

UMSL

Assessing the utility of Hyb-Seq for phylogenomics in rapid angiosperm radiations: a pilot study in *Burmeistera* bellflowers

Justin C. Bagley^{1,2}, Simon Uribe-Convers³, Monica Carlsen⁴, and Nathan Muchhal¹¹ University of Missouri-St. Louis, St. Louis, MO; ² Virginia Commonwealth University, Richmond, VA;³ Phylagen, San Francisco, CA; ⁴ Missouri Botanical Garden, St. Louis, MO

✉ bagleyj@umsl.edu
 @justincbagley
 justinbagley.org
 justincbagley



Introduction

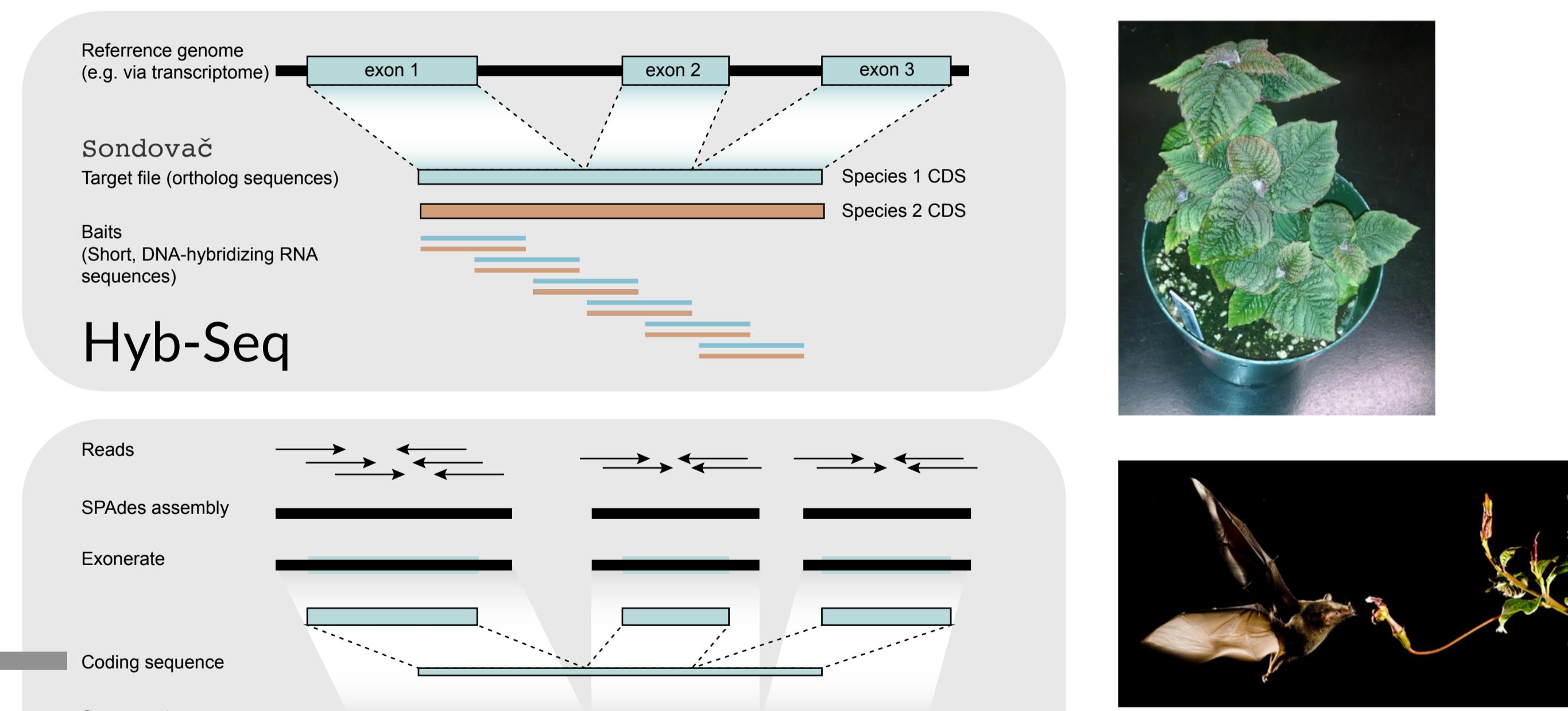
Genome reduction approaches to high-throughput sequencing (HTS) are revolutionizing our ability to generate large-scale phylogenomic datasets with DNA sequences from 100s to 1000s of loci (e.g. Cronn et al. 2012; Lemmon and Lemmon 2013). Targeted sequence capture has emerged as one of the most promising approaches, and Hyb-Seq (Weitmeier et al. 2014) is ideal for plant taxa, as it combines targeted exon sequence capture and 'genome-skimming' to obtain intergenic sequences flanking targeted exons (Straub et al. 2012; Johnson et al. 2014). Applications to angiosperms showcase the feasibility of this approach. However, rapid angiosperm radiations (<10–15 Myr) pose significant challenges for phylogenetic inference (e.g. ILS, phylogenetic noise) and the ability of nuclear loci from Hyb-Seq to overcome these challenges and resolve species trees for rapid, recent angiosperm radiations remains relatively poorly studied.

Objectives

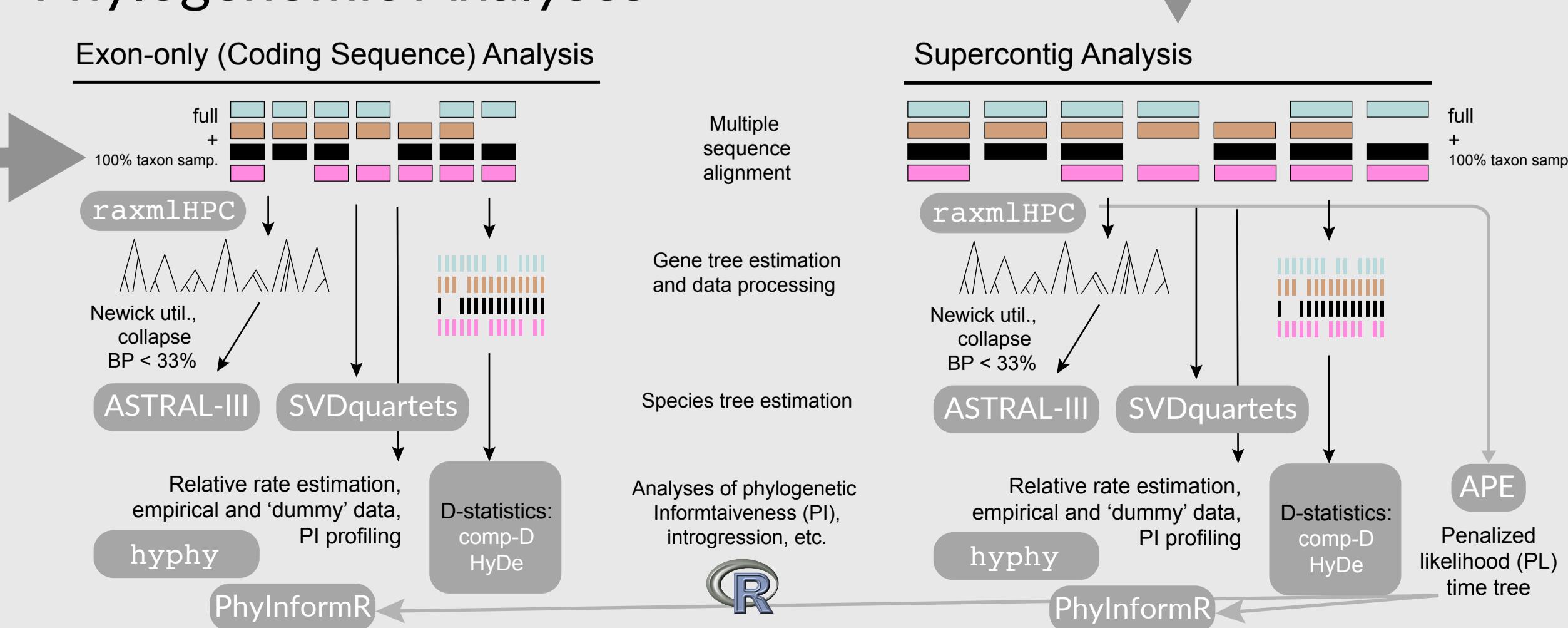
- To test utility of Hyb-Seq to resolve phylogeny & introgression patterns in rapid angiosperm radiations, using *Burmeistera* (Campanulaceae) as an exemplar of high Andean speciation rates.
- Use Hyb-Seq to target 745 low-copy nuclear loci from across the genome (1.35 Mb) and compare phylogenetic results from 'supercontig' versus exon-only loci.
- Evaluate potential effects of phylogenetic signal and noise, missing taxa, compositional bias, and introgression on our inferences.

Materials & methods

- Transcriptomes → 745 target nuclear loci and probes.
- DNA extraction, SE/PE Illumina sequencing (46 *Burmeistera*, 10 outgroups). → HybPiper
- raxmlHPC 'supermatrix' concatenation vs. coalescent ASTRAL-III and SVDquartets.
- Penalized likelihood (PL) time tree (APE chronos function, 2 hard calibrations).
- Phylogenetic informativeness (PI) profiles (hyphy, PhyInforR).
- D-statistic and D_{FOIL} (comp-D) plus HyDe introgression tests on sympatric species.

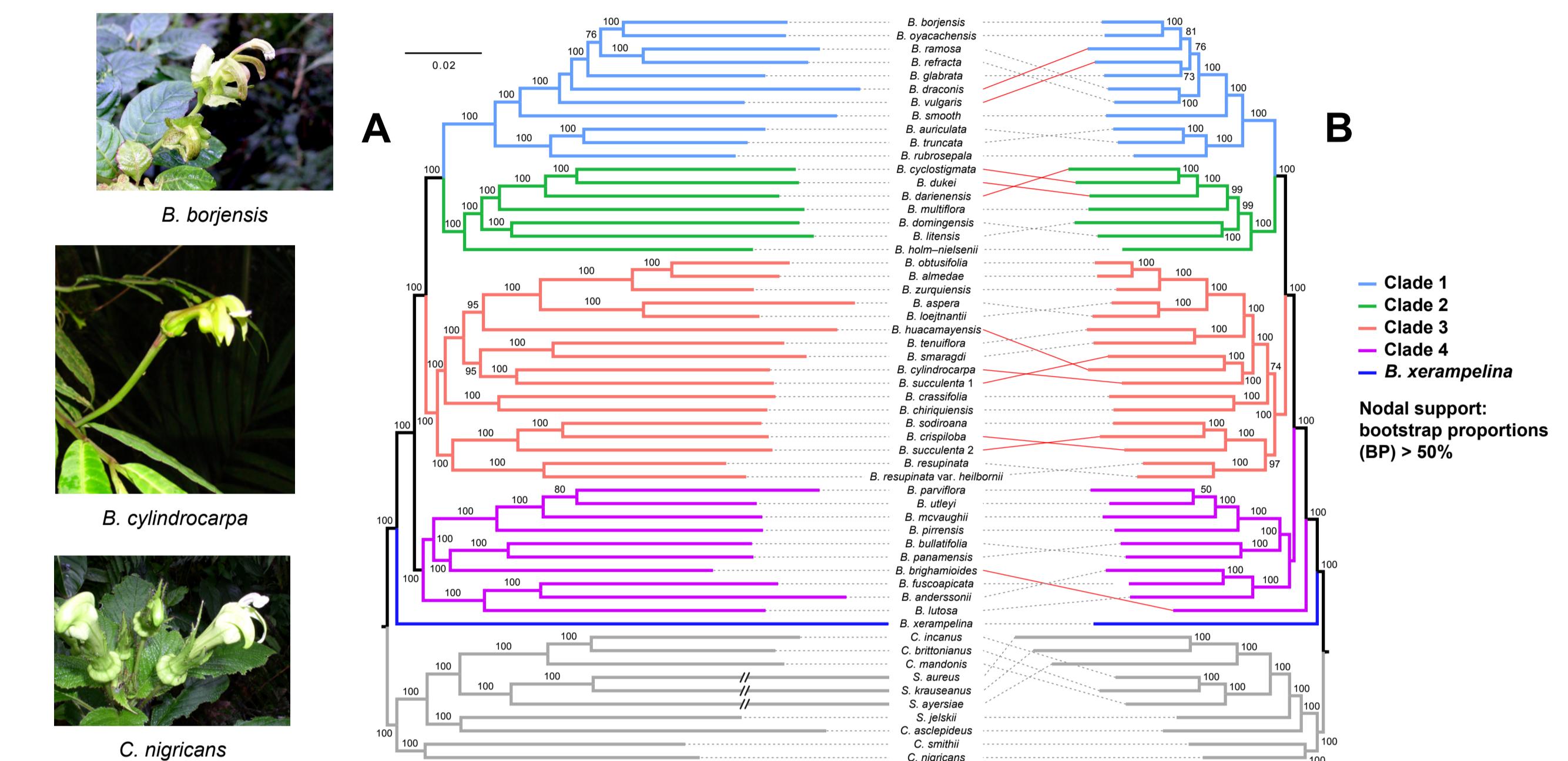


Phylogenomic Analyses



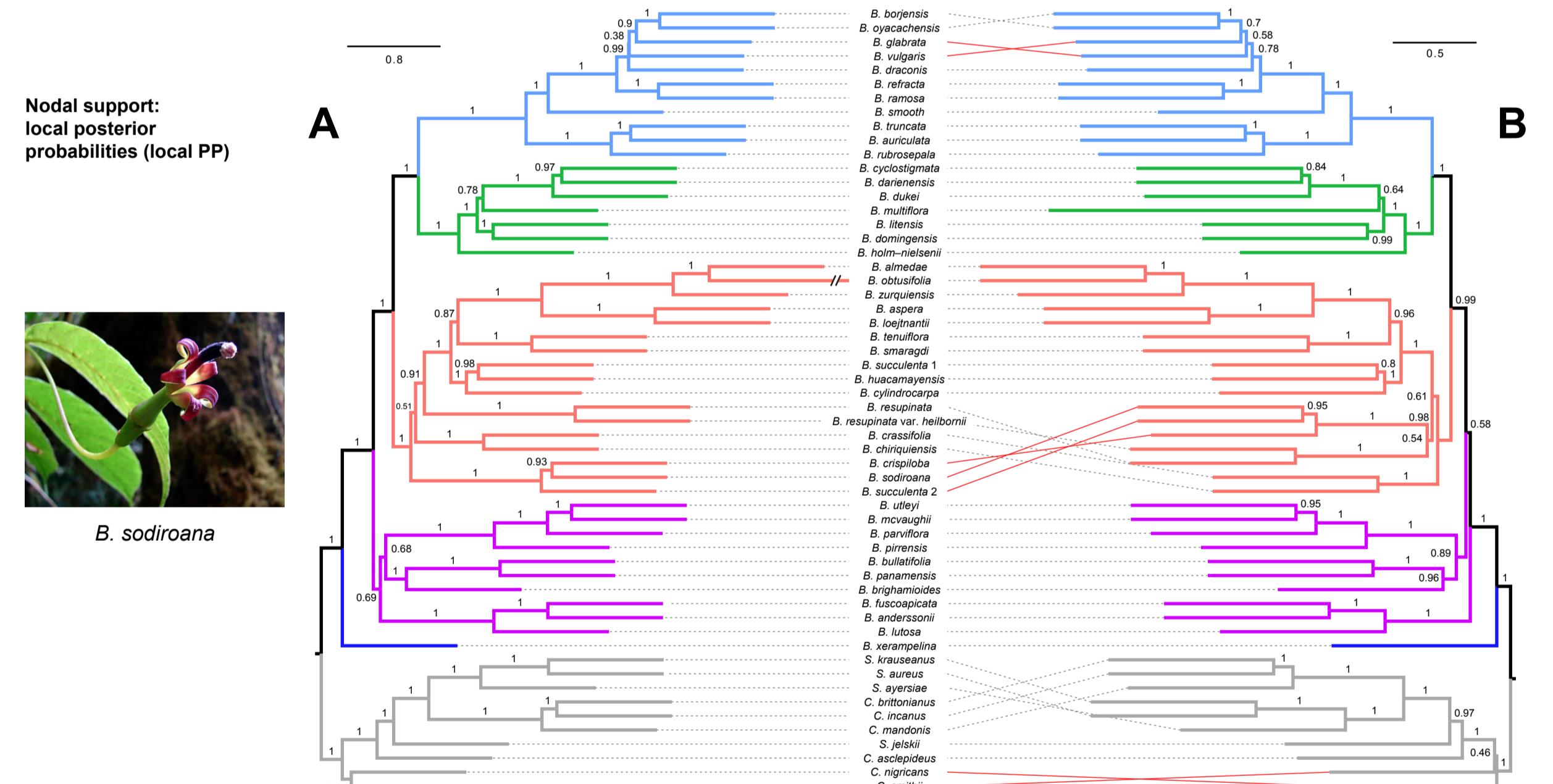
Results

Model-based supermatrix approach in RAxML



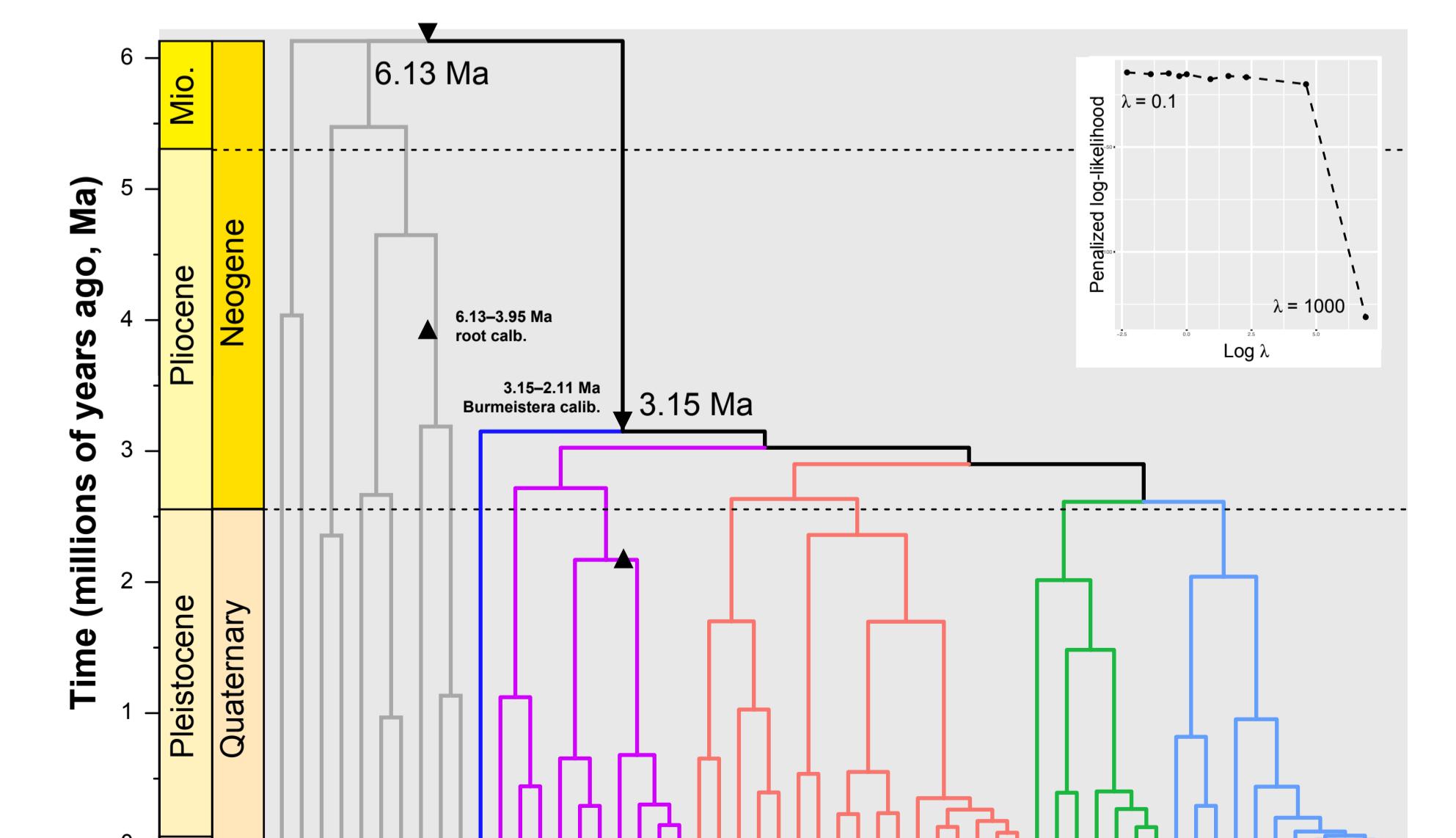
Tanglegram comparison of tree topologies from concatenated maximum-likelihood analyses of the Hyb-Seq (A) full 'supercontig' dataset (542 loci) and (B) full exon-only dataset (546 loci).

Coalescent-based gene tree summarization in ASTRAL-III



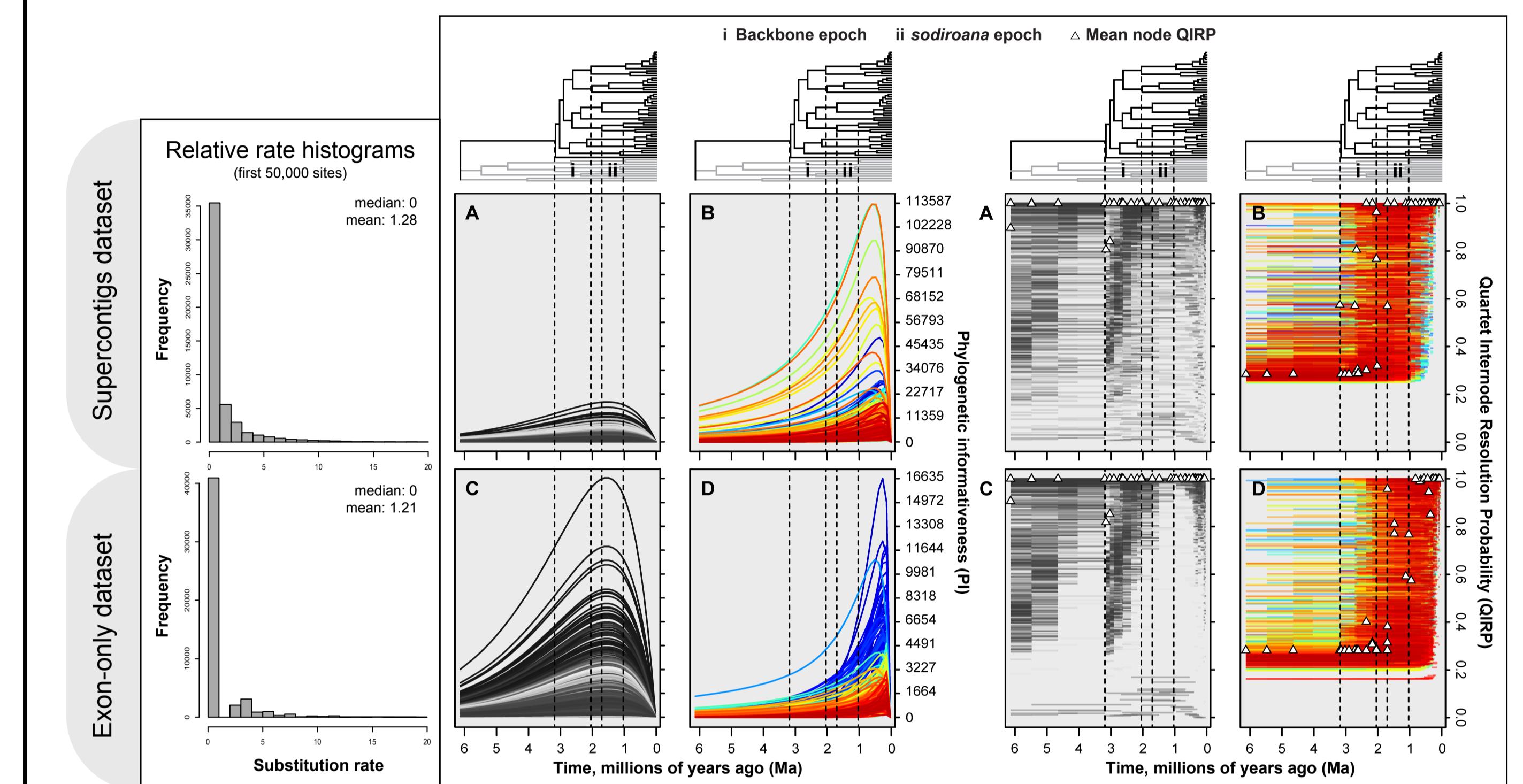
Tanglegram comparison of trees from coalescent-consistent analyses of the Hyb-Seq (A) full 'supercontig' dataset (542 loci) and (B) full exon-only dataset (546 loci) in ASTRAL-III.

Penalized likelihood (PL) chronogram of *Burmeistera* diversification



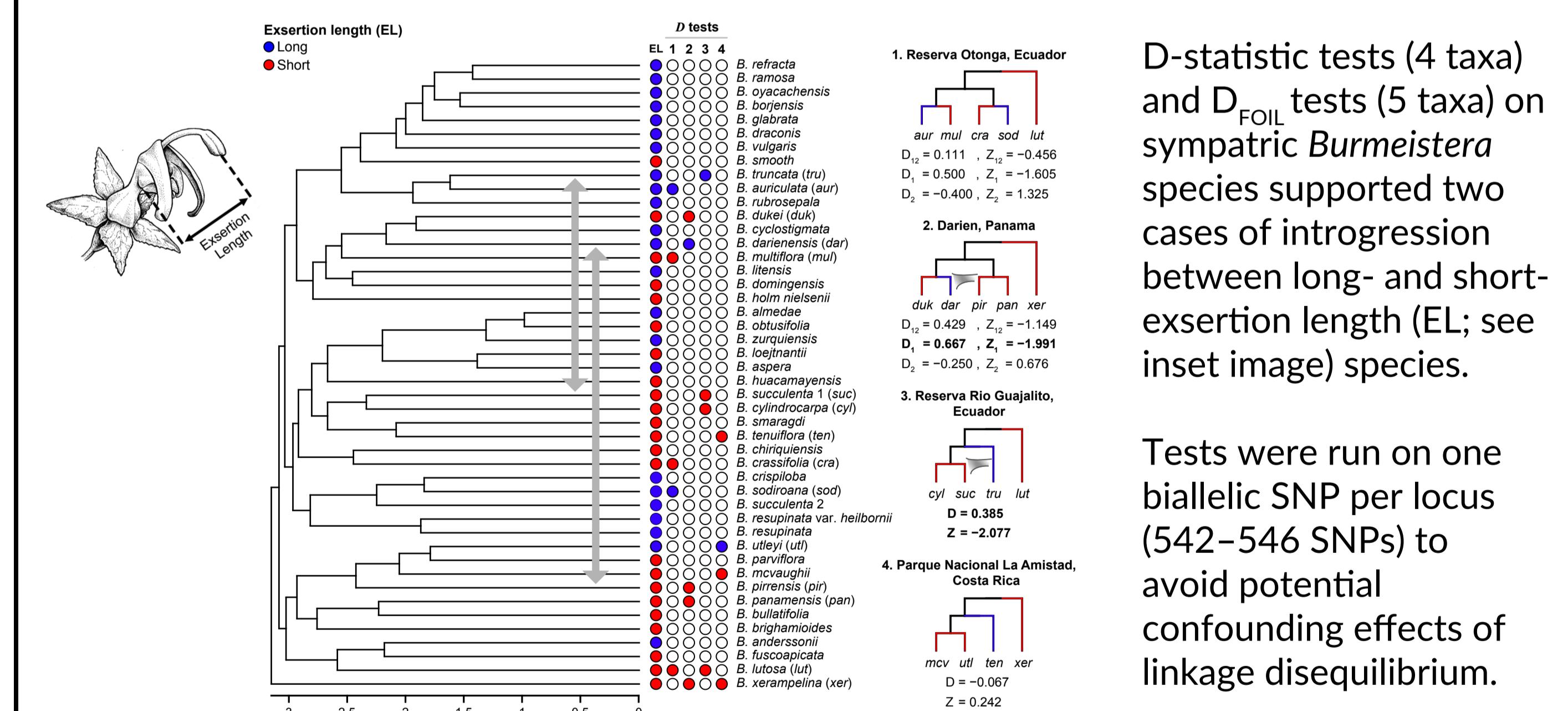
Time-calibrated phylogeny of *Burmeistera* and outgroups, derived from PL (Sanderson 2002) analysis in APE (2 sec. calibrations; Lagomarsino et al. 2016) on the best ML tree from our RAxML analysis of the full supercontig concatenated supermatrix.

Phylogenetic informativeness (PI) profiling



Compilation of site-rate patterns, phylogenetic informativeness (PI; Townsend 2007), and quartet internode resolution probabilities (QIRP; Townsend et al. 2012) for loci in the (top) full-supercontig and (bottom) full exon-only datasets. To evaluate PI and QIRP profiles against a null expectation of equal site rates, we generated 'dummy' rates set to 1.0 for each site mimicking non-gap sites in all loci, using MEGA-CC, then ran identical PI and QIRP analyses on the dummy data (grayscale lines, panels A, C). Empirical results for each locus (B, D) are shown with different colors, and colors match across rows (e.g. both B panels). Two main time epochs: (i) phylogeny backbone, and (ii) internode of 'sodiroana' clade that exhibited incongruence.

Testing for introgression among sympatric *Burmeistera*



D-statistic tests (4 taxa) and D_{FOIL} tests (5 taxa) on sympatric *Burmeistera* species supported two cases of introgression between long- and short-exsertion length (EL; see inset image) species.

Tests were run on one biallelic SNP per locus (542–546 SNPs) to avoid potential confounding effects of linkage disequilibrium.

Conclusions

Hyb-Seq loci supported *Burmeistera* as monophyletic, and gene/species trees yielded well-supported relationships among taxa as well as insights into introgression in sympatry. Results from supercontig vs. exon-only datasets, and variable taxon completeness (not shown), were largely congruent. Also, relationships were highly similar between concatenation & coalescent methods. Yet, short internodes received lower local PP support (ASTRAL-III) but high bootstrap support in concatenation runs (RAxML). Despite a majority of site rates near 1 and high power to resolve longer internodes towards tips of the tree (mean QIRP), PI was variable and showed lower, declining PI during rapid Plio-Pleistocene backbone divergences. Thus lower support appears driven in part by homoplasy and increased nucleotide saturation during rapid divergence. Concatenated or very long loci are most likely to maintain high phylogenetic signal over varying internodes. We recommend increased taxon sampling to break up short internodes in the future.



Take a picture to view the References and download the poster PDF.