New Phytologist

***New Phytologist* Supporting Information**

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

Authors: Oscar M. Vargas, Edgardo M. Ortiz and Beryl B. Simpson

Article acceptance date: 19 February 2016

The following Supporting Information is available 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% (purple), 75% (blue), and 90% (black) consensus datasets.

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

Fig. S14. Chronogram of the ITS matrix.

Table S1 (separate xls file) 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 (separate xls file) 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

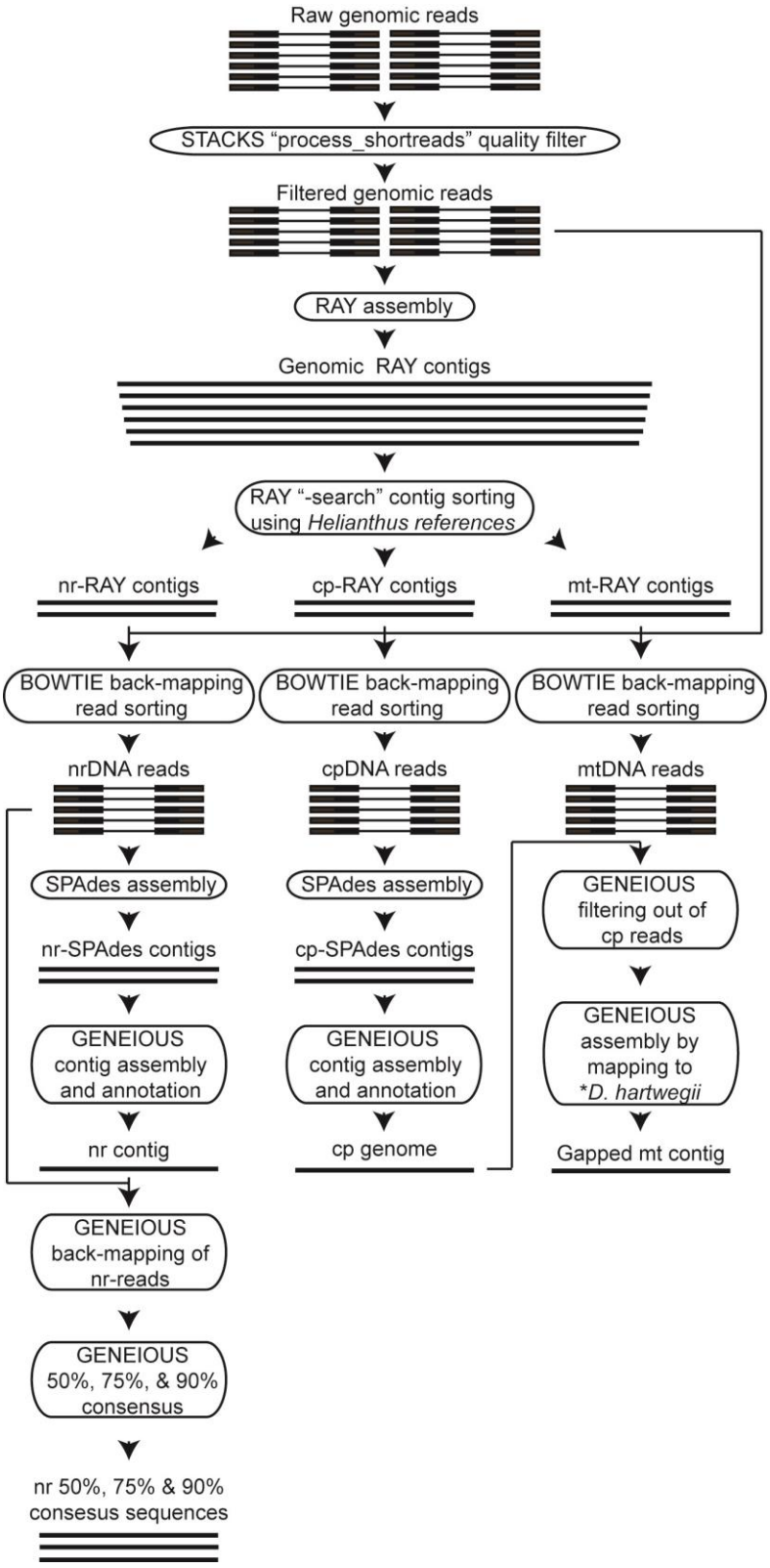

Fig. S1 Diagram representing the sequence assembly pipeline for the genome skimming dataset.

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

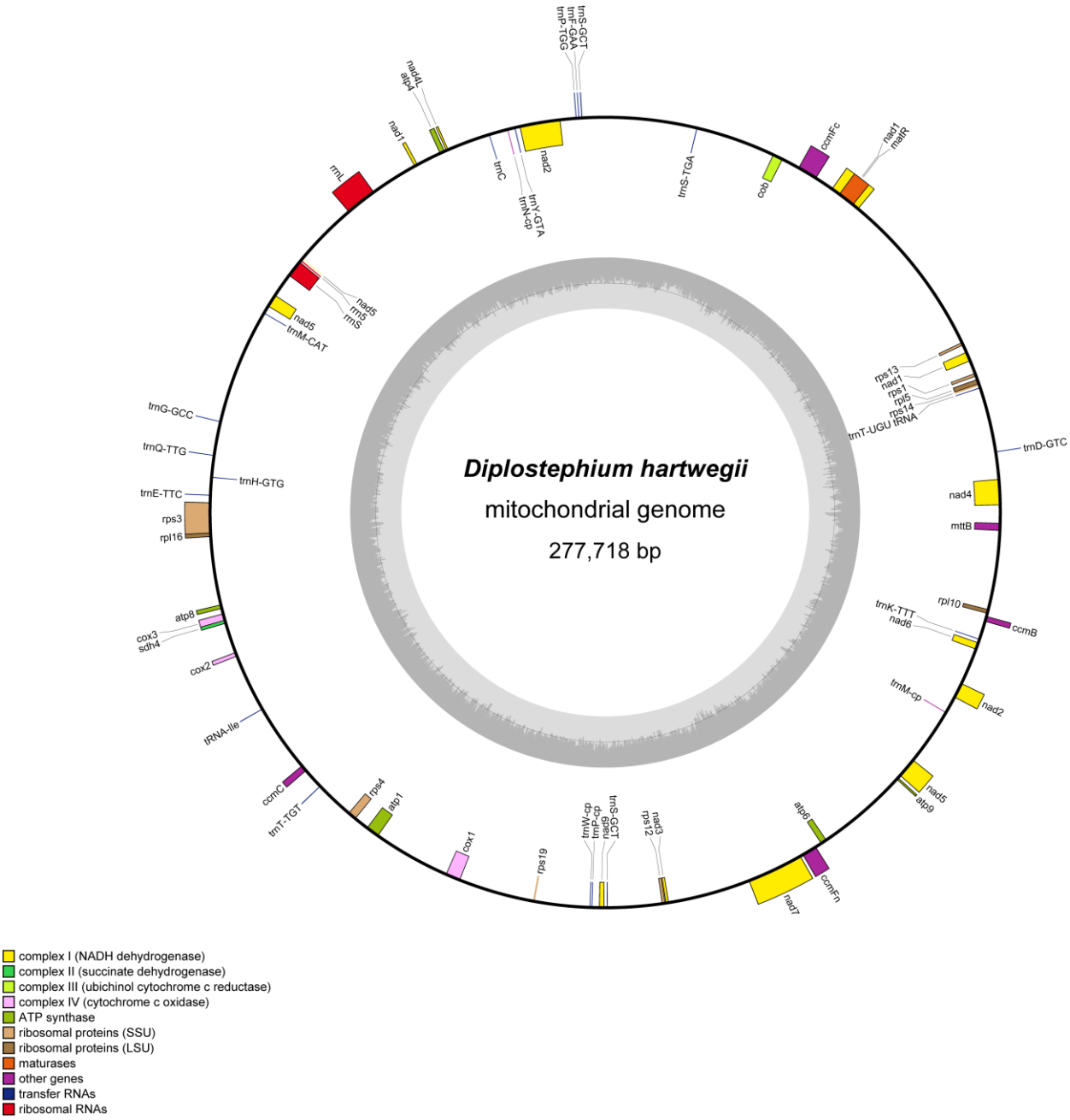

Fig. S3 Circular representation of the mitochondrial genome of *Diplostephium hartwegii*.

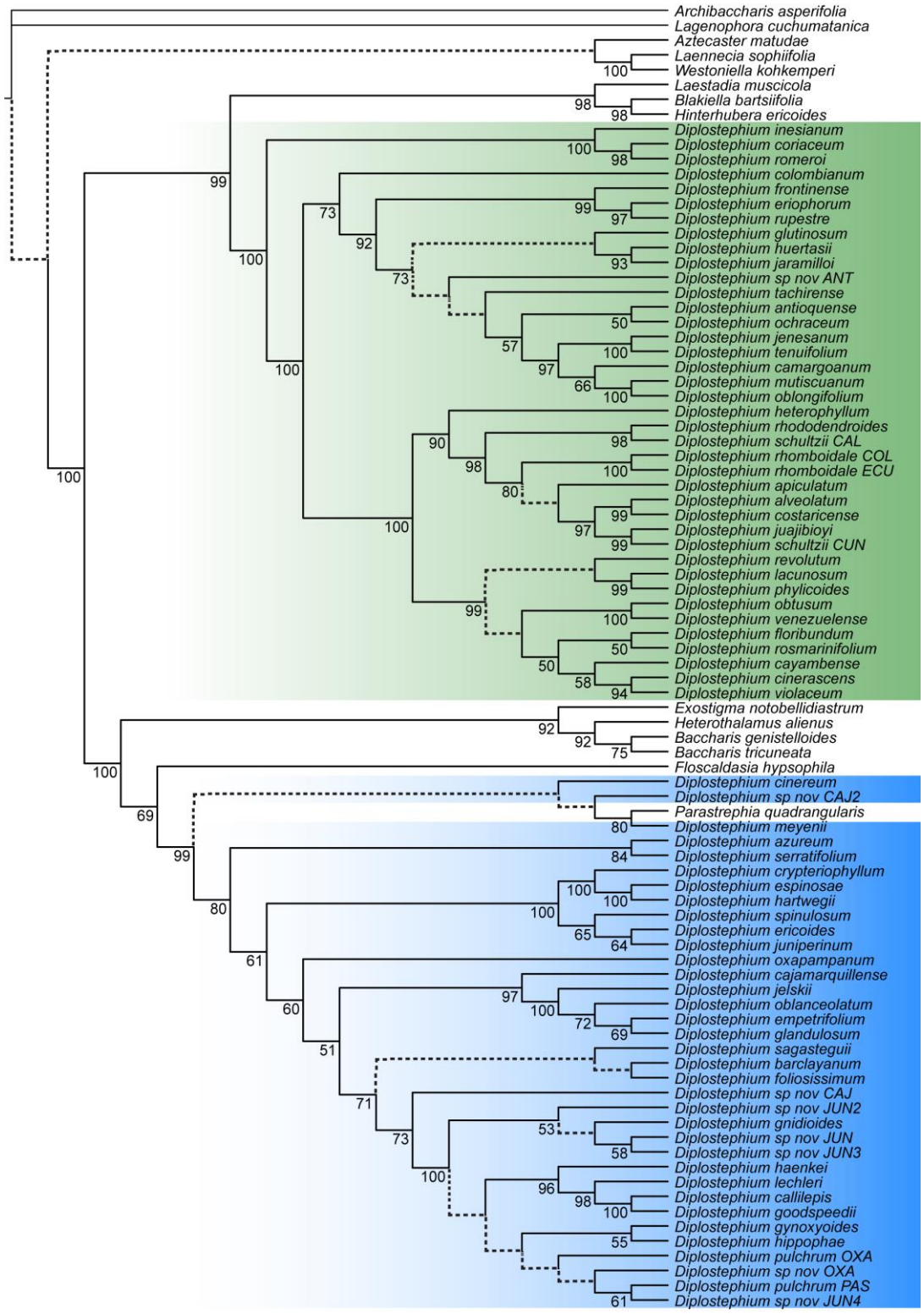

Fig. S4 Nuclear-ddRAD tree obtained with SVD quartets. Numbers below the branches indicate bootstrap support (BS). Branches with low support (BS<50) are dashed and their support is not shown. Color boxes follow the convention of Fig. 1.

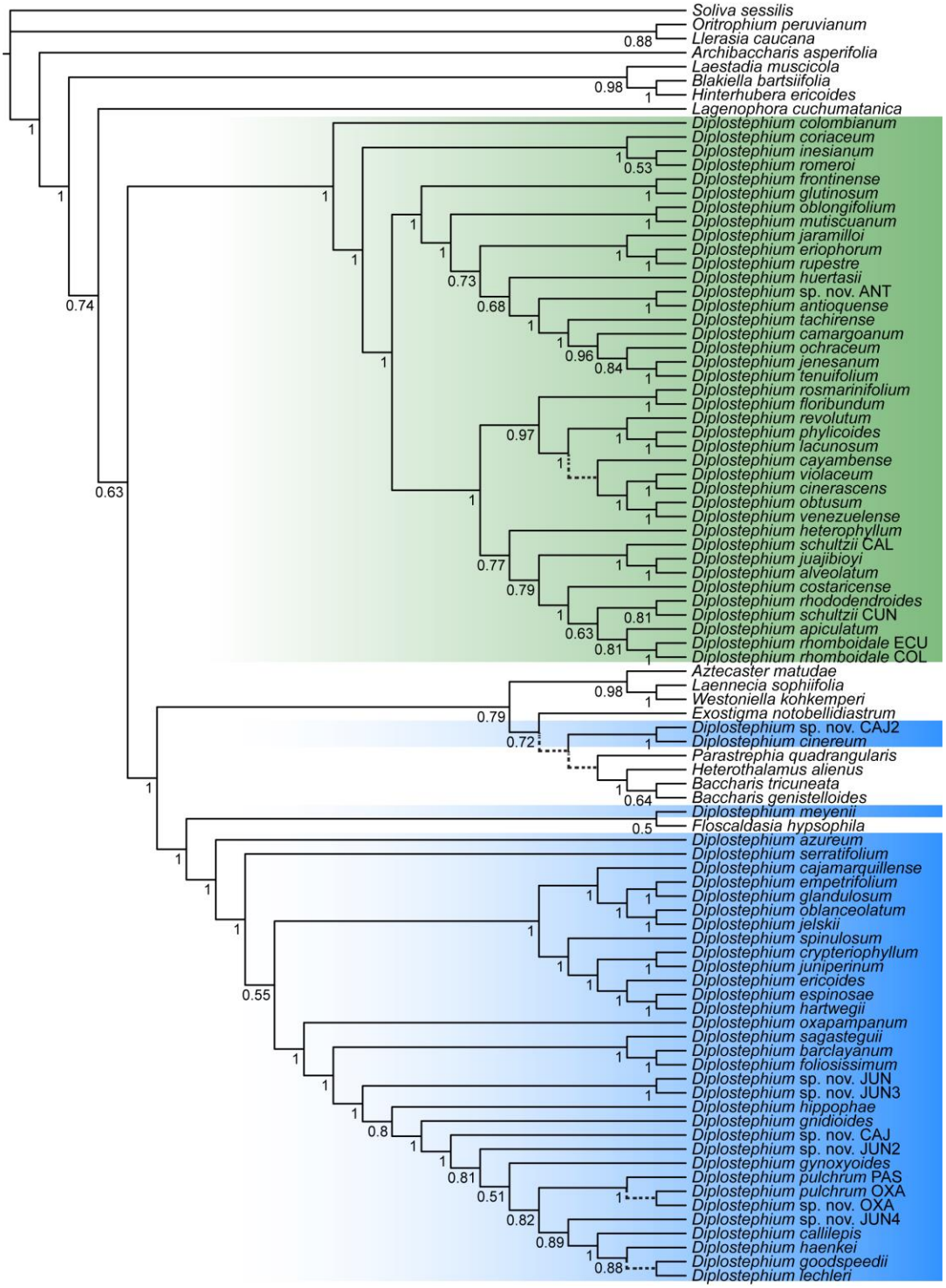

Fig. S5 Nuclear ribosomal maximum clade credibility tree obtained by Bayesian Inference with the 50% consensus matrix. Numbers below the branches indicate the Bayesian posterior probability (BPP). Branches with low support (BPP<0.5) are dashed and their support is not shown. Color boxes follow the convention of Fig. 1.

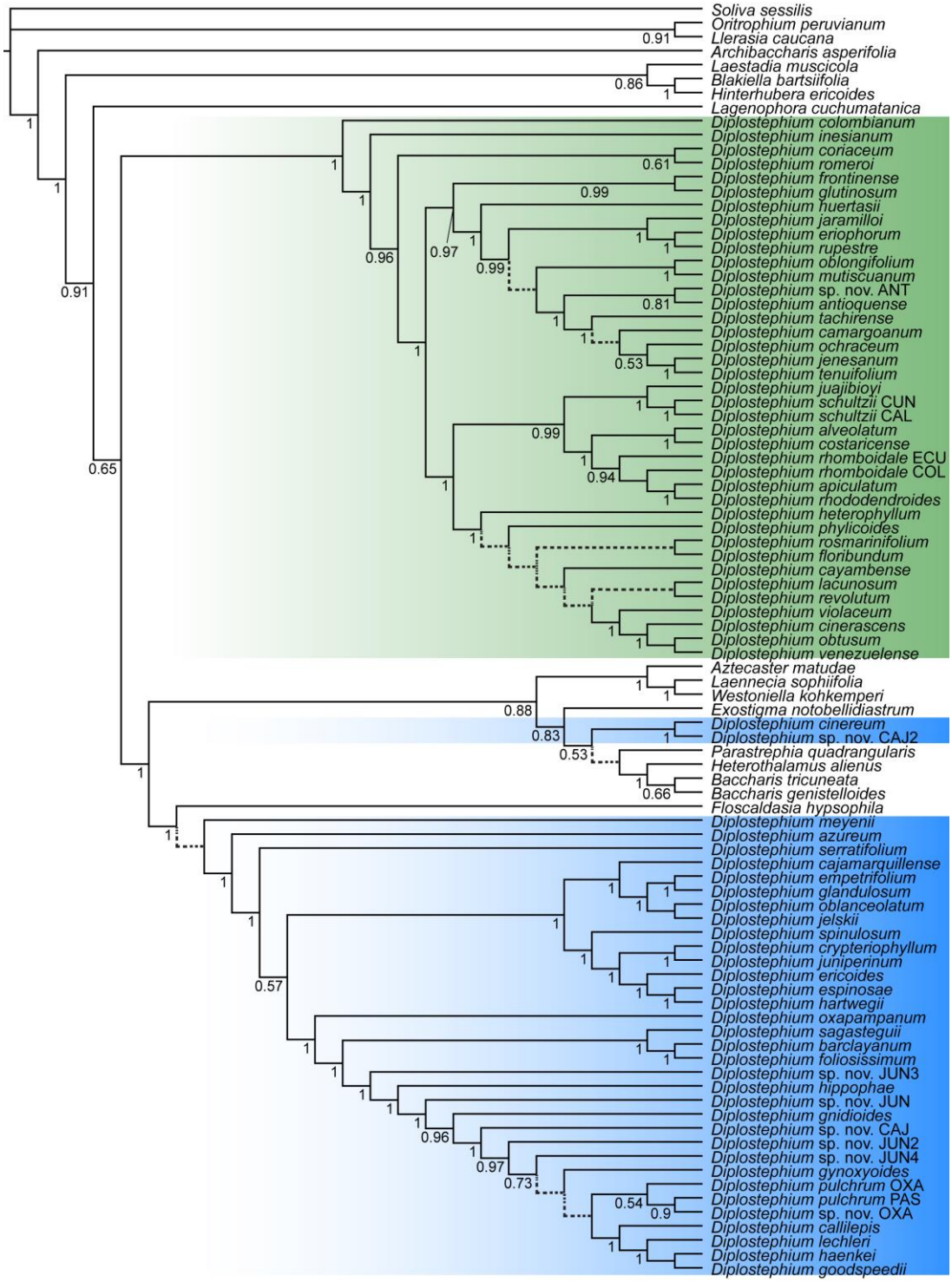

Fig. S6 Nuclear ribosomal maximum clade credibility tree obtained by Bayesian Inference with the 75% consensus matrix. Numbers below the branches indicate the Bayesian posterior probability (BPP). Branches with low support (BPP<0.5) are dashed and their support is not shown. Color boxes follow the convention of Fig. 1.

Fig. S7 Nuclear ribosomal best tree obtained by Maximum Likelihood with the 50% consensus matrix. Numbers below the branches indicate bootstrap support (BS). Branches with low support (BS<50) are dashed and their support is not shown. Color boxes follow the convention of Fig. 1.

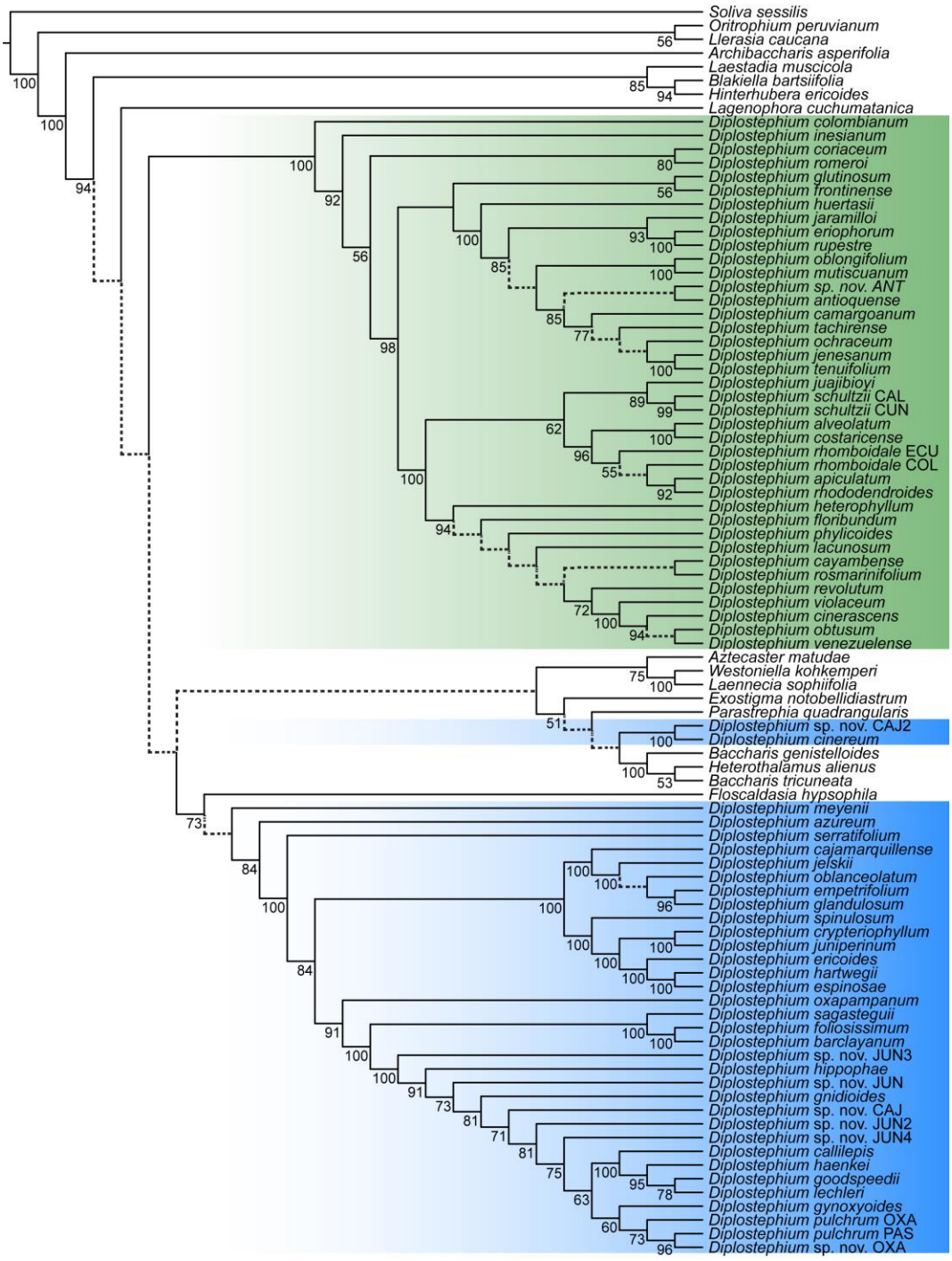

Fig. S8 Nuclear ribosomal best tree obtained by Maximum Likelihood with the 75% consensus matrix. Numbers below the branches indicate bootstrap support (BS). Branches with low support (BS<50) are dashed and their support is not shown. Color boxes follow the convention of Fig. 1.

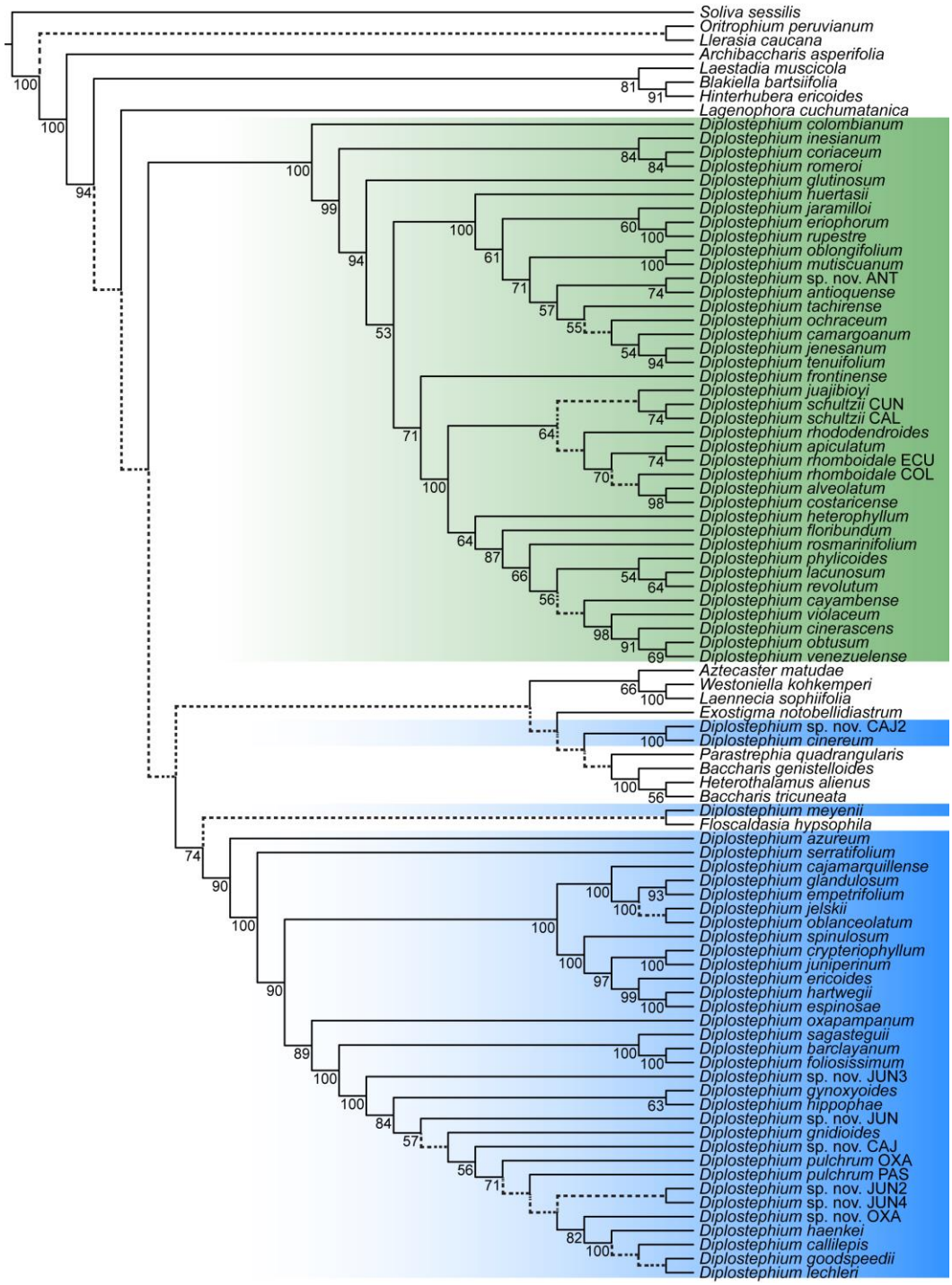

Fig. S9 Nuclear ribosomal best tree obtained by Maximum Likelihood with the 90% consensus matrix. Numbers below the branches indicate bootstrap support (BS). Branches with low support (BS<50) are dashed and their support is not shown. Color boxes follow the convention of Fig. 1.

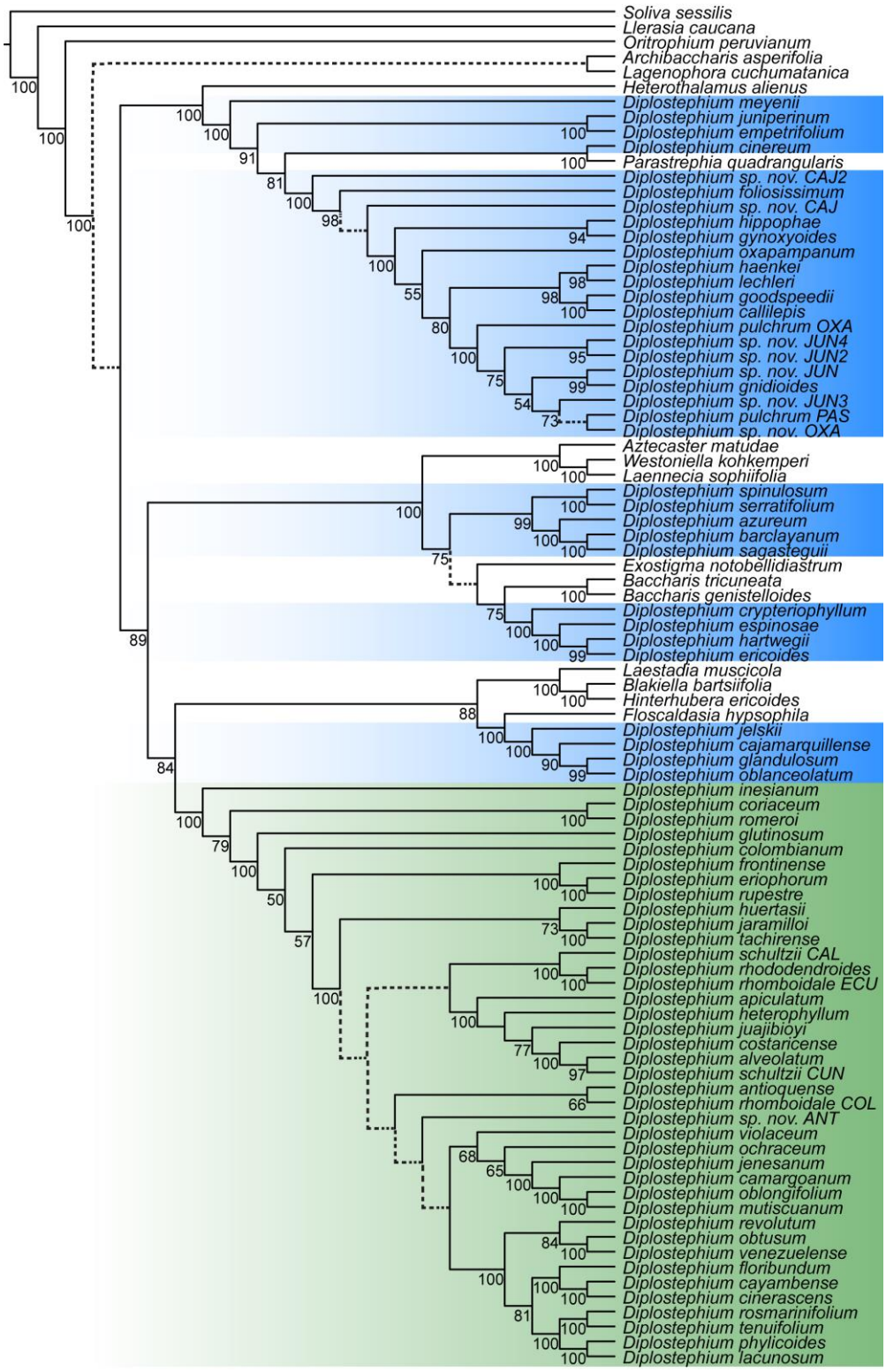

Fig. S10 Chloroplast best tree obtained by Maximum Likelihood. Numbers below the branches indicate bootstrap support (BS). Branches with low support (BS<50) are dashed and their support is not shown. Color boxes follow the convention of Fig. 1.

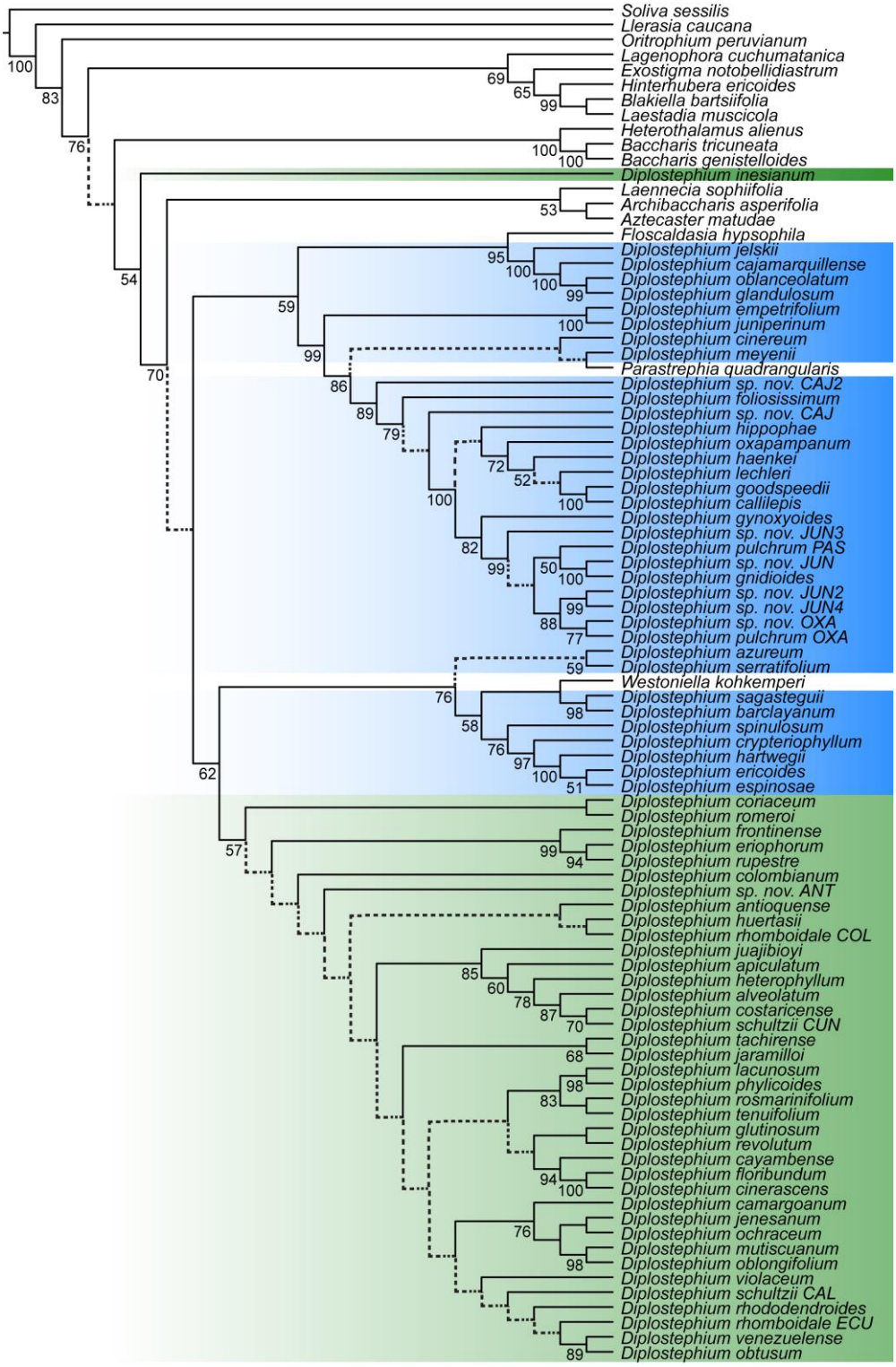

Fig. S11 Mitochondrial best tree obtained by Maximum Likelihood. Numbers below the branches indicate bootstrap support (BS). Branches with low support (BS<50) are dashed and their support is not shown. Color boxes follow the convention of Fig. 1.

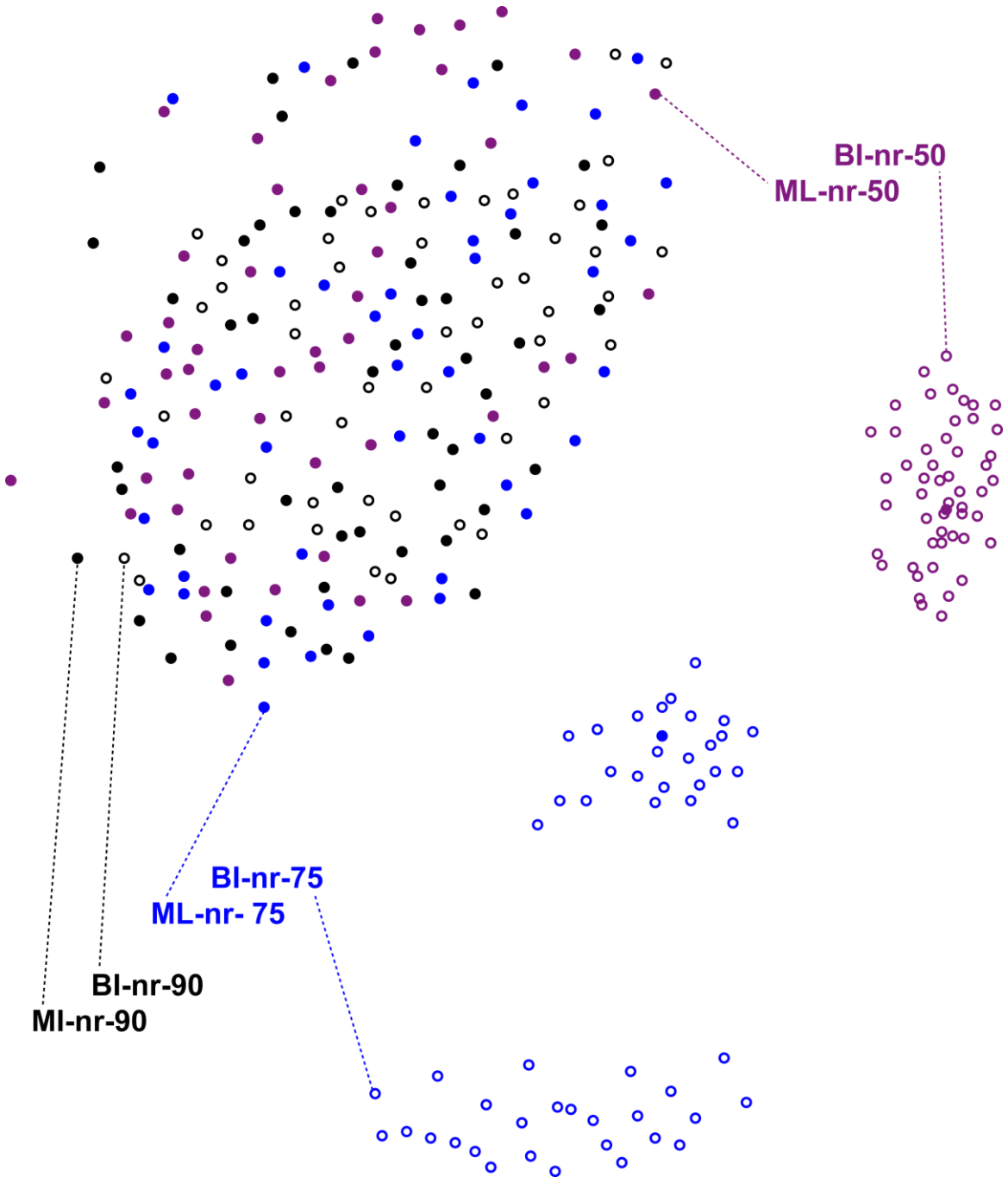

Fig. S12 Robinson-Foulds tree distances visualization among the nuclear ribosomal 50% (purple), 75% (blue), and 90% (black) consensus datasets. Bayesian inference (outline dots) tree subsets are composed by the maximum clade credibility tree and 50 random trees sample from the Markov Chain Monte Carlo runs after a 0.25 burnin. Maximum likelihood (full dots) tree subsets are composed by the best tree and 50 random trees sampled from the RAxML bootstrap.

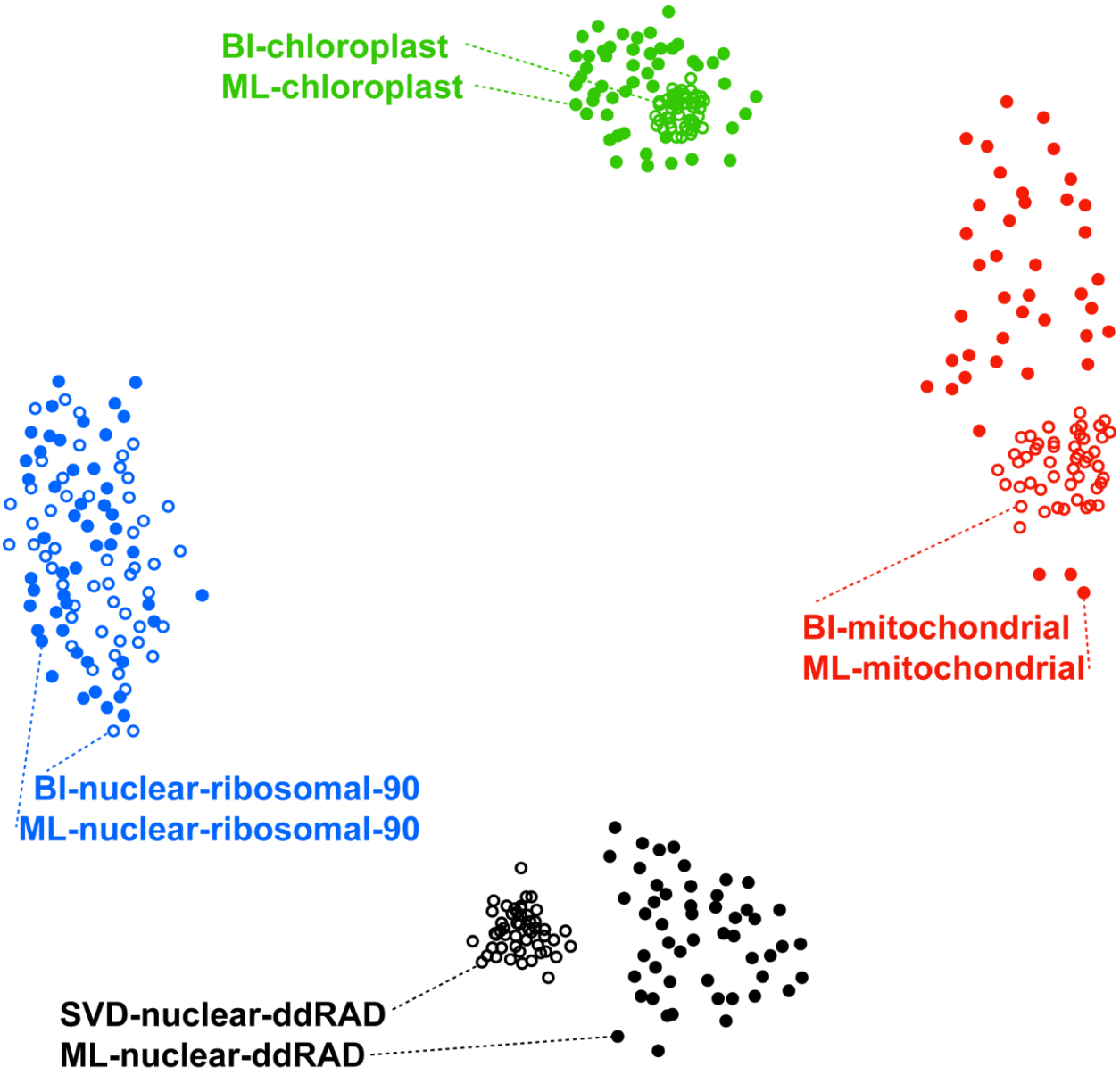

Fig. S13 Robinson-Foulds tree distances visualization among the ddRAD (black), nuclear ribosomal 90% (blue), chloroplast (green), and mitochondrial (red) datasets. Bayesian inference (BI) tree subsets are composed by the maximum clade credibility tree and 50 random trees sample from the Markov Chain Monte Carlo runs after a 0.25 burnin. Maximum likelihood (ML) and SVD quartets (SVDq) tree subsets are composed by the best tree and 50 random trees sampled from bootstrap replicates. Outlined dots represent BI topologies; full dots represent ML and SVDq (ddRAD only) topologies.

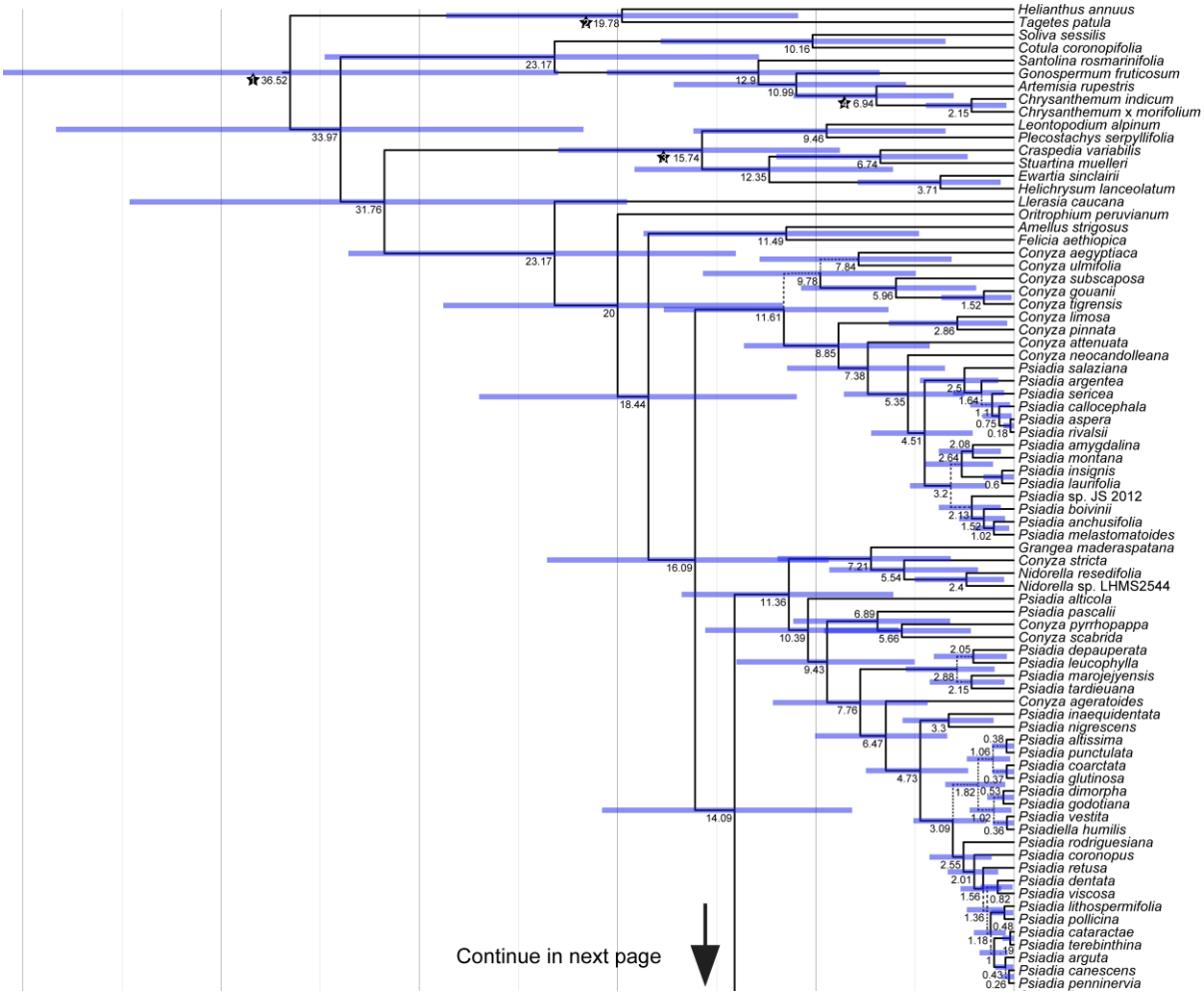

Fig S14 (part). Chronogram of the ITS matrix.

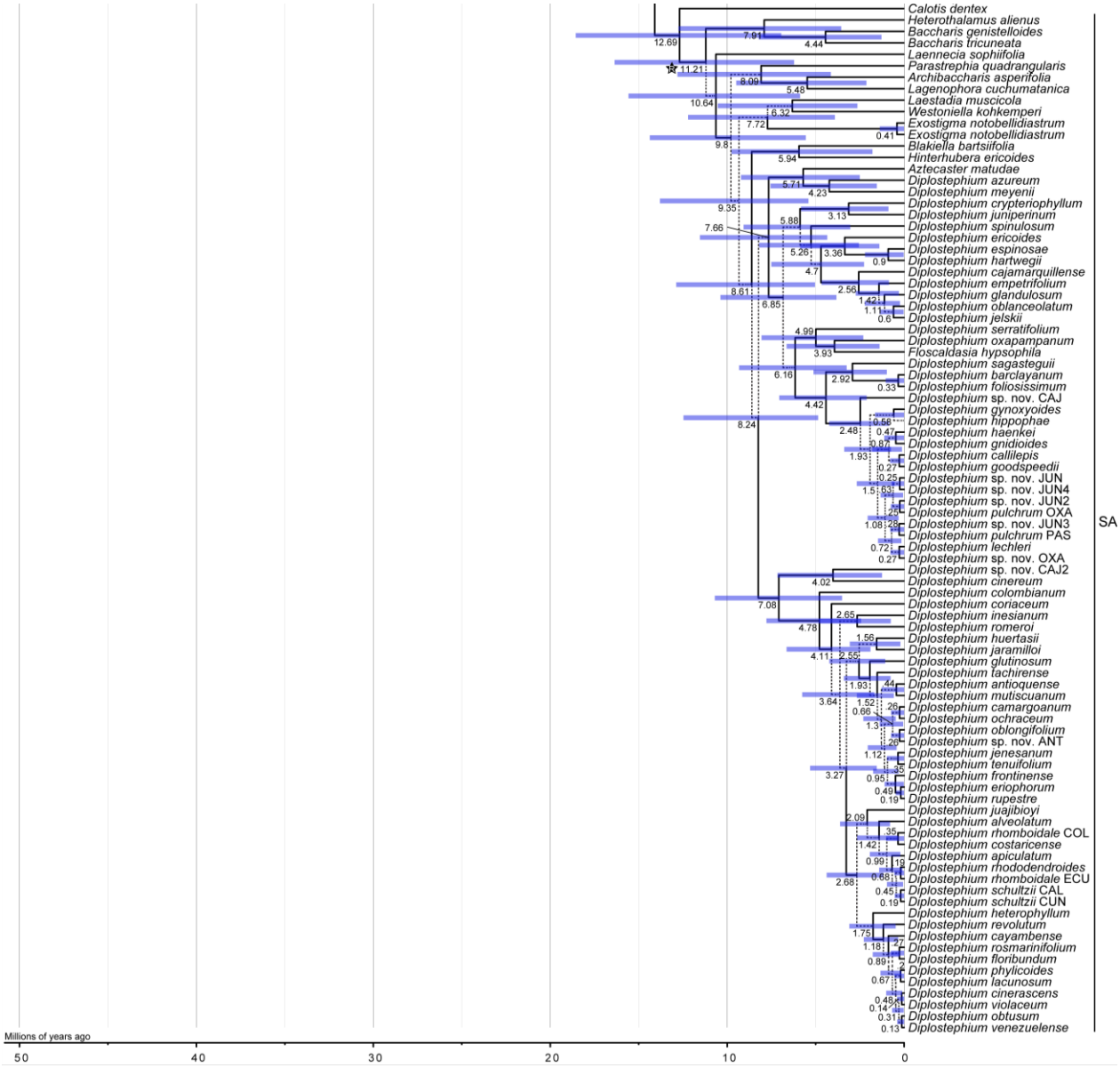

Fig S14. Chronogram of the ITS matrix. Numbers at nodes indicate the average estimated ages while bars show their 95% confidence interval. Low supported branches (Bayesian posterior probability <0.50) are indicated by dashed lines. Stars 1–4 indicate calibration points, R represents the calibration point extracted from this chronogram to calibrate the nuclear dataset. SA indicates the "South American lineages."

Table S2 Descriptive statistics of the matrices obtained. PICs: parsimony informative sites (calculated excluding outgroups). TS: transcribed spacers (including ETS, ITS1, and ITS2). NTS: non-transcribed spacer. nr: nuclear ribosomal. n-ddRAD: nuclear-DDrad.

Dataset	Sites	Missing data %	Ambiguities (% = ambiguities/ (sites x #samples))	PICs (%)	Coding PICs (%)	Non-coding PICs (%, TS+NTS% if applicable)
nr50	13,362	0.00%	25 (0.003)	1,425 (10.66)	114(8.00)	1,311 (34.32+57.68)
nr75	13,362	0.00%	3201 (0.384)	1,273 (9.53)	93(7.31)	1,180 (34.01+58.68)
nr90	13,362	0.00%	5012 (0.602)	1,203 (9.00)	86(7.15)	1,117 (33.83+59.02)
chloroplast	135,440	0.02%	103 (0.022)	2,169 (1.60)	709 (32.69)	1,460 (67.31)
mitochondrial	209,392	1.50%	0	1,124 (0.54)	40(3.56)	1,084 (96.44)
n-ddRAD	244,255	74.2%	251,092 (1.168)	96,421(39.5)	NA	NA

Table S3 Bayesian factor comparison between the two partition models calculated with MrBayes in each genome skimming dataset. M0: unpartitioned model. M1: partitioned model. nr: nuclear ribosomal matrix. cp: chloroplast matrix. mt: mitochondrial matrix.

Region	Harmonic mean Marginal ln*L*		Bayes Factor	Stepping-stone Marginal ln*L*		Bayes Factor
	M0	M1	M0 – M1	M0	M1	M0 – M1
nr50	-72698.76	-68046.52	-4652.24	-73447.88	-68805.2	-4642.68
nr75	-67156.72	-62591.73	-4564.99	-67934.08	-63360.31	-4573.77
nr90	-65102.02	-60592.81	-4509.21	-65858.43	-61379.04	-4479.39
cp	-293607.91	-290782.09	-2825.82	-294399.45	-291582.59	-2816.86
mt	-410381.23	-408912.98	-1468.25	-411267.49	-409841.04	-1426.45

Table S4 Models of evolution inferred by jModelTest and MrBayes for the genome skimming datasets. The final MrBayes analysis used the nucleotide substitution model inferred by MrBayes plus the +Γ and +I parameters if suggested by jModeltest. nr: nuclear ribosomal matrix. cp: chloroplast matrix. mt: mitochondrial matrix.

Dataset	Partition	AICc model	-lnL	AICc score	MrBayes model	Posterior Probability
nr75	coding	GTR+I+Γ	10236.83329	20867.81749	122345	0.285
	transcribed spacers	GTR+I+Γ	17071.14411	34554.87915	123451	0.336
	non-coding (NTS)	012232+Γ+F	34224.70918	68836.35287	121121	0.375
cp	coding	012313+I+Γ+F	140860.8997	282098.7494	123145	0.672
	non-coding	012310+I+Γ+F	149090.4473	298558.0724	123342	0.289
mt	coding	001102+I+Γ+F	47170.9926	94718.24837	122211	0.188
	non-coding	TPM1uf+I+Γ	360919.2953	722212.9854	123321	0.821

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

```
------- ipyrad params file (v.0.5.1)-------------------------------------------
ast-FW100-85-m2              ## [0] [assembly_name]: Assembly name. Used to name output directories for assembly
./                           ## [1] [project_dir]: Project dir (made in curdir if not present)
                             ## [2] [raw_fastq_path]: Location of raw non-demultiplexed fastq files
                             ## [3] [barcodes_path]: Location of barcodes file
./ast-PE100_fastqs/*_R1_.fq.gz ## [4] [sorted_fastq_path]: Location of demultiplexed/sorted fastq files
denovo                       ## [5] [assembly_method]: Assembly method (denovo, reference, denovo+ref, denovo-ref)
                             ## [6] [reference_sequence]: Location of reference sequence file
ddrad                        ## [7] [datatype]: Datatype (see docs): rad, gbs, ddrad, etc.
,                            ## [8] [restriction_overhang]: Restriction overhang (cut1,) or (cut1, cut2)
5                            ## [9] [max_low_qual_bases]: Max low quality base calls (Q<20) in a read
26                           ## [10] [phred_Qscore_offset]: phred Q score offset (33 is default and very standard)
6                            ## [11] [mindepth_statistical]: Min depth for statistical base calling
6                            ## [12] [mindepth_majrule]: Min depth for majority-rule base calling
9999999999                   ## [13] [maxdepth]: Max cluster depth within samples
0.85                         ## [14] [clust_threshold]: Clustering threshold for de novo assembly
0                            ## [15] [max_barcode_mismatch]: Max number of allowable mismatches in barcodes
2                            ## [16] [filter_adapters]: Filter for adapters/primers (1 or 2=stricter)
35                           ## [17] [filter_min_trim_len]: Min length of reads after adapter trim
2                            ## [18] [max_alleles_consens]: Max alleles per site in consensus sequences
8                            ## [19] [max_Ns_consens]: Max N's (uncalled bases) in consensus (R1, R2)
14                           ## [20] [max_Hs_consens]: Max Hs (heterozygotes) in consensus (R1, R2)
2                            ## [21] [min_samples_locus]: Min # samples per locus for output
29                           ## [22] [max_SNPs_locus]: Max # SNPs per locus (R1, R2)
100, 100                     ## [23] [max_Indels_locus]: Max # of indels per locus (R1, R2)
1.0                          ## [24] [max_shared_Hs_locus]: Max # heterozygous sites per locus (R1, R2)
0, 0                         ## [25] [edit_cutsites]: Edit cut-sites (R1, R2) (see docs)
0, 0, 0, 0                   ## [26] [trim_overhang]: Trim overhang (see docs) (R1>, <R1, R2>, <R2)
*                            ## [27] [output_formats]: Output formats (see docs)
                             ## [28] [pop_assign_file]: Path to population assignment file
```

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

```
==== parameter inputs for shotgun2rad version 2.0  ============================================
pyRAD                   ## 1.: command (or path) to call pyRAD
bwa                     ## 2.: command (or path) to call bwa
12                      ## 3.: number of threads for bwa (def. 1)
samtools                ## 4.: command (or path) to call samtools
ref/ast-FW100-85.fasta  ## 5.: path and prefix of bwa-indexed loci of reference
0.8                     ## 6.: minOverlap: of read mapped against ref. locus (def. 0.9)
30                      ## 7.: minMAPQ: minimum mapping quality (0-60, def. 60)
                        ## 8.: clustSize: reads to retain per cluster (def. = mindepth)
================================================================================================
```

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

```
==** parameter inputs for pyRAD version 3.0.66  **======================== affected step ==
./                    ## 1. Working directory                          (all)
                      ## 2. Loc. of non-demultiplexed files (if not line 18)  (s1)
                      ## 3. Loc. of barcode file (if not line 18)           (s1)
vsearch                  ## 4. command (or path) to call vsearch (or usearch)    (s3,s6)
muscle                   ## 5. command (or path) to call muscle                  (s3,s7)
                      ## 6. Restriction overhang (e.g., C|TGCAG -> TGCAG)     (s1,s2)
48                     ## 7. N processors (parallel)                      (all)
6                      ## 8. Mindepth: min coverage for a cluster             (s4,s5)
5                      ## 9. NQual: max # sites with qual < 20 (or see line 20)(s2)
.85                    ## 10. Wclust: clustering threshold as a decimal        (s3,s6)
gbs                    ## 11. Datatype: rad,gbs,pairgbs,pairddrad,(others:see docs)(all)
44                     ## 12. MinCov: min samples in a final locus             (s7)
88                     ## 13. MaxSH: max inds with shared hetero site          (s7)
c85d6mj2m44p88out         ## 14. Prefix name for final output (no spaces)         (s7)
==== optional params below this line ===================================  affected step ==
                      ## 15.opt.: select subset (prefix* only selector)           (s2-s7)
OUT_Ssess_19_12,OUT_Operu_14_04,OUT_Lcauc_18_19 ## 16.opt.: add-on (outgroup) taxa (list or prefix*)         (s6,s7)
                      ## 17.opt.: exclude taxa (list or prefix*)                  (s7)
                      ## 18.opt.: loc. of de-multiplexed data                     (s2)
                      ## 19.opt.: maxM: N mismatches in barcodes (def= 1)         (s1)
                      ## 20.opt.: phred Qscore offset (def= 33)                   (s2)
                      ## 21.opt.: filter: def=0=NQual 1=NQual+adapters. 2=strict   (s2)
0.00304,0.01820        ## 22.opt.: a priori E,H (def= 0.001,0.01, if not estimated) (s5)
13                     ## 23.opt.: maxN: max Ns in a cons seq (def=5)              (s5)
12                     ## 24.opt.: maxH: max heterozyg. sites in cons seq (def=5)   (s5)
                      ## 25.opt.: ploidy: max alleles in cons seq (def=2;see docs) (s4,s5)
56                     ## 26.opt.: maxSNPs: (def=100). Paired (def=100,100)        (s7)
99,99                  ## 27.opt.: maxIndels: within-clust,across-clust (def. 3,99) (s3,s7)
                      ## 28.opt.: random number seed (def. 112233)             (s3,s6,s7)
                      ## 29.opt.: trim overhang left,right on final loci, def(0,0) (s7)
a,s                    ## 30.opt.: output formats: p,n,a,s,v,u,t,m,k,g,* (see docs) (s7)
2                      ## 31.opt.: maj. base call at depth>x<mindepth (def.x=mindepth) (s5)
                      ## 32.opt.: keep trimmed reads (def=0). Enter min length.    (s2)
9999999999             ## 33.opt.: max stack size (int), def= max(500,mean+2*SD)    (s3)
                      ## 34.opt.: minDerep: exclude dereps with <= N copies, def=1 (s3)
                      ## 35.opt.: use hierarchical clustering (def.=0, 1=yes)      (s6)
                      ## 36.opt.: repeat masking (def.=1='dust' method, 0=no)      (s3,s6)
                      ## 37.opt.: vsearch max threads per job (def.=6; see docs)   (s3,s6)
==== optional: list group/clade assignments below this line (see docs) ==================
```