Hybrid capture of 964 nuclear genes resolves evolutionary relationships in the mimosoid legumes and reveals the polytomous origins of a large pantropical radiation

American Journal of Botany

Hybrid capture of 964 nuclear genes resolves evolutionary relationships in the mimosoid legumes and reveals the polytomous origins of a large pantropical radiation

Erik J. M. Koenen[1,12], Catherine Kidner[2,3], Élvia R. de Souza[5], Marcelo F. Simon[6], João R. Iganci[7], James A. Nicholls[2,4], Gillian K. Brown[8], Luciano P. de Queiroz[5], Melissa Luckow[9], Gwilym P. Lewis[10], R. Toby Pennington[3,11], and Colin E. Hughes[1]

Manuscript received 15 May 2020; revision accepted 7 August 2020.

[1] Department of Systematic and Evolutionary Botany, University of Zurich, Zollikerstrasse 107, Zurich CH-8008, Switzerland

[2] School of Biological Sciences, University of Edinburgh, King's Buildings, Mayfield Road, Edinburgh, UK

[3] Royal Botanic Gardens Edinburgh, 20a Inverleith Row, Edinburgh EH3 5LR, UK

[4] Australian National Insect Collection, CSIRO, Clunies Ross St, Acton ACT 2601, Australia

[5] Departamento Ciências Biológicas, Universidade Estadual de Feira de Santana, Avenida Transnordestina s/n—Novo Horizonte, 44036-900, Feira de Santana, Brazil

[6] Embrapa Recursos Genéticos e Biotecnologia, Parque Estação Biológica (PqEB), Avenida W5 norte, 70770-917, Brasília, Brazil

[7] Instituto de Biologia, Universidade Federal de Pelotas, Campus Universitário Capão do Leão, Travessa André Dreyfus s/n, Capão do Leão 96010-900, Rio Grande do Sul, Brazil

[8] Queensland Herbarium, Brisbane Botanic Gardens, Mount Coot-tha, Mt Coot-tha Road, Toowong 4066, Queensland, Australia

[9] L.H. Bailey Hortorium, Department of Plant Biology, Cornell University, 412 Mann Library Building, Ithaca, New York 14853, USA

[10] Comparative Plant and Fungal Biology Department, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AE, UK

[11] Geography, University of Exeter, Amory Building, Rennes Drive, Exeter EX4 4RJ, UK

[12] Author for correspondence (e-mail: erik.koenen@systbot.uzh.ch)

Citation: Koenen, E. J. M., C. Kidner, É. R. de Souza, M. F. Simon, J. R. Iganci, J. A. Nicholls, G. K. Brown, L. P. de Queiroz, M. Luckow, G. P. Lewis, R. T. Pennington, and C. E. Hughes. 2020. Hybrid capture of 964 nuclear genes resolves evolutionary relationships in the mimosoid legumes and reveals the polytomous origins of a large pantropical radiation. *Am J Bot*. 107(12): 1710–1735.

doi:10.1002/ajb2.1568

PREMISE: Targeted enrichment methods facilitate sequencing of hundreds of nuclear loci to enhance phylogenetic resolution and elucidate why some parts of the "tree of life" are difficult (if not impossible) to resolve. The mimosoid legumes are a prominent pantropical clade of ~3300 species of woody angiosperms for which previous phylogenies have shown extensive lack of resolution, especially among the species-rich and taxonomically challenging ingoids.

METHODS: We generated transcriptomes to select low-copy nuclear genes, enrich these via hybrid capture for representative species of most mimosoid genera, and analyze the resulting data using de novo assembly and various phylogenomic tools for species tree inference. We also evaluate gene tree support and conflict for key internodes and use phylogenetic network analysis to investigate phylogenetic signal across the ingoids.

RESULTS: Our selection of 964 nuclear genes greatly improves phylogenetic resolution across the mimosoid phylogeny and shows that the ingoid clade can be resolved into several well-supported clades. However, nearly all loci show lack of phylogenetic signal for some of the deeper internodes within the ingoids.

CONCLUSIONS: Lack of resolution in the ingoid clade is most likely the result of hyperfast diversification, potentially causing a hard polytomy of six or seven lineages. The gene set for targeted sequencing presented here offers great potential to further enhance the phylogeny of mimosoids and the wider Caesalpinioideae with denser taxon sampling, to provide a framework for taxonomic reclassification, and to study the ingoid radiation.

KEY WORDS Caesalpinioideae; Fabaceae; Leguminosae; hard polytomy; hybrid capture; incomplete lineage sorting; ingoid clade; lack of phylogenetic signal; mimosoid clade; phylogenomics.

The field of molecular plant phylogenetics has had tremendous impacts on botanical studies and taxonomic classification, macroevolution and biogeography, ever since the pioneering studies of Chase et al. (1993) based on DNA sequence data. While those early studies used just a single locus, the plastid gene *rbcL*, modern studies often employ hundreds to several thousands of genes to infer phylogenetic relationships (e.g., Lee et al., 2011; Wen et al., 2013; Wickett et al., 2014; Yang et al., 2015; Zeng et al., 2017). Targeted enrichment via hybrid capture is now one of the most widely used methods for phylogenomics (e.g., Mandel et al., 2014; Weitemier et al., 2014; Nicholls et al., 2015; Sass et al., 2016; Johnson et al., 2018; Couvreur et al., 2019; Ojeda et al., 2019). Several methods for selecting genes (e.g., Johnson et al., 2018; Vatanparast et al., 2018) and assembling and analyzing the captured DNA sequence data have recently been developed. A number of pipelines are available to assemble gene matrices from the captured loci (Yang and Smith, 2014, with modifications described here; Johnson et al., 2016; Moore et al., 2017).

At the same time, it has become clear that many parts of the "tree of life" that are difficult to resolve are rife with conflicting gene tree histories, resulting in polytomies in species tree inference. Gene tree conflict can be caused by lack of phylogenetic signal (Salichos and Rokas, 2013; Shen et al., 2017), incomplete lineage sorting (ILS), intragenic recombination (Scornavacca and Galtier, 2017; Smith et al., 2020), hybridization and/or horizontal gene transfer, or combinations of these (Rokas et al., 2003; Salichos and Rokas, 2013; Suh et al., 2015; Copetti et al., 2017; Moore et al., 2017; Walker et al., 2018; Koenen et al., 2020a), and can be aggravated by gene tree estimation errors (Richards et al., 2018). Furthermore, ancient whole-genome duplications (WGDs), and gene duplications more generally, can complicate orthology assessment and contribute to the difficulties of resolving phylogenetic relationships (Koenen et al., 2020b). Detailed analyses of phylogenetic signal and conflict across a large number of gene trees can shed light on what factors are causing lack of resolution and determine whether they should be represented, in extreme cases, as candidate hard polytomies (Suh, 2016) (i.e., episodes of nearly instantaneous speciation of three or more lineages).

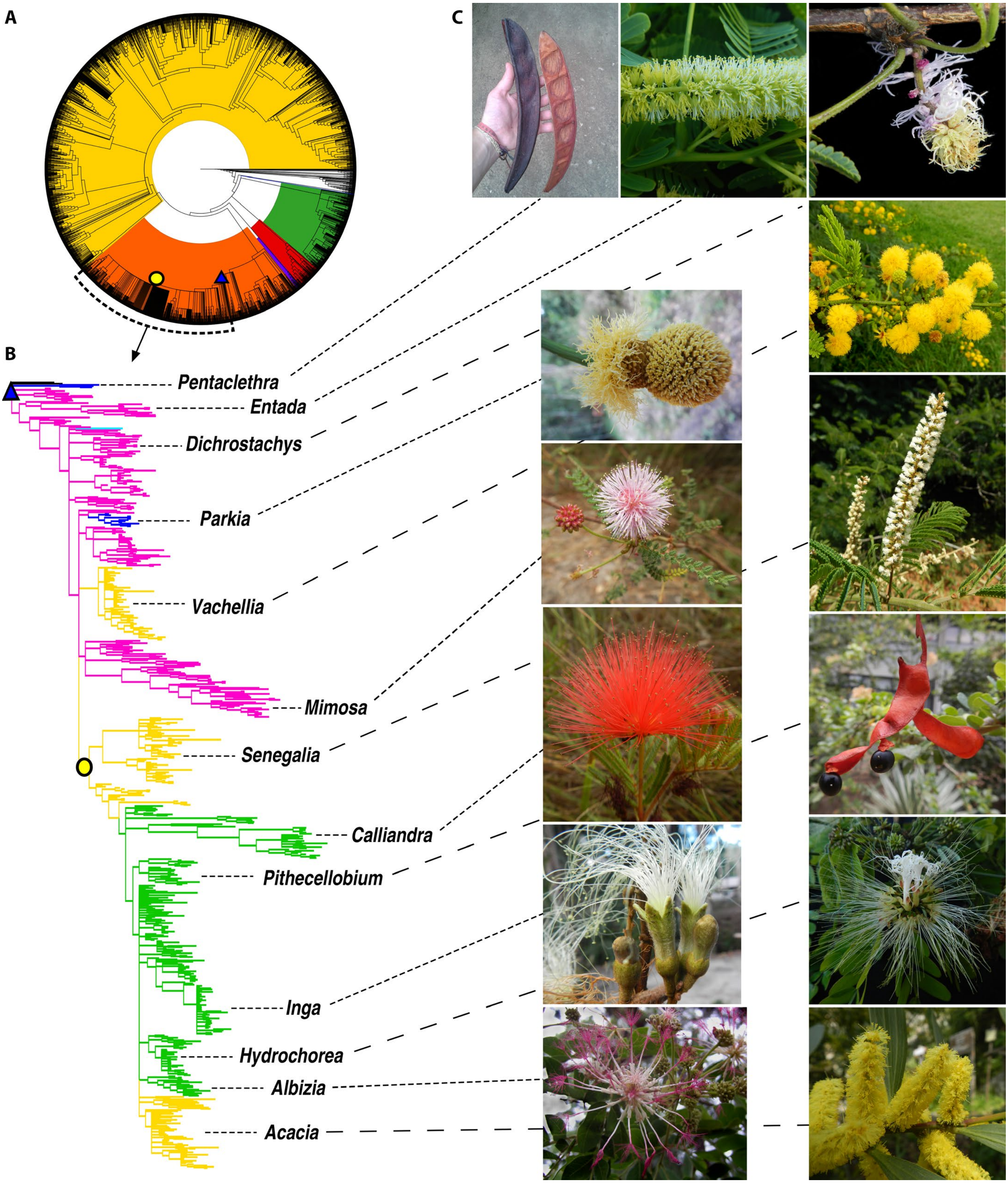
In the present study, we used hybrid capture to enrich a set of 964 putative low-copy genes, with the goal of inferring a robust generic backbone phylogeny for the mimosoid legumes, which include the large ingoid clade that has been particularly recalcitrant to phylogenetic resolution.

The mimosoid clade (LPWG, 2017), formerly subfamily Mimosoideae, comprises ~3300 species in ~87 genera of trees, shrubs, geoxyles, and lianas. Highly typical of the clade, though also found in other members of subfamily Caesalpinioideae, are bipinnate leaves (with few exceptions, most notably the once-pinnate leaves of the genus *Inga* and the phyllodes of *Acacia* s.s.; note that taxonomic authorities of all mimosoid genera are included in Table 2) that show extensive quantitative variation in size and numbers of leaflets and pinnae, and usually bear extrafloral nectaries on the petiole, rachis, and/or pinnae (Marazzi et al., 2019). Furthermore, many mimosoids have some form of armature (i.e., stipular spines, spinescent shoots, or prickles). Also highly characteristic of the clade is the diversity of inflorescence types composed of many small flowers in which the often colorful stamens are the most conspicuous floral whorl, and the whole inflorescence acts as the unit of pollinator attraction. Pollen characteristics are diverse and, notably, pollen is aggregated into tetrads or often in larger (up to 48-celled) polyads in many genera (Guinet, 1981). By contrast, floral morphology is relatively uniform across mimosoids, all species having radially symmetric flowers with valvate petal aestivation, showing mainly quantitative variation in sizes of organs, numbers of floral parts per whorl, and the degree of fusion within whorls.

Based on a few conspicuous floral characters, the clade has been divided into three large tribes (Elias, 1981; Lewis et al., 2005): Mimoseae Bronn (≤10 free stamens per flower), Acacieae Benth. (usually >30 free stamens, but sometimes slightly fused at the base), and Ingeae Benth. (usually >30 stamens partly fused into a tube), which have all been shown to be non-monophyletic (Fig. 1; Luckow et al., 2003, 2005; LPWG, 2013). The smaller tribe Parkieae (Wright & Arn.) Benth. is also non-monophyletic, and *Parkia* itself is nested within Mimoseae (Luckow et al., 2003), as is the monospecific tribe Mimozygantheae Burkart (Luckow et al., 2005). With a dysfunctional tribal classification, generic affinities have increasingly been referred to informally named clades (e.g., Hughes et al., 2003) and informal generic groups (Lewis et al., 2005) or alliances (Barneby and Grimes, 1996). Generic delimitation remains frustrated by what appears to be extensive morphological homoplasy and lack of phylogenetic resolution, and many genera remain poorly defined and have been suspected or shown to be non-monophyletic; examples include *Archidendron* (Brown et al., 2008), *Prosopis* (Catalano et al., 2008), *Abarema* (Iganci et al., 2016), *Stryphnodendron* (Simon et al., 2016), and *Zygia* (Ferm et al., 2019). This has been especially the case for tribe Ingeae, for which different authors have proposed starkly discordant generic systems (e.g., Nielsen, 1981; Lewis and Rico Arce, 2005; Barneby and Grimes, 1996; reviewed by Brown, 2008). In particular, the genus *Albizia* is poorly defined and its delimitation remains one of the most challenging taxonomic problems in the legume family. Indeed, *Albizia* is now considered the main "dustbin" genus, following the narrower circumscription of *Pithecellobium*, which was previously the dumping ground for difficult taxa (Nielsen, 1981; Barneby and Grimes, 1996; Brown, 2008).

Most species of mimosoids occur in the tropics, with major centers of diversity in Central and South America, Australia, Africa, and Madagascar. The ability of most mimosoids to fix atmospheric nitrogen through nodulation (Sprent, 2007) means they are important in tropical agroforestry, and their nitrogen- and protein-rich leaves and fruits are often used as animal fodder and green manure, among many other human uses including for timber, ornamentals, food, and hallucinogens (Lewis et al., 2005). Mimosoids occur in virtually every lowland tropical biome or vegetation type. They are abundant and diverse in evergreen rainforests, in particular in Africa and the Americas; form some of the most prominent groups in the woody flora of tropical grasslands in Brazil, Africa, and Australia; and dominate seasonally dry tropical forests and woodlands (SDTFs *sensu* Pennington et al., 2000, 2009; SDTFWs *sensu* Queiroz et al., 2017;

FIGURE 1. Mimosoid phylogeny, classification, and diversity. (A) Majority-rule bootstrap consensus tree from 1000 bootstrap replicates of the matK phylogeny from LPWG (2017), indicating the position of the mimosoid clade (crown node indicated by a blue triangle) within subfamily Caesalpinioideae (shaded orange) and showing that the ingoid clade (crown node indicated by a yellow circle) is the least resolved portion of the legume phylogeny. (B) Majority-rule Bayesian consensus tree for the mimosoid clade, extracted from the matK phylogeny of LPWG (2017), highlighting the non-monophyly of mimosoid tribes Parkieae (dark blue), Mimoseae (pink), Acacieae (yellow), and Ingeae (green). The monotypic Mimozygantheae (light blue) is nested in Mimoseae. (C) From left to right, top to bottom: pod valves of *Pentaclethra macrophylla* Benth., spicate inflorescence of *Entada chrysostachys* Drake, heteromorphic inflorescences of *Dichrostachys akataensis* Villiers and likewise for *Parkia bahiae* H.C.Hopkins, compound inflorescence of *Vachellia karroo* (Hayne) Banfi & Galasso, capitate inflorescence of *Mimosa blanchetii* Benth, spicate inflorescence of *Senegalia ataxacantha* (DC.) Kyal. & Boatwr., capitate inflorescence of *Calliandra fuscipila* Harms, dehisced fruit with seeds suspended on arillodia in *Pithecellobium diversifolium* Benth., flowers of *Inga subnuda* Salzm. Ex Benth., dimorphic inflorescence with enlarged central flowers of *Hydrochorea corymbosa* (Rich.) Barneby & J.W.Grimes and likewise for *Albizia grandibracteata* Taub., spicate inflorescences of *Acacia longifolia* Paxton. All photos by E. J. M. Koenen.


A
B
C
Pentaclethra
Entada
Dichrostachys
Parkia
Vachellia
Mimosa
Senegalia
Calliandra
Pithecellobium
Inga
Hydrochorea
Albizia
Acacia

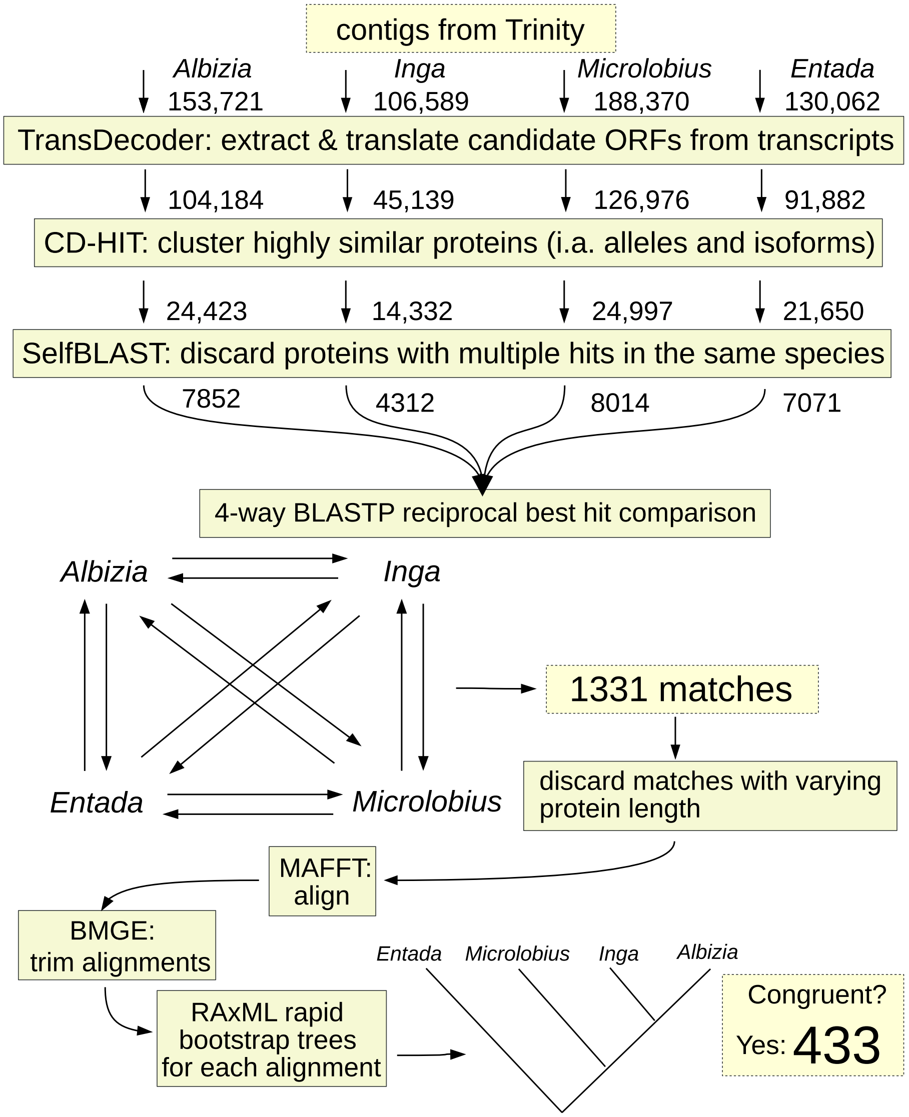

contigs from Trinity
Albizia 153,721
Inga 106,589
Microlobius 188,370
Entada 130,062
TransDecoder: extract & translate candidate ORFs from transcripts
104,184
45,139
126,976
91,882
CD-HIT: cluster highly similar proteins (i.a. alleles and isoforms)
24,423
14,332
24,997
21,650
SelfBLAST: discard proteins with multiple hits in the same species
7852
4312
8014
7071
4-way BLASTP reciprocal best hit comparison
Albizia
Inga
Entada
Microlobius
1331 matches
discard matches with varying protein length
MAFFT: align
BMGE: trim alignments
RAxML rapid bootstrap trees for each alignment
Entada
Microlobius
Inga
Albizia
Congruent?
Yes: 433

FIGURE 2. Target gene selection workflow, indicating the number of sequences and loci retained at each step for the RBH4 gene set (see text).

or the succulent biome *sensu* Schrire et al., 2005; Gagnon et al., 2019; Ringelberg et al., 2020) in Mexico, Central America, the Caribbean, Northeast Brazil, the Horn of Africa, and Madagascar. Because of this prominence across tropical lowland biomes, the mimosoid clade offers an excellent study system to investigate adaptation along the gradient from ever-wet to seasonally dry and arid tropical climates, as well as the extent of phylogenetic biome conservatism vs. biome shifting, which are the focus of forthcoming studies. However, a well-resolved species tree for comparative analyses is lacking. Lack of resolution is particularly stark in the large Ingeae + Acacieae p.p. clade (Luckow et al., 2003; Miller et al., 2003; Brown et al., 2008; Bouchenak-Khelladi et al., 2010; LPWG, 2017), hereafter referred to as the ingoid clade (Fig. 1A, B). This clade includes some 2000 species in ~38 genera, but the relationships among these genera are uncertain because of lack of phylogenetic resolution, even though all were sampled in the most densely sampled legume phylogeny to date based on the chloroplast gene *matK* (Fig. 1; LPWG, 2017). In fact, this clade appears to represent the least resolved part of the whole legume *matK* phylogeny (Fig. 1A).

Here, we present a complete phylogenomics project using hybrid capture, from generating transcriptome data and selecting targeted genes to assembling and analyzing the captured DNA sequence data. The targeted genes were selected using a custom pipeline, which has recently also been used to select loci for other groups (Couvreur et al., 2019; Ojeda et al., 2019) and which is potentially useful across all taxonomic groups. Using these genome-scale data, we generated a robust generic backbone phylogeny for the mimosoid clade, which forms the foundation for expanded taxon sampling to address biogeographical and macroecological questions. Here, however, we focus especially on the large, poorly resolved ingoid clade, to try to understand why inferred relationships in this clade have been so contentious. To address this, we also quantified conflicting signals across gene trees and used phylogenetic network approaches to assess whether the evolution of the ingoid clade is tree-like or polytomous.

MATERIALS AND METHODS

The workflow for selecting targeted genes from transcriptome data is presented in Figure 2. The workflow for assembly and phylogenetic analysis of the captured DNA sequence data is presented in Figure 3.

RNAseq to generate genomic resources

With no fully sequenced genome for mimosoids available when we started this study, we generated transcriptome data for four mimosoid genera to select nuclear markers for targeted enrichment. For the species *Albizia julibrissin* Durazz., *Entada abyssinica* Steud. ex A.Rich., and *Microlobius foetidus* (Jacq.) M.Sousa & G.Andrade, seedlings were grown at the Botanic Garden of the University of Zurich, and RNA was extracted from young leaves and shoot tips, as well as roots (*A. julibrissin*), using the RNeasy Plant Mini kit (Qiagen, Venlo, The Netherlands). Libraries for sequencing were produced using the TruSeq RNA Library Prep kit (Illumina, San Diego, California, USA) and sequenced 3-plex on an Illumina HiSeq-2000 sequencer, at the Functional Genomics Center in Zurich. Raw data were cleaned with prinseq-lite (Schmieder and Edwards, 2011), and transcriptomes assembled using Trinity (Grabherr et al., 2011; Haas et al., 2013) with default settings. In addition, transcriptome data for three species of *Inga* were generated at the Royal Botanic Garden Edinburgh using similar sequencing and assembly methods (described in Nicholls et al., 2015). Since the comparisons among

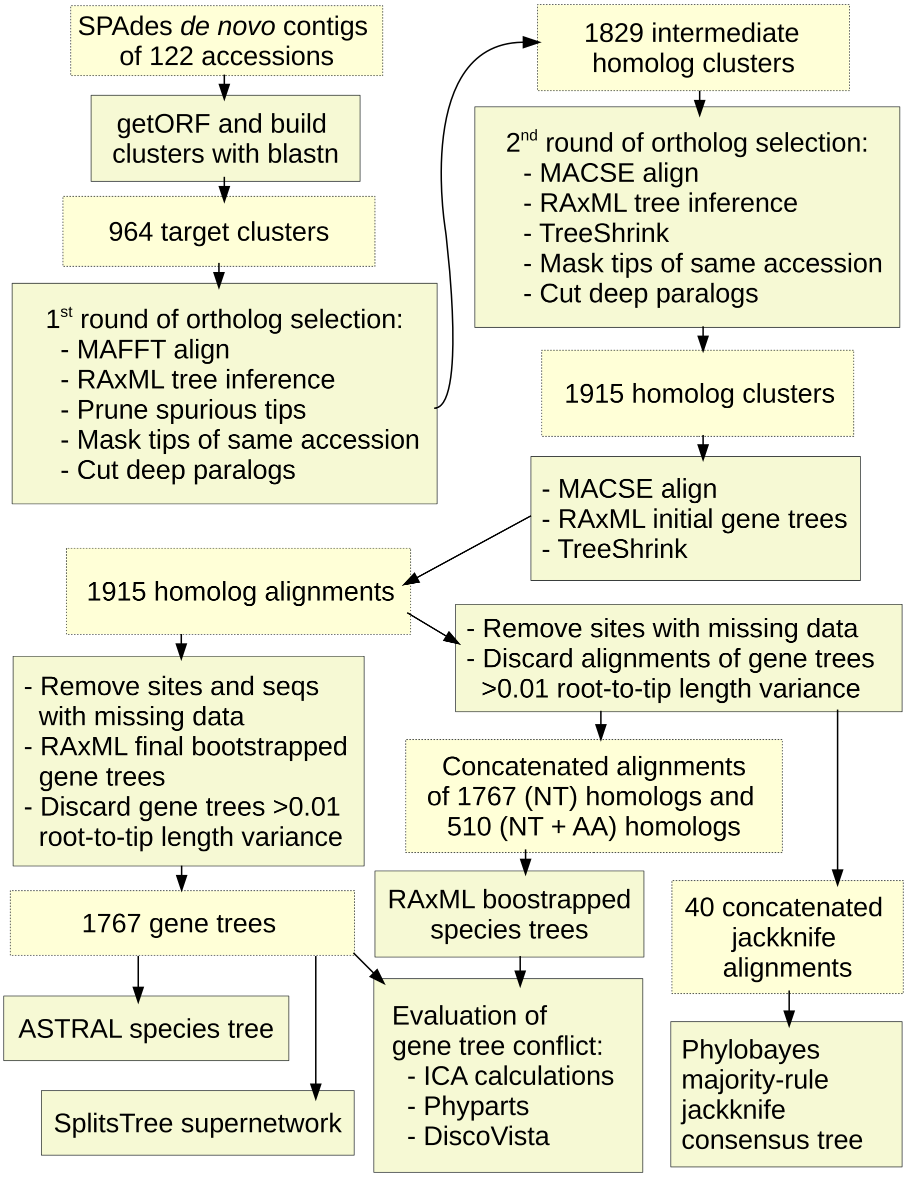

SPAdes de novo contigs of 122 accessions
getORF and build clusters with blastn
964 target clusters
1st round of ortholog selection:
- MAFFT align
- RAxML tree inference
- Prune spurious tips
- Mask tips of same accession
- Cut deep paralogs
1829 intermediate homolog clusters
2nd round of ortholog selection:
- MACSE align
- RAxML tree inference
- TreeShrink
- Mask tips of same accession
- Cut deep paralogs
1915 homolog clusters
- MACSE align
- RAxML initial gene trees
- TreeShrink
1915 homolog alignments
- Remove sites with missing data
- Discard alignments of gene trees >0.01 root-to-tip length variance
- Remove sites and seqs with missing data
- RAxML final bootstrapped gene trees
- Discard gene trees >0.01 root-to-tip length variance
Concatenated alignments of 1767 (NT) homologs and 510 (NT + AA) homologs
1767 gene trees
RAxML boostrapped species trees
40 concatenated jackknife alignments
ASTRAL species tree
Evaluation of gene tree conflict:
- ICA calculations
- Phyparts
- DiscoVista
Phylobayes majority-rule jackknife consensus tree
SplitsTree supernetwork

FIGURE 3. Workflow for phylogenetic ortholog selection and gene tree and species tree analyses.

transcriptomes (see below) are effectively carried out at genus level, we generated a more comprehensive transcriptome for the genus *Inga* by combining the separate assemblies for the three *Inga* species into a nonredundant set of transcripts. This was done by running BLAST searches of the largest *Inga* transcriptome assembly against the second-largest assembly and adding all transcripts without a significant hit (e-value cutoff 1e-10) in the latter. This procedure was then repeated for the third species.

Selecting putative single-copy genes

From the four transcriptome data sets, putative single- or low-copy nuclear genes were selected, using a procedure inspired by Wu et al. (2006) (Fig. 2). This procedure was recently used by Couvreur et al. (2019) and Ojeda et al. (2019), but because it was first designed for the mimosoid bait set, it is described in more detail here. First, for each of the four transcriptome data sets, TransDecoder (https://github.com/TransDecoder/TransDecoder) was used to predict open reading frames (ORFs) and translate those to protein sequences, using default settings. Highly similar proteins were removed to reduce redundancy (i.e., keeping only one protein sequence per gene and removing multiple alleles and isoforms) with CD-HIT (Li and Godzik, 2006). This was repeated with four cutoff values (90%, 95%, 97%, and 99% identity), to avoid either clustering paralogous sequences with relatively low divergence or keeping alleles and isoforms with relatively high divergence. This means that the following steps were each repeated four times, and from each repetition only the putative orthologs that were more divergent among and less divergent within taxa were kept. In other words, for each repetition, sequences with higher identity among taxa than the cutoff values were removed. For each transcriptome, we performed a BLAST search of the CD-HIT output against itself ("self-BLAST") with an e-value cutoff of 1e-10, and sequences with multiple hits within the same transcriptome were removed to eliminate gene families. Next, a reciprocal best hit (RBH) algorithm was implemented in a custom Python script (available from https://github.com/erikkoenen/mimobaits/), to compare the four transcriptome data sets after removing redundancy and gene families. This is an extension of the RBH triangulation method of Wu et al. (2006), where a set of four sequences are considered as a putative ortholog if all possible pairwise reciprocal BLAST searches among the four transcriptomes yield the same RBH (Fig. 2). This works as follows: first, we take the first sequence of the transcriptome that we want to design the baits from (in our case *Albizia*) and run a BLAST search against one of the other transcriptomes; the best hit from the second transcriptome is then used as a query for a BLAST search against the first transcriptome, and when the original sequence that we started with is recovered as the best hit, this is considered an RBH. This is repeated for all combinations of transcriptomes by taking the sequence of the previous RBH and running a BLAST search against another transcriptome. This procedure was repeated for all sequences from the first transcriptome and then for sequences from the four transcriptomes that gave an RBH across all pairwise BLAST searches, and these were then written to separate FASTA files for each putative ortholog. Putative orthologs in which sequence length varied by >5% were discarded as an additional quality-control step. From the resulting FASTA files, we also performed a phylogenetic congruence test similar to that of Wu et al. (2006). Orthologs were aligned with MAFFT using the G-INS-i algorithm (Katoh et al., 2005), alignments trimmed with BMGE with default settings (Criscuolo and Gribaldo, 2010) and rapid bootstrap analyses carried out with RAxML under the PROTCATLGF model (Stamatakis, 2014). If the resulting 95% bootstrap consensus topology was incongruent with previously known, and well established, relationships among the four taxa (Fig. 2), the putative ortholog was discarded. After running these procedures for each of the four different CD-HIT cutoff values, the resulting ortholog sets were combined as the “RBH4 set.”

Additionally, an “RBH3 set” was generated by comparing just the three largest transcriptomes (*Albizia*, *Entada*, and *Microlobius*) but omitting the phylogenetic congruence test (because a minimum of four taxa are needed to infer a phylogeny). A third set of putative orthologs was generated by running RBH comparisons among the two largest transcriptomes (*Albizia* and *Microlobius*), to sets of genes found by De Smet et al. (2013) to be strictly or mostly single-copy across 20 angiosperm genomes (using sequences of *Arabidopsis thaliana* (L.) Heynh.). This third set is split into two subsets referred to as “SSC” (strictly single-copy) and “MSC” (mostly single-copy), following terminology from De Smet et al. (2013).

Bait design

For bait design, the sequences of the *Albizia julibrissin* transcriptome were used, because the genus *Albizia* and allies are the focus of an ongoing project in Zurich, and this will increase successful capture for these taxa. Intron-exon boundaries were predicted for all transcripts in the four ortholog sets (RBH4, RBH3, SSC, and MSC), by running BLAST searches against a custom database combining the *Arabidopsis thaliana* (Lamesch et al., 2011), *Medicago truncatula* Gaertn. (Young et al., 2011), and *Glycine max* (L.) Merr. (Schmutz et al., 2010) genomes. For the genome database, gene models including introns were used, and the coordinates to which our transcripts aligned were used to partition sequences for each predicted exon to avoid designing baits spanning intron-exon boundaries. This step is not essential but is likely to increase the efficiency of the capture. In addition to coding sequences, we also included 120 bp of the 3′-UTR and 240 bp of the 5′-UTR, but sequences obtained for these regions are not analyzed further here. Furthermore, additional target genes were added that included functionally interesting genes and genes targeted for separate studies in *Inga* (Nicholls et al., 2015), but again, none of these genes are analyzed here, as we focus on the low-copy loci selected for phylogenetic analysis. Final bait design was carried out by Mycroarray (now Arbor Biosciences, Ann Arbor, Michigan, USA), with 3× tiling, and RNA baits were synthesized as part of the myBaits Custom Target Capture kit.

DNA extraction, library preparation, hybrid capture, and sequencing

We extracted DNA for 122 accessions, partly from tissue preserved in silica gel and partly from herbarium specimens, representing 75 of the ~87 currently recognized mimosoid genera and six closely related genera of non-mimosoid Caesalpinioideae (voucher details in Appendix 1), using the DNeasy Plant Mini Kit (Qiagen). Sequencing libraries were prepared with the NEBNext Ultra DNA Library Prep kit for Illumina (New England Biolabs, Ipswich, Massachusets, USA), in combination with the NEBNext Multiplex Oligos for Illumina (both single and dual index kits). Libraries were quantified using qPCR and pooled prior to hybrid capture. Pools consisted of 8–21 libraries based on approximate evolutionary distances to the species from which the baits were designed (thus, species of *Albizia* were pooled together, species of closely related genera were pooled together in another pool, and species from more distantly related genera were pooled in yet another pool, etc.). This was done to avoid more distantly related accessions being underrepresented in the postcapture pools because it is expected that DNA molecules with higher sequence similarity hybridize more efficiently. The different pools were then enriched for the targeted regions in separate reactions with the myBaits Custom Target Capture kit. Enriched pools were quantified and pooled into a single library that was sequenced on Illumina HiSeq 2000 at the Functional Genomics Center in Zurich.

Assembly of sequence data and aligned matrices

After demultiplexing, raw reads were processed with Trimmomatic (Bolger et al., 2014) to remove adapter sequence artifacts and trim or remove low-quality reads (using the settings MAXINFO:40:0.1 LEADING:20 TRAILING:20), and PEAR (Zhang et al., 2013) to merge overlapping read pairs (after removing adapter artifacts but before trimming). Resulting fastq files of quality-filtered merged, paired, and unpaired reads were used in a de novo assembly for each accession using the SPAdes assembler (Bankevich et al., 2012). From the resulting scaffolds, we extracted all ORFs of ≥300 bp long between two stop codons with getorf (using the option -find 2) from the Emboss software suite. We reduced redundancy in the set of ORFs found for each accession with cd-hit, using an identity cutoff of 0.99. For all ORFs from each accession, a BLAST search was carried out against the target sequences and for each target a multifasta file was created. Each ORF for each accession was added to the target multifasta file for which it received the best BLAST hit under an e-value cutoff of 1e-10, resulting in multifasta files for each target with potentially multiple sequences per accession included, which we refer to hereafter as “clusters.”

Numbers of reads on target were estimated by mapping the untrimmed reads to the bait sequences with BLAT (Kent, 2002), using a minimum sequence identity threshold of 70%. Numbers of recovered loci were estimated with BLASTX, using protein sequences for the 964 targeted genes as the database and the SPAdes contigs as the query sequences, with an e-value cutoff of 1e-10.

Using the fasta_to_tree.py script of Yang and Smith (2014), each cluster was aligned with MAFFT, sites with excessive missing data were removed (with a minimum column occupancy of 0.3), and a tree was inferred for each cluster with RAxML. We used other scripts of Yang and Smith (2014) to trim outlier long tips (with relative and absolute cutoffs of 0.1 and 0.3, respectively), mask monophyletic and paraphyletic clusters belonging to the same taxon, and cut deep paralogs (cutting internal branches >0.3 and keeping subtrees of ≥25 accessions). From the resulting trimmed subtrees, new multifasta files were created for a second round of tree inference, trimming, and masking. However, for the second round we used MACSE (Ranwez et al., 2018), instead of MAFFT, to obtain more accurate alignments and TreeShrink (with quantile of trees to remove set to $q = 0.1$; Mai and Mirarab, 2018) to trim tips, instead of relative and absolute cutoffs. Finally, after cutting deep paralogs again, we extracted all non-overlapping subclusters with ≥25 accessions using the maximum inclusion (MI) method of Yang and Smith (2014).

Besides analyzing the targeted nuclear genes, we also extracted off-target reads with a BLAST hit against a reference set of chloroplast genomes of *Inga leiocalycina* Benth. (Dugas et al., 2015; GenBank accession KT428296), *Leucaena trichandra* Urb. (Dugas et al., 2015; GenBank accession KT428297), and *Erythrophleum fordii* Oliv. (Huang et al., 2018; GenBank accession MG644609), assembled chloroplast sequences for all accessions, and extracted the coding sequences gene by gene using a custom Python script with BLAST searches, confirming that sequence data for the chloroplast genome can be efficiently extracted and analyzed from off-target reads in hybrid capture experiments as shown by Weitemier et al. (2014). The *clpP* gene was discarded because it shows accelerated evolution (Williams et al., 2015; Dugas et al., 2015) and yields a tree that strongly conflicts with those inferred using the other chloroplast genes. The *accD* gene has been lost from the chloroplast genome in several papilionoids, is highly variable in others (Magee et al., 2010), and is difficult to align across mimosoids, so we also removed this gene for phylogenetic analysis. The remaining 72 plastid genes were aligned with MACSE and concatenated with the pxcat program of the phyx package (Brown et al., 2017).

Phylogenetics

The MI subclusters were aligned with MACSE (Ranwez et al., 2018) to yield codon alignments, codons with >95% missing data were removed using pxclsq from the phyx package, and initial gene trees were inferred with RAxML. Using TreeShrink with a relatively high quantile cutoff ($q = 0.25$), we removed outlier long tips, to ensure a low error rate in the alignments. The drawback of this is that outgroup taxa and other taxa outside the "core mimosoids" (as defined in Appendix 2) also get pruned relatively frequently from these loci. Given that the mimosoid phylogeny in those parts is already well characterized from previous work (Luckow et al., 2003; Bouchenak-Khelladi et al., 2010), this is unlikely to be problematic.

For gene tree inference, codons with ambiguous or missing sites for >75% of accessions were removed from the alignments, after which sequences <300 bp and at the same time occupying <50% of the total aligned length were removed. Gene trees were inferred with RAxML under the GTRGAMMA model with 200 rapid bootstrap replicates. Using pxlstr from the phyx package, root-to-tip variance was estimated to discover outlier gene trees that might have originated from poor orthology inference or alignment artifacts. After inspecting a subset of gene trees, we decided to discard all those with a root-to-tip variance >0.01. Gene trees were used to calculate internode certainty all (ICA) values using RAxML (Kobert et al., 2016), for species tree analysis using ASTRAL-III (Zhang et al., 2018), and for phylogenetic supernetwork analysis. ASTRAL-III analyses were done on the best maximum likelihood (ML) gene trees, and subsets of gene trees with >25% or >50% of the accessions present to check if the analyses are sensitive to including gene trees with a lot of missing data. We also ran the polytomy test in ASTRAL-III (Sayyari and Mirarab, 2018) to see for which nodes a polytomy null model could not be rejected.

Another way to analyze conflicting signals across gene trees is to infer a filtered Z-closure supernetwork (Whitfield et al., 2008). For deciding which splits to consider, we used the "mintrees" parameter, which allowed us to infer multiple networks, including rarer splits or only fewer, more commonly observed, and therefore better-supported splits. For phylogenetic supernetwork analysis, we pruned all gene trees to a selection of taxa from the ingoid clade representing its main lineages that were present in high proportions of the gene trees, yielding a total of 878 gene trees in which more than half of the selected ingoid taxa were represented (≥6 out of 11). All pruned gene trees with less than half of the selected taxa present were discarded. Phylogenetic supernetworks were constructed using Splitstree version 4 (Huson, 1998), using different cutoffs for the MinTrees setting, representing 2.5%, 5%, 7.5%, and 10% of the total number of gene trees.

For phylogenetic analyses of the concatenated alignments, codons with missing data for >90% of the accessions were removed. Both nucleotide and translated peptide alignments of loci with more than half of the taxa present were concatenated with pxcat of the phyx package. Loci for which the gene tree had a root-to-tip variance >0.01 were discarded prior to concatenation. Concatenated alignments, including the chloroplast alignment, were analyzed with RAxML, using the GTRCAT model for DNA sequences and the PROTGAMMALG4X model for protein sequences (Le et al., 2008), running 200 rapid bootstrap replicates for each. In addition, we carried out a gene jackknifing analysis with Phylobayes (Lartillot et al., 2013) using the CATGTR model, by dividing the loci randomly over four relatively equally sized concatenated protein sequence alignments with 10 replicates, running a total of 40 analyses for 1000 cycles. For faster convergence, the ML estimate of the concatenated analysis in RAxML was provided as a starting tree for the chains. The first 500 cycles of each replicate were discarded as burn-in prior to summarizing a majority-rule consensus tree over all replicates.

Visualizing gene tree discordance

Numbers of supporting and conflicting bipartitions for each node were extracted from gene trees with more than half the accessions present, using Phyparts (Smith et al., 2015). For this, gene trees were first rooted using pxrr from the phyx package, with a list of outgroup taxa outside the "core mimosoids" ranked by their relative divergence from Ingeae/Acacieae. Additionally, we visualized proportions of supporting and rejecting gene trees for selected clades with DiscoVista (Sayyari et al., 2018), from the same set of gene trees for which at least half the accessions are present. Clades for these visualizations were selected based on results from the ASTRAL polytomy test (described above).

RESULTS

Transcriptome sequencing, gene selection, and bait design

Transcriptome sequencing statistics are in Table 1; data are available on the National Center for Biotechnology Information (NCBI) databases under BioProjects PRJEB8722 and PRJNA574148; FASTQ files with raw read data are available on the Sequence Read Archive (SRA), under accession nos. SRX6901075 (*Albizia julibrissin*), SRX6901076 (*Entada abyssinica*), ERX719658 (*Inga spectabilis* (Vahl) Willd.), ERX719681 (*Inga umbellifera* (Vahl) Steud. ex DC.), ERX719690 (*Inga sapindoides* Willd.), and SRX6901077 (*Microlobius foetidus*); assembled transcripts are available on Dryad (for the *Inga* spp., https://doi.org/10.5061/dryad.r9c12) and through the Transcriptome Shotgun Assembly (TSA) database, accession nos. GHWM00000000 (*Albizia julibrissin*), GHWN00000000 (*Entada abyssinica*), and GHWO00000000 (*Microlobius foetidus*). Results from the gene selection procedure for the RBH4 gene set

TABLE 1. Transcriptome sequencing and assembly.

Taxon	Total no. of reads	Quality filtered reads	Trinity contigs	Predicted ORFs
Albizia julibrissin	65,129,217	Left: 60,128,377 Right: 57,345,882	153,721	104,184
Entada abyssinica	65,006,875	Left: 59,821,838 Right: 56,882,422	130,062	91,882
Microlobius foetidus	97,515,912	Left: 89,669,576 Right: 85,024,338	188,370	126,976
Inga (three species, nonredundant set)	NA[a]	NA[a]	106,589	45,139

[a]See Nicholls et al. (2015) for sequencing results of the three *Inga* transcriptomes used here.

are summarized in Figure 2. After running the pipeline with four different similarity cutoffs in CD-HIT, we found 433 RBH4 and 334 RBH3 target genes. We recovered 320 MSC and 19 SSC genes, of which 134 and eight genes, respectively, were already included in the RBH sets. Combining all gene sets we obtained a total of 964 low-copy nuclear genes for enrichment. The complete coding sequences from the *Albizia julibrissin* transcriptome for these targeted genes are in Appendix S1. The bait design included 24,856 probes at 3× tiling. Target sequences and baits are also available at https://github.com/erikkoenen/mimobaits/.

Targeted sequencing and data assembly

Sequencing and de novo assembly statistics for targeted sequencing for all accessions are presented in Appendix 1, including full species names with taxonomic authorities, and sequence reads have been submitted to the European Nucleotide Archive (study no. PRJEB38138). Accessions were enriched and sequenced in three separate batches, with different levels of multiplexing, which explains some of the variation observed in numbers of total reads and reads on target. Total reads per accession varied from 1,360,502 to 70,271,424. For the largest batch of samples, the enrichment was less efficient, with number of reads on target between 3.81% and 17.77%, while for the two smaller batches it varied between 69.00% and 85.27%. The percentage of reads on target is particularly low for taxa most distantly related to *Albizia julibrissin* on which the bait sequences are based. Highly divergent sequences are not expected to be captured, but even so, these percentages of reads on target may be underestimated if the targeted sequences are highly divergent (<70% sequence identity to the baits) given the mapping threshold that we employed. Despite the variable enrichment efficiency, we were able to reconstruct at least partial sequences for the large majority of loci across almost all taxa (Appendix 1), with the number of target loci that were at least partially recovered, as determined by BLASTX searches of the scaffolds, ranging from 644 to 957.

After ortholog detection, a total of 1915 gene alignments were recovered (Fig. 3), representing 767 of the targeted genes. Clusters representing the remaining 197 targeted genes were discarded because orthologous subclusters contained too few accessions, which may in turn be caused by poor phylogenetic resolution. For 279 targets, only a single gene alignment was recovered (i.e., they are putatively single-copy). For the remainder of the gene alignments, it is sometimes difficult to establish whether the multiple alignments represent paralogous copies, multiple exon alignments for the same gene that became separated during phylogenetic ortholog detection, or gene alignments that were split into two-taxon sets because of long internal branches. Using BLAST searches of the longest sequence of each gene alignment against the target sequences, it became clear that many of these do indeed represent different non-overlapping fragments (most likely exons) of the same gene. Furthermore, some of the multiple alignments for the same gene do not have any overlapping accessions, which suggests they represent orthologous sequences for two distinct groups of taxa. It is thus not straightforward to accurately determine the precise number of paralog copies among the targeted genes.

The number of accessions per gene alignment ranged from 13 to 121 (Fig. 4A), and aligned length per gene alignment varied from 282 to 2526 bp (Fig. 4B). Taxon occupancy per locus shows about a fourfold difference, with generally higher occupancy for members of the ingoid clade compared to more divergent taxa (Fig. 4C). However, even the least represented accession (*Acaciella villosa*) is still present in 274 gene alignments, which is likely sufficient to resolve its placement in the phylogeny, at least in concatenated analyses. Numbers of distinct alignment patterns, an indication of the phylogenetic informativeness of an alignment, show an uneven distribution across gene alignments. This suggests there are relatively few highly informative genes in the data set, but also few that are relatively uninformative (Fig. 4D). However, this does not indicate whether certain genes are particularly informative for deeper nodes or for more recent ones.

Gene and species tree inference

Gene trees for 148 of the genes had relatively high root-to-tip variances (>0.01; Fig. 4E). This marked branch length variation suggests they are not suitable for phylogenetic reconstruction, and inspection of these gene trees made it clear (based on our understanding of mimosoid phylogeny) that many of the inferred relationships were spurious. Apart from genuine variation in substitution rates, it is also likely that missing data (e.g., complete exons missing in some unrelated taxa) could lead to such relationships being inferred. These gene trees were discarded and not analyzed further. After excluding these loci, the remaining 1767 were aligned, giving a total aligned length of 861,525 bp, with 450,375 alignment patterns and 62.12% missing data. A second concatenated alignment for only those loci with at least half of the accessions included (510 genes/exons) has a total aligned length of 254,250 bp, or 84,750 amino acids with 176,713 or 73,179 alignment patterns, respectively, and 34.89% missing data. Jackknife alignments consist of between 127 and 129 genes with total aligned lengths of 19,949 to 22,218 amino acids. The chloroplast alignment is 60,321 bp long, contains 16,589 alignment patterns, and has 17.33% missing data.

The concatenated ML and ASTRAL species tree analyses yielded highly supported and similar topologies, except for a

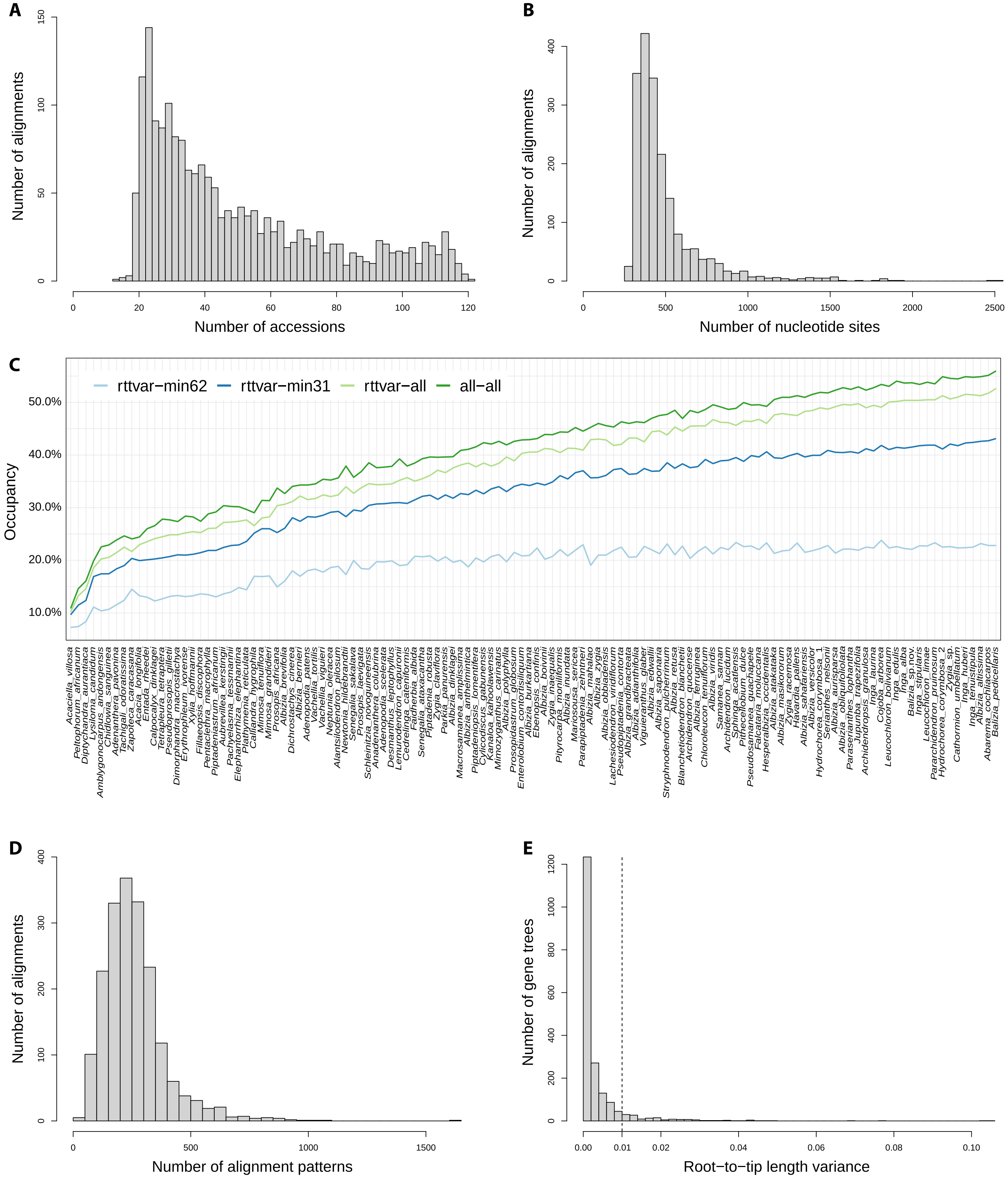


FIGURE 4. Statistics for recovered loci. (A) Number of accessions per locus, (B) aligned length per locus, (C) taxon occupancy per locus (all vs. rttvar = all loci or only those with <0.01 root-to-tip variance in the gene trees; all vs. min31 vs. min62 = without or with minimum taxon cutoffs of 25% or 50%, respectively), (D) number of alignment patterns per locus, and (E) root-to-tip variance in the inferred gene trees, with the dashed line at 0.01 indicating the cutoff for retaining or discarding loci.

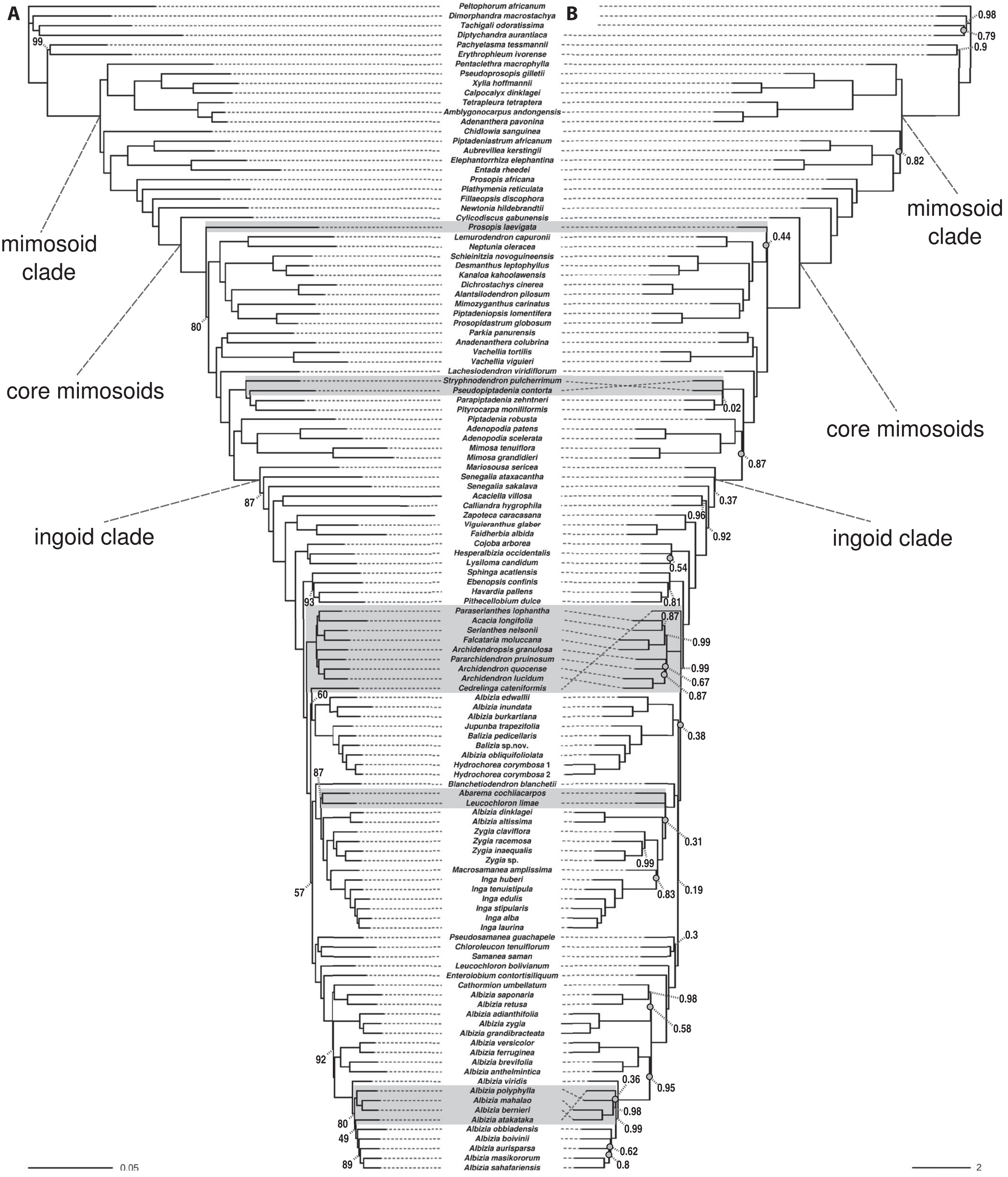


FIGURE 5. Generic backbone phylogeny of mimosoid legumes. Comparison between the concatenated ML and ASTRAL species trees, with gray shading indicating topological differences. (A) RAxML tree inferred from the full concatenated alignment (1767 loci) with bootstrap support indicated for internodes that received <100%, and branch lengths in number of substitutions per site. (B) ASTRAL species tree inferred from 1229 loci with more than a quarter of the accessions present, with branch lengths in coalescent units. Local posterior probability is indicated for internodes that received <1.00 pp; circles on nodes indicate those nodes for which a polytomy could not be rejected. Terminal branch lengths in the ASTRAL tree are set at 1 (instead of 0) for better visualization.

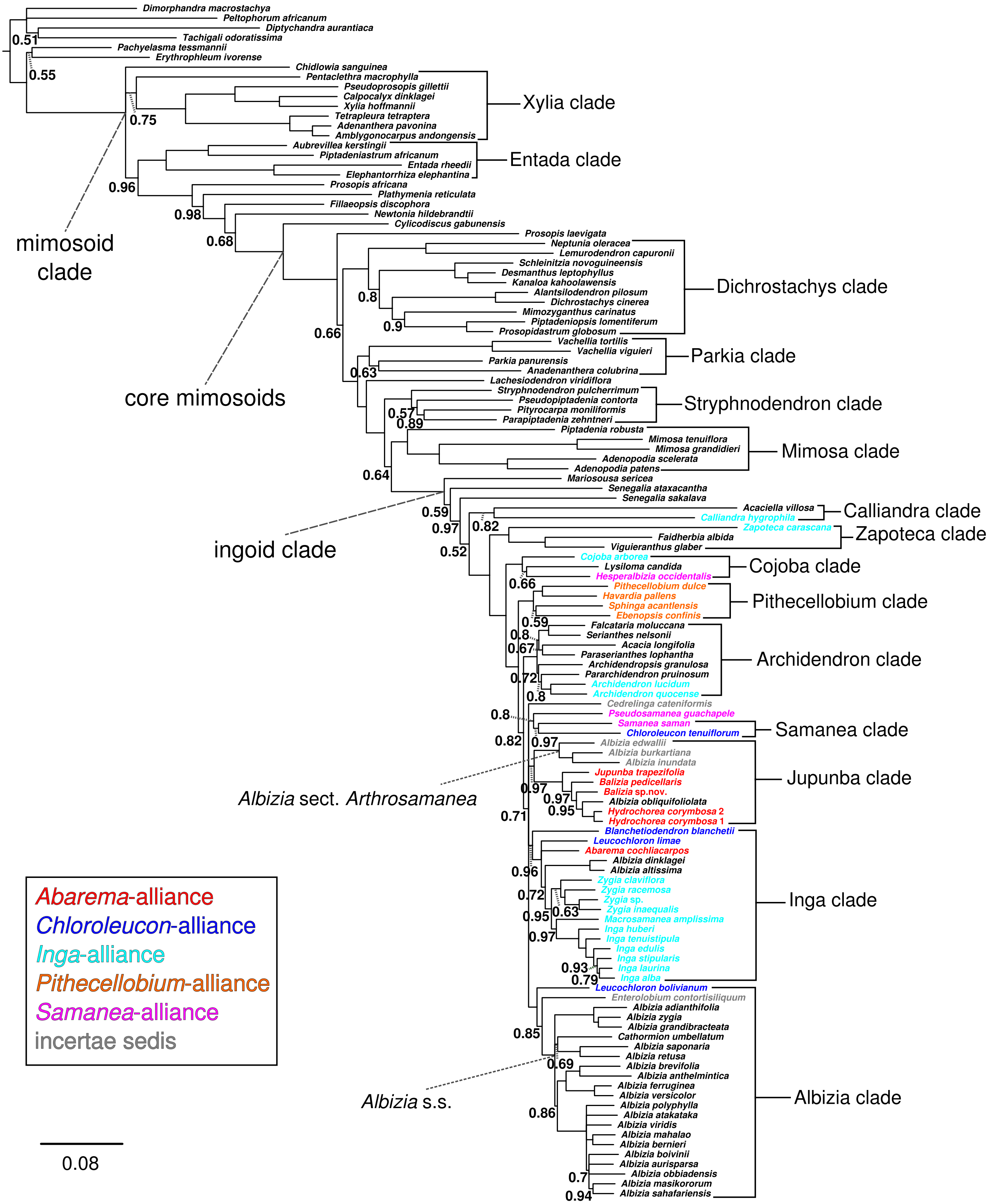


FIGURE 6. Robustly supported clades in the mimosoid phylogeny. Clades are annotated on the Bayesian jackknife majority-rule consensus tree, with posterior probability values for internodes with <1.00 pp indicated. Colored taxon names indicate non-monophyly of all but one of the alliances recognized by Barneby and Grimes (1996), as per the legend. Terminal names in black were not included in any alliance by Barneby and Grimes (1996) because they did not include genera outside tribe Ingeae and did not comprehensively treat the genera of Ingeae that do not occur in the Americas.

TABLE 2. Higher- and lower-level clades informally recognized in this study.

Higher-level clades	No. of genera[a]	No. of species[b]
Mimosoid clade	~87	~3300
Core mimosoids	~72	~3220
Ingoid clade	~43	~2000

Lower-level clades	List of genera[c]
Xylia clade	*Adenanthera* L., *Amblygonocarpus* Harms, *Calpocalyx* Harms, *Pentaclethra* Benth., *Pseudoprosopis* Harms, *Tetrapleura* Benth., *Xylia* Benth.
Entada clade	*Aubrevillea* Pellegr., *Elephantorrhiza* Benth., *Entada* Adans., *Piptadeniastrum* Brenan
Dichrostachys clade	*Alantsilodendron* Villiers, *Calliandropsis* H.M. Hern. & P. Guinet*, *Desmanthus* Willd, *Dichrostachys* (DC.) Wight & Arn., *Gagnebina* Neck. ex DC.*, *Kanaloa* Lorence & K. R. Wood, *Lemurodendron* Villiers & Guinet, *Leucaena* Benth.*, *Mimozyganthus* Burkart, *Neptunia* Lour., *Piptadeniopsis* Burkart, *Prosopidastrum* Burkart, *Schleinitzia* Warb.
Parkia clade	*Anadenanthera* Speg., *Parkia* R.Br., *Vachellia* Wight & Arn.
Stryphnodendron clade	*Microlobius* C.Presl*, *Parapiptadenia* Brenan, *Pityrocarpa* (Benth.) Britton & Rose, *Pseudopiptadenia* Rauschert, *Stryphnodendron* Mart.
Mimosa clade	*Adenopodia* C.Presl, *Mimosa* L., *Piptadenia* Benth.
Calliandra clade	*Acaciella* Britton & Rose, *Afrocalliandra* E.R.Souza & L.P.Queiroz*, *Calliandra* Benth.
Zapoteca clade	*Faidherbia* A.Chev., *Sanjappa* E.R.Souza & M.V.Krishnaraj*, *Thailentadopsis* Kosterm.*, *Viguieranthus* Villiers, *Zapoteca* H.M.Hern.
Cojoba clade	*Cojoba* Britton & Rose, *Hesperalbizia* Barneby & J.W.Grimes, *Lysiloma* Benth.
Pithecellobium clade	*Ebenopsis* Britton & Rose, *Havardia* Small, *Painteria* Britton & Rose*, *Pithecellobium* Mart., *Sphinga* Barneby & J.W.Grimes
Archidendron clade	*Acacia* Mill., *Archidendron* F.Muell., *Archidendropsis* I.C.Nielsen, *Falcataria* (I.C.Nielsen) Barneby & J.W.Grimes, *Pararchidendron* I.C.Nielsen, *Paraserianthes* I.C.Nielsen, *Serianthes* Benth., *Wallaceodendron* Koord.*
Samanea clade	*Chloroleucon* (Benth.) Britton & Rose, *Samanea* (Benth.) Merr.
Jupunba clade	*Albizia* sect. *Arthrosamanea* (Britton & Rose) Barneby & J.W.Grimes (species of the Americas), *Balizia* Barneby & J.W.Grimes, *Jupunba* Britton & Rose, *Hydrochorea* Barneby & J.W.Grimes, *Punjuba* Britton & Rose, *Albizia obliquifoliata* De Wild.
Inga clade	*Abarema* Pittier, *Blanchetiodendron* Barneby & J.W.Grimes, *Inga* Mill., *Leucochloron* Barneby & J.W.Grimes**, *Macrosamanea* Britton & Rose ex Britton & Killip, *Albizia dinklagei* Harms, *Albizia altissima* Hook.f.
Albizia clade	*Albizia* Durazz.** *sensu stricto* (Asian, African, Madagascan and Pacific species; incl. *Cathormion* Hassk.), *Enterolobium* Mart., *Leucochloron bolivianum* C.E.Hughes & Atahuachi
Incertae sedis	
In paraphyletic grade or isolated phylogenetic position	*Cylicodiscus* Harms, *Fillaeopsis* Harms, *Lachesiodendron* P.G.Ribeiro, L.P.Queiroz & Luckow, *Mariosousa* Seigler & Ebinger, *Newtonia* Baill., *Parasenegalia* Seigler & Ebinger*, *Plathymenia* Benth., *Prosopis* L.**, *Pseudosenegalia* Seigler & Ebinger*, *Senegalia* Raf.**, *Xerocladia* Harv.*
Phylogenetic position uncertain	*Cedrelinga* Ducke, *Chidlowia* Hoyle, *Indopiptadenia* Brenan, *Pseudosamanea* Harms

[a] Numbers of genera remain tentative pending resolution of generic delimitation issues caused by generic non-monophyly across the mimosoid clade.
[b] Numbers of species are approximate but most likely underestimated pending the description and further discovery of species new to science.
[c] Including lineages/species pending transfer to newly described segregate genera.
*Not sampled here but placement inferred from previous studies.
**Note that genus is non-monophyletic.

relatively small number of internodes (Fig. 5). ML analyses of the concatenated alignment of 510 loci (Appendices S2 and S3) show higher support and almost identical topologies. The Bayesian jackknife consensus tree (Fig. 6) shows a polytomy at the base of the mimosoid clade, involving the position of *Chidlowia* and several polytomies within the ingoid clade, including a large one along the backbone of that clade. The chloroplast phylogeny (Appendix S4) differs in some places from the species trees inferred from nuclear gene data. For example, there is notable cytonuclear discordance in relation to the monophyly of *Senegalia* (see below). The chloroplast phylogeny is less robustly supported than the nuclear species tree, particularly within the ingoid clade. A tanglegram comparing the chloroplast phylogeny with the ASTRAL species tree (Appendix S5) shows only minor differences outside the ingoid clade, but rather different relationships across the base of the ingoid clade, as expected due to low support in that portion of the tree (see below). Generally, apart from the well-supported discordance related to *Senegalia*, the differences between the chloroplast and ASTRAL phylogenies do not appear to be beyond what could be expected, based on the observed gene tree incongruence among nuclear genes, and coincide mostly with poorly supported nodes. Alignments and trees are included in Appendices S6–S11 and available on TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S26316).

Characterization of well-supported clades

The species trees provide a robust framework for recognizing two higher-level and 15 lower-level clades within the mimosoid clade (*sensu* LPWG, 2017; i.e., former subfamily Mimosoideae plus *Chidlowia*, which is here confirmed as a member of the mimosoid clade, as suggested by Manzanillo and Bruneau, 2012; see Fig. 6 and Table 2) that receive high support in (almost) all analyses, and that are mostly also well supported across gene trees (Fig 7A). These clades serve as an informal classification for communicating about, and navigating across, the mimosoid phylogeny. Following the long tradition of using informal group or clade names in legume systematics (Polhill and Raven, 1981; Lewis et al., 2005; LPWG, 2013), lower-level clades are named after a characteristic genus within each clade and provide monophyletic groupings of genera to replace previously defined informal groups or alliances (Barneby and Grimes, 1996; Lewis et al., 2005), almost all of which are now shown to be non-monophyletic. However, not all genera are included in a named clade, because of the imbalanced topology, which includes a few paraphyletic grades. These include a grade of *Prosopis africana* and the genera *Plathymenia*, *Fillaeopsis*, *Newtonia*, *Cylicodiscus*, and *Prosopis laevigata*; the senegalioid grade that includes *Mariosousa*, *Senegalia* and its recent segregates *Pseudosenegalia* and *Parasenegalia* (Miller et al., 2017; Seigler et al., 2017; neither of which is sampled here); as well as several genera in isolated positions with deep-branching stem lineages (e.g., *Cedrelinga*, *Chidlowia*, and *Lachesiodendron*). The named clades and unplaced genera are listed in Table 2, and clade definitions are included in Appendix 2 with notes about notable characteristics.

Evaluation of support for inferred relationships

The ASTRAL topology differs in only five places from the ML topology (Fig. 5): (1) *Prosopis laevigata* is sister to the Dichrostachys clade with 0.44 pp, instead of to the rest of the core mimosoids with 80% BS; (2) *Stryphnodendron pulcherrimum* and *Pseudopiptadenia contorta* have swapped positions, with 0.02 pp in the ASTRAL tree, while the alternative relationship in the ML tree has full support; (3) *Cedrelinga cateniformis* is sister to a large clade composed of several subclades of the ingoid clade with 0.38 pp, instead of being sister to the Jupunba clade with 60% BS; (4) *Abarema cochliacarpos* and *Leucochloron limae* are not sister taxa, with full support, while they are in the ML tree with 87% BS; and (5) *Albizia atakataka* is in a different position in the two trees, with 49% BS vs. 0.36 pp.

Support along the backbone of the phylogeny, with the exception of the ingoid clade, is generally high in concatenated analyses (Figs. 5 and 6) but is known to be overestimated in large data sets (Salichos and Rokas, 2013). Taking into account conflicting signals across gene trees, levels of support are less robust, with many internodes receiving relatively low ICA support (<0.5; Appendix S12) suggesting significant conflict at those nodes. In a few cases, ICA values below zero indicate that the most common conflicting bipartitions are more prevalent than the supporting ones (Fig. 7). Comparing proportions of gene tree bipartitions supporting an internode, in relation to the most common conflicting bipartions, all other conflicting bipartitions, and uninformative gene trees (including those with missing data; e.g., most strikingly for the Calliandra clade, due to poor representation of *Acaciella villosa* across gene alignments), it is clear that the majority of gene trees are either uninformative or contain an infrequent conflicting bipartition (Fig. 7 and Appendix S13). This strongly suggests that the majority of the gene trees lack phylogenetic signal, especially across the ingoid backbone.

The ASTRAL polytomy test showed that for several nodes, the null model of a polytomy could not be rejected (given a *p*-value threshold of 0.05; Fig. 4B and Appendix S14). We quantified gene tree conflict in more detail for three questionable deeper nodes along the backbone of the phylogeny (Fig. 7B–D): (1) placement of *Chidlowia*, (2) placement of *Prosopis laevigata*, and (3) mono/paraphyly of the Piptadenia group *sensu* Lewis et al. (2005) but excluding *Parkia*, *Anadenanthera*, and the recently segregated *Lachesiodendron* (Ribeiro et al., 2018). The same was done for the backbone of the ingoid clade (Fig. 7E–H), where polytomies could not be ruled out for several nodes. This shows that the placement of *Chidlowia* as sister to all other mimosoids excluding the Xylia clade (as in Fig. 5) is preferred slightly over the two alternative hypotheses (Fig. 7B). For *Prosopis laevigata*, a sister-group relationship with the rest of the core mimosoids (i.e., being sister to the Dichrostachys clade + the remaining core mimosoids except *Cylicodiscus,* as in Figs. 5A and 6) is equally or slightly better supported across gene trees than the two alternatives (Fig. 7C). For the Piptadenia group, paraphyly is slightly more often supported across gene trees than monophyly, with the Mimosa clade as the most likely sister group of the ingoid clade (Figs. 5, 6, and 7D). Within the ingoid clade, there is a notable lack of resolution especially in the clade that includes *Cedrelinga* and *Pseudosamanea* plus the Archidendron, Jupunba, Inga, Samanea, and Albizia clades. The phylogenetic placement of the monotypic *Cedrelinga* appears to be unstable, with hardly any gene tree support for any of its possible placements (Fig. 7E). There are some weakly supporting gene trees showing a sister-group relationship of *Cedrelinga* with *Pseudosamanea*, but that taxon is more likely related to *Chloroleucon* and *Samanea* (Fig. 7F), and one of the other three possible placements (Fig. 7E) is probably more likely. There are no gene trees strongly supporting *Cedrelinga* as sister to the rest, and for the other two options there is just one gene tree strongly in support of each. A sister-group relationship between the Samanea (including *Pseudosamanea*) and Albizia clades has minimal support across gene trees, even though it is found in the ML and ASTRAL species tree analyses (Fig. 5) and remains the most likely possibility in relation to alternatives (Fig. 7G).

These results suggest that *Cedrelinga* and *Pseudosamanea*, and perhaps also the (other) two genera of the Samanea clade, are potentially causing lack of resolution in the ingoid clade, acting as “rogue taxa,” for example due to lack of phylogenetic signal for the placement of these taxa or long branch attraction (LBA) artifacts, particularly for *Cedrelinga*. Another possibility is that ancient hybridization or (allo)polyploidization has occurred, giving rise to (some of) these rogue lineages. ML analyses on the concatenated alignment of 510 genes omitting these taxa do indeed increase support along the ingoid backbone (compare Appendices S3 and S15–S17). To investigate this further, we evaluated support for all possible groupings of the Archidendron, Jupunba, Inga, Samanea, and Albizia clades as sister clades, triplets, and quartets across gene trees with *Cedrelinga* and *Pseudosamanea* removed. This shows that the sister-group relationship of the Albizia and Samanea clades is more likely than any other conflicting relationship (Fig. 7H) and that the Jupunba and Inga clades are likely to be sister clades. No well-supported triplets are found, while the quartet that unites the Jupunba, Inga, Samanea, and Albizia clades is better supported than all other possible quartets (Fig. 7H). Taken together, this would suggest a branching order of (Archidendron((Jupunba,Inga),(Samanea,Albizia))) for these clades. However, none of the possible relationships among these clades, nor the placements of *Cedrelinga* and *Pseudosamanea*, appear in many gene trees with strong support, and it is striking that there are many more strongly conflicting gene trees for most of these (Fig. 7H).

Phylogenetic supernetwork analysis

At the lowest mintrees setting (n = 22, ~2.5% of the total number of trees; Fig. 8A, B), there appears to be little signal. Increasing to n = 44 or n = 66 (Fig. 8C–F), the network becomes somewhat more treelike and shows more or less the same relationships among clade representatives as the gene tree support summarization (Fig. 7E–H). However, increasing mintrees to n = 88 causes that resolution to collapse (Fig. 8G), showing that just limited phylogenetic signal hints at a resolved topology. In other words, taking into account more of the uncommon splits across gene trees (Fig. 8A, B), discordance is too high to reveal phylogenetic signal, while a stricter approach taking into account only splits that are more commonly found across gene trees (Fig. 8G) results in no phylogenetic signal being observed.

DISCUSSION

We found that targeted enrichment via hybrid capture is a powerful and efficient way to reconstruct the phylogeny of a challenging taxonomic group, in line with findings across a rapidly growing number of other groups (e.g., Mandel et al., 2014; Weitemier et al., 2014; Nicholls et al., 2015; Sass et al., 2016; Johnson et al., 2018; Couvreur et al., 2019; Ojeda et al., 2019). The phylogenetic resolution and statistical support obtained here offer a significant improvement over previous mimosoid phylogenies (Luckow et al., 2003; Bouchenak-Khelladi et al., 2010; LPWG, 2017), yielding a robust generic backbone tree for the mimosoid clade (Figs. 5 and 6).

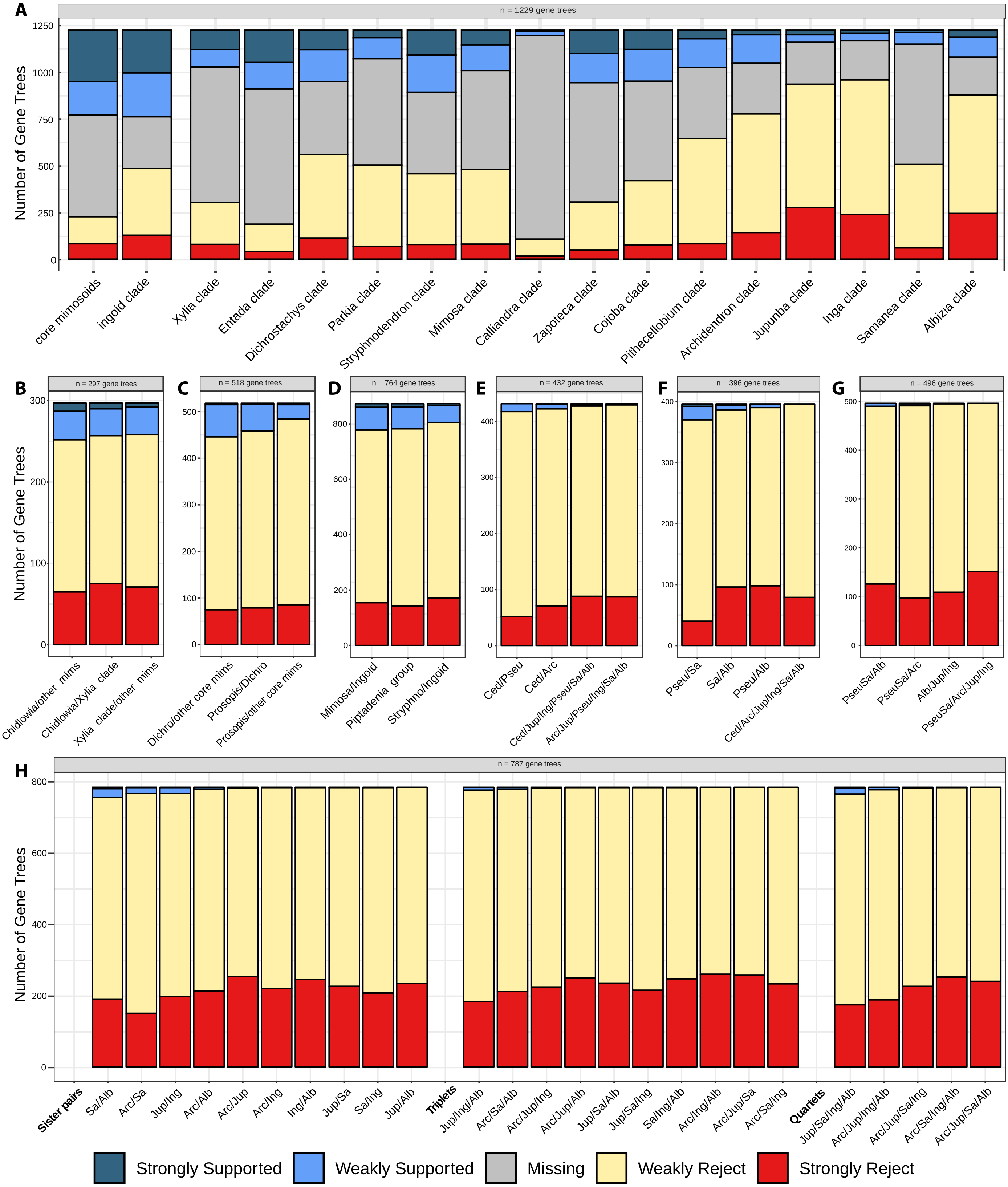
Nevertheless, relationships among well-supported clades within the ingoid clade appear to be impossible to resolve with our data set (Fig. 7E–H), which is surprising given the large number of genes deployed here and the general robustness of the mimosoid phylogeny that was recovered using these genes. Therefore, this lack of resolution is probably not caused by insufficient data, but is instead most likely the result of extremely rapid speciation leading to a lack of phylogenetic signal as implied by lack of resolution across nearly all gene trees (Fig. 7H). While evaluation of supporting gene trees and the filtered supernetworks (Fig. 8) suggest some clade relationships as more likely than others, this may simply be an exercise in extracting the least conflicting signal from a data set where there is virtually no signal to begin with. In any case, there appear to be many conflicting bipartitions among the set of gene trees (Fig. 7E–H), and hardly any that strongly support any of the possible relationships among the ingoid subclades.

Gene tree conflict is often attributed to ILS, as found in the initial radiation of the Neoaves clade of birds (Suh et al., 2015), which provides one of the most convincing examples of a hard polytomy documented so far (Suh, 2016). Suh et al. (2015) used retroposon insertion sites that are virtually free from homoplasy as strong evidence for ILS. While such evidence is lacking here, part of the ingoid backbone appears similarly unresolvable based on 964 nuclear genes. In other cases, such as mammals, ILS has been shown to be only a minor cause of gene tree conflict (Scornavacca and Galtier, 2017), suggesting that such conflict could equally be caused by gene tree estimation errors due to lack of phylogenetic signal, homoplasy, alignment errors, and/or poor model fit (Richards et al., 2018). Across the ingoid clade, the majority of conflicting gene tree bipartitions appear to be rare and most of them are only weakly conflicting (Fig. 7E–H). This suggests that most of the conflicting bipartitions stem from lack of phylogenetic signal, with gene tree estimation errors accounting in part for the strongly conflicting bipartitions (Richards et al., 2018). Other reasons for poor gene tree estimation include alignment errors, homoplasy, poor model fit, and LBA artifacts. We have attempted to minimize alignment errors by using MACSE (Ranwez et al., 2018), which simultaneously aligns coding sequences and the amino acid translations, yielding considerably better alignments than MAFFT and making the additional computational time worthwhile. The interrelated issues of homoplasy, poor model fit, and LBA artifacts are less easily tackled and could be the main sources of gene tree estimation errors in our data set. In that case, this conflict would constitute phylogenetic noise rather than genuine conflicting signal, and such noise is present across much of the tree (Appendix S13). However, even though the number of conflicting bipartitions for many nodes across the tree far outnumber the most prevalent bipartition, the second most prevalent (green part of pie charts in Appendix S13) is not close to equally prevalent in parts of the species tree where resolution and support are consistently high. Within the ingoid radiation there is simply not enough signal to override this noise.

Apart from gene tree estimation errors and ILS, ancient hybridization during the radiation of the ingoid clade could offer an alternative explanation for the large number of strongly conflicting gene tree topologies. The strong conflicting gene tree support for the placement of *Pseudosamanea*, in particular, could be indicative of hybridization, although it is also possible that LBA artifacts could be causing an apparent sister-group relationship with *Cedrelinga* in some gene trees.

For the Neoaves clade of birds, lack of treelike structure in phylogenetic supernetworks is similar to that found in networks generated from simulated random topologies, suggesting that this clade is indeed best considered a hard polytomy (Suh, 2016). Together with the lack of gene tree support (Fig. 7E–H), our supernetworks (Fig. 8) also suggest that the ingoid radiation perhaps constitutes a hard polytomy. With intermediate mintrees parameter settings (Fig. 8C–F) the networks show some structure. However, given that this resolution collapses at the higher mintrees setting (Fig. 8G), this is likely driven by a very small number of gene trees, while conflicting gene trees largely outnumber the few supporting ones (Fig. 7E–H), in line with the idea that many contentious relationships are supported by just a handful of genes (Shen et al., 2017). Our network at the lowest mintrees setting is similar to that of a simulated hard polytomy (cf. Fig. 8A, B, with Suh, 2016: fig. 4E). We therefore conclude, pending enhanced taxon sampling and eventually completely sequenced genomes, that there is potentially a hard polytomy embedded in the backbone of the ingoid clade, from which derives a large pantropical radiation that includes an estimated 1750 extant species. This putative hard polytomy involves six or seven lineages and is resistant to resolution, even using sequences from hundreds of nuclear genes. With complete exomes and positional

FIGURE 7. Evaluation of gene tree support for selected nodes. (A) Bar graphs of supporting and rejecting gene trees for the two higher-level and 15 lower-level clades identified in this study and for alternative topologies involving (B) placement of *Chidlowia*, (C) placement of *Prosopis*, (D) monophyly or paraphyly of the Piptadenia group, (E) placement of *Cedrelinga*, (F) placement of *Pseudosamanea*, (G) affinities of the Samanea clade, and (H) all possible sister pairs, clade triplets, and quartets within the polytomous portion of the ingoid clade after pruning *Cedrelinga* and the Samanea clade from the gene trees. Note that for panels B–H, the bars for each graph are sorted from most to least supported. Abbreviations: mims = mimosoids, Stryphno = Stryphnodendron clade, Ced = *Cedrelinga cateniformis*, Pseu = *Pseudosamanea guachapele*, Arc = Archidendron clade, Jup = Jupunba clade, Ing = Inga clade, Sa = Samanea clade, Alb = Albizia clade, PseuSa = *Pseudosamanea guachapele* + Samanea clade.


A
n = 1229 gene trees
Number of Gene Trees
0
250
500
750
1000
1250
core mimosoids
ingoid clade
Xylia clade
Entada clade
Dichrostachys clade
Parkia clade
Stryphnodendron clade
Mimosa clade
Calliandra clade
Zapoteca clade
Cojoba clade
Pithecellobium clade
Archidendron clade
Jupunba clade
Inga clade
Samanea clade
Albizia clade
B
n = 297 gene trees
Number of Gene Trees
0
100
200
300
Chidlowia/other mims
Chidlowia/Xylia clade
Xylia clade/other mims
C
n = 518 gene trees
0
100
200
300
400
500
Dichro/other core mims
Prosopis/Dichro
Prosopis/other core mims
D
n = 764 gene trees
0
200
400
600
800
Mimosa/Ingoid
Piptadenia group
Stryphno/Ingoid
E
n = 432 gene trees
0
100
200
300
400
Ced/Pseu
Ced/Arc
Ced/Jup/Ing/Pseu/Sa/Alb
Arc/Jup/Pseu/Ing/Sa/Alb
F
n = 396 gene trees
0
100
200
300
400
Pseu/Sa
Sa/Alb
Pseu/Alb
Ced/Arc/Jup/Ing/Sa/Alb
G
n = 496 gene trees
0
100
200
300
400
500
PseuSa/Alb
PseuSa/Arc
Alb/Jup/Ing
PseuSa/Arc/Jup/Ing
H
n = 787 gene trees
Number of Gene Trees
0
200
400
600
800
Sister pairs
Sa/Alb
Arc/Sa
Jup/Ing
Arc/Alb
Arc/Jup
Arc/Ing
Ing/Alb
Jup/Sa
Sa/Ing
Jup/Alb
Triplets
Jup/Ing/Alb
Arc/Sa/Alb
Arc/Jup/Ing
Arc/Jup/Alb
Jup/Sa/Alb
Jup/Sa/Ing
Sa/Ing/Alb
Arc/Ing/Alb
Arc/Jup/Sa
Arc/Sa/Ing
Quartets
Jup/Sa/Ing/Alb
Arc/Jup/Ing/Alb
Arc/Jup/Sa/Ing
Arc/Sa/Ing/Alb
Arc/Jup/Sa/Alb
Strongly Supported
Weakly Supported
Missing
Weakly Reject
Strongly Reject

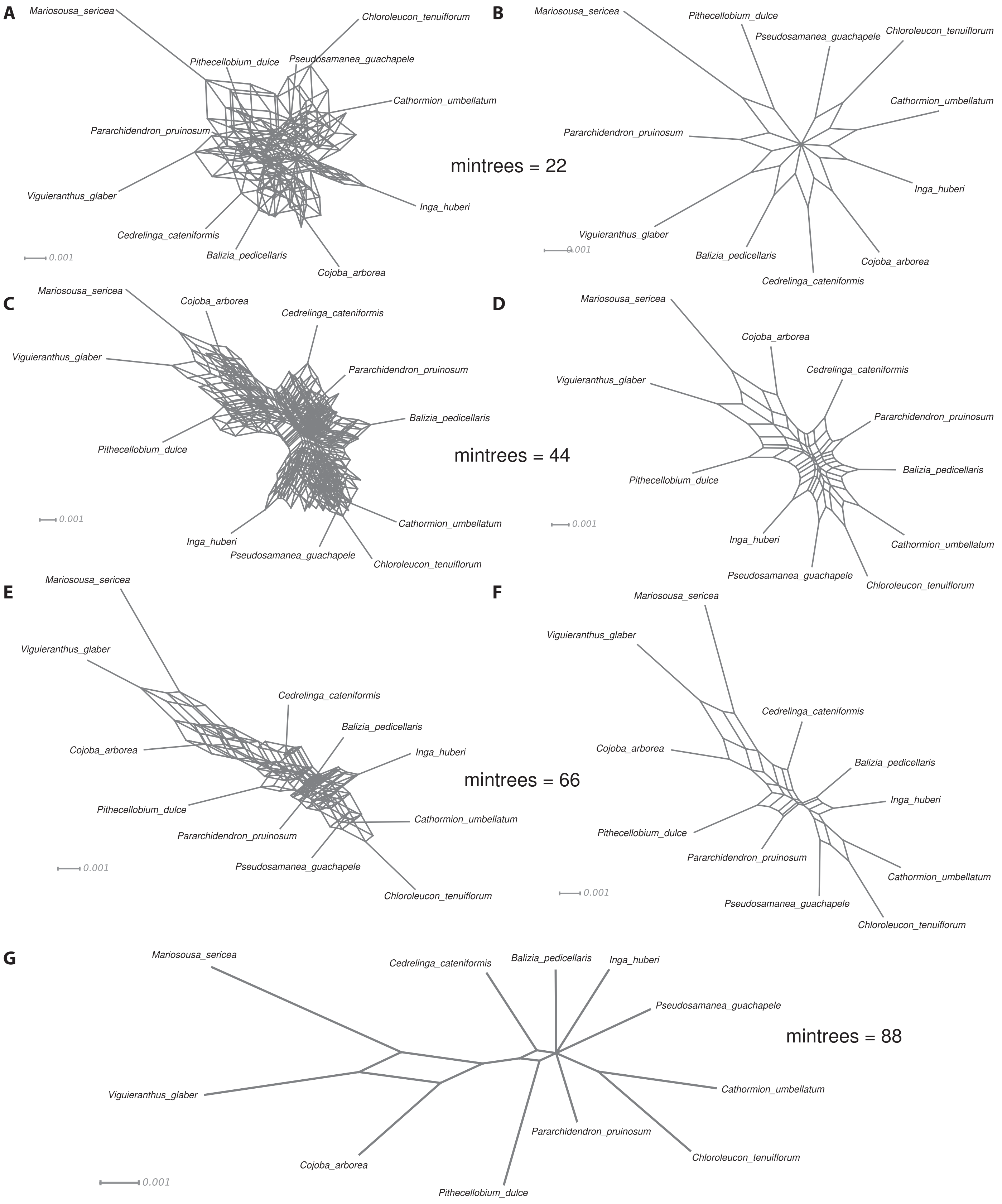

A
Mariosousa_sericea
Chloroleucon_tenuiflorum
Pithecellobium_dulce
Pseudosamanea_guachapele
Cathormion_umbellatum
Pararchidendron_pruinosum
Viguieranthus_glaber
Inga_huberi
Cedrelinga_cateniformis
Balizia_pedicellaris
Cojoba_arborea
0.001
mintrees = 22
B
Mariosousa_sericea
Pithecellobium_dulce
Pseudosamanea_guachapele
Chloroleucon_tenuiflorum
Cathormion_umbellatum
Pararchidendron_pruinosum
Inga_huberi
Viguieranthus_glaber
0.001
Balizia_pedicellaris
Cojoba_arborea
Cedrelinga_cateniformis
C
Mariosousa_sericea
Cojoba_arborea
Cedrelinga_cateniformis
Viguieranthus_glaber
Pararchidendron_pruinosum
Balizia_pedicellaris
Pithecellobium_dulce
0.001
Cathormion_umbellatum
Inga_huberi
Pseudosamanea_guachapele
Chloroleucon_tenuiflorum
mintrees = 44
D
Mariosousa_sericea
Cojoba_arborea
Viguieranthus_glaber
Cedrelinga_cateniformis
Pararchidendron_pruinosum
Balizia_pedicellaris
Pithecellobium_dulce
0.001
Inga_huberi
Cathormion_umbellatum
Pseudosamanea_guachapele
Chloroleucon_tenuiflorum
E
Mariosousa_sericea
Viguieranthus_glaber
Cedrelinga_cateniformis
Balizia_pedicellaris
Cojoba_arborea
Inga_huberi
Pithecellobium_dulce
Cathormion_umbellatum
Pararchidendron_pruinosum
Pseudosamanea_guachapele
0.001
Chloroleucon_tenuiflorum
mintrees = 66
F
Mariosousa_sericea
Viguieranthus_glaber
Cedrelinga_cateniformis
Cojoba_arborea
Balizia_pedicellaris
Inga_huberi
Pithecellobium_dulce
Pararchidendron_pruinosum
Cathormion_umbellatum
Pseudosamanea_guachapele
0.001
Chloroleucon_tenuiflorum
G
Mariosousa_sericea
Cedrelinga_cateniformis
Balizia_pedicellaris
Inga_huberi
Pseudosamanea_guachapele
mintrees = 88
Viguieranthus_glaber
Cathormion_umbellatum
Pararchidendron_pruinosum
Chloroleucon_tenuiflorum
Cojoba_arborea
0.001
Pithecellobium_dulce

FIGURE 8. Phylogenetic Z-closure filtered supernetworks for a selection of taxa representing the main lineages of the ingoid clade with mintrees parameter set at 22, drawn (A) with and (B) without the Convex Hull algorithm, and the same for (C, D) mintrees setting at 44, (E, F) mintrees setting at 66, and (G) the mintrees setting at 88 (the networks with and without the Convex Hull method are identical, indicating that not many splits are included under this parameter setting).

homology data, it will be possible to investigate the sorting of different unlinked exons, genes, or other genomic elements (e.g., retroelement insertions; Doronina et al., 2015; Suh et al., 2015; Suh, 2016; Springer et al., 2020) across lineages within the ingoid clade in greater detail to shed light on the underlying treelike structure of the phylogeny, or the lack thereof, ultimately confirming or rejecting the hard polytomy suggested here.

We note that our analyses employ only protein-coding sequence data, but noncoding data flanking the targeted exons (UTRs, introns) were also captured to some extent, and may be more variable and phylogenetically informative. We chose to use only the coding data because we consider them superior to noncoding data, being less saturated with multiple substitutions and more reliably alignable, especially with an alignment program that takes protein translations into account (Ranwez et al., 2018). The resulting alignments can also be translated and analyzed with models of protein evolution such as LG4X (Le et al., 2008) and CAT (Lartillot and Philippe, 2004), which are more realistic and less prone to LBA artifacts than DNA nucleotide substitution models (Lartillot et al., 2007; Philippe et al., 2011). It may nonetheless be interesting to explore noncoding regions from our data set in the future.

Furthermore, it is possible that fragmentation of exons from the same gene could have contributed to lack of resolution across gene trees. While it is well suited for distinguishing orthologs from paralogs, fragmentation of exons is a limitation of the modified Yang and Smith (2014) ortholog selection pipeline used here. Other available pipelines can potentially deal with this issue and hence improve individual gene trees and thereby allow more accurate evaluation of alternative topologies. However, these pipelines have other limitations. For example, the Hybpiper pipeline (Johnson et al., 2016) could potentially reconstruct longer gene sequences than the pipeline used here, but does not automatically sort different paralogs into separate gene alignments. Similarly, Moore et al.'s (2017) method to classify exons by their respective paralog gene copies offers a promising approach, but relies on having initial backbone gene family trees for all loci. Furthermore, recombination may also take place in between different exons of the same gene, suggesting that exons could be better evolutionary units for phylogenetic analysis than full gene sequences (Scornavacca and Galtier, 2017), thereby potentially mitigating this limitation of the Yang and Smith (2014) pipeline.

It has been established that multiple WGD events occurred during the early evolution of the legume family (Cannon et al., 2015; Koenen et al., 2020b), including one that affected subfamily Caesalpinioideae, in which the mimosoid clade is embedded. This WGD has most likely contributed to the number of genes for which multiple copies were found in our study, raising the possibility that paralogy could have contributed to conflicting topologies among gene trees. However, given that paleopolyploidy most likely occurred before the crown group divergence of Caesalpinioideae (Koenen et al., 2020b), and the fact that the mimosoid crown group diverged substantially later than that, paralog copies derived from this WGD are expected to have sufficiently divergent sequences across mimosoids for correct separation of paralogous sequences into separate alignments using the robust orthology assessment pipeline of Yang and Smith (2014) for most of these genes. Paleopolyploidy is also known to have occurred within the mimosoid clade (e.g., *Leucaena*, Govindarajulu et al., 2011; *Mimosa*, Dahmer et al., 2011), but chromosome count data suggest that these events were restricted to a few genera (Santos et al., 2012) and that polyploidy is most likely rare in the ingoid clade. Furthermore, this also suggests that a WGD event shared by a larger clade within mimosoids can probably be ruled out. Because the Yang and Smith (2014) pipeline also efficiently removes paralogous copies found within single accessions to which a WGD is restricted, the relationships inferred here are also unlikely to be affected by WGD events within the mimosoid clade.

Conceptually, a hard polytomy may seem problematic, because it may appear unlikely that multiple populations would become instantaneously and simultaneously isolated from each other. However, several processes could explain a hard polytomy. First, the scenario of a paraphyletic "mother" species (Naciri and Linder, 2015) would most likely mean that population-level processes would mitigate against inferring a branching order. Additionally, or alternatively, when the spread of "daughter" populations outpaces the rate at which genetic mutations become fixed within these populations, the result would also be a lack of phylogenetic signal across nearly all loci. These scenarios are especially likely when a species rapidly expands its range, followed by isolation and differentiation of subpopulations. Given that several extant mimosoid species are widespread and among the world's most notorious invasive plants, this hypothesis could provide a possible explanation for the ingoid polytomy.

Implications for the taxonomic classification of mimosoids

In this study, we advance our understanding of the evolutionary relationships among mimosoid legumes, in particular for the ingoid clade, moving forward from a soft polytomy that included almost all of the ~43 ingoid genera, to identify a potentially hard polytomy that involves six or seven highly supported monophyletic lineages. These lineages provide a robust framework for recognizing a set of informally named clades, replacing the previously defined informal groups and alliances, most of which are now shown to be non-monophyletic (Fig. 6).

This framework provides the first step toward a new tribal (Linnean) and clade-based (Phylocode) classification of mimosoids and the wider Caesalpinioideae. Achieving this will require expanded taxon sampling of all potentially non-monophyletic and missing genera within mimosoids, as well as wider sampling of genera across subfamily Caesalpinioideae as a whole, sampling that is currently being undertaken using the gene set employed here (J. J. Ringelberg, E. J. M. Koenen, et al., unpublished data). However, it is already clear that establishing a Linnean classification of tribes within mimosoids would require recognition of a large number of monogeneric tribes because of the strong imbalance across the generic backbone phylogeny (Figs. 5 and 6), which does not serve the purpose of hierarchical rank-based classification. Therefore, recognition of the mimosoid clade as a single tribe, Mimoseae (the oldest tribal name; see Polhill and Raven, 1981), within Caesalpinioideae is more fit-for-purpose, complemented with a Phylocode classification to formally name and describe clades within mimosoids along the lines informally outlined here (Appendix 2), once they are better characterized with denser taxon sampling.

The absence of a fully bifurcating topology for the ingoid clade could have led to incomplete sorting of morphological characters across the clade and the consequent difficulties associated with delimiting genera, the discordant generic systems of different authors (reviewed by Brown, 2008), and the non-monophyly of previous generic groupings (e.g., Barneby and Grimes, 1996), which were entirely morphologically based. For example, lomentaceous fruits that break up into one-seeded articles occur in at least six different lineages scattered across the Albizia, Inga, and Jupunba clades plus *Cedrelinga cateniformis* (Barneby and Grimes, 1996, 1997; E. J. M. Koenen, personal observation). Dimorphic capitate inflorescences with an enlarged central nectar-producing flower are similarly phylogenetically scattered across genera in the Albizia, Jupunba, and Samanea clades and in *Blanchetiodendron blanchetii* and *Calliandra* (Barneby and Grimes, 1996, 1997; Barneby, 1998; E. J. M. Koenen, personal observation). While reconstructing the evolution of pollination and seed dispersal syndromes across the ingoid clade would undoubtedly be illuminating in this regard, it remains unclear to what extent this will be possible in the face of lack of phylogenetic resolution.

At generic level within the mimosoids, it has been clear for some time that despite significant progress, further generic re-delimitation is needed to account for the non-monophyly of several genera (Luckow et al., 2003; Brown et al., 2008; Iganci et al., 2016; Ferm et al., 2019; É. R. de Souza et al., unpublished data). Our results add to this tally of non-monophyletic mimosoid genera. For example, while the non-monophyly of *Albizia* has long been suspected, we demonstrate robust support for two separate main evolutionary lineages currently ascribed to the genus: *Albizia* s.s., which includes species from Africa, Madagascar, and Asia; and the Neotropical *Albizia* sect. *Arthrosamanea*. Nielsen (1992:143) considered *Cathormion* to be a monotypic genus restricted to Asia, preferring to assign the African and American species to *Albizia*. Barneby and Grimes (1996) subsequently referred the American species of *Cathormion* to *Albizia*, *Chloroleucon*, and *Hydrochorea*. Lewis et al. (2005) followed Nielsen (1992) and Barneby and Grimes (1996), but the inclusion of the African species of *Cathormion* in *Albizia* has not been universally accepted, with some of these being referred to *Samanea* (e.g., Hawthorne and Jongkind, 2006). We show here that *Cathormion* should be considered a synonym of *Albizia*, as its type species *C. umbellatum* is nested within that genus (Figs. 5 and 6), and hence we make the new combination *Albizia umbellata* (Vahl) E.J.M. Koenen comb.nov.(see Appendix 2). However, the African species previously referred to *Cathormion* are not included in *Albizia*: *A. altissima* (syn. *Cathromion altissimum*) and *A. dinklagei* (syn. *Samanea dinklagei* and *Cathormion dinklagei*) are here resolved as sister taxa within the Inga clade and will need to be ascribed to a new genus. Furthermore, *Albizia obliquifoliolata* (syn. *Cathormion obliquifoliolatum*) appears to be most closely related to the Neotropical genus *Hydrochorea* in the Jupunba clade. Our results also show that *Balizia* is non-monophyletic with respect to *A. obliquifoliolata* and *Hydrochorea*, providing further evidence that the genera of the Abarema alliance of Barneby and Grimes need to be re-delimited (Iganci et al., 2016). Finally, the non-monophyly of *Senegalia* (beyond the recent segregation of *Parasenegalia* and *Pseudosenegalia*; Miller et al., 2017; Seigler et al., 2017) identified in all nuclear data analyses here (Figs. 5 and 6; with 100% BS or 1.00 pp) is unexpected, given that Boatwright et al. (2015) showed Malagasy species of *Senegalia* grouping with the rest of the genus based on three chloroplast regions. Notably, in our chloroplast phylogeny, the two species of *Senegalia* form a sister pair (Appendix S4; 100% BS), suggesting that the evolutionary history of chloroplast genomes of *Senegalia* conflicts with the nuclear-based species tree due to ILS or introgression (e.g., chloroplast capture or hybridization). This probable non-monophyly of *Senegalia*, with the two species sampled potentially representing the two main clades recovered for *Senegalia* (Kyalangalilwa et al., 2013; Boatwright et al., 2015; Terra et al., 2017), would imply that yet another segregate genus of *Acacia* s.l. may need to be erected to accommodate a large subset of species currently placed in *Senegalia*.

Outlook for Caesalpinioideae phylogenomics

In this study, we developed a gene set for targeted enrichment via hybrid capture in the mimosoid clade. The resulting bait design (available at https://github.com/erikkoenen/mimobaits/) can be used for phylogenomic studies across mimosoids and beyond. Further work in our lab has shown the utility of this gene set across the whole of Caesalpinioideae (J. J. Ringelberg and E. J. M. Koenen et al., unpublished data) and at species level within the genus *Albizia* (E. J. M. Koenen et al., unpublished data). As taxon sampling is increased across the Caesalpinioideae and more studies are carried out using this gene set on individual genera, eventually a large and densely sampled phylogeny for the subfamily can be inferred and used for taxonomic reclassification and to study the evolution of this prominent tropical woody plant clade.

ACKNOWLEDGMENTS

The authors thank R. Stalder, M. Meierhofer, and M. Knabe for greenhouse assistance; the K, L, NY, and WAG herbaria for provision of leaf samples; P. Ribeiro, M. Morim, and F. Bonadeu in Brazil and the staff of the Kew Madagascar Conservation Centre and the Botanical and Zoological Garden of Tsimbazaza, Madagascar, for support during fieldwork; the Functional Genomics Centre Zurich (FGCZ), especially Catherine Aquino, for lab support and high-throughput sequencing; the S3IT of the University of Zurich for the use of the ScienceCloud computational infrastructure and P.-A. Christin, A. Liston, and two anonymous reviewers for constructive feedback that greatly improved the manuscript.

FUNDING INFORMATION

This work was supported by the Swiss National Science Foundation (grants 31003A_135522 and 31003A_182453 to C.E.H.), the U.K. Natural Environment Research Council (Grant NE/I027797/1 to R.T.P.), the Claraz Schenkung Foundation, and the Department of Systematic and Evolutionary Botany, University of Zurich. Field trips in Brazil were partially funded by FAPESB (PTX0004 and APP0096 to L.P.d.Q.), CNPq (480530/2012-2 and PROTAX 440487 to J.R.I and L.P.d.Q.), and Energia Sustentável do Brasil. The use

of DNA from Brazilian plant species is authorized by SISGEN n° R4CAAB3 and n° R0AAA9E.

AUTHOR CONTRIBUTIONS

E.J.M.K., C.K., R.T.P., and C.E.H. designed the study. E.J.M.K. and C.K. carried out the targeted gene selection, E.J.M.K. did the labwork, analyses, and wrote the draft manuscript. M.F.S., L.P.d.Q., M.L., and G.P.L. contributed tissue samples for sequencing, C.K., J.A.N., and R.T.P. contributed data. E.J.M.K., E.R.d.S., M.F.S., J.R.I., L.P.d.Q., M.L., and C.E.H. carried out the fieldwork. All coauthors contributed to interpretation of the results and writing of the final version of the manuscript.

DATA AVAILABILITY

Transcriptome data are available on the National Center for Biotechnology Information (NCBI) databases under BioProjects PRJEB8722 and PRJNA574148; FASTQ files with raw read data are available on the Sequence Read Archive (SRA), under accession nos. SRX6901075 (*Albizia julibrissin*), SRX6901076 (*Entada abyssinica*), ERX719658 (*Inga spectabilis* (Vahl) Willd.), ERX719681 (*Inga umbellifera* (Vahl) Steud. ex DC.), ERX719690 (*Inga sapindoides* Willd.), and SRX6901077 (*Microlobius foetidus*); assembled transcripts are available on Dryad (for the *Inga* spp., https://doi.org/10.5061/dryad.r9c12) and through the Transcriptome Shotgun Assembly (TSA) database, accession nos. GHWM00000000 (*Albizia julibrissin*), GHWN00000000 (*Entada abyssinica*), and GHWO00000000 (*Microlobius foetidus*). Hybrid capture sequence reads are deposited in the European Nucleotide Archive (https://www.ebi.ac.uk/ena) under study no. PRJEB38138. Targeted sequences, alignments, and gene trees are included as Appendices S1–S7 and are made available in TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S26316). The bait design and scripts used in this study are available at https://github.com/erikkoenen/mimobaits/.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

APPENDIX S1. Complete ORFs for the 964 target genes used for bait design are included in the file Albizia_target_ORFs.fa, with sequences derived from the transcriptome of *Albizia julibrissin*.

APPENDIX S2. ML tree of the concatenated amino acid alignment of the 510 gene alignments with more than half of the accessions present, inferred with the LG4X model.

APPENDIX S3. ML tree of the concatenated nucleotide alignment of the 510 gene alignments with more than half of the accessions present, inferred with the GTRCAT model.

APPENDIX S4. ML phylogeny of 72 protein coding genes from the chloroplast genome inferred with the GTRCAT model.

APPENDIX S5. Tanglegram comparing the ASTRAL topology with the chloroplast ML topolgy.

APPENDIX S6. The concatenated alignment of all gene alignments with <0.01 root-to-tip length variance.

APPENDIX S7. The concatenated alignment of all gene alignments with <0.01 root-to-tip length variance and more than half the taxa present.

APPENDIX S8. The concatenated amino acid alignment of all gene alignments with <0.01 root-to-tip length variance and more than half the taxa present.

APPENDIX S9. The 1915 gene alignments, with sequences both <300 bp long and <50% of the aligned length removed, from which final gene trees were inferred.

APPENDIX S10. The 1915 gene trees inferred from the alignments of Appendix S9, with bootstrap support indicated.

APPENDIX S11. The concatenated alignment of chloroplast genes.

APPENDIX S12. ML topology of concatenated alignment of 1767 gene alignments, with ICA values indicated as branch labels.

APPENDIX S13. ML topology of the concatenated alignment of the 510 gene alignments with more than half of the accessions present, with number of concordant and conflicting gene trees from the same set of 510 alignments written above and below internodes, respectively.

APPENDIX S14. ASTRAL tree with polytomy test results indicated.

APPENDIX S15. ML tree of the concatenated nucleotide alignment of the 510 gene alignments with more than half of the accessions present, but with *Cedrelinga cateniformis* removed, inferred with the GTRCAT model.

APPENDIX S16. ML tree of the concatenated nucleotide alignment of the 510 gene alignments with more than half of the accessions present, but with *Cedrelinga cateniformis* and Pseudosamanea guachapele removed, inferred with the GTRCAT model.

APPENDIX S17. ML tree of the concatenated nucleotide alignment of the 510 gene alignments with more than half of the accessions present, but with *Cedrelinga cateniformis, Pseudosamanea guachapele* and the Samanea clade removed, inferred with the GTRCAT model.

LITERATURE CITED

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455–477.

Barneby, R. C. 1998. Silk tree, guanacaste, monkey's earring: a generic system of the synandrous Mimosaceae of the Americas. Part III. Calliandra. *Memoirs of the New York Botanical Garden* 74(3).

Barneby, R. C., and J. Grimes. 1996. Silk tree, guanacaste, monkey's earring: a generic system of the synandrous Mimosaceae of the Americas. Part I. *Abarema*, *Albizia*, and allies. *Memoirs of the New York Botanical Garden* 74(1).

Barneby, R. C., and J. Grimes. 1997. Silk tree, guanacaste, monkey's earring: a generic system of the synandrous Mimosaceae of the Americas. Part II. *Pithecellobium*, *Cojoba*, and *Zygia*. *Memoirs of the New York Botanical Garden* 74(2).

Boatwright, J. S., O. Maurin, and M. van der Bank. 2015. Phylogenetic position of Madagascan species of *Acacia* s.l. and new combinations in *Senegalia* and *Vachellia* (Fabaceae, Mimosoideae, Acacieae). *Botanical Journal of the Linnean Society* 179: 288–294.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

Bouchenak-Khelladi, Y., O. Maurin, J. Hurter, and M. Van der Bank. 2010. The evolutionary history and biogeography of Mimosoideae (Leguminosae): an emphasis on African acacias. *Molecular Phylogenetics and Evolution* 57: 495–508.

Brown, G. K. 2008. Systematics of the tribe Ingeae (Leguminosae-Mimosoideae) over the past 25 years. *Muelleria* 26: 27–42.

Brown, G. K., D. J. Murphy, J. T. Miller, and P. Y. Ladiges. 2008. *Acacia* s.s. and its relationship among tropical legumes, tribe Ingeae (Leguminosae: Mimosoideae). *Systematic Botany* 33: 739–751.

Brown, J. W., J. F. Walker, and S. A. Smith. 2017. Phyx: phylogenetic tools for unix. *Bioinformatics* 33: 1886–1888.

Cannon, S. B., M. R. McKain, A. Harkess, M. N. Nelson, S. Dash, M. K. Deyholos, Y. Peng, et al. 2015. Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. *Molecular Biology and Evolution* 32: 193–210.

Catalano, S. A., J. C. Vilardi, D. Tosto, and B. O. Saidman. 2008. Molecular phylogeny and diversification history of *Prosopis* (Fabaceae: Mimosoideae). *Biological Journal of the Linnean Society* 93: 621–640.

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, R. A. Price, et al. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. *Annals of the Missouri Botanical Garden* 80: 528–580.

Copetti, D., A. Búrquez, E. Bustamante, J. L. Charboneau, K. L. Childs, L. E. Eguiarte, S. Lee, et al. 2017. Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. *Proceedings of the National Academy of Sciences* 114: 12003–12008.

Couvreur, T. L., A. J. Helmstetter, E. J. Koenen, K. Bethune, R. D. Brandão, S. A. Little, H. Sauquet, and R. H. Erkens. 2019. Phylogenomics of the major tropical plant family Annonaceae using targeted enrichment of nuclear genes. *Frontiers in Plant Science* 9: 1941.

Criscuolo, A., and S. Gribaldo. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology* 10: 210.

Dahmer, N., M. F. Simon, M. T. Schifino-Wittmann, C. E. Hughes, S. T. S. Miotto, and J. C. Giuliani. 2011. Chromosome numbers in the genus *Mimosa* L.: cytotaxonomic and evolutionary implications. *Plant Systematics and Evolution* 291: 211–220.

De Smet, R., K. L. Adams, K. Vandepoele, M. C. Van Montagu, S. Maere, and Y. Van de Peer. 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proceedings of the National Academy of Sciences* 110: 2898–2903.

De Souza, É. R., G. P. Lewis, F. Forest, A. S. Schnadelbach, C. van den Berg, and L. P. de Queiroz. 2013. Phylogeny of *Calliandra* (Leguminosae: Mimosoideae) based on nuclear and plastid molecular markers. *Taxon* 62: 1200–1219.

De Souza, É. R., M. V. Krishnaraj, and L. P. de Queiroz. 2016. *Sanjappa*, a new genus in the tribe Ingeae (Leguminosae, Mimosoideae) from India. *Rheedea* 26: 1–12.

Doronina, L., G. Churakov, J. Shi, J. Brosius, R. Baertsch, H. Clawson, and J. Schmitz. 2015. Exploring massive incomplete lineage sorting in arctoids (Laurasiatheria, Carnivora). *Molecular Biology and Evolution* 32: 3194–3204.

Dugas, D. V., D. Hernandez, E. J. Koenen, E. Schwarz, S. Straub, C. E. Hughes, R. K. Jansen, et al. 2015. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in *clpP*. *Scientific Reports* 5: 16958.

Elias, T. S. 1981. Mimosoideae. *In* R. M. Polhill and P. H. Raven. [eds.], Advances in legume systematics, part 1, 143–152. Royal Botanic Gardens, Kew, UK.

Ferm, J., P. Korall, G. P. Lewis, and B. Ståhl. 2019. Phylogeny of the Neotropical legume genera *Zygia* and *Marmaroxylon* and close relatives. *Taxon* 68: 661–672.

Gagnon, E., J. J. Ringelberg, A. Bruneau, G. P. Lewis, and C. E. Hughes. 2019. Global Succulent Biome phylogenetic conservatism across the pantropical Caesalpinia Group (Leguminosae). *New Phytologist* 222: 1994–2008.

Govindarajulu, R., C. E. Hughes, P. J. Alexander, and C. D. Bailey. 2011. The complex evolutionary dynamics of ancient and recent polyploidy in *Leucaena* (Leguminosae; Mimosoideae). *American Journal of Botany* 98: 2064–2076.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, et al. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology* 29: 644.

Guerra, E., M. P. Morim, and J. R. V. Iganci. 2016. A new species of *Abarema* (Fabaceae) from Brazil. *Phytotaxa* 289: 77–82.

Guinet, P. 1981. Mimosoideae: the characters of their pollen grains. *In* R. M. Polhill and P. H. Raven. [eds.], Advances in legume systematics, part 2, 835–857. Royal Botanic Gardens, Kew, UK.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8: 1494.

Hawthorne, W. D., and C. C. Jongkind. 2006. Woody plants of Western African forests, A guide to the forest trees, shrubs and lianes from Senegal to Ghana. Royal Botanic Gardens, Kew.

Huang, S., W. Wu, Z. Chen, Q. Zhu, W. L. Ng, and Q. Zhou. 2018. Characterization of the chloroplast genome of *Erythrophleum fordii* (Fabaceae). *Conservation Genetics Resources* 11: 165–167.

Hughes, C. E., C. D. Bailey, S. Krosnick, and M. A. Luckow. 2003. Relationships among genera of the informal Dichrostachys and Leucaena groups (Mimosoideae) inferred from nuclear ribosomal ITS sequences. *In* B. B. Klitgaard and A. Bruneau [eds.], Advances in Legume Systematics, part 10, Higher Level Systematics, 221–238. Royal Botanic Gardens, Kew, UK.

Huson, D. H. 1998. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14: 68–73.

Iganci, J. R., M. V. Soares, E. Guerra, and M. P. Morim. 2016. A preliminary molecular phylogeny of the Abarema alliance (Leguminosae) and implications for taxonomic rearrangement. *International Journal of Plant Sciences* 177: 34–43.

Jobson, R. W., and M. Luckow. 2007. Phylogenetic study of the genus *Piptadenia* (Mimosoideae: Leguminosae) using plastid trnL-F and trnK/matK sequence data. *Systematic Botany* 32: 569–575.

Johnson, M. G., E. M. Gardner, Y. Liu, R. Medina, B. Goffinet, A. J. Shaw, N. J. Zerega, and N. J. Wickett. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* 4: 1600016.

Johnson, M. G., L. Pokorny, S. Dodsworth, L. R. Botigué, R. S. Cowan, A. Devault, W. L. Eiserhardt, et al. 2018. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology* 68: 594–606.

Katoh, K., K. I. Kuma, H. Toh, and T. Miyata. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* 33: 511–518.

Kent, W. J. 2002. BLAT—the BLAST-like alignment tool. *Genome Research* 12: 656–664.

Kobert, K., L. Salichos, A. Rokas, and A. Stamatakis. 2016. Computing the internode certainty and related measures from partial gene trees. *Molecular Biology and Evolution* 33: 1606–1617.

Koenen, E. J., D. I. Ojeda, R. Steeves, J. Migliore, F. T. Bakker, J. J. Wieringa, C. Kidner, et al. 2020a. Large-scale genomic sequence data resolve the deepest divergences in the legume phylogeny and support a near-simultaneous evolutionary origin of all six subfamilies. *New Phytologist* 225: 1355–1369.

Koenen, E. J., D. I. Ojeda, F. T. Bakker, J. J. Wieringa, C. Kidner, O. J. Hardy, R. T. Pennington, et al. 2020b. The origin of the legumes is a complex paleopolyploid phylogenomic tangle closely associated with the Cretaceous-Paleogene (K-Pg) mass extinction event. *Systematic Biology* syaa041.

Kyalangalilwa, B., J. S. Boatwright, B. H. Daru, O. Maurin, and M. van der Bank. 2013. Phylogenetic position and revised classification of *Acacia* sl (Fabaceae:

Mimosoideae) in Africa, including new combinations in *Vachellia* and *Senegalia*. *Botanical Journal of the Linnean Society* 172: 500–523.

Lamesch, P., T. Z. Berardini, D. Li, D. Swarbreck, C. Wilks, R. Sasidharan, R. Muller, et al. 2011. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* 40: D1202–D1210.

Lartillot, N., and H. Philippe. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution* 21: 1095–1109.

Lartillot, N., H. Brinkmann, and H. Philippe. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evolutionary Biology* 7: S4.

Lartillot, N., N. Rodrigue, D. Stubbs, and J. Richer. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Systematic Biology* 62: 611–615.

Le, S. Q., N. Lartillot, and O. Gascuel. 2008. Phylogenetic mixture models for proteins. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363: 3965–3976.

Lee, E. K., A. Cibrian-Jaramillo, S. O. Kolokotronis, M. S. Katari, A. Stamatakis, M. Ott, J. C. Chiu, et al. 2011. A functional phylogenomic view of the seed plants. *PLoS Genetics* 7: e1002411.

Lewis, G. P. and T. S. Elias. 1981. Tribe 3. Mimoseae Bronn (1822). *In* R. M. Polhill and P. H. Raven [eds.], Advances in legume systematics, part 1, 155–168. Royal Botanic Gardens, Kew, UK.

Lewis, G. P., and M. L. Rico Arce. 2005. Tribe Ingeae. *In* G. Lewis, B. Schrire, B. Mackinder, and M. Lock. [eds.], Legumes of the world, 193–213, Royal Botanic Gardens, Kew, UK.

Lewis, G., B. Schrire, B. Mackinder, and M. Lock. 2005. Legumes of the World, Royal Botanic Gardens, Kew, UK.

Li, W., and A. Godzik. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658–1659.

LPWG. 2013. Legume phylogeny and classification in the 21st century: progress, prospects and lessons for other species–rich clades. *Taxon* 62: 217–248.

LPWG. 2017. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny: The Legume Phylogeny Working Group (LPWG). *Taxon* 66: 44–77.

Luckow, M., J. T. Miller, D. J. Murphy, and T. Livshultz. 2003. A phylogenetic analysis of the Mimosoideae (Leguminosae) based on chloroplast DNA sequence data. *In* B. B. Klitgaard and A. Bruneau [eds.], Advances in Legume Systematics, part 10, Higher Level Systematics, 197–220. Royal Botanic Gardens, Kew, UK.

Luckow, M., R. H. Fortunato, S. Sede, and T. Livshultz. 2005. The phylogenetic affinities of two mysterious monotypic mimosoids from southern South America. *Systematic Botany* 30: 585–602.

Magee, A. M., S. Aspinall, D. W. Rice, B. P. Cusack, M. Sémon, A. S. Perry, S. Stefanović, et al. 2010. Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Research* 20: 1700–1710.

Mai, U., and S. Mirarab. 2018. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics* 19: 272.

Mandel, J. R., R. B. Dikow, V. A. Funk, R. R. Masalia, S. E. Staton, A. Kozik, R. W. Michelmore, et al. 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: an example from the Compositae. *Applications in Plant Sciences* 2: 1300085.

Manzanilla, V., and A. Bruneau. 2012. Phylogeny reconstruction in the Caesalpinieae grade (Leguminosae) based on duplicated copies of the sucrose synthase gene and plastid markers. *Molecular Phylogenetics and Evolution* 65: 149–162.

Marazzi, B., A. M. Gonzalez, A. Delgado-Salinas, M. A. Luckow, J. J. Ringelberg, and C. E. Hughes. 2019. Extrafloral nectaries in Leguminosae: phylogenetic distribution, morphological diversity and evolution. *Australian Systematic Botany* 32: 409–458.

Miller, J. T., J. W. Grimes, D. J. Murphy, R. J. Bayer, and P. Y. Ladiges. 2003. A phylogenetic analysis of the Acacieae and Ingeae (Mimosoideae: Fabaceae) based on *trnK*, *matK*, *psbA-trnH*, and *trnL/trnF* sequence data. *Systematic Botany* 28: 558–567.

Miller, J. T., V. Terra, C. Riggins, J. E. Ebinger, and D. S. Seigler. 2017. Molecular phylogenetics of *Parasenegalia* and *Pseudosenegalia* (Fabaceae: Mimosoideae). *Systematic Botany* 42: 465–469.

Moore, A. J., J. M. D. Vos, L. P. Hancock, E. Goolsby, and E. J. Edwards. 2017. Targeted enrichment of large gene families for phylogenetic inference: phylogeny and molecular evolution of photosynthesis genes in the portullugo clade (Caryophyllales). *Systematic Biology* 67: 367–383.

Naciri, Y., and H. P. Linder. 2015. Species delimitation and relationships: the dance of the seven veils. *Taxon* 64: 3–16.

Nicholls, J. A., R. T. Pennington, E. J. Koenen, C. E. Hughes, J. Hearn, L. Bunnefeld, K. G. Dexter, et al. 2015. Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). *Frontiers in Plant Science* 6: 710.

Nielsen, I. C. 1981. Tribe 5. Ingeae. *In* R. M. Polhill and P. H. Raven [eds.], Advances in legume systematics, part 1, 173–190. Royal Botanic Gardens, Kew, UK.

Nielsen, I. C., P. Guinet, and T. Baretta-Kuipers. 1983. Studies in the Malesian, Australian and Pacific Ingeae (Leguminosae-Mimosoideae): the genera *Archidendropsis*, *Wallaceodendron*, *Paraserianthes*, *Pararchidendron* and *Serianthes*, part I. *Bulletin du Muséum national d'histoire naturelle. Section B, Adansonia* 5: 303–329.

Nielsen, I. C.1992. Flora Malesiana-Series I, Volume 11, part I: Mimosaceae (Leguminosae-Mimosoideae). Rijksherbarium/Hortus Botanicus, Leiden University, Leiden, The Netherlands.

Ojeda, D. I., E. Koenen, S. Cervantes, M. de la Estrella, E. Banguera-Hinestroza, S. B. Janssens, J. Migliore, et al. 2019. Phylogenomic analyses reveal an exceptionally high number of evolutionary shifts in a florally diverse clade of African legumes. *Molecular Phylogenetics and Evolution* 137: 156–167.

Pennington, R. T., D. E. Prado, and C. A. Pendry. 2000. Neotropical seasonally dry forests and Quaternary vegetation changes. *Journal of Biogeography* 27: 261–273.

Pennington, R. T., M. Lavin, and A. Oliveira-Filho. 2009. Woody plant diversity, evolution, and ecology in the tropics: perspectives from seasonally dry tropical forests. *Annual Review of Ecology, Evolution, and Systematics* 40: 437–457.

Philippe, H., H. Brinkmann, D. V. Lavrov, D. T. J. Littlewood, M. Manuel, G. Wörheide, and D. Baurain. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology* 9: e1000602.

Polhill, R. M., and P. H. Raven. 1981. Advances in legume systematics, part 1. Royal Botanic Gardens, Kew, UK.

Queiroz, L. P., D. Cardoso, M. F. Fernandes, and M. Moro. 2017. Diversity and evolution of flowering plants of the Caatinga domain. *In* J. C. Silva, I. Leal, and M. Tabarelli [eds.], Caatinga: the largest tropical dry forest region in South America, 23–63. Springer, Cham, Switzerland.

Ranwez, V., E. J. Douzery, C. Cambon, N. Chantret, and F. Delsuc. 2018. MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Molecular Biology and Evolution* 35: 2582–2584.

Ribeiro, P. G., M. Luckow, G. P. Lewis, M. F. Simon, D. Cardoso, É. R. de Souza, A. P. Conceicao Silva, et al. 2018. *Lachesiodendron*, a new monospecific genus segregated from *Piptadenia* (Leguminosae: Caesalpinioideae: mimosoid clade): Evidence from morphology and molecules. *Taxon* 67: 37–54.

Richards, E. J., J. M. Brown, A. J. Barley, R. A. Chong, and R. C. Thomson. 2018. Variation across mitochondrial gene trees provides evidence for systematic error: How much gene tree variation is biological? *Systematic Biology* 67: 847–860.

Ringelberg, J. J., N. E. Zimmermann, A. Weeks, M. Lavin, and C. E. Hughes. 2020. Biomes as evolutionary arenas: Convergence and conservatism in the trans-continental succulent biome. *Global Ecology and Biogeography* 00: 1–14.

Rokas, A., B. L. Williams, N. King, and S. B. Carroll. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425: 798–804.

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

Santos, E. C. X. R., R. Carvalho, E. M. Almeida, and L. P. Felix. 2012. Chromosome number variation and evolution in Neotropical Leguminoseae (Mimosoideae) from northeastern Brazil. *Genetics and Molecular Research* 11: 2451–2475.

Sass, C., W. J. Iles, C. F. Barrett, S. Y. Smith, and C. D. Specht. 2016. Revisiting the Zingiberales: using multiplexed exon capture to resolve ancient and recent phylogenetic splits in a charismatic plant lineage. *PeerJ* 4: e1584.

Sayyari, E., and S. Mirarab. 2018. Testing for polytomies in phylogenetic species trees using quartet frequencies. *Genes* 9: 132.

Sayyari, E., J. B. Whitfield, and S. Mirarab. 2018. DiscoVista: Interpretable visualizations of gene tree discordance. *Molecular Phylogenetics and Evolution* 122: 110–115.

Schmieder, R., and R. Edwards. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27: 863–864.

Schmutz, J., S. B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D. L. Hyten, et al. 2010. Genome sequence of the palaeopolyploid soybean. *Nature* 463: 178–183.

Scornavacca, C., and N. Galtier. 2017. Incomplete lineage sorting in mammalian phylogenomics. *Systematic Biology* 66: 112–120.

Shen, X. X., C. T. Hittinger, and A. Rokas. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nature Ecology & Evolution* 1: 0126.

Seigler, D. S., J. E. Ebinger, C. W. Riggins, V. Terra, and J. T. Miller. 2017. *Parasenegalia* and *Pseudosenegalia* (Fabaceae): new genera of the Mimosoideae. *Novon* 25: 180–205.

Schrire, B. D., M. Lavin, and G. P. Lewis. 2005. Global distribution patterns of the Leguminosae: insights from recent phylogenies. *Biologiske Skrifter* 55: 375–422.

Simon, M. F., J. F. B. Pastore, A. F. Souza, L. M. Borges, V. R. Scalon, P. G. Ribeiro, J. Santos-Silva, et al. 2016. Molecular phylogeny of *Stryphnodendron* (Mimosoideae, Leguminosae) and generic delimitations in the Piptadenia group. *International Journal of Plant Sciences* 177: 44–59.

Smith, S. A., M. J. Moore, J. W. Brown, and Y. Yang. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology* 15: 150.

Smith, S. A., N. Walker-Hale, and J. F. Walker 2020. Intragenic conflict in phylogenomic datasets. *Molecular Biology and Evolution* msaa170.

Sprent, J. I. 2007. Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. *New Phytologist* 174: 11–25.

Springer, M. S., E. K. Molloy, D. B. Sloan, M. P. Simmons, and J. Gatesy. 2020. ILS-Aware Analysis of Low-Homoplasy Retroelement Insertions: Inference of Species Trees and Introgression Using Quartets. *Journal of Heredity* esz076.

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

Suh, A., L. Smeds, and H. Ellegren. 2015. The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS Biology* 13: e1002224.

Suh, A. 2016. The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. *Zoologica Scripta* 45: 50–62.

Terra, V., F. C. Garcia, L. P. de Queiroz, M. van der Bank, and J. T. Miller. 2017. Phylogenetic relationships in *Senegalia* (Leguminosae-Mimosoideae) emphasizing the south American lineages. *Systematic Botany* 42: 458–464.

Vatanparast, M., A. Powell, J. J. Doyle, and A. N. Egan. 2018. Targeting legume loci: A comparison of three methods for target enrichment bait design in Leguminosae phylogenomics. *Applications in Plant Sciences* 6: e1036.

Walker, J. F., J. W. Brown, and S. A. Smith. 2018. Analysing contentious relationships and outlier genes in phylogenomics. *Systematic Biology* 67: 916–924.

Weitemier, K., S. C. Straub, R. C. Cronn, M. Fishbein, R. Schmickl, A. McDonnell, and A. Liston. 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 2: 1400042.

Wen, J., Z. Xiong, Z. L. Nie, L. Mao, Y. Zhu, X. Z. Kan, S. M. Ickert-Bond, et al. 2013. Transcriptome sequences resolve deep relationships of the grape family. *PLoS One* 8: e74394.

Whitfield, J., S. A. Cameron, D. Huson, and M. Steel. 2008. Filtered Z-closure supernetworks for extracting and visualizing recurrent signal from incongruent gene trees. *Systematic Biology* 57: 939–947.

Wickett, N. J., S. Mirarab, N. Nguyen, T. Warnow, E. Carpenter, N. Matasci, S. Ayyampalayam, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences, USA* 111: E4859–E4868.

Williams, A. V., L. M. Boykin, K. A. Howell, P. G. Nevill, and I. Small. 2015. The complete sequence of the *Acacia ligulata* chloroplast genome reveals a highly divergent *clpP1* gene. *PLoS One* 10: p.e0125768.

Wu, F., L. A. Mueller, D. Crouzillat, V. Pétiard, and S. D. Tanksley. 2006. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. *Genetics* 174: 1407–1420.

Yang, Y., and S. A. Smith. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Molecular Biology and Evolution* 31: 3081–3092.

Yang, Y., M. J. Moore, S. F. Brockington, D. E. Soltis, G. K. S. Wong, E. J. Carpenter, Y. Zhang, et al. 2015. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Molecular Biology and Evolution* 32: 2001–2014.

Young, N. D., F. Debellé, G. E. Oldroyd, R. Geurts, S. B. Cannon, M. K. Udvardi, V. A. Benedito, et al. 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480: 520–524.

Zeng, L., N. Zhang, Q. Zhang, P. K. Endress, J. Huang, and H. Ma. 2017. Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. *New Phytologist* 214: 1338–1354.

Zhang, J., K. Kobert, T. Flouri, and A. Stamatakis. 2013. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30: 614–620.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19: 153.

APPENDIX 1. Voucher details, repository accession numbers and sequencing results for the 122 accessions used in this study.

Taxon	Voucher	ENA accession number	Total number of reads	Reads on target	No. of targets recovered	No. of gene alignments
Abarema cochliacarpos (Gomes) Barneby & J.W.Grimes	L.P. de Queiroz 15538 (HUEFS)	ERS4812838	23779774	19277048 (81.06%)	940 (97.51%)	1078 (56.29%)
Acacia longifolia (Andrews) Willd,	E. Koenen 182 (Z)	ERS4812840	4142738	288693 (6.97%)	830 (86.10%)	520 (27.15%)
Acaciella villosa (Sw.) Britton & Rose	C.E. Hughes 2635 (FHO)	ERS4812841	3393924	129207 (3.81%)	644 (66.80%)	274 (14.31%)
Adenanthera pavonina L.	Ambriansyah & Arifin AA295 (K)	ERS4812842	4812194	251953 (5.24%)	879 (91.18%)	606 (31.64%)
Adenopodia patens (Hook. & Arn.) J.R.Dixon ex Brenan	Sandoval MS343 (K)	ERS4812843	7863140	791520 (10.07%)	919 (95.33%)	769 (40.16%)
Adenopodia scelerata (A. Chev.) Brenan	C. Jongkind 10602 (WAG)	ERS4812844	10094360	1357624 (13.45%)	950 (98.55%)	902 (47.10%)
Alantsilodendron pilosum Villiers	E. Koenen 203 (Z)	ERS4812845	3468930	364779 (10.52%)	933 (96.78%)	806 (42.09%)
Albizia adianthifolia (Schumach.) W.Wight	J.J. Wieringa 6278 (WAG)	ERS4812846	10128880	1151955 (11.37%)	945 (98.03%)	976 (50.97%)
Albizia altissima Hook.f.	C. Jongkind 10709 (WAG)	ERS4812847	23622566	18312807 (77.52%)	943 (97.82%)	1097 (57.28%)
Albizia anthelmintica Brongn.	O. Maurin 0363 (JRAU)	ERS4812848	5576920	754863 (13.54%)	934 (96.89%)	894 (46.68%)

(Continued)

APPENDIX 1. (Continued)

Taxon	Voucher	ENA accession number	Total number of reads	Reads on target	No. of targets recovered	No. of gene alignments
Albizia atakataka Capuron	E. Koenen 229 (Z)	ERS4812849	47608874	33141289 (69.61%)	940 (97.51%)	1010 (52.74%)
Albizia aurisparsa (Drake) R.Vig.	E. Koenen 230 (Z)	ERS4812850	15816078	2127848 (13.45%)	952 (98.76%)	1047 (54.67%)
Albizia bernieri E. Fourn. ex Villiers	E. Koenen 354 (Z)	ERS4812851	3752342	331728 (8.84%)	898 (93.15%)	692 (36.14%)
Albizia boivinii E. Fourn.	E. Koenen 270 (Z)	ERS4812852	3677634	453433 (12.33%)	925 (95.95%)	882 (46.06%)
Albizia brevifolia Schinz	O. Maurin 0826 (JRAU)	ERS4812853	2819456	327264 (11.61%)	878 (91.08%)	678 (35.40%)
Albizia burkartiana Barneby & J.W.Grimes	Stival-Santos 678 (RB)	ERS4812854	6094776	465465 (7.64%)	927 (96.16%)	882 (46.06%)
Albizia dinklagei Harms	C. Jongkind 7359 (WAG)	ERS4812855	2074148	144269 (6.96%)	912 (94.61%)	802 (41.88%)
Albizia edwallii (Hoehne) Barneby & J.W.Grimes	Dalmaso 272 (RB)	ERS4812856	3339160	359353 (10.76%)	931 (96.58%)	947 (49.45%)
Albizia ferruginea (Guill. & Perr.) Benth.	C. Jongkind 10762 (WAG)	ERS4812857	6722992	1035687 (15.41%)	936 (97.10%)	968 (50.55%)
Albizia grandibracteata Taub.	E. Koenen 159 (WAG)	ERS4812858	38109552	27944549 (73.33%)	946 (98.13%)	950 (49.61%)
Albizia inundata (Mart.) Barneby & J.W.Grimes	J.R.I. Wood 26530 (K)	ERS4812859	35775492	25641806 (71.67%)	942 (97.72%)	965 (50.39%)
Albizia mahalao Capuron	E. Koenen 216 (Z)	ERS4812860	70271424	55605048 (79.13%)	946 (98.13%)	906 (47.31%)
Albizia masikororum R.Vig.	E. Koenen 237 (Z)	ERS4812861	12599676	1562117 (12.40%)	953 (98.86%)	1024 (53.47%)
Albizia obbiadensis (Chiov.) Brenan	Thulin 4163 (UPS)	ERS4812862	5614760	735383 (13.10%)	940 (97.51%)	937 (48.93%)
Albizia obliquifoliolata De Wild.	J.J. Wieringa 6519 (WAG)	ERS4812863	13303218	10816943 (81.31%)	941 (97.61%)	1047 (54.67%)
Albizia polyphylla E.Fourn.	E. Koenen 256 (Z)	ERS4812864	3215066	434008 (13.50%)	932 (96.68%)	843 (44.02%)
Albizia retusa Benth.	Hyland 2732 (L)	ERS4812865	11996368	1476589 (12.31%)	948 (98.34%)	1004 (52.43%)
Albizia sahafariensis Capuron	E. Koenen 405 (Z)	ERS4812866	12994846	1600201 (12.31%)	945 (98.03%)	1011 (52.79%)
Albizia saponaria (Lour.) Blume	Jobson 1041 (BH)	ERS4812867	39202190	27805263 (70.93%)	944 (97.93%)	998 (52.11%)
Albizia versicolor Welw. ex Oliv,	O. Maurin 560 (JRAU)	ERS4812868	66547258	53198730 (79.94%)	945 (98.03%)	1045 (54.57%)
Albizia viridis E.Fourn.	Du Puy M251 (K)	ERS4812869	7260284	870430 (11.99%)	934 (96.89%)	988 (51.59%)
Albizia zygia (DC.) J.F.Macbr.	J.J. Wieringa 5915 (WAG)	ERS4812870	8003478	793032 (9.91%)	941 (97.61%)	977 (51.02%)
Amblygonocarpus andongensis (Welw. ex Oliv.) Exell & Torre	Sokpon 1451 (WAG)	ERS4812871	5307456	263884 (4.97%)	843 (87.45%)	569 (29.71%)
Anadenanthera colubrina (Vell.) Brenan	L.P. de Queiroz 15685 (HUEFS)	ERS4812872	4286504	491557 (11.47%)	929 (96.37%)	841 (43.92%)
Archidendron lucidum (Benth.) I.C.Nielsen	Wang and Lin 2534 (L)	ERS4812873	6285326	658012 (10.47%)	939 (97.41%)	972 (50.76%)
Archidendron quocense (Pierre) I.C.Nielsen	Newman 2094 (E)	ERS4812874	55446888	41045031 (74.03%)	947 (98.24%)	972 (50.76%)
Archidendropsis granulosa (Labill.) I.C.Nielsen	McKee 38353 (L)	ERS4812875	13150138	1492706 (11.35%)	947 (98.24%)	1047 (54.67%)
Aubrevillea kerstingii (Harms) Pellegr.	Nimba Botanic Team JR957 (WAG)	ERS4812876	6327042	343767 (5.43%)	936 (97.10%)	770 (40.21%)
Balizia pedicellaris (DC.) Barneby & J.W.Grimes	L.P. de Queiroz 15529 (HUEFS)	ERS4812877	28193862	22668050 (80.40%)	941 (97.61%)	1104 (57.65%)
Balizia sp.nov.	M.P. Morim 577 (RB)	ERS4812878	21239644	16903890 (79.59%)	936 (97.10%)	1071 (55.93%)
Blanchetiodendron blanchetii (Benth.) Barneby & J.W.Grimes	L.P. de Queiroz 15616 (HUEFS)	ERS4812879	6639992	780827 (11.76%)	936 (97.10%)	965 (50.39%)
Calliandra hygrophila Mackinder & G.P.Lewis	L.P. de Queiroz 15542 (HUEFS)	ERS4812880	4127232	483827 (11.72%)	910 (94.40%)	732 (38.22%)
Calpocalyx dinklagei Harms	J.J. Wieringa 6094 (WAG)	ERS4812881	11391816	614443 (5.39%)	929 (96.37%)	671 (35.04%)
Cathormion umbellatum Kosterm.	Jobson 1037 (BH)	ERS4812882	26129888	20718828 (79.29%)	944 (97.93%)	1118 (58.38%)
Cedrelinga cateniformis (Ducke) Ducke	T.D. Pennington 17761 (K)	ERS4812883	4070738	406653 (9.99%)	919 (95.33%)	803 (41.93%)
Chidlowia sanguinea Hoyle	J.J. Wieringa 4338 (WAG)	ERS4812884	9263792	438049 (4.73%)	888 (92.12%)	584 (30.50%)
Chloroleucon tenuiflorum (Benth.) Barneby & J.W.Grimes	L.P. de Queiroz 15514 (HUEFS)	ERS4812885	7301118	779106 (10.67%)	945 (98.03%)	1031 (53.84%)
Cojoba arborea (L.) Britton & Rose	M.F. Simon 1545 (CEN)	ERS4812886	9948972	1062718 (10.68%)	954 (98.96%)	1095 (57.18%)
Cylicodiscus gabunensis Harms	M. Sosef 645A (WAG)	ERS4812887	6792968	649666 (9.56%)	951 (98.65%)	943 (49.24%)
Desmanthus leptophyllus Kunth	C.E. Hughes 2035 (FHO)	ERS4812888	4816620	392291 (8.14%)	923 (95.75%)	816 (42.61%)
Dichrostachys cinerea (L.) Wight & Arn.	O. Maurin 256 (JRAU)	ERS4812889	4876856	416124 (8.53%)	935 (96.99%)	822 (42.92%)
Dimorphandra macrostachya Benth.	J.R. Iganci 877 (RB)	ERS4812890	6731034	248839 (3.70%)	935 (96.99%)	689 (35.98%)
Diptychandra aurantiaca Tul.	J.R.I. Wood 26513 (K)	ERS4812891	8520962	117138 (1.37%)	881 (91.39%)	400 (20.89%)
Ebenopsis confinis (Standl.) Britton & Rose	C.E. Hughes 1539 (FHO)	ERS4812892	5779758	654578 (11.33%)	936 (97.10%)	927 (48.41%)
Elephantorrhiza elephantina (Burch.) Skeels	KMS198 (JRAU)	ERS4812893	7379446	717080 (9.72%)	946 (98.13%)	765 (39.95%)
Entada rheedei Spreng.	E. Koenen 496 (Z)	ERS4812894	8695656	531548 (6.11%)	948 (98.34%)	661 (34.52%)
Enterolobium contortisiliquum (Vell.) Morong	L.P. de Queiroz 15579 (HUEFS)	ERS4812895	2729658	240130 (8.80%)	919 (95.33%)	868 (45.33%)
Erythrophleum ivorense A.Chev.	J.J. Wieringa 5487 (WAG)	ERS4812896	11500640	485354 (4.22%)	947 (98.24%)	719 (37.55%)
Faidherbia albida (Delile) A.Chev.	O. Maurin 3495 (JRAU)	ERS4812897	6376338	734941 (11.53%)	945 (98.03%)	946 (49.40%)
Falcataria moluccana (Miq.) Barneby & J.W.Grimes	Ambri & Arifin W826A (K)	ERS4812898	7669018	815087 (10.63%)	946 (98.13%)	991 (51.75%)

(Continued)

APPENDIX 1. (Continued)

Taxon	Voucher	ENA accession number	Total number of reads	Reads on target	No. of targets recovered	No. of gene alignments
Fillaeopsis discophora Harms	J.J. Wieringa 5498 (WAG)	ERS4812899	2259316	111269 (4.92%)	816 (84.65%)	597 (31.17%)
Havardia pallens (Benth.) Britton & Rose	C.E. Hughes 2138 (FHO)	ERS4812900	6521266	726457 (11.14%)	943 (97.82%)	1056 (55.14%)
Hesperalbizia occidentalis (Brandegee) Barneby & J.W.Grimes	C.E. Hughes 1296 (FHO)	ERS4812901	5403622	788809 (14.60%)	947 (98.24%)	1032 (53.89%)
Hydrochorea corymbosa (Rich.) Barneby & J.W.Grimes [1]	F. Bonadeu 655 (RB)	ERS4812902	39645090	27356455 (69.00%)	943 (97.82%)	1028 (53.68%)
Hydrochorea corymbosa (Rich.) Barneby & J.W.Grimes [2]	J.R. Iganci 862 (RB)	ERS4812903	19909090	15987983 (80.30%)	944 (97.93%)	1071 (55.93%)
Inga alba (Sw.) Willd.	P.D. Coley & T.A. Kursar TAKPDC1677 (UT)	ERR776844	1658880	1363817 (82.21%)	942 (97.72%)	1062 (55.46%)
Inga edulis Mart,	P.D. Coley & T.A. Kursar TAKPDC1719 (UT)	ERR776838	1617410	1324567 (81.89%)	934 (96.89%)	1076 (56.19%)
Inga huberi Ducke	P.D. Coley & T.A. Kursar TAKPDC1755 (UT)	ERR776810	1555208	1291086 (83.02%)	937 (97.20%)	1085 (56.66%)
Inga laurina (Sw.) Willd.	K.G. Dexter 398 (E)	ERR776816	1612110	1374610 (85.27%)	944 (97.93%)	1053 (54.99%)
Inga stipularis DC.	P.D. Coley & T.A. Kursar TAKPDC1856 (UT)	ERR776821	1692290	1393432 (82.34%)	940 (97.51%)	1055 (55.09%)
Inga tenuistipula Ducke	K.G. Dexter 110 (E)	ERR776831	1388002	1125394 (81.08%)	938 (97.30%)	1077 (56.24%)
Jupunba trapezifolia Moldenke	M.F. Simon 1600 (CEN)	ERS4812839	16357084	13117719 (80.20%)	945 (98.03%)	1042 (54.41%)
Kanaloa kahoolawensis Lorence & K.R.Wood	Lorence 7380 (PTBG)	ERS4812904	12222002	1915460 (15.67%)	956 (99.17%)	933 (48.72%)
Lachesiodendron viridiflorum (Kunth) P.G.Ribeiro, L.P.Queiroz & Luckow	L.P. de Queiroz 15614 (HUEFS)	ERS4812905	18632852	2381616 (12.78%)	957 (99.27%)	973 (50.81%)
Lemurodendron capuronii Villiers & P.Guinet	E. Koenen 435 (Z)	ERS4812906	7108042	881933 (12.41%)	947 (98.24%)	1000 (52.22%)
Leucochloron bolivianum C.E. Hughes & Atahuachi	C.E. Hughes 2608 (FHO)	ERS4812907	7946434	1218355 (15.33%)	950 (98.55%)	1046 (54.62%)
Leucochloron limae Barneby & J.W.Grimes	MWC8250 (K)	ERS4812908	7767490	965594 (12.43%)	949 (98.44%)	1078 (56.29%)
Lysiloma candidum Brandegee	B. Marazzi 300 (ASU)	ERS4812909	2030974	102461 (5.04%)	753 (78.11%)	428 (22.35%)
Macrosamanea amplissima (Ducke) Barneby & J.W.Grimes	Bonadeu 663 (RB)	ERS4812910	2360238	217690 (9.22%)	920 (95.44%)	824 (43.03%)
Mariosousa sericea (M.Martens & Galeotti) Seigler & Ebinger	MWC18949 (K)	ERS4812911	8160316	1450135 (17.77%)	951 (98.65%)	1011 (52.79%)
Mimosa grandidieri Baill.	E. Koenen 207 (Z)	ERS4812912	7792272	717042 (9.20%)	951 (98.65%)	795 (41.51%)
Mimosa tenuiflora (Willd.) Poir.	L.P. de Queiroz 15498 (HUEFS)	ERS4812913	6210710	475738 (7.66%)	944 (97.93%)	799 (41.72%)
Mimozyganthus carinatus (Griseb.) Burkart	C.E. Hughes 2476 (FHO)	ERS4812914	8148502	817441 (10.03%)	943 (97.82%)	944 (49.30%)
Neptunia oleracea Lour.	E. Koenen 283 (Z)	ERS4812915	10836680	1176757 (10.86%)	945 (98.03%)	861 (44.96%)
Newtonia hildebrandtii (Vatke) Torre	O. Maurin 2457 (JRAU)	ERS4812916	8663120	826146 (9.54%)	948 (98.34%)	914 (47.73%)
Pachyelasma tessmannii (Harms) Harms	J.J. Wieringa 5229 (WAG)	ERS4812917	11845384	793886 (6.70%)	954 (98.96%)	766 (40.00%)
Parapiptadenia zehntneri (Harms) M.P.Lima & H.C.Lima	L.P. de Queiroz 15692 (HUEFS)	ERS4812918	4446508	378119 (8.50%)	932 (96.68%)	939 (49.03%)
Pararchidendron pruinosum (Benth.) I.C.Nielsen	Jobson 1039 (BH)	ERS4812919	7647352	738506 (9.66%)	952 (98.76%)	1052 (54.93%)
Paraserianthes lophantha (Willd.) I.C.Nielsen	M. van Slageren & R. Newton MSRN648 (K)	ERS4812920	6378910	751573 (11.78%)	950 (98.55%)	1048 (54.73%)
Parkia panurensis Benth. ex H.C.Hopkins	J.R. Iganci 842 (RB)	ERS4812921	2640302	231835 (8.78%)	907 (94.09%)	814 (42.51%)
Peltophorum africanum Sond.	E. Koenen 601 (Z)	ERS4812922	2910944	145733 (5.01%)	717 (74.38%)	367 (19.16%)
Pentaclethra macrophylla Benth.	Galeuchet & Balthazar 10 (Z)	ERS4812923	18158278	776900 (4.28%)	949 (98.44%)	734 (38.33%)
Piptadenia robusta Pittier	M. Luckow 4633 (BH)	ERS4812924	3486554	371485 (10.65%)	938 (97.30%)	898 (46.89%)
Piptadeniastrum africanum (Hook.f.) Brenan	E. Koenen 152 (WAG)	ERS4812925	8894316	514787 (5.79%)	948 (98.34%)	741 (38.69%)
Piptadeniopsis lomentifera Burkart	M. Luckow 4505 (BH)	ERS4812926	6399676	826642 (12.92%)	947 (98.24%)	926 (48.36%)
Pithecellobium dulce (Roxb.) Benth.	B. Marazzi 309 (ASU)	ERS4812927	6485068	881345 (13.59%)	954 (98.96%)	1061 (55.40%)
Pityrocarpa moniliformis (Benth.) Luckow & R.W. Jobson	J.R.I. Wood 26516 (K)	ERS4812928	6003692	449263 (7.48%)	938 (97.30%)	951 (49.66%)
Plathymenia reticulata Benth.	L.P. de Queiroz 15688 (HUEFS)	ERS4812929	2477330	209417 (8.45%)	920 (95.44%)	757 (39.53%)
Prosopidastrum globosum (Gillies ex Hook. & Arn.) Burkart	M. Luckow sn (BH)	ERS4812930	6211352	691922 (11.14%)	946 (98.13%)	924 (48.25%)
Prosopis africana (Guill. & Perr.) Taub.	Essou 2110 (WAG)	ERS4812931	11459252	1374601 (12.00%)	953 (98.86%)	860 (44.91%)
Prosopis laevigata (Humb. & Bonpl. ex Willd.) M.C.Johnst.	C.E. Hughes 2058 (FHO)	ERS4812932	3353428	246735 (7.36%)	920 (95.44%)	843 (44.02%)

(Continued)

APPENDIX 1. (Continued)

Taxon	Voucher	ENA accession number	Total number of reads	Reads on target	No. of targets recovered	No. of gene alignments
Pseudopiptadenia contorta (DC.) G.P.Lewis & M.P.Lima	L.P. de Queiroz 15582 (HUEFS)	ERS4812933	7625306	719618 (9.44%)	949 (98.44%)	1009 (52.69%)
Pseudoprosopis gilletii (De Wild.) Villiers	J.J. Wieringa 6021 (WAG)	ERS4812934	5958100	264874 (4.45%)	931 (96.58%)	700 (36.55%)
Pseudosamanea guachapele (Kunth) Harms	C.E. Hughes 1198 (FHO)	ERS4812935	7396824	1018670 (13.77%)	944 (97.93%)	1015 (53.00%)
Samanea saman (Jacq.) Merr.	C.E. Hughes 421 (FHO)	ERS4812936	3344450	515562 (15.42%)	943 (97.82%)	1027 (53.63%)
Schleinitzia novoguineensis (Warb.) Verdc.	Chaplin 57/84	ERS4812937	16565732	2799712 (16.90%)	956 (99.17%)	881 (46.01%)
Senegalia ataxacantha (DC.) Kyal. & Boatwr.	C. Jongkind 10603 (WAG)	ERS4812938	11987764	1423870 (11.88%)	951 (98.65%)	941 (49.14%)
Senegalia sakalava (Drake) Boatwr.	E. Koenen 215 (Z)	ERS4812939	12102414	990240 (8.18%)	946 (98.13%)	863 (45.07%)
Serianthes nelsonii Merr.	P. Moore 1241 (L)	ERS4812940	7283252	597846 (8.21%)	943 (97.82%)	1038 (54.20%)
Sphinga acatlensis (Benth.) Barneby & J.W.Grimes	C.E. Hughes 2112 (FHO)	ERS4812941	9238996	1054313 (11.41%)	950 (98.55%)	1035 (54.05%)
Stryphnodendron pulcherrimum (Willd.) Hochr.	L.P. de Queiroz 15482 (HUEFS)	ERS4812942	11852118	1440760 (12.16%)	954 (98.96%)	1007 (52.58%)
Tachigali odoratissima (Spruce ex Benth.) Zarucchi & Herend.	M.P. Morim 562 (RB)	ERS4812943	7900532	193104 (2.44%)	925 (95.95%)	622 (32.48%)
Tetrapleura tetraptera (Schumach. & Thonn.) Taub.	E. Koenen 155 (WAG)	ERS4812944	4276206	310727 (7.27%)	933 (96.78%)	707 (36.92%)
Vachellia tortilis (Forssk.) Galasso & Banfi	E. Koenen 603 (Z)	ERS4812945	5519408	614954 (11.14%)	930 (96.47%)	830 (43.34%)
Vachellia viguieri (Villiers & Du Puy) Boatwr.	E. Koenen 199 (Z)	ERS4812946	4782514	572917 (11.98%)	945 (98.03%)	902 (47.10%)
Viguieranthus glaber Villiers	E. Koenen 325 (Z)	ERS4812947	10569806	1179380 (11.16%)	950 (98.55%)	1004 (52.43%)
Xylia hoffmannii (Vatke) Drake	E. Koenen 402 (Z)	ERS4812948	7511352	464326 (6.18%)	944 (97.93%)	713 (37.23%)
Zapoteca caracasana (Jacq.) H.M.Hern.	C.E. Hughes 3071 (FHO)	ERS4812949	5236294	475921 (9.09%)	912 (94.61%)	608 (31.75%)
Zygia claviflora (Spruce ex Benth.) Barneby & J.W.Grimes	J.R. Iganci 841 (RB)	ERS4812950	2422758	195154 (8.06%)	910 (94.40%)	784 (40.94%)
Zygia inaequalis (Humb. & Bonpl. ex Willd.) Pittier	J.R. Iganci 832 (RB)	ERS4812951	5622222	494386 (8.79%)	931 (96.58%)	899 (46.95%)
Zygia racemosa (Ducke) Barneby & J.W.Grimes	M.F. Simon 1658 (CEN)	ERS4812952	10769766	850638 (7.90%)	946 (98.13%)	1012 (52.85%)
Zygia sp.	P.D. Coley & T.A. Kursar Tip917 (UT)	ERR776824	1360502	1069298 (78.60%)	938 (97.30%)	1091 (56.97%)

APPENDIX 2. Definitions and notable features of informally recognized clades and taxonomy of *Cathormion*.

Clade definitions

Higher-level clades

Core mimosoids are here defined as the clade that includes the most recent common ancestor of *Cylicodiscus gabunensis*, *Prosopis laevigata*, *Dichrostachys cinerea*, and *Inga edulis*, and all of its decendants. This clade includes the bulk of mimosoid species, all of the larger genera, and all of the armed mimosoids (i.e., those genera and species with stipular spines, spinescent shoots, and/or prickles). It is subtended by a particularly long internode (Fig. 6), prompting us to recognize the clade here.

The **ingoid clade** is well supported in all analyses and is defined as the clade that includes all genera of tribe Ingeae plus *Acacia* and all its segregates except *Vachellia* (Fig. 6), and approximately two-thirds of all mimosoid species. Like the core mimosoids, it is subtended by a relatively long internode. All taxa in the clade share the feature of flowers with >10 stamens, which is otherwise present only in *Vachellia*. Fusion of the stamens into a tube is exclusively found in this clade and characterizes most of the genera (and is not found in *Vachellia*). In this study we are able to recognize, for the first time, several well-supported subclades within this taxonomically notoriously difficult and poorly resolved clade.

Lower-level clades

The **Xylia clade** is defined as the clade that includes all genera from the monophyletic Adenanthera group (Lewis et al., 2005) plus its sister group *Pentaclethra*. The clade includes two distinctive subclades: (1) the subclade of *Xylia*, *Pseudoprosopis*, and *Calpocalyx*, which is restricted to Africa and Madagascar and characterized by sickle-shaped explosively dehiscent fruits; and (2) the subclade comprising *Adenanthera*, *Tetrapleura*, and *Amblygonocarpus*, with the former genus largely restricted to Southeast Asia and characterized by indehiscent or nonexplosively dehiscent fruits. Since *Pentaclethra* (sister to these two subclades; see Fig. 6) also has explosively dehiscent fruit, this is likely the ancestral fruit state of the Xylia clade.

The **Entada clade** includes the genera *Entada*, *Elephantorrhiza*, *Piptadeniastrum*, and *Aubrevillea* and has its center of diversity in, and three of the four genera restricted to, continental Africa. *Entada* is more widespread, with several species in Madagascar and a few species with drift-seeds, which have attained large pantropical distributions following trans-oceanic dispersal.

The **Dichrostachys clade** includes the informal Dichrostachys and Leucaena groups (Hughes et al., 2003; Lewis et al., 2005), as well as *Mimozyganthus*, *Piptadeniopsis*, and *Prosopidastrum* (Luckow et al., 2005), and the genera *Neptunia* and *Lemurodendron*. Most taxa in this clade are found in the succulent biome (Ringelberg et al., 2020) comprising seasonally dry tropical forest and thorn scrub (SDTFs *sensu* Pennington et al., 2000, 2009; STDFWs *sensu* Queiroz et al., 2017), with centers of diversity in Mexico and Central America (the Leucaena group) and Madagascar (the Dichrostachys group). *Lemurodendron* is monotypic, narrowly endemic in northwest Madagascar, known until recently from only a handful of herbarium collections from the 1960s, and of previously unknown phylogenetic affinities. We re-collected it in 2014, allowing us to include it here. The sister-group relationship of *Lemurodendron* with *Neptunia* is surprising at first sight, given their disparate morphologies, but arguably *Neptunia* is morphologically unlike any other mimosoid because of its (semi-)aquatic lifestyle.

While not restricted to it, nor universal within it, the presence of heteromorphic inflorescences with showy staminodes at the base is highly characteristic of this clade.

The **Parkia clade** comprises the genera *Vachellia*, *Anadenanthera*, and *Parkia*. While these genera do not share any conspicuous morphological features, the clade is well supported in all analyses.

The informal Piptadenia group (*sensu* Lewis and Elias, 1981; Jobson and Luckow, 2007; minus the genus *Anadenanthera*) is paraphyletic and resolved into two well-supported clades, the **Stryphnodendron clade** and the **Mimosa clade**. The former includes *Parapiptadenia*, *Pityrocarpa*, *Pseudopiptadenia*, *Stryphnodendron*, and *Microlobius* (not sampled here) (Simon et al., 2016). The Mimosa clade includes *Adenopodia*, *Mimosa*, and *Piptadenia*. The monotypic genus *Lachesiodendron*, recently segregated from *Piptadenia* (Ribeiro et al., 2018), is placed outside both these clades and instead forms the sister group of the remainder of the Piptadenia group and the ingoid clade.

The **Calliandra clade** includes the two Neotropical genera *Acaciella* and *Calliandra* and the recently segregated African sister group of the latter, *Afrocalliandra* (de Souza et al., 2013; not sampled here). Typical for this clade is pollen aggregated in eight-celled polyads, whereas otherwise ingoid genera typically have 16-celled polyads.

The **Zapoteca clade** includes the genera *Faidherbia*, *Viguieranthus*, and *Zapoteca*, as well as *Sanjappa* and *Thailentadopsis* (de Souza et al., 2016), which are not sampled here. The clade has a pantropical distribution. Typical for this clade are the fruits elastically dehiscent from the apex, similar to those of *Calliandra* and lacking only in *Faidherbia*. Spinescent stipules are found in *Faidherbia*, *Sanjappa*, and *Thailentadopsis*.

The **Pithecellobium clade** is identical to the Pithecellobium alliance of Barneby and Grimes (1996). The clade is native to the Americas, has its center of diversity in Mexico and Central America, and is characterized by spinescent stipules.

The **Cojoba clade** includes, besides *Cojoba* itself, *Lysiloma* and *Hesperalbizia* and is also native to the Americas, centered in Mexico, Central America, and the Caribbean.

The **Archidendron clade** includes the large Southeast Asian and Papuasian genus *Archidendron*, as well as *Archidendropsis*, *Falcataria*, *Pararchidendron*, *Paraserianthes*, *Serianthes*, and the largest genus of mimosoids, the predominantly Australian *Acacia* s.s.; *Wallaceodendron* (not sampled here) is also tentatively included in this clade (Nielsen et al., 1983). The clade is widespread across and almost entirely restricted to the Indomalayan and Australasian regions, with centers of diversity in Malesia, New Guinea, and Australia, making it biogeographically distinct from all other lower-level clades.

The **Jupunba clade** is largely composed of the Abarema alliance of Barneby and Grimes (1996), with the exclusion of the type species of *Abarema* (*A. cochliacarpos* which is nested in the Inga clade), and hence we name the clade after the genus *Jupunba*, which is to be reinstated to accommodate most *Abarema* species (Iganci et al., 2016; M.V. Soares et al., submitted). A smaller number of *Abarema* species that are characterized by spicate inflorescences will be accommodated in a reinstated *Punjuba* (M.V. Soares et al., submitted), which is also included in this clade. The Neotropical species of the genus *Albizia*, placed in section *Arthrosamanea* by Barneby and Grimes (1996), are included here and form the sister group to the rest of the clade. Apart from the exclusively Neotropical Abarema alliance and *Albizia* sect. *Arthrosamanea*, the African *Albizia obliquifoliolata* (syn. *Cathormion obliquifoliolatum*) is also nested in this clade, being most closely related to *Hydrochorea*. Together with *Albizia rhombifolia* (syn. *Cathormion rhombifolium*; not sampled, also African), *A. obliquifoliolata* is morphologically very similar to Neotropical *Hydrochorea*, with the water-dispersed seeds presumably having facilitated transatlantic dispersal. Dimorphic flowers are found in the majority of species across the clade.

The **Inga clade** includes the Neotropical genera *Abarema*, *Blanchetiodendron*, *Inga*, *Leucochloron* (except *L. bolivianum*), *Macrosamanea*, and *Zygia* and is particularly diverse in lowland rainforests. *Blanchetiodendron* is a highly distinctive genus and is sister to the rest of the clade. While most species of *Abarema* are placed in the Jupunba clade and are being transferred to two reinstated genera (Iganci et al., 2016; M.V. Soares et al., submitted), *Abarema* s.s., which comprises the type species of the genus and the recently described *A. diamantina* (Guerra et al., 2016), is included in this clade. Two African taxa, *Albizia altissima* (syn. *Cathormion altissimum*) and *A. dinklagei* (syn. *Samanea dinklagei* and *C. dinklagei*), are also nested in this clade, for which a new genus name will be needed. While not sampled here, *Enterolobium* sect. *Robrichia* is also included in this clade (É. R. de Souza, unpublished data). Two of the three species of section Robrichia of *Enterolobium*, *Blanchetiodendron*, and the African *A. altissima/A. dinklagei* group are the only taxa in the Inga clade with dimorphic flowers.

The **Samanea clade** includes *Chloroleucon* and *Samanea*, is restricted to the Neotropics, and all genera have species with dimorphic flowers. *Pseudosamanea* can likely be included in this clade, but the support for that relationship is considered insufficient at the moment. This clade could also potentially be merged with the Albizia clade if the two are well supported as sister clades (Fig. 5), but this is not supported in the Bayesian jackknife analysis (Fig. 6).

The **Albizia clade** includes *Albizia* s.s. (i.e., the African, Malagasy, and Southeast Asian species currently accommodated in *Albizia*), plus Neotropical *Enterolobium* sect. *Enterolobium* (É. R. de Souza, unpublished data) and *Leucochloron bolivianum*, which is shown to be unrelated to the rest of *Leucochloron* (Figs. 5 and 6; É. R. de Souza, unpublished data). Furthermore, *Cathormion umbellatum*, the type species of the genus *Cathormion* s.s. (Lewis et al., 2005), is nested within *Albizia* and is therefore synonymized with that genus (see below). Both *Albizia* and *Enterolobium* include species with dimorphic flowers.

Taxonomy

Cathormion Hassk. **syn.nov.**, Retzia 1 (1855) 231 = ***Albizia*** Durazz., Mag. Tosca. 3 (1772) 11.

Albizia umbellata (Vahl) E.J.M.Koenen **comb.nov.** Basionym: *Mimosa umbellata* Vahl, Symb. Bot. 2 (1790) 103; — *Inga umbellata* (Vahl) Willd., Sp. Pl. ed. 4, 4 (1806) 1027; — *Pithecellobium umbellatum* (Vahl) Benth., Lond. J. Bot. 3 (1844) 202; — *Feuilleea umbellata* (Vahl) O.Kuntze, Rev. Gen. Pl. 1 (1891) 188; — *Cathormion umbellatum* (Vahl) Kosterm., Bull. Organ. Natuurw. Onderz. Indon. 20 (1954) 12.

Inga concordiana DC., Prod. 2 (1825) 441; — *Mimosa concordiana* Roxb., Fl. Ind. ed. 2, 2 (1832) 556.

Pithecellobium malayanum Pierre, Fl. Cochinch. 5 (1899) t. 394A.

Note: Two subspecies of ***Cathormion umbellatum*** have been recognized, subsp. ***umbellatum*** and subsp. ***moniliforme*** (DC.) Brummitt, which differ in the number of pinnae per leaf, number of leaflets per pinna, leaflet dimensions, leaflet venation, and shape of the seeds (Nielsen, 1992:143). Moreover, they have non-overlapping geographic ranges, with the former occurring in mainland Asia while subsp. *moniliforme* is Malesian and Australian. It may therefore be preferable to recognize these as distinct species. The name *Albizia moniliforme* (DC.) F.Muell. is already available. [correction added after the original online publication on 16 December 2020, the appendix 2 has been corrected].