BOTANY LETTERS
https://doi.org/10.1080/23818107.2024.2414981

Taylor & Francis
Taylor & Francis Group

Evidence for a Miocene pulse of diversification of the tropical American clade of the Brazil nut family (Lecythidaceae)

Jérôme Chave[a], Charles Pouchon[b,c], Uxue Suescun[a], Sébastien Lavergne[d], Christopher Dick[e,f], Oscar M. Vargas[g] and Myriam Heuertz[h]

[a]Centre de Recherche sur la Biodiversité et l'Environnement (CRBE) UMR5300, Université de Toulouse, CNRS, IRD, Toulouse INP, Toulouse, France; [b]Conservatoire et Jardin botaniques de la Ville de Genève, Geneva, Switzerland; [c]Plant Systematics and Biodiversity, Department of Plant Sciences, Université de Genève, Geneva, Switzerland; [d]Laboratoire d'Ecologie, Alpine, University Grenoble Alpes, University Savoie Mont Blanc, CNRS, LECA, Grenoble, France; [e]Department of Ecology & Evolutionary Biology and Herbarium, University of Michigan, Ann Arbor, MI, USA; [f]Smithsonian Tropical Research Institute, Panama, Republic of Panama; [g]Department of Biological Sciences, California State Polytechnic University, Arcata, CA, USA; [h]BIOGECO, INRAE, University Bordeaux, Cestas, France

ABSTRACT

The tropical American species of the Brazil nut family, Lecythidaceae, are abundant in the tree flora, with some economically important species. Yet, the phylogenetic relationships and pace of diversification within this family are understudied. Here, we used shotgun sequencing data for 86 of the 228 currently accepted species in the exclusively tropical American subfamily Lecythidoideae in a phylogenomic context. For each sampled species, we built the full plastid DNA sequence, and also extracted nuclear DNA for 571 regions using the new bioinformatics pipeline REFMAKER, which we used to produce a time-calibrated phylogenetic hypothesis. Our analysis shows that phylogenetic inference from plastid and nuclear DNA alignments resulted in different topologies. The nuclear DNA topology strongly suggests that genus *Lecythis* should be split into at least four genera. Samples of the genus *Eschweilera* formed a monophyletic group, with the exception of one sampled species (*Eschweilera amazoniciformis* S.A.Mori). The Bertholletia clade, which contains the majority of the Lecythidoideae species, and all *Lecythis* and *Eschweilera* species, started diversifying around 27.5 Ma, with an accelerated rate of diversification starting in the middle Miocene (c. 12 Ma). The clade sister to *Bertholletia* (including *Corythophora*, *Eschweilera*, and Corrugata, Chartacea and Poiteaui clades) includes at least 124 species and it has diversified less than 10 Ma. This time frame of diversification coincides with major changes in tropical American landscapes and climate associated with the Andean uplift.

ARTICLE HISTORY
Received 31 May 2024
Accepted 27 September 2024

KEYWORDS
Andean uplift; diversification; Lecythidaceae; phylogeny; tropical America

Introduction

Few plant families are as emblematic of the tropics as the Brazil nut family, Lecythidaceae. The family is best known for the edible seeds of the Brazil nut tree *Bertholletia excelsa* Bonpl., a worldwide staple, with over 30,000 metric tons produced in 2019, mostly from Bolivia, for a supply value of 260 million USD (Bertwell et al. 2018). The splendid ornamental cannonball tree (*Couroupita guianensis* Aubl.) displays spectacular red flowers on the trunks and has been planted throughout the tropics. In the moist regions of tropical America, the Lecythidaceae species are an important component of the tree flora. In Amazonia, an estimated 6.8% of tree stems belong to Lecythidaceae, second only to Fabaceae and Arecaceae (Ter Steege et al. 2019). *Eschweilera* is the most abundant Amazonian tree genus, and *Eschweilera coriacea* (DC.) S.A.Mori, the most abundant tree species in Amazonian forests, with recorded presence in 932 of the 1985 plots of the Amazon Tree Diversity Network (Ter Steege et al. 2019). *Gustavia augusta* L. is the 29th most abundant species, *Eschweilera albiflora* Miers is 43rd, and *Eschweilera grandiflora* (Aubl.) Sandwith is 49th. Flowers and fruits are important resources for wildlife, many seeds are edible, and some species produce valuable timber (*Cariniana* spp), or non-wood products (Mori et al. 2017).

A pantropical family of 24 genera and 373 species (The World Flora Online 2023), the Lecythidaceae is characterized by "great almond-like seeds and alternate, often serrated, non-punctate leaves" (Morton et al. 1998). As currently delimited, the family is monophyletic and includes five subfamilies (Morton et al. 1998; Prance and Jongkind 2015; Huang et al. 2015): Napoleonaeoideae (21 species, two genera: *Napoleonaea* and *Crateranthus*), Scytopetaloideae (25 species, six genera), Barringtonioideae (80 species, five genera), Foetidioideae (19 species, one genus), and Lecythidoideae (228 species, ten genera). Molecular analyses including all subfamilies place Napoleonaeoideae as the earliest diverging clade,

CONTACT Jérôme Chave jerome.chave@univ-tlse3.fr Centre de Recherche Sur la Biodiversité Et l'Environnement (CRBE) UMR5300, Université de Toulouse, CNRS, IRD, Toulouse INP, 118 Route de Narbonne, Toulouse 31062, France

Published online 18 Oct 2024

followed by Scytopetaloideae, and the clade (Barringtonoideae + Foetidioideae) is sister to Lecythidoideae (Morton et al. 1998; Mori et al. 2007; Zuntini et al. 2024). Lecythidoideae, with almost two-thirds of the total species number in the family, is strictly restricted to tropical America, and only one species of the other subfamilies is found in tropical America, *Asteranthos brasiliensis* Desf. (Scytopetaloideae). All molecular evidence so far suggests that subfamily Lecythidoideae is monophyletic, with 10 recognized genera: *Grias*, *Gustavia*, *Couroupita*, *Allantoma*, *Cariniana*, *Couratari*, *Bertholletia*, *Lecythis*, *Eschweilera*, and *Corythophora* (Mori et al. 2007, 2017; Huang et al. 2008, 2015).

Past inferences based on one or a few plastid DNA regions have greatly advanced the understanding of Lecythidaceae. Early work on the phylogenetic reconstruction of Lecythidaceae was based on Sanger sequencing of the *rbc*L gene, plus the *trn*L intron for 22 species in the family (Morton et al. 1998). Later, Mori et al. (2007) published another study also based on two plastid DNA regions (*ndh*F and the *trn*L-*trn*F intergenic spacer) including 77 of the tropical American species (Mori et al. 2007). Huang et al. (2015) expanded the sampling to 84 species and to two other DNA sequence regions, the *psb*A-*trn*H intergenic spacer and the ITS region of the ribosomal cluster. The addition of more species resulted in reduced branch support in the Huang et al. (2015) tree. The most recent analysis by Vargas and Dick (2020) reanalyzed the Huang et al. (2015) data, complemented with more species (118 species in total) and 16 plastid regions resulting in a DNA matrix of 12,726 aligned nucleotides, and 61% missing data. These analyses agree that the early-diverging group is *Gustavia*+*Grias* in the tropical American clade, both genera with actinomorphic flowers. *Couratari*, *Allantoma*, *Cariniana* and *Couroupita* differ in their position between the analyses, but one clade containing *Bertholletia*, *Lecythis*, *Corythophora* and *Eschweilera* is strongly supported (Mori et al. 2007; Huang et al. 2015; Vargas and Dick 2020). This "Bertholletia clade" contains about 133 of the tropical American species. Huang et al. (2015) concluded that *Corythophora* is monophyletic, *Bertholletia excelsa* is nested within *Lecythis*, *Lecythis* is formed of five clades and *Eschweilera* of three clades, therefore both *Lecythis* and *Eschweilera* were predicted to be polyphyletic. Vargas and Dick (2020) confirmed the monophyly of *Corythophora*, found the Bertholletia clade also monophyletic, and found that *Lecythis* formed three clades and *Eschweilera* four clades. Thus, in both Huang et al. (2015) and Vargas and Dick (2020), the monophyly of *Lecythis* and *Eschweilera* was strongly called into question. Mori et al. (2017) believed that the relationships between these clades were not resolved. The only option to resolve these controversies was to analyze nuclear DNA, a step that Vargas et al. (2019) took.

Vargas et al. (2019) used an *in silico* capture experiment to extract 354 nuclear gene sequences for 24 species previously analyzed using shotgun sequencing to produce completely sequenced plastid genomes by Thomson et al. (2018). In both plastid and nuclear analyses, the actinomorphic clade (*Grias*+*Gustavia*) was confirmed as the earliest diverging in the tropical American clade, followed by *Couroupita*. The plastid DNA analysis predicted a clade including *Allantoma*, *Cariniana*, and *Couratari* while the nuclear DNA analysis predicted a clade with *Allantoma* and *Cariniana*, and *Couratari* sister to the Bertholletia clade. The Bertholletia clade was confirmed in both analyses, but the relationships within the clade were widely divergent. The nuclear DNA analysis lent support to the hypothesis that *Lecythis* should be split into at least four groups. Vargas et al. (2019) recovered a monophyletic *Eschweilera* in their nuclear topology (Integrifolia and Parvifolia clades of e.g. Huang et al. 2015); however, their limited sampling did not include taxa from the Tetrapeta clade, which was close to the Corrugata clade of genus *Lecythis* (e.g. Huang et al. 2015). It is clear from these studies that plastid DNA alone is not sufficient to infer the evolutionary history of the Brazil nut family, and one important question is whether increased species sampling will provide enhanced support and resolution of phylogenetic relationships within the Leythidoideae subfamily.

From the standpoint of plant diversification, tropical American Lecythidaceae are hypothesized to follow a pattern similar to several other plant families (Koenen et al. 2015; Pirie et al. 2018; Chave et al. 2020), with a rapid diversification from the early Miocene due to the geologic shifts of the American continent and drastic climate shifts during this period (Hoorn et al. 2023). It is relevant to ask whether the diversification of Lecythidaceae in tropical America has followed the same tempo. Vargas and Dick (2020) provided some of the first insights into this question. They dated the Lecythidoideae crown clade in the middle Eocene at 46 Ma (36.5–56 Ma 95% confidence intervals) and the Bertholletia clade around 28 Ma (22–36 Ma) in the mid Oligocene. This pattern is suggestive of the lineage through time diversification pattern of Chrysobalanaceae (Hoorn et al. 2023). Tree reconstructions based on better species and character sampling may generate different topologies. Also, the fossil record of Lecythidaceae is limited, and there is a risk of missing important constraints and of underestimating the patterns of diversification.

Here, we reassess this question based on an increased sampling of Lecythidaceae of tropical America. We newly sequenced 75 accessions using the Illumina high-throughput sequencing technology in addition to those published by Thomson et al.

(2018), which we reanalyzed. We used a new bioinformatic software, REFMAKER (Pouchon and Boluda 2023), that mines nuclear DNA regions in Illumina short-read sequences and were able to assemble a phylogenetic tree based on nuclear DNA for 96 of these accessions. This is a four-fold increase over the work of Vargas et al. (2019) based on nuclear DNA information and is comparable in taxon sampling to the work of Vargas and Dick (2020), but based on nuclear DNA and with much increased character sampling. We also mined plastid DNA regions of 101 accessions to create a high-quality alignment to produce a plastid DNA phylogenetic tree. By increasing the quantity and coverage of genomic material, we hope to elucidate the diversification of Lecythidaceae, paving the way for more detailed research on the ecology and evolution of this fascinating plant family. Specifically, we ask the following three questions: (i) What insights does the nuclear DNA tree offer to the systematics of Lecythidaceae? (ii) can phylogenetic reconstructions based on plastid and nuclear DNA be reconciled? (iii) what are the key periods of diversification in the evolutionary history of Lecythidoideae?

Added note: This study is a tribute to Dr Scott A. Mori, who passed away in August of 2020. Scott Mori was a champion of the study of the Lecythidaceae, and a distinguished botanist. He has trained a generation of tropical botanists from Central America to Brazil, and he has devoted much of his research coordinating the Flora of Central French Guiana, with the now classic contribution *Guide to the Vascular Plants of Central French Guiana* (Mori et al. 2002). He took every opportunity to share his knowledge with colleagues, early-career botanists, and amateurs (Mori et al. 2011). More detailed accounts of Scott Mori's career and outstanding achievements have been published elsewhere (Boom 2020; Prance et al. 2021).

Methods

Species sampling and sequencing

We used sampling tissue for 103 Lecythidoideae, representing 87 species or 38% of the Lecythidoideae, and one outgroup, *Barringtonia edulis* Seem., taken from both fresh silica dried tissue and from herbarium specimen. The sampling included three accessions of *Couroupita guianensis* Aubl., *Lecythis congestiflora* (Benoist) Eyma, and *Eschweilera coriacea*, and two accessions each of *Cariniana estrellensis* (Raddi) Kuntze, *Cariniana ianeirensis* R.Knuth, *Corythophora amapaensis* Pires ex S.A.Mori & Prance, *Couratari macrosperma* A.C.Sm., *Couratari stellata* A.C.Sm., *Eschweilera micrantha* Miers, *Eschweilera wachenheimii* Sandwith, *Gustavia hexapetala* Sm., and *Lecythis corrugata* Poit. The complete list of species and provenances is provided in the Appendix. DNA was extracted from leaf tissue material, then sheared to obtain DNA fragments. High-throughput sequencing libraries were prepared following the specifications of Chave et al. (2020), except for the 24 DNA libraries already described in Thomson et al. (2018) and Vargas et al. (2019). Libraries were tag-labeled and pooled by 48, before being sequenced on Illumina HiSeq 3000 or Nova-Seq high-throughput sequencers, yielding pair-ended DNA reads of 150-nucleotides.

DNA sequence assembly and matrix construction

We captured the CDS and rRNA plastid genes in the shotgun libraries using ORTHOSKIM (Pouchon et al. 2022). Global assemblies were first conducted using SPAdes (Bankevich et al. 2012) with the default parameters (i.e. kmer size of 55, minimal kmer coverage ≥ 3, minimal contig size ≥ 500 bp). Reference sequences for targeted plastid genes, consisting of 79 CDS and 4 rRNA genes, were obtained from 169 libraries produced within the PhyloAlps project for Ericales taxa (Lavergne et al. in prep.). Only the exonic regions were targeted. Genes were considered as successfully captured when the size of the captured sequence covered at least 50% of the reference length, and when the longest ORF of the captured sequence covered at least 80% of the captured sequence. Captured sequences were next aligned and trimmed within ORTHOSKIM using MAFFT (Katoh and Standley 2013) and trimAL (Capella-Gutiérrez et al. 2009). Paralogs and errors were tagged and removed using the filtering option with the default parameters, resulting in a DNA alignment referred to as the cpDNA matrix. We also performed a full plastid genome reconstruction to explore the role of non-coding cpDNA sequences, but this compartment did not bring in significantly more information, so this procedure is described in the Supplementary Information.

We also extracted genomic regions from the mitochondrial DNA (mtDNA) and ribosomal DNA (rDNA) regions. The mtDNA and the rDNA trees were poorly supported, and their topologies were highly dependent on reconstruction assumptions. Since the number of informative characters was low in mtDNA and rDNA analyses, only nuDNA and cpDNA trees are discussed in Results (see Supplementary Information for further details about mtDNA and rDNA compartments).

Next we extracted low-copy nuclear regions using REFMAKER (Pouchon and Boluda 2023). REFMAKER builds a reference set of nuclear loci from low-coverage genome skimming libraries (<1X). For that, a global and *de novo* assembly of each genome skimming library is first performed using SPADES (Bankevich et al. 2012), and larger contigs (ie. meta-contigs) are next built from these

sets of individual contigs to construct a reference catalog. As ORTHOSKIM was first run, we used the contigs sets previously built as input to build the REFMAKER catalog while removing all contigs for which cpDNA, mtDNA and rDNA targets were, respectively, captured. The REFMAKER catalog was built and filtered using default parameters (ie. kmers of 31, 51, 71 and 91; similarity threshold of 80%; minimal contig length of 250bp; 25% of overlapping between meta-contigs). This resulted in a catalog of 9116 putative nuclear loci. Raw reads were next mapped on this catalog using BWA (Li and Durbin 2009), variable sites were called using BCFTOOLS (Danecek et al. 2021) and consensus sequences were built for each library within REFMAKER. We set a minimal mapping quality of 60 and a minimal coverage of four reads for variant calling. Multispecies alignments were next filtered as follows: a maximum frequency of heterozygous sites of 0.05 per sample sequence, a frequency of samples sharing heterozygous sites of 1.0 to detect putative paralogs, we allowed a maximum frequency of 35% missing data allowed per sequence, we allowed a maximum frequency of missing data allowed of 75% per sample across all the putative nuclear loci, a window size of 20 nucleotides to detect paralogs or errors, a maximal of five polymorphic sites allowed within this window, a minimal locus length of 400 bp and minimal frequency of samples sharing a locus of 80%. This final cleaning step resulted in a partitioned concatenated alignment of 1,256,743 nucleotides for 96 of the 103 genome skimming libraries, corresponding to 571 nuclear regions (referred to nuDNA matrix).

Phylogenetic tree reconstruction

For the nuDNA dataset, we used IQTREE-2 to construct the most likely topology using a concatenated matrix of all the nuclear regions recovered (Minh et al. 2020). IQTREE-2 computes a maximum-likelihood topology with two metrics of branch support: the first is inferred from a likelihood-based method, the Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT) support value (Anisimova et al. 2011), and the second is obtained by bootstrapping the matrix columns, using an ultrafast method (Hoang et al. 2018). A consensus tree was also provided with an ultrafast bootstrap support value based on 1000 trees.

Species trees were also inferred for the nuDNA dataset using ASTRAL-III (Zhang et al. 2018) and SVD-Quartets (Chifman and Kubatko 2014). We first inferred gene trees for each nuclear locus using IQTREE-2 and used the best trees as input for ASTRAL-III. ASTRAL analysis was run with the default setting values. Node support values were calculated by the local posterior probability (LPP) estimated from the normalized quartet support (Sayyari and Mirarab 2016). SVD-Quartets was run on the nuclear partitioned concatenated supermatrix (nuDNA) using 500,000 random quartets. Node values were estimated from 500 bootstrap replicates.

Phylogenetic trees were also generated independently for each of the DNA matrices (cpDNA, mtDNA and rDNA). We did not concatenate these matrices, because the genomic compartments could be under separate evolutionary constraints, such as introgression, hybridization, or other processes. We used the ModelFinder software to test which combination of CDS optimizes the tree reconstruction (Chernomor et al. 2016; Kalyaanamoorthy et al. 2017). This procedure concatenates the CDS with similar models of molecular evolution to reduce the number of free parameters. We treated all CDS as independent partitions in the input of ModelFinder, 83 for the cpDNA matrix, 36 for the mtDNA matrix, and three for the rDNA matrix (18S, 5.8S and 26S). Using the best partitions and models of molecular evolution, we generated topologies again based on IQTREE-2. A consensus tree was also provided with a bootstrap support value based on 1000 trees.

Phylogenetic tree space exploration

Phylogenetic landscape of trees was examined to assess the level of topological discordances found among the trees. We first computed a pairwise topological distance matrix between all nuclear gene trees inferred with IQTREE-2. Topological distances were also estimated for the nuclear species trees (inferred with IQTREE-2, ASTRAL-III and SVD-Quartets) and the mtDNA, cpDNA and rDNA trees. Each pair of trees were rooted and pruned to the same taxa list using "ape" R package v.5.6.1 (Paradis et al. 2004). Topological distances between the trees were estimated by the normalized Robinson-Foulds distance implemented in "phangorn" R package v.2.8.1 (Schliep 2011). A hierarchical clustering approach was next performed to identify clusters of gene trees by using the "heatmap" function of the R package stats v.3.6.3. We also performed a principal coordinate analysis of the topological distances within the "ade4" R package v.3.6.2 (Dray and Dufour 2007), to project each gene tree within the phylogenetic landscape.

Phylogenetic network inference

We used SplitsTreeCE v.6.1.16 (https://github.com/husonlab/splitstree6) to estimate reticulation in the evolution of the Lecythidaceae. A phylogenetic network was computed from the NeighborNet distance transformation with the uncorrected P distance (Albrecht et al. 2012), using the concatenated nuDNA dataset with 300 bootstrap replicates.

Phylogenetic tree dating

The age of Lecythidaceae remains poorly known due to its scarce fossil record. To reconstruct a dated phylogenetic tree, we searched for a reliable fossil record in the Lecythidaceae. Wheeler et al. (2017) reassessed fossil woods of the Deccan Intertrappean Beds of India, which, based on $^{40}Ar/^{39}Ar$ dates, are due to volcanic activity that occurred from 67.5 to 63 Ma, with the bulk of the eruption around 65 Ma. The wood fossil of *Barringtonioxylon deccanense* Shallom (Shallom 1960) dated at c. 66 Ma (95%CI: 60–72 Ma) was reassessed as "probably Lecythidaceae". Fortunately, at least two other samples of the same formation have been attributed to genus *Barringtonioxylon*, namely *Barringtonioxylon eopterocarpum* (Prakash and Dayal 1964) and *Barringtonioxylon mandlaense* Bande & Khatri (Bande and Khatri 1980). Therefore, we can confidently use the range 67.5 to 63 Ma as a minimal age for the stem age of Lecythioideae. This range is consistent with the indication of the presence of possible Lecythioideae wood fossils in the lowermost Bagua Formation, Northern Peru, which could be older than 66 Ma (Mourier et al. 1988).

We also used two internal calibrations. We used the seed fossil of *Lecythidospermum bolivarensis* (Berry) Pons found close to the city of Planeta Rica, Colombia, on the road to Monteria, and dated from the late Oligocene and the early Miocene, i.e. from 20 to 29 Ma (Pons 1983). We also used the wood of *Cariniana valverdei* Woodcock, found in the Pietra Chamana fossil forest in Northern Peru (Woodcock et al. 2017) and dated based on $^{40}Ar/^{39}Ar$ dates as 39.35 ± 0.21 Ma (Allen et al. 2023).

We did not use the Oligocene fossils of *Barringtonia* (Srivastava and Mehrotra 2018) and *Careya* (Mehrotra and Srivastava 2017), both from the Makum Coalfield (Northeast India) and wood fossils from the middle Miocene (12–16 Ma, Pons and De Franceschi 2007), and from the late Miocene (Kloster et al. 2017), because they would not have further constrained tree dating.

We first time-calibrated the maximum likelihood tree produced by the IQTREE-2 analysis. We also time-calibrated each of the 1000 trees produced by the ultrafast bootstrap analysis for both the nuDNA and cpDNA matrices. A single consensus maximum clade credibility tree across all 1000 time-calibrated IQTREE-2 trees was produced using TreeAnnotator v.2.7.4 (Heled and Bouckaert 2013).

To perform tree dating, we used a penalized likelihood method (Sanderson 1997, 2002). Several options have been proposed in the past, such as strict maximum likelihood inference, so that the branch lengths representing substitution rates in the original tree are fitted to the branch lengths of the chronogram using a Poisson likelihood (Sanderson 2002). An improvement over this method has been to penalize against large shifts in substitution rates in consecutive branches, the "rate smoothing" approach (Sanderson 1997). In mathematical terms, if the log-likelihood function to be maximized is denoted $\ln L$ and the function that minimizes shifts in substitution rates is called P (for penalty), a penalized likelihood function is denoted $PL = \ln L - \lambda P$. Here, λ is a smoothing parameter. This method was implemented in r8s (Sanderson 2002), treePL (Smith and O'Meara 2012) and in the chronos() function available with "ape" v.5.8 in R (Paradis et al. 2004; Paradis 2013). We performed the analyses with chronos, which is equivalent to treePL, but more general and flexible. This routine has three implemented models: correlated substitution rates (as in r8s), fully uncorrelated substitution rates (relaxed clock model) or a discrete set of substitution rates (as in Bayesian phylogenetic inference, Yang 2007). After an extensive set of comparisons and tests based on Paradis (2013) information criterion ΦIC, we retained the default value of a model with correlated substitution rates and the default smoothing parameter of $\lambda = 1$.

We dated the trees using the following three time constraints: the root had minimal and maximal ages of 63 and 67.5 Ma, the node subtending all *Lecythis* had minimal and maximal ages of 20 and 35 Ma (*Lecythidospermum bolivarensis* seed fossil), and the node subtending *Cariniana* and the Bertholletia clade had minimal and maximal ages of 39 and 47 Ma (*Cariniana valverdei* wood fossil). We lacked nuDNA material for *Barringtonia*, so we constrained the root (the split between *Grias*+*Gustavia* and the rest of the family) with the range obtained in the cpDNA tree, i.e. minimal and maximal ages of 55.4 and 59.7 Ma. For the nuDNA tree, we kept the same two internal constraints as for the cpDNA tree (from the *Lecythidospermum* fossil and the *Cariniana valverdei* fossil).

Diversification analysis

The dated phylogenetic tree was used to explore how rates of diversification changed across geological periods. We first simply plotted the lineage through time plot using the ape R package (function ltt.plot()). One limitation of this approach is that clades are unevenly sampled, resulting in biased estimations of absolute rates of diversification. The Bayesian analysis of macroevolutionary mixture (BAMM, Rabosky and Kolokotronis 2014; Shi and Rabosky 2015) includes the possibility of better estimating rates by including clade sampling intensity as a prior information. We ran BAMM with the reverse-jump MCMC simulation for 10 million iterations to ensure convergence, which was assessed with the EffectiveSize() function of the BAMMtools package in R (Rabosky et al. 2014).

Results

Phylogenetic relationships in the Lecythidoideae inferred from nuclear DNA

The aligned matrix of nuclear DNA sequences (nuDNA) had a total length of 1,256,743 nucleotides, corresponding to 571 regions. The mean fraction of missing data in this nuDNA matrix was 24.6%. Five accessions could not be included in the nuclear DNA matrix due to too low coverage: *Gustavia hexapetala* (one of the two accessions of this species), *Grias cauliflora* L., *Lecythis confertiflora* (A.C.Sm.) S.A.Mori, *Lecythis praeclara* (Sandwith) S.A.Mori ex Molino & Sabatier, and *Couratari oblongifolia* Ducke & R.Knuth. The final nuDNA matrix contained 96 accessions belonging to 83 species.

The time-calibrated nuDNA tree based on IQTREE-2 is reported in Figure 1. We recovered *Gustavia* as sister to *Grias*, and the actinomorphic clade (*Grias*+*Gustavia*) sister to the rest of Lecythidoideae. Our analysis dates the actinomorphic clade at around 29 Ma, in the Oligocene.

A strong support was found for *Couroupita* being the earliest diverging lineage of the clade sister to *Grias* +*Gustavia*. The next clade was strongly supported and formed of *Cariniana* and *Allantoma* the two being sister genera, with *Couratari* being the next clade, with a crown age around 28 Ma. Two subgenera were clearly apparent in *Couratari*, one including *C. macrosperma*, and *C. stellata* (section *Echinata* sensu Mori and Prance 1990), the other including *C. guianensis*, *C. gloriosa*, *C. calycina*, and *C. multiflora* (sections *Microcarpa* and *Couratari* sensu Mori and Prance 1990). *Couratari* was sister to the Bertholletia clade, which contains *Lecythis*, *Eschweilera*, *Corythophora* and *Bertholletia excelsa* (Mori and Prance 1990; Mori et al. 2007; Huang et al. 2015), and with a crown age at 27 Ma.

The earliest diverging clade in the Bertholletia clade is the Pisonis clade, followed by the Ollaria clade, and then *Bertholletia excelsa*. The rest of the Bertholletia clade, here referred to as the "core Bertholletia clade", has a crown age of 8.1 Ma, and it includes the rest of the *Lecythis* clades, *Corythophora*, and *Eschweilera*. The earliest diverging clade comprises the Poiteaui and Chartacea clades followed by a clade comprising *Eschweilera amazoniciformis* and *Corythophora*, which forms a clade sister to all remaining *Eschweilera*. The crown age of *Eschweilera* is dated at 4.8 Ma. In the *Eschweilera* clade, the Parvifolia and Integrifolia clades are sister clades.

Species trees inferred by the ASTRAL and SVD-Quartets coalescent-based phylogenetic reconstruction softwares support the same clades as those found with the concatenated tree (Figure 2), with the exception of *Allantoma* and *Cariniana* clades. In both methods, *Allantoma* was nested within the *Cariniana* clade. In the ASTRAL tree, one accession of *C. domestica* falls outside of this clade. Both trees also highlight within-clade discordances in comparison with the concatenated tree. This is consistent with a relatively low support inferred within the clades, in particular for relationships within the *Lecythis* "Chartaceae" and *Eschweilera* "Parvifolia" clades.

The SplitsTree4 network is consistent with the other methods, showing a separation of the same main clades (Figure 2). Interestingly, *C. domestica* segregated clearly from the other *Cariniana* lineages. Likewise, *E. coriacea*/*E. sagotiana* segregated from the Parvifolia clade. This analysis also reveals levels of phylogenetic inconsistencies within these clades, evidenced by interconnected edges, which may be due to recombination events.

An exploration of the phylogenetic tree space reveals two main clusters of tree topologies and three outlier trees. The first cluster, including 346 nuclear trees, the ribosomal tree and the organelle trees, includes the more similar trees, as shown by lower normalized Robinson-Foulds distances. In contrast, the second cluster is composed of 228 dissimilar nuclear trees. This can be explained by a lower proportion of informative sites and a shorter size of the loci found in the second cluster, with a lack of phylogenetic signal. The ASTRAL tree inferred on nuclear trees of the first cluster was consistent with the one inferred on the whole nuclear dataset. This analysis shows a high level of topological discordance, which can be related to a lack of phylogenetic signal in some loci.

Comparing nuclear and plastid DNA tree reconstructions

The cpDNA tree was constructed with IQTREE-2 from an aligned matrix with 63,429 nucleotides. The mean fraction of missing data in this cpDNA matrix was 5.8% (Figure 3). The nuDNA and cpDNA trees generally displayed similar grouping of the sampled species as the groups discussed in Mori et al. (2017) and Vargas et al. (2019). However, the Pisonis, Ollaria and Integrifolia clades had a different position in the cpDNA and nuDNA trees, resulting in different topologies overall, confirming cytonuclear incongruence in Lecythidaceae.

Patterns of diversification in the Lecythidoideae

The analysis of rates of diversification within the Lecythidoideae displayed a marked increase around the Middle Miocene (Figure 4). The inference of

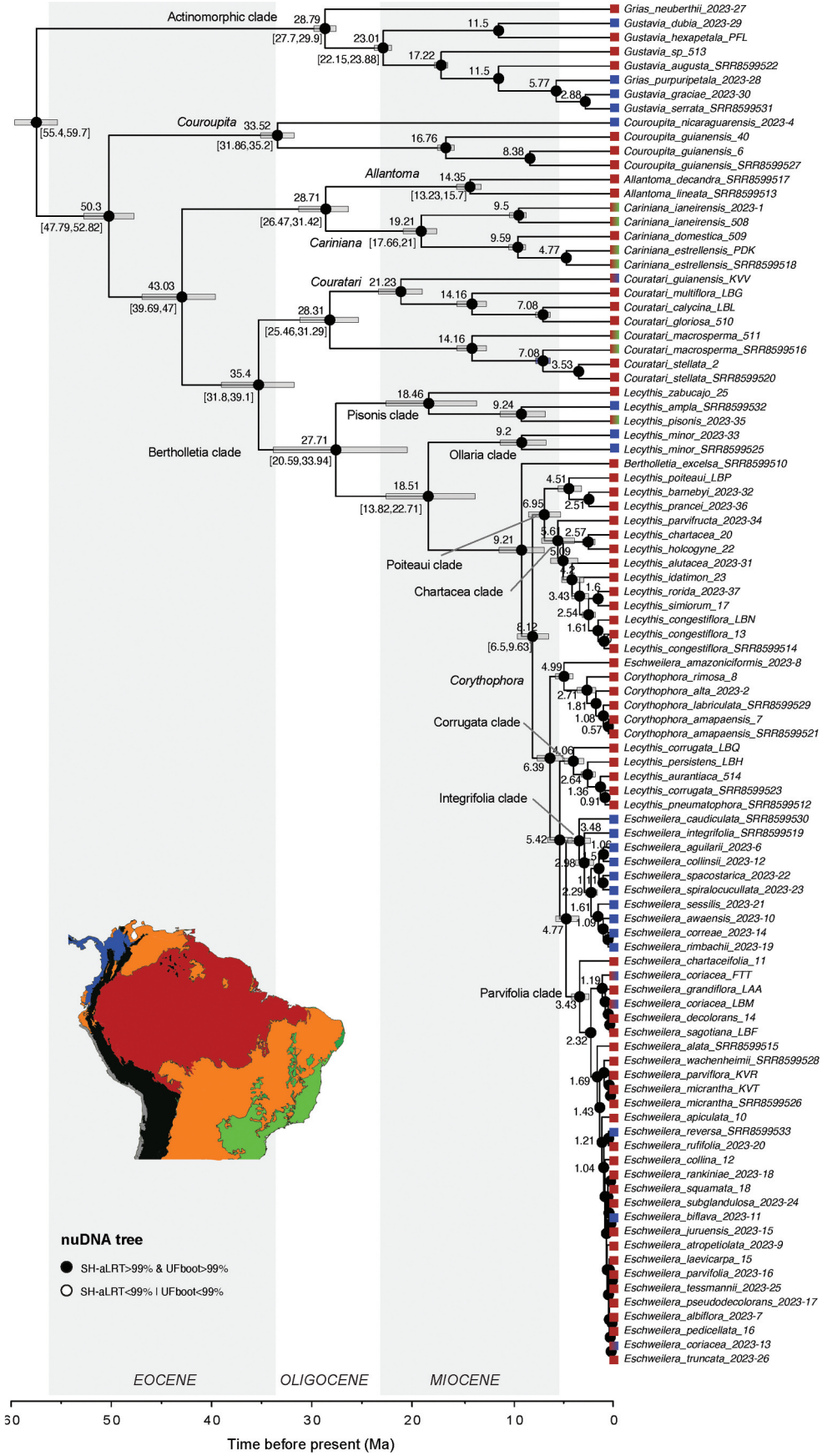

Figure 1. Time-calibrated phylogenetic tree obtained with IQTREE-2 from a nuclear DNA dataset of 571 genes for 96 accessions and 83 species and dated by a penalized likelihood method. Node ages and error bars are reported against geological epochs. All branches had a high branch support for both the Shimodaira-Hasegawa approximate likelihood ratio test (sh-aLRT) and the ultrafast bootstrapping method (Ufboot). Each species was color coded depending on the biogeographic region where they occur: Central America and forests west of the Andes (blue), Amazon (red), Atlantic Forest of Brazil (green), and open woodland and dry forest areas (orange). See Figure 2 for the rest of the clade.

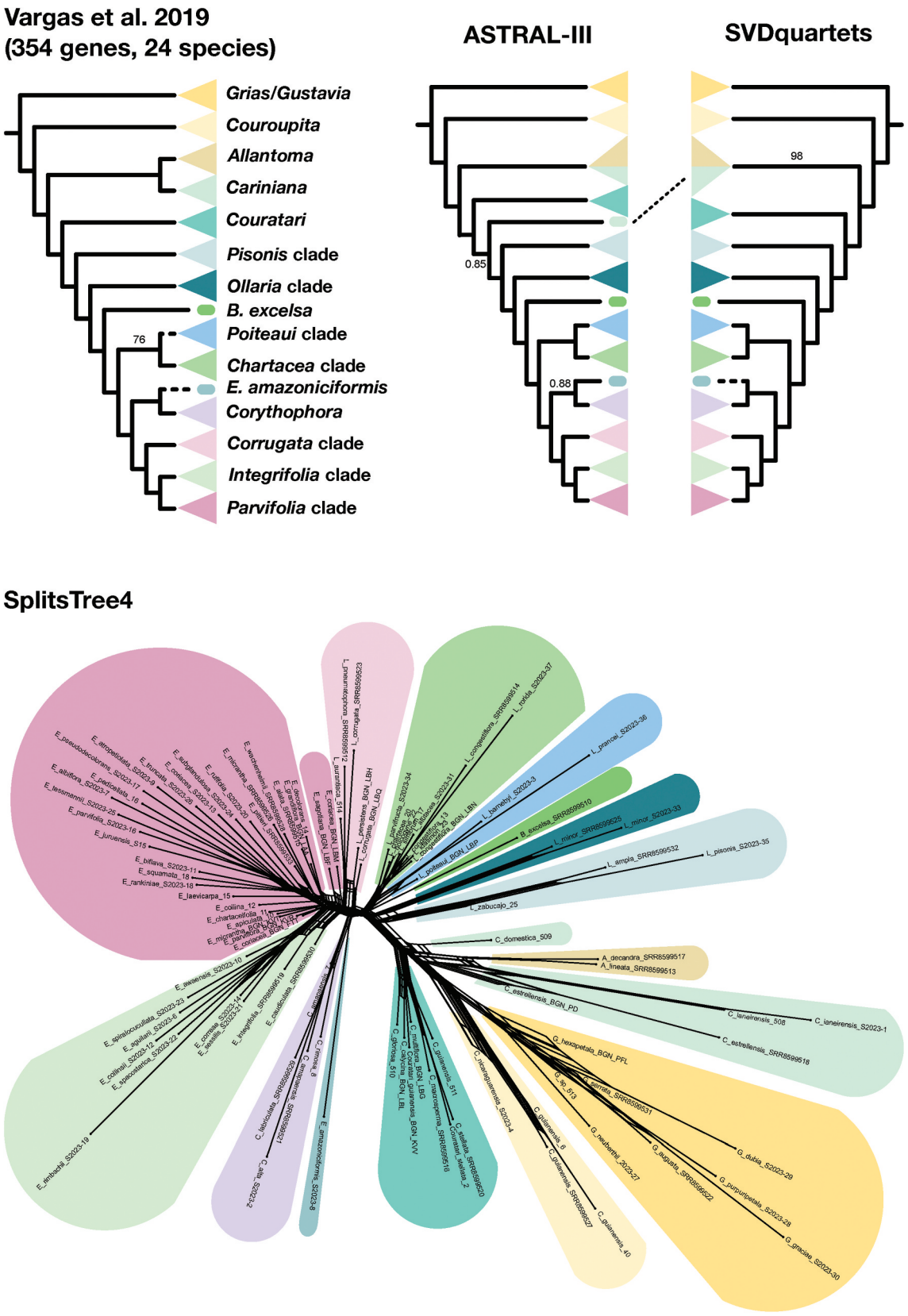

Figure 2. Comparison of the ASTRAL-III, SVD-Quartets coalescent-based methods with the tree published by Vargas et al. (2019) and with the topology obtained by the SplitsTree4 method.

speciation rates by BAMM demonstrated a threefold increase from less than 0.1 lineages/My around 15 Ma (Middle Miocene) to almost 0.3 lineages/My around 1 Ma (Pleistocene), and a twofold increase in diversification rates. The BAMM analysis also reveals that a single main shift in diversification rate occurred in core Bertholletia clade (*Bertholletia*, *Corythophora*, *Eschweilera*, and Corrugata, Chartacea and Poiteaui clades), which has a crown age of 9.2 Ma (Figure 4).

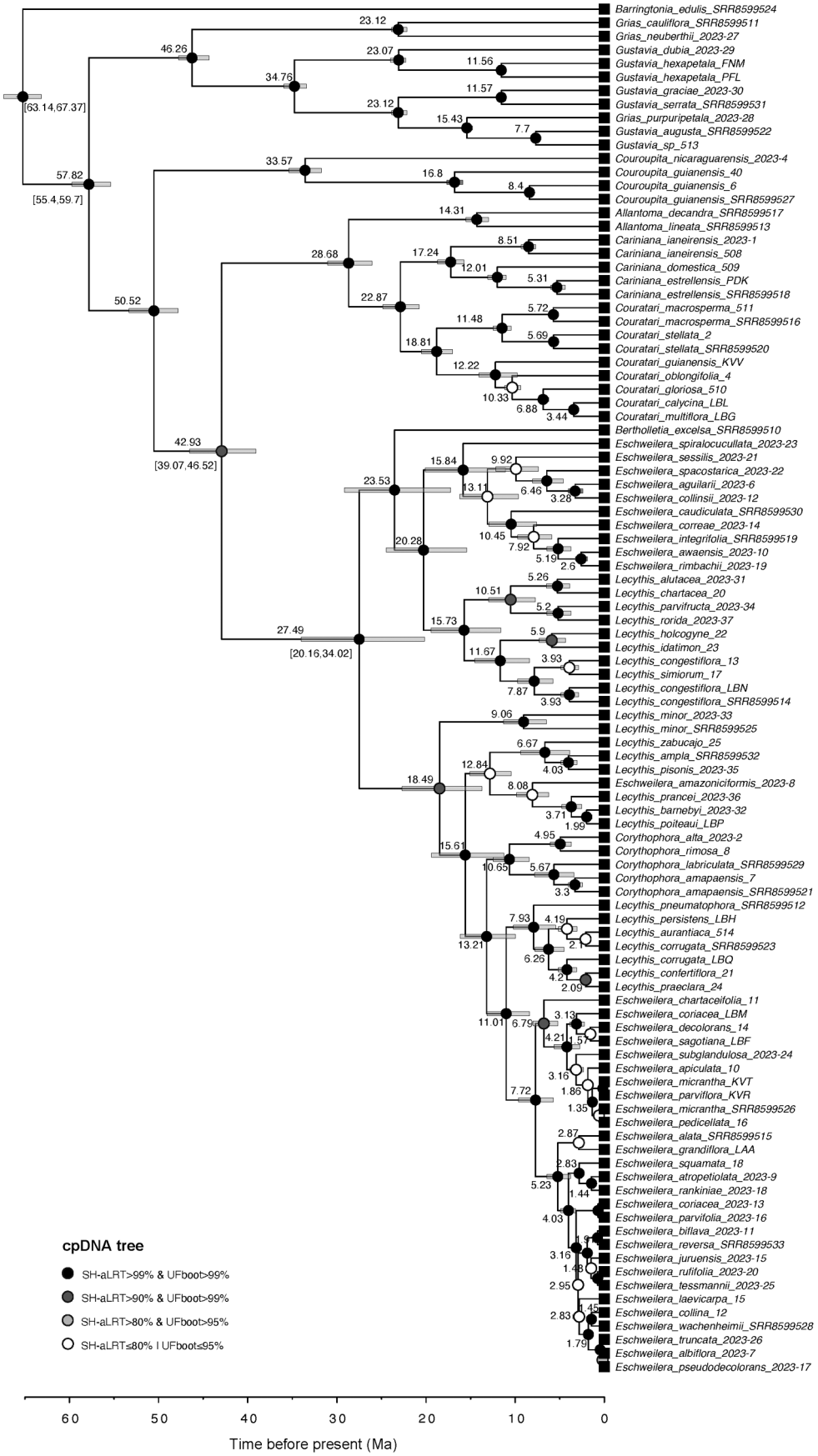

Figure 3. Time-calibrated phylogenetic tree obtained with IQTREE-2 from a plastid DNA dataset of all coding DNA sequences genes for 101 accessions. Most branches had a high branch support as reported by the shimodaira-hasegawa approximate likelihood ratio test (sh-aLRT) and by the ultrafast bootstrapping method (ufboot). The split between (*Gustavia+Grias*) and the rest of Lecythidoideae inferred in this analysis was used to date the root of the nuclear DNA tree reported in Figure 1.

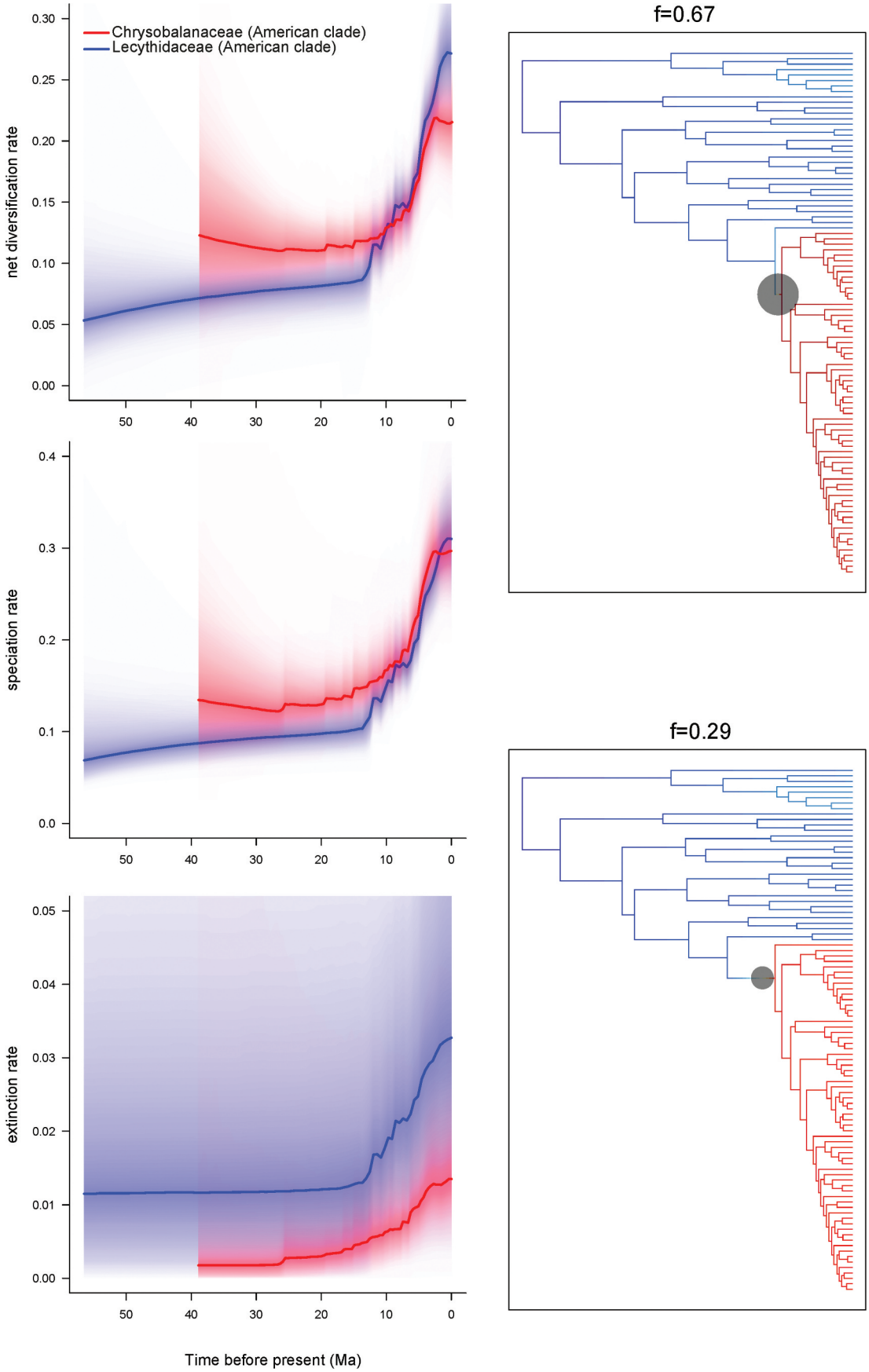

Figure 4. Diversification through time in American Lecythidaceae and chrysobalanaceae. Left: diversification rates (top), speciation rates (middle) and extinction rates (bottom) and their confidence intervals estimated from the BAMM software. The Chrysobalanaceae analysis is taken from Chave et al. (2020). Right: the two most credible shifts in diversification in Lecythidoideae: the first most credible shift corresponds to a clade including *Bertholletia* and its sister clade (core bertholletia clade).

Discussion

Phylogenetic reconstruction and genomic compartments

The present analysis confirms that organelle genomes do not accurately represent the evolutionary history of the Lecythidoideae clade. This is not surprising in light of the large literature on cytonuclear discordance in flowering plants that could be caused by a range of evolutionary processes, such as incomplete lineage sorting, horizontal gene transfer, hybridization, introgression, recombination or convergent molecular evolution (Rieseberg and Soltis 1991; Degnan and Rosenberg 2009; Steenwyk et al. 2023). Examples of such cytonuclear discordance are found in many plant families, e.g. in Asteraceae (Pelser et al. 2010; Vargas et al. 2017; Pouchon et al. 2021), or Vitaceae (Liu et al. 2021). Larson et al. (2021) found genomic evidence for hybridization among common species of *Eschweilera* in the Parvifolia clade, including within *E. coriacea*. Here, we did not attempt to evaluate the potential causes of phylogenetic discordance, and assumed that the nuDNA compartment better reflected the evolutionary history of the family.

The fact that low-depth genome skimming sequencing data can be used to extract phylogenetically relevant information from the nuclear DNA compartment is an important result (Pouchon and Boluda 2023). It was especially important to test this hypothesis on a clade for which independent phylogenetic evidence was available based on transcriptome sequencing of nuclear DNA (Vargas et al. 2019). Our nuDNA tree is consistent with the tree of Vargas et al. (2019), albeit based on a much larger species sampling and a different methodology. Future studies could explore the diversification of other plant groups without having to implement expensive nuclear genome or transcriptome sequencing projects and genome-wide phylogenetic reconstructions. Of course, some rapid diversification events will remain difficult to decipher no matter how much genomic material is available (see discussion below about the Parviflora clade), but this method should help gain confidence in the phylogenetic hypothesis of many plant clades, especially in the tropics. Interestingly, the use of low-depth genome skimming allows all genomic compartments to be recovered at once, since sufficient depth is obtained for reads from nuclear DNA.

Advances and limitations of the Lecythidoideae phylogeny

The phylogenetic tree recovered in the present analysis furthers our knowledge on Lecythidoideae but also has limitations. The earliest diverging clades for Neotropical Lecythidaceae are consistent with that retrieved by Huang et al. (2015) and Vargas et al. (2019). Only 6 of the 44 described species of *Gustavia* are included in the present analysis and 2 of the 12 described species of *Grias*, so further research is needed to confirm the monophyly of both genera, as well as intrageneric relationships. The recently described species *Grias purpuripetala* was found to fall within the *Gustavia* clade. We cannot exclude errors in manipulation of the material from sampling to sequencing for this sample.

The sister relationship of *Cariniana* and *Allantoma* was already suggested by Huang et al. (2008) based on morphology. This prediction, together with the position of *Couratari* and *Couroupita*, is also fully consistent with that of Vargas et al. (2019). *Cariniana* stands out as a genus with an accelerated evolution: the long branch subtending *C. ianeirensis* especially suggests an accelerated rate of molecular evolution in this species.

Our finding that the two *Lecythis* clades Ollaria and Pisonis are basal to the *Bertholletia* clade contrasts with the results of Huang et al. (2015) but is consistent with Vargas et al. (2019). Our analysis shows that *Lecythis pisonis* forms a clade with *Lecythis ampla*, and that *Lecythis zabucajo* is sister to that clade, suggesting an Amazonia origin of the Pisonis clade. It would be interesting to test this hypothesis by including the other two species in this clade (*L. lanceolata* and *L. marcgraaviana*, both from the Atlantic rainforest). The Ollaria clade contains only one (*L. minor*) of the three species assigned to this clade (*L. ollaria*, *L. tuyrana*), so it would be important to further explore its phylogenetic position based on more species.

Our analysis is consistent with the hypothesis that *Eschweilera* is a monophyletic genus, with clades Integrifolia and Parvifolia being positioned as sister clades. The exception to this statement is the segregate species *Eschweilera amazoniciformis*, which has four-petaled flowers, a rare morphology in *Eschweilera*. The only other four-petaled flower species are the seven known species of the Tetrapetala clade sensu Huang et al. (2015), six from Southeast Brazil, plus *E. nana* from Central Brazil, and *E. perumbonata*, from Venezuela. Further work is needed to clarify the phylogenetic position of *E. amazoniciformis* relative to the other four-petaled species in the genus (see Vargas and Dick 2020 and below).

The Integrifolia clade of *Eschweilera* is predominantly of west Andean affinity, with only *E. andina* and *E. ovalifolia* attaining Amazonia (Figure 3b), as first proposed by Huang et al. (2015). For this clade, nuDNA and cpDNA provide different phylogenetic hypotheses. The nuDNA tree suggests *E. caudiculata* as sister to the rest of the clade, followed by *E. integrifolia*, the rest of the Integrifolia clade being split between two subclades of four species each. In contrast, the cpDNA tree predicts *E. spiralocucullata* sp. ined. S.A. Mori & Cornejo as sister to the rest of the

Integrifolia clade and two different subclades of four and five species. The fact that the Integrifolia clade is inferred to be relatively recent (4.6 Ma), with many species sharing the same biogeographical areas, likely explains these cytonuclear incongruences, as also observed in other genera (*Linochilus* and *Diplostephium*, Astereae, Asteraceae; Vargas et al. 2017, Espeletiinae, Asteraceae; Pouchon et al. 2021).

Because of its recency and species richness, the Parviflora clade in *Eschweilera* is phylogenetically challenging, and it is unclear whether the nuclear genomic material retrieved in our study is sufficient to provide support for the phylogenetic relationships in this clade. Evidence of this fact may be gained from a comparison of a reconstruction using IQTREE-2 as presented in Figure 2 and the coalescent-based ASTRAL software (Zhang et al. 2018). The topology obtained using ASTRAL (Figure 2) shows that support for the clades within the Parviflora clade is limited. We therefore refrain from reading too much into the tree within the Parviflora clade, except for noticing that the three accessions of the common *E. coriacea* do not cluster together, which may be an indication that either this corresponds to several species, introgression with several taxa, presence of incomplete lineage sorting or evidence of hybridization, as was documented by Larson et al. (2021).

Finally, three of the *Lecythis* clades within the Bertholletia clade (Corrugata, Poiteaui, and Chartacea) were recognized as distinct clades by Mori et al. (2007) and Huang et al. (2015). Our analysis points to the likely sister relationship between the Poiteaui and Chartacea clades. This is a novel result given that the Poiteaui clade was missing in the recent analysis by Vargas et al. (2019). Here, we found that the Poiteaui clade includes three species: *L. poiteaui*, *L. barnebyi*, and *L. prancei*, consistent with Huang et al. (2015). In the Corrugata clade, the two accessions of *L. corrugata* do not group near the Poiteaui and Chartaceae clades. Also, *Lecythis idatimon*, previously listed in the Corrugata clade, was retrieved in the Chartacea clade, which may be due to a problem with the sample. In the Chartacea clade, support for the interspecific relationships was limited, and did not match between the IQTREE-2 and the ASTRAL analyses.

After the submission of this study, a systematic revision of the Lecythidoideae was published (Vargas et al. 2024), where the resurrected genus name *Pachylecythis* Ledoux is proposed for the Pisonis clade as described above, the resurrected genus *Chytroma* Miers for the clade forming the grouping of the Poiteaui and the Chartacea clades (sister clades as shown above), the new genus *Guaiania* O.M. Vargas & C.W. Dick for the Corrugata clade, the new genus *Waimiria* C.W. Dick and O.M. Vargas for the *Eschweilera amazoniciformis* S.A. Mori clade, and the new genus *Scottmoria* Cornejo for the Integrifolia clade. They also propose that the Tetrapetala clade be renamed *Imbiriba* O.M. Vargas, M. Ribeiro & C.W. Dick.

Diversification of Lecythidaceae and insights into the history of the tropical American forest flora

The emerging picture about the evolution of plants in tropical America is that it largely coincides with the orogeny of the Andes, for which geological evidence has much improved in recent years (Boschman 2021). The uplift of the Andes is thought to have contributed to massive transformation of regional climate, drainage patterns and nutrient cycling, and ultimately extensive landscape changes in South American plains east of the Andes, especially during the Miocene (Hughes et al. 2013; Jaramillo et al. 2017; Hoorn et al. 2023). Such substantial geologic and climatic changes have certainly generated a number of ecological and geographical drivers for plant speciation, and it would be particularly interesting to disentangle different speciation scenarios (e.g. Pouchon et al. 2021) to further understand how plant diversification occurred within tropical South American forests.

Recently, Hoorn et al. (2023) performed a diversification rate analysis for eight important plant clades for which time-calibrated phylogenetic trees had been previously obtained (*Andira*: Fabaceae, *Anemopaegma*: Bignoniaceae, *Bactris*: Arecaceae, *Brownea*: Fabaceae, Chrysobalanaceae, *Eschweilera*: Lecythidaceae, Protiae: Burseraceae, *Tynanthus*: Bignoniaceae). The results showed a strikingly dissimilar pattern of speciation rates, with *Andira*, *Brownea* and Chrysobalanaceae showing an increase in speciation rates during the Miocene, while the other five groups did not. According to this analysis, which used the time-calibrated plastid DNA phylogeny of Vargas and Dick (2020), speciation rates declined steadily in *Eschweilera*.

This finding differs strikingly with our result, where we found a rapid acceleration of net diversification rates in Lecythidoideae (Figure 4). When our diversification rates are compared with those in neotropical Chrysobalanaceae (Figure 4) (clade sister to *Geobalanus*; Chave et al. 2020) we find a remarkable similarity, with a marked increase in both net diversification and speciation rates during the Middle Miocene. We also found that the rapid rate of diversification in genus *Eschweilera* is sufficient to explain this trend.

One conclusion is that attempts to detect shifts in speciation rates should be based on sufficiently well-sampled, and carefully dated phylogenetic trees. Also, we conclude that the tropical American clade of Lecythidaceae does contribute to the mounting evidence that major changes in the tropical American

landscape and its climate during the Miocene have caused a pulse in diversification rate in many plant families. Lecythidaceae are tightly dependent on biological interactions for their reproduction since they are animal pollinated and dispersed, so it would be important to further explore the role of biological interactions in the patterns of species diversification within this clade.

Acknowledgements

We thank the New York Botanical Garden DNA collections and the Herbier de Guyane IRD in Cayenne for making collections available. This work would have not been possible without the fieldwork of many collectors, to whom we are grateful. We also thank Pierre-Olivier Antoine and Camila Martinez Aguilon for useful discussion on the Lecythidaceae fossil record. Support for sequencing has been provided by Genoscope, Evry (PhyloAlps project) and by ANR Investissement d'Avenir grants: CEBA (ANR-10-LABEX-0025) and TULIP (ANR-10-LABX-0041).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the Agence Nationale de la Recherche [ANR-10-LABEX-0025, ANR-10-LABX-0041].

Author contributions

JC conceived the idea and wrote the manuscript, CP and JC performed statistical analyses, US performed lab analyses, SL, CD, OV and MH contributed data; all authors edited the manuscript."

Data availability statement

Alignments, partitions, trees and metadata are available at https://doi.org/10.5281/zenodo.13365268.

References

Albrecht B, Scornavacca C, Cenci A, Huson DH. 2012. Fast computation of minimum hybridization networks. Bioinformatics. 28(2):191–197. doi: 10.1093/bioinformatics/btr618.

Allen SE, Martin KD, Meyer HW, Woodcock DW. 2023. An Eocene leaf flora from the northern Peruvian andes. Rev Palaeobot Palynol. 313:104889. doi: 10.1016/j.revpalbo.2023.104889.

Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Systemat Biol. 60(5):685–699. doi: 10.1093/sysbio/syr041.

Bande MB, Khatri SK. 1980. Some more fossil woods from the Deccan Intertrappean beds of Mandla District, Madhya Pradesh, India. Palaeontogr Abteilung B Paläophytologie. 173(4–6):147–165.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD. 2012. Spades: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477. doi: 10.1089/cmb.2012.0021.

Bertwell TD, Kainer KA, Cropper WP, Jr, Staudhammer CL, de Oliveira Wadt LH. 2018. Are Brazil nut populations threatened by fruit harvest? Biotropica. 50(1):50–59. doi: 10.1111/btp.12505.

Boom BM. 2020. Scott Alan Mori (1941—2020): an appreciation. Bot Rev. 86(3–4):149–179. doi: 10.1007/s12229-020-09232-0.

Boschman LM. 2021. Andean mountain building since the late cretaceous: a paleoelevation reconstruction. Earth-Sci Rev. 220:103640. doi: 10.1016/j.earscirev.2021.103640.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25(15):1972–1973. doi: 10.1093/bioinformatics/btp348.

Chave J, Sothers C, Iribar A, Suescun U, Chase MW, Prance GT. 2020. Rapid diversification rates in Amazonian Chrysobalanaceae inferred from plastid genome phylogenetics. Bot J Linn Soc. 194(3):271–289. doi: 10.1093/botlinnean/boaa052.

Chernomor O, Von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. Syst Biol. 65(6):997–1008. doi: 10.1093/sysbio/syw037.

Chifman J, Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics. 30(23):3317–3324. doi: 10.1093/bioinformatics/btu530.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM. 2021. Twelve years of SAMtools and BCFtools. Gigascience. 10(2):giab008. doi: 10.1093/gigascience/giab008.

Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol & Evol. 24(6):332–340. doi: 10.1016/j.tree.2009.01.009.

Dray S, Dufour A. 2007. The ade4 package: implementing the duality diagram for ecologists. J Stat Soft. 22(4):1–20. doi: 10.18637/jss.v022.i04.

Heled J, Bouckaert RR. 2013. Looking for trees in the forest: summary tree from posterior samples. BMC Evol Biol. 13(1):1–11. doi: 10.1186/1471-2148-13-221.

Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 35(2):518–522. doi: 10.1093/molbev/msx281.

Hoorn C, Lohmann LG, Boschman LM, Condamine FL. 2023. Neogene history of the Amazonian Flora: a perspective based on geological, palynological, and molecular phylogenetic data. Annu Rev Earth Planet Sci. 51(1):419–446. doi: 10.1146/annurev-earth-081522-090454.

Huang YY, Mori SA, Kelly LM. 2015. Toward a phylogenetic-based generic classification of neotropical Lecythidaceae—I. Status of Bertholletia, Corythophora, Eschweilera and Lecythis. Status Bertholletia, Corythophora, Eschweilera Lecythis Phytotaxa. 203(2):85–121. doi: 10.11646/phytotaxa.203.2.1.

Huang YY, Mori SA, Prance GT. 2008. A phylogeny of Cariniana (lecythidaceae) based on morphological and

anatomical data. Brittonia. 60(1):69–81. doi: 10.1007/s12228-008-9014-3.
Hughes CE, Pennington RT, Antonelli A. 2013. Neotropical plant evolution: assembling the big picture. Bot J Linn Soc. 171(1):1–18. doi: 10.1111/boj.12006.
Jaramillo C, Romero I, D'Apolito C, Bayona G, Duarte E, Louwye S, Escobar J, Luque J, Carrillo-Briceño JD, Zapata V. 2017. Miocene flooding events of western Amazonia. Sci Adv. 3(5):e1601693. doi: 10.1126/sciadv.1601693.
Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 14(6):587–589.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol and Evol. 30(4):772–780.
Kloster A, Gnaedinger S, 1dami-Rodrigues K. 2017. Nuevos registros de Maderas Del Mioceno De La Formación Solimões, Cuenca de Acre, Amazonia, Brasil. Rev Bras Paleontol. 20(3):383–404. doi: 10.4072/rbp.2017.3.09.
Koenen EJ, Clarkson JJ, Pennington TD, Chatrou LW. 2015. Recently evolved diversity and convergent radiations of rainforest mahoganies (meliaceae) shed new light on the origins of rainforest hyperdiversity. New Phytol. 207(2):327–339.
Larson DA, Vargas OM, Vicentini A, Dick CW. 2021. Admixture may be extensive among hyperdominant amazon rainforest tree species. New Phytol. 232(6):2520–2534. doi: 10.1111/nph.17675.
Li H, Durbin R. 2009. Fast and accurate short read alignment with burrows–Wheeler transform. Bioinformatics. 25(14):1754–1760. doi: 10.1093/bioinformatics/btp324.
Liu BB, Ma ZY, Ren C, Hodel RG, Sun M, Liu XQ, Liu G-N, Hong D-Y, Zimmer EA, Wen J. 2021. Capturing single-copy nuclear genes, organellar genomes, and nuclear ribosomal DNA from deep genome skimming data for plant phylogenetics: a case study in vitaceae. J Sytematics Evol. 59(5):1124–1138. doi: 10.1111/jse.12806.
Mehrotra RC, Srivastava G. 2017. In situ lecythidaceae wood from the Oligocene of Makum Coalfield, Northeast India. Iawa J. 38(2):162–169. doi: 10.1163/22941932-20170165.
Minh BQ, Hahn MW, Lanfear R, Rosenberg M. 2020. New methods to calculate concordance factors for phylogenomic datasets. Mol Biol Evol. 37(9):2727–2733. doi: 10.1093/molbev/msaa106.
Mori SA, Berkov A, Gracie CA, Hecklau EF. Editors. 2011. Tropical plant collecting – from the field to the internet. 332. Tcc Editora.
Mori SA, Cremers G, Gracie CA, de Granville JJ, Heald SV, Hoff M, Mitchell JD. 2002. Guide to the vascular plants of Central French Guiana. (NY): New York Botanical Garden Press.
Mori SA, Kiernan EA, Smith NP, Kelly LM, Huang YY, Prance GT, Thiers BM. 2017. *Observations on the phytogeography of the lecythidaceae clade (Brazil nut family). Phytoneuron 2017-30: 1–85.*
Mori SA, Prance GT. 1990. Lecythidaceae–part II. The zygomorphic-flowered New World genera. Flora Neotrop Monogr. 21(2):1–376.
Mori SA, Tsou CH, Wu CC, Cronholm B, Anderberg AA. 2007. Evolution of Lecythidaceae with an emphasis on the circumscription of neotropical genera: information from combined ndhF and trnL-F sequence data. Am J Botany. 94(3):289–301. doi: 10.3732/ajb.94.3.289.
Morton CM, Prance GT, Mori SA, Thorburn LG. 1998. Recircumscription of the Lecythidaceae. Taxon. 47(4):817–827. doi: 10.2307/1224186.
Mourier T, Bengtson P, Bonhomme M, Buge E, Cappetta H, Crochet JY, Taquet P, Hirsch KF, Jaillard E, Laubacher G. 1988. The upper cretaceous-lower tertiary marine to continental transition in the Bagua basin, northern Perú. Newsl Stratigraphy. 19(3):143–177. doi: 10.1127/nos/19/1988/143 .
Paradis E. 2013. Molecular dating of phylogenies by likelihood methods: a comparison of models and a new information criterion. Mol Phylogenet Evol. 67(2):436–444. doi: 10.1016/j.ympev.2013.02.008.
Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 20(2):289–290. doi: 10.1093/bioinformatics/btg412.
Pelser PB, Kennedy AH, Tepe EJ, Shidler JB, Nordenstam B, Kadereit JW, Watson LE. 2010. Patterns and causes of incongruence between plastid and nuclear senecioneae (Asteraceae) phylogenies. Am J Botany. 97(5):856–873. doi: 10.3732/ajb.0900287.
Pirie MD, Maas PJ, Wilschut RA, Melchers-Sharrott H, Chatrou LW. 2018. Parallel diversifications of Cremastosperma and Mosannona (annonaceae), tropical rainforest trees tracking neogene upheaval of South America. R Soc Open Sci. 5(1):171561.
Pons D. 1983. *Lecythidospermum bolivarensis* (Berry) nov. comb. graine fossile de Lecythidaceae (Angiospermae) du Tertiaire de la Colombie. Ann de Paléontologie. 69:1–12.
Pons D, De Franceschi DARIO. 2007. Neogene woods from western Peruvian Amazon and palaeoenvironmental interpretation. Bull Geosci. 82(4):343–354. doi: 10.3140/bull.geosci.2007.04.343.
Pouchon C, Boluda CG. 2023. REFMAKER: make your own reference to target nuclear loci in low coverage genome skimming libraries. Phylogenomic application in Sapotaceae. Mol Phylogenet Evol. 186:107826. doi: 10.1016/j.ympev.2023.107826.
Pouchon C, Boyer F, Roquet C, Denoeud F, Chave J, Coissac E, Wincker P, Lavergne S. 2022. ORTHOSKIM: in silico sequence capture from genomic and transcriptomic libraries for phylogenomic and barcoding applications. Mol Ecol Resour. 22(5):2018–2037. doi: 10.1111/1755-0998.13584.
Pouchon C, Lavergne S, Fernández Á, Alberti A, Aubert S, Mavárez J. 2021. Phylogenetic signatures of ecological divergence and leapfrog adaptive radiation in *Espeletia*. Am J Botany. 108(1):113–128. doi: 10.1002/ajb2.1591.
Prakash U, Dayal R. 1964. *Barringtonioxylon eopterocarpum* sp. nov. a fossil wood of Lecythidaceae from the Deccan intertrappean beds of Mahurzari. J Palaeosciences. 13(1–3):25–29. doi: 10.54991/jop.1964.675.
Prance G, Naczi R, Lohmann LG. 2021. In remembrance of: Scott A. Mori (1941–2020), tropical botanist extraordinaire. Biotropica. 53(1):339–346. doi: 10.1111/btp.12920.
Prance GT, Jongkind CC. 2015. A revision of African Lecythidaceae. Kew Bull. 70(1):1–68. doi: 10.1007/s12225-014-9547-4.
Rabosky DL, Grundler M, Anderson C, Title P, Shi JJ, Brown JW, Huang H, Larson JG. 2014. BAMM tools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. Methods Ecol Evol. 5(7):701–707. doi: 10.1111/2041-210X.12199.
Rabosky DL, Kolokotronis S-O. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PLOS ONE. 9(2):e89543. doi: 10.1371/journal.pone.0089543.

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

Sanderson MJ. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol Biol Evol. 14(12):1218–1231. doi: 10.1093/oxfordjournals.molbev.a025731.

Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol. 19(1):101–109. doi: 10.1093/oxfordjournals.molbev.a003974.

Sayyari E, Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol Biol and Evol. 33(7):1654–1668.

Schliep K. 2011. Phangorn: phylogenetic analysis in R. Bioinformatics. 27(4):592–593. doi: 10.1093/bioinformatics/btq706.

Shallom LJ. 1960. Fossil dicotyledonous wood of Lecythidaceae from the Deccan intertrappean beds of Mahurzari. J Indian Bot Soc. 39:198–203.

Shi JJ, Rabosky DL. 2015. Speciation dynamics during the global radiation of extant bats. Evolution. 69 (6):1528–1545. doi: 10.1111/evo.12681.

Smith SA, O'Meara BC. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics. 28(20):2689–2690. doi: 10.1093/bioinformatics/bts492.

Srivastava G, Mehrotra RC. 2018. *Barringtonia* Forster & Forster (Lecythidaceae) leaf from the late Oligocene of Assam, India. J Palaeosciences. 67(1–2):139–145. doi: 10.54991/jop.2018.54.

Steenwyk JL, Li Y, Zhou X, Shen XX, Rokas A. 2023. Incongruence in the phylogenomics era. Nat Rev Genet. 24(12):834–850. doi: 10.1038/s41576-023-00620-x.

Ter Steege H, Mota de Oliveira S, Pitman NC, Sabatier D, Antonelli A, Guevara Andino JE, Salomão RP, Salomão RP. 2019. Towards a dynamic list of Amazonian tree species. Sci Rep. 9(1):3501. doi: 10.1038/s41598-019-40101-y.

Thomson AM, Vargas OM, Dick CW. 2018. Complete plastome sequences from *Bertholletia excelsa* and 23 related species yield informative markers for Lecythidaceae. Appl Plant Sci. 6(5):e01151. doi: 10.1002/aps3.1151.

Vargas OM, Dick CW. 2020. Diversification history of neotropical lecythidaceae, an ecologically dominant tree family of amazon rain forest. Neotrop Diversif: Patterns and Processes; p. 791–809.

Vargas OM, Heuertz M, Smith SA, Dick CW. 2019. Target sequence capture in the Brazil nut family (Lecythidaceae): Marker selection and in silico capture from genome skimming data. Mol Phylogenet Evol. 135:98–104. doi: 10.1016/j.ympev.2019.02.020.

Vargas OM, Larson DA, Batista J, Cornejo X, Luize BG, Medellín- Zabala D, Ribeiro M, Smith NP, Smith SA, Vicentini A, et al. 2024. Reclassification of the bertholletia clade of the Brazil nut family (lecythidaceae) based on a phylogenetic analysis of Plastome and target sequence capture data. Harvard Papers in Botany. 29(1):159–179.

Vargas OM, Ortiz EM, Simpson BB. 2017. Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). New Phytol. 214 (4):1736–1750. doi: 10.1111/nph.14530.

Wheeler EA, Srivastava R, Manchester SR, Baas P, Wiemann M. 2017. Surprisingly modern latest cretaceous–earliest Paleocene woods of India. Iawa J. 38 (4):456–542. doi: 10.1163/22941932-20170174.

Woodcock DW, Meyer HW, Prado Y. 2017. The piedra chamana fossil woods (Eocene, Peru). Iawa J. 38 (3):313–365. doi: 10.1163/22941932-20170175.

World Flora Online Plant, Alan E, Roger H, William U, Mark W, Gregory A, Giovani carlos A, Marco A, Wisnu handoyo A, Hannah A, Joh A., et al.2023. World Flora Online Plant List December 2023. (online database). doi: 10.5281/zenodo.10425161.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol and Evol. 24(8):1586–1591.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III : polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinf. 19(6):15–30. doi: 10.1186/s12859-018-2129-y.

Zuntini AR, Carruthers T, Maurin O, Bailey PC, Leempoel K, Brewer GE, Knapp S. 2024. Phylogenomics and the rise of the angiosperms. Nature 629: 843–850.