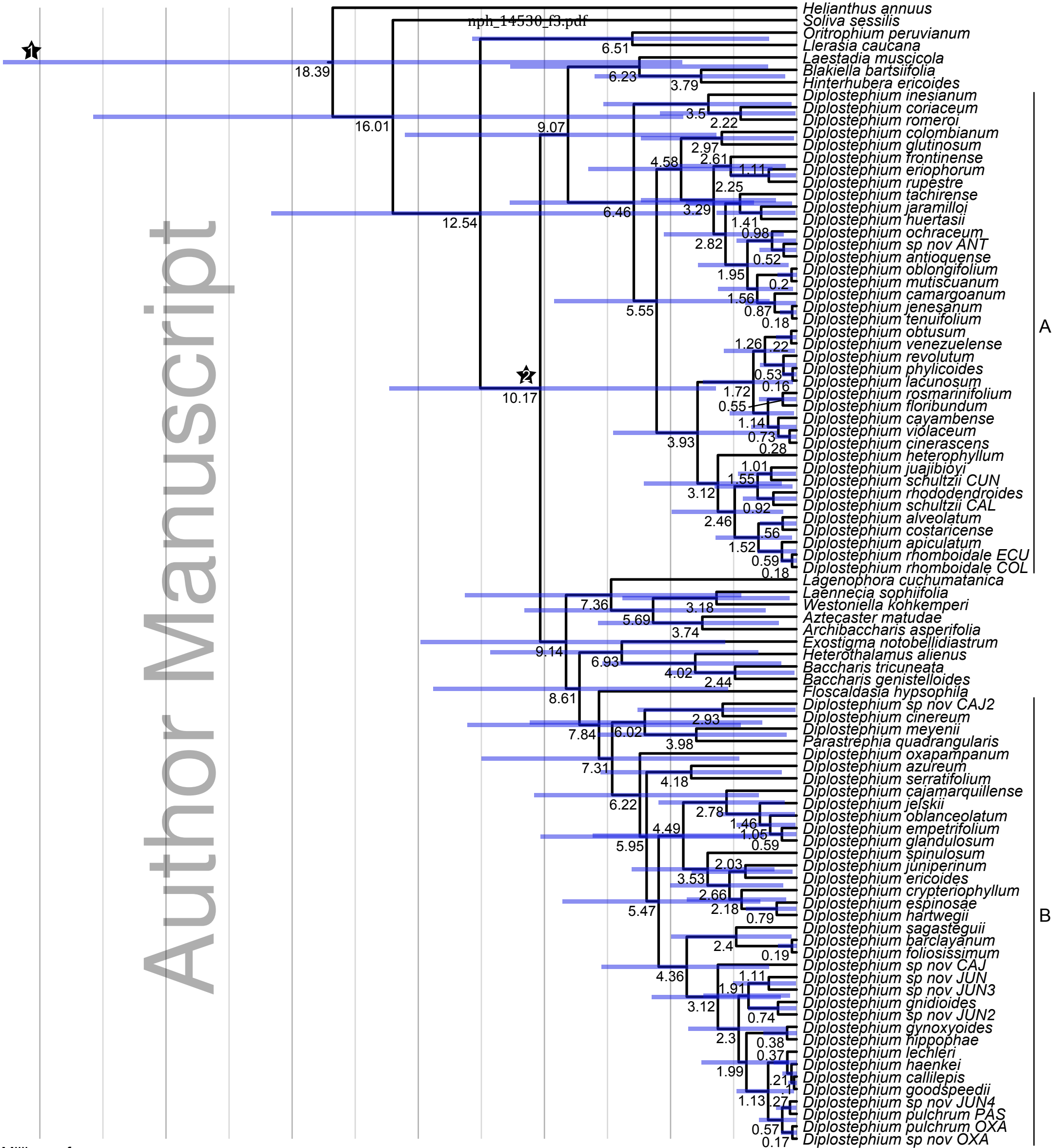
(d) Mitochondrial

(e)

South America



nph_14530_f3.pdf



nph_14530_f4.tif