RESEARCH ARTICLE

AMERICAN JOURNAL OF BOTANY

Unraveling the biogeographical history of Chrysobalanaceae from plastid genomes[1]

Léa Bardon[2], Cynthia Sothers[3], Ghillean T. Prance[3], Pierre-Jean G. Malé[4], Zhenxiang Xi[5], Charles C. Davis[5], Jerome Murienne[2], Roosevelt García-Villacorta[6], Eric Coissac[7], Sébastien Lavergne[7], and Jérôme Chave[2,8]

PREMISE OF THE STUDY: The complex geological and climatic history of the Neotropics has had major implications on the diversification of plant lineages. Chrysobalanaceae is a pantropical family of trees and shrubs with 75% of its 531 species found in the Neotropics, and a time-calibrated phylogeny of this family should shed light on the tempo of diversification in the Neotropical flora. Previously published phylogenetic hypotheses of this family were poorly supported, and its biogeography remains unclear.

METHODS: We assembled the complete plastid genome of 51 Chrysobalanaceae species, and increased taxon sampling by Sanger-sequencing of five plastid regions for an additional 88 species. We generated a time-calibrated tree including all 139 Chrsyobalanaceae species and 23 outgroups. We then conducted an ancestral area reconstruction analysis and estimated diversification rates in the family.

KEY RESULTS: The tree generated with the plastid genome alignment was almost fully resolved. It supports the polyphyly of *Licania* and *Hirtella*. The family has diversified starting around the Eocene-Oligocene transition. An ancestral area reconstruction confirms a Paleotropical origin for Chrysobalanaceae with several transoceanic dispersal events. The main Neotropical clade likely resulted from a single migration event from Africa around 28 mya ago, which subsequently underwent rapid diversification.

CONCLUSIONS: Given the diverse ecologies exhibited by extant species, we hypothesize that the rapid diversification of Chrysobalanaceae following the colonization of the Neotropics was triggered by habitat specialization during the complex geological and paleoclimatic history of the Neotropics.

KEY WORDS biogeography; Chrysobalanaceae; diversification; phylogeny; plastid genomes

Studies in Neotropical biogeography have a long history (Cracraft, 1973, 1985; Prance, 1974; Raven and Axelrod, 1974; Simpson, 1980; Gentry, 1982), and recent investigations applying phylogenetic approaches to this area have greatly stimulated this field. Unraveling the causes of diversification in the Neotropics is complex because the study area is vast, and its geological and paleoclimatic history remains controversial (Bush, 1994). We now know that the Neogene uplift of the Andes has spurred drastic changes in atmospheric circulation, local climates, and hydrology (Gregory-Wodzicki, 2000; Wesselingh et al., 2002; Insel et al., 2009; Hoorn et al., 2010). Novel environments arose both in the highlands and the lowlands, which probably caused extensive opportunities for species diversification. For example, Andean plant diversification has been rapid and spectacular (Hughes and Eastwood, 2006; Nürk et al., 2013; Luebert and Weigend, 2014), and novel biomes have also developed in dry-climate areas in the Neotropics. Such widespread continental-scale changes have lead to rapid diversification in tropical plant lineages (Simon et al., 2009; Lohmann et al., 2013; Hernandez-Hernandez et al., 2014; Willis et al., 2014; Weeks et al., 2014; Koenen et al., 2015).

In addition to in situ diversification, migration has also played a role in generating high species diversity in the lowland Neotropics. Lineages have migrated into and out of South America since the breakup of Gondwana (80–100 million years ago [mya]), and until the great biotic interchange between North and South America

[1] Manuscript received 3 November 2015; revision accepted 4 May 2016.

[2] Laboratoire Evolution et Diversité Biologique UMR 5174 CNRS, ENFA, Université Paul Sabatier 31062 Toulouse, France;

[3] Herbarium, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, United Kingdom;

[4] Department of Ecology and Evolutionary Biology, University of Toronto, Ontario, Canada;

[5] Department of Organismic and Evolutionary Biology, Harvard University Herbaria, 22 Divinity Avenue, Cambridge, Massachusetts 02138, USA;

[6] Royal Botanic Garden Edinburgh, 20a Inverleith Row Edinburgh EH3 5LR, United Kingdom; and

[7] Université Grenoble Alpes, CNRS, UMR 5553 LECA, F-38000 Grenoble, France.

[8] Author for correspondence (jerome.chave@univ-tlse3.fr)

doi:10.3732/ajb.1500463

(3–10 mya; Gentry, 1982; Cody et al., 2010; Sepulchre et al., 2014; Willis et al., 2014; Bacon et al., 2015). Multiple migration pathways along past land corridors have been suggested, including the North Atlantic land bridge (Tiffney, 1985; Davis et al., 2002; Antonelli et al., 2009; Couvreur et al., 2011), Beringia (Manns et al., 2012), or the Antarctic continent (Bartish et al., 2011). Transoceanic long-distance dispersal may also have played a key role (Renner, 2004). It is only through the accumulation of multiple biogeographical scenarios for unrelated taxa that the relative contribution of these dispersal pathways can be properly assessed.

The current study focuses on the cocoa plum family (Chrysobalanaceae), a medium-sized family of trees and shrubs (20 genera, 531 species; Prance and Sothers, 2003; Sothers and Prance, 2014; Sothers et al., 2014). Chrysobalanaceae are an important component of the lowland Neotropical flora. The majority of the described species are in the Neotropics (422 species), while only 66 species occur in tropical Africa and 43 in tropical Asia and Oceania. Chrysobalanaceae inhabit a range of lowland habitats including tropical forests, seasonally dry woodlands, and white-sand forests, including coastal areas (*Chrysobalanus* spp.). Within the Malpighiales, Chrysobalanaceae belong to the chrysobalanoid clade (Litt and Chase, 1999; Xi et al., 2012), which also includes Balanopaceae (9 spp. from Oceania), Dichapetalaceae (163 spp., pantropical), Trigoniaceae (23 spp., South America, Madagascar, and Southeast Asia), and Euphroniaceae (3 spp., restricted to north Amazonia; Litt and Chase, 1999), the latter being the sister clade of Chrysobalanaceae (Litt and Chase, 1999). Two Chrysobalanaceae genera, *Maranthes* and *Parinari*, are described as pantropical (Prance and White, 1988; Prance and Sothers, 2003; Sothers and Prance, 2014; Sothers et al., 2014); while *Chrysobalanus* and *Hirtella* occur in both tropical Africa and the Neotropics (Prance and White, 1988). The 16 remaining genera have a distribution restricted to a single continent (5 in southeastern Asia, 6 in Africa, and 5 in South America).

Recent studies have used Sanger DNA sequencing to address the phylogeny and biogeography of the family (Yakandawala et al., 2010; Bardon et al., 2013; Sothers and Prance, 2014; Sothers et al., 2014), but a comprehensive generic-level phylogeny is still lacking. Bardon et al. (2013) published a biogeographical scenario suggesting that the family originated in the Paleotropics around 80 mya and then migrated into the Neotropics for the first time during the mid-Eocene. They also suggested that Chrysobalanaceae have experienced faster rates of diversification in the Neotropics than in the Paleotropics, however, both results were based on a limited number of species, few genetic markers, and limited information about the fossil record. Jud et al. (2016) have provided a comprehensive review of the fossil record for Chrysobalanaceae. They show that one fossil dating back to the Eocene cannot be reliably assigned to Chrysobalanaceae. Because this calibration point predates the other fossils by almost 30 mya, their conclusion may have dramatic consequences on the time-calibration of the tree, and on the resulting biogeographic interpretations. Thus, there is a need for a critical re-examination of the evolutionary history of this important Neotropical plant clade.

To this end, we increased both character sampling and taxon sampling for this family. Using new fossil evidence, we propose a new phylogenetic hypothesis and biogeographical scenario for Chrysobalanaceae, and compare the diversification rates among the main clades. We then discuss the results in light of the climatic and geologic history of South America.

MATERIALS AND METHODS

Study species—Our DNA sequence data set covers 19 of the 20 currently described genera for the family (Appendix S1; see Supplemental Data with the online version of this article). We were unable to access tissue material for the monotypic genus *Bafodeya* Prance ex. F White, which is only known from West Africa. However, one *rbc*L sequence from this genus was produced and is available for analysis (Yakandawala et al., 2010). We added 23 outgroup species representative of the major lineages in Malpighiales, plus one outgroup from the sister order Oxalidales (*Oxalis latifolia* Kunth, Oxalidaceae, Moore et al., 2010).

DNA extraction—Approximately 20 mg of frozen leaves were homogenized in a TissueLyser II (Qiagen, Courtaboeuf, France) with three glass beads (4 mm in diameter) in a 2 mL microtube for three rounds of 45 s at 30 Hz separated by a 45 s pause to prevent overheating the samples. Total DNA was extracted with a Biosprint 15 auto-extractor (Qiagen), following the manufacturer's protocol. For *Euphronia guianensis*, total DNA was extracted from silica dried leaf tissue using DNeasy Plant Mini Kit (Qiagen) following the manufacturer's protocol.

Library preparation and high-throughput sequencing—We selected a subsample of 51 Chrysobalanaceae species for the sequencing of the plastid genome (Table 1). The selection of these species was made to include all major lineages in the family, and if possible, more than one species per lineage and per genus. We also included nine species in the Malpighiales to complement a previous analysis (Xi et al., 2012). We sheared total DNA into small fragments by sonication, and the fragments were end-repaired. Unlike previous studies, we did not perform any enrichment in plastid DNA (Straub et al., 2012; Xi et al., 2012). After this, purified fragments were ligated to sequencing indexed adapters. Fragments with an insert size of approximately 400 bp were gel-extracted and enriched with 10 cycles of PCR before library quantification and validation. One library was constructed for each species using the Illumina TruSeq DNA Sample Prep v2 kit following the protocol recommended by the supplier (Illumina, San Diego, CA), and library products were multiplexed by 24, and hybridized to an HiSEq. 2000 flow cell using the Illumina TruSeq PE Cluster Kit v3. High-throughput sequencing generated approximately 10 million pair-end reads of 101 nucleotides per species. We sequenced the plastid genome of *Euphronia guianensis* with a slightly different protocol. Illumina TruSeq PE libraries were prepared using the Apollo 324 system and its associated PrepX ILM DNA Library Protocol (WaferGen Biosystems), and sequenced using an HiSEq. 2500 sequencing platform. This yielded approximately 21 million pair-end reads of 150 nucleotides for *E. guianensis*. The nine outgroup species were sequenced on a HiSEq. 2500 platform. For all species, the average sequencing depth of the plastid genome reads ranged from ×90 to over ×300.

Plastid genome assembly and alignment—High-throughput sequencing reads were de-novo assembled using an iterative search based on the Velvet assembler v1.2.07 (Zerbino and Birney, 2008; Appendix S2a). We assembled the complete plastid genomes for all 51 Chrysobalanaceae species and for an additional 9 outgroup species (Table 1). Genomes were annotated by transferring the annotations of already published Chrysobalanaceae genomes using Geneious v9.0.5 (Kearse et al., 2012; Malé et al., 2014). These annotations

were manually edited when necessary, and checked using the DOGMA online annotation portal (Wyman et al., 2004).

Protein-coding sequences were analyzed separately. In addition to the newly generated plastid genomes, we also downloaded protein-coding sequences for 14 species listed in Table 1, using accessions reported in Xi et al. (2012) and Moore et al. (2010). This resulted in a 74-species data set. A total of 76 protein-coding regions were considered (Appendix S2b). They were aligned one by one in the forward sense, and checked visually using the procedure described in Ruhfel et al. (2014) and Stull et al. (2015). The concatenated protein-coding alignment had a total of 66597 nucleotides, and on average had 7.3% missing data (the two extremes were 28.6% missed reads for *Trigonia* sp. and 37.5% missed reads for *Hieronyma alchorneoides*).

For the 51 Chrysobalanaceae plastid genomes, noncoding regions were also aligned using the FFT-NS-i algorithm of MAFFT v7.222 (Katoh and Standley, 2013) keeping the default gap open penalty, and the default offset values. We separately aligned intergenic regions, gene introns, and tRNA introns, and subsequently concatenated them. We then trimmed the highly variable regions using Gblocks v0.91b (Castresana, 2000; Talavera and Castresana, 2007; Appendix S2c). We did not consider noncoding regions of outgroup species because these are highly divergent with possible issues of saturation. The final noncoding alignment had 55,333 nucleotides.

Sanger sequencing—To increase species sampling in Chrysobalanaceae, we complemented the above data set with Sanger sequences for 88 additional species (Appendix S2d). With the plastid genome alignment, we explored variable regions using PhyDesign (López-Giráldez and Townsend, 2011). Five regions were selected: two noncoding regions (*rps*4-*trn*T and *psa*J-*rpl*33), the *clp*P intron 1, and two regions in the *ycf*1 gene (see Koenen et al., 2015; and references therein). We designed primer pairs in Geneious (Appendix S2e). DNA amplification was performed using standard PCR protocols. PCR products were purified and then sequenced on an ABI3730XL automated DNA sequencer. Each sequence was assembled and manually edited in Geneious. Our data set included 92–100% of the species, depending on the locus (Appendix S2f). The concatenated alignment included 2246 nucleotides.

Phylogenetic inference—We inferred the topology of the phylogenetic tree using RAxML 8.2.4 (Stamatakis, 2006; 2014), with a general time reversible model with gamma rate (GTR+Γ). We inferred the topology using several partitioning strategies (Moore et al., 2010; Xi et al., 2012): a single partition for all protein-coding sequences, two partitions: one for the first two codon positions and the other for the third codon position, one partition per protein-coding sequence, and a single partition for amino-acids (with the Dayhoff model of substitutions). Topological node support was evaluated using the rapid bootstrapping algorithm with the 'autoMRE' automatic stopping criterion, as implemented in RAxML (Stamatakis et al., 2008). We then performed a phylogenetic analysis using the full 162-species alignment with RAxML 8.2.4. For this analysis, we assumed only three partitions: the first two codon positions, the third codon position, and the noncoding sequences.

The species *Bafodeya benna* was treated separately, given that only one *rbc*L sequence was available (NCBI accession number GQ424475). For this species, we ran a separate RAxML analysis after aligning the sequence to the 74-species plastid genome matrix.

Fossil record and time calibration of the phylogeny—We estimated a time-calibrated phylogenetic tree using BEAST v2.3.2 (Drummond and Rambaut, 2007; Drummond et al., 2012; Bouckaert et al., 2014) from the 162-species alignment, assuming a GTR+Γ model of molecular evolution and a Yule branching process prior. Because of the size of the matrix (121,930 × 162), and its sparseness, initialization with a random tree prior did not converge even after 100 million generations, as checked with Tracer v.1.6.0 (Drummond and Rambaut, 2007). We instead prescribed the best RAxML tree as a tree prior within the BEAST input file and fixed the topology by removing the tree model operators from the BEAST input file. The starting tree was time-calibrated with the same calibration constraints as those used in the main BEAST analysis (see below), using the chronos function in the R package APE v.3.4 (Paradis et al., 2004). Post–burn-in trees were merged into a maximum clade credibility tree using TreeAnnotator v.2.3.2 (Drummond and Rambaut, 2007).

Two fossils were used for calibration within Chrysobalanaceae. Fossils of endocarps and of wood were assigned to *Parinari* in a geological formation dating to the early Miocene (Jud et al., 2016). The determination of the endocarps excludes a possible confusion with the sister genus *Neocarya* on the basis of shape, position of the germination plugs, and presence of a beak at the base of the endocarp (Tiffney et al., 1994; Jud et al., 2016). We therefore constrained the stem node of *Parinari* by dating the most common recent ancestor (MRCA) of *Parinari-Neocarya* at 19 mya. We also used the fossil flower and fruits assigned to *Licania dominicensis* (Poinar et al., 2008; Chambers and Poinar, 2010). This fossil preserved in the Dominican amber dates back to at least the early Miocene, and we used 15–20 mya to date the MRCA between *Licania michauxii* and the rest of the Neotropical clade sensu Bardon et al., (2013). Note that the fossil dating to 49 Ma and assigned to *Chrysobalanus* by Wodehouse (1932) was used in the calibration of Xi et al. (2012) and Bardon et al. (2013). However, it does not clearly belong to this family (Jud et al., 2016) and was therefore excluded from the present analysis.

We constrained the tree with six calibration times (Appendix S2g). *Balanops caledonica* was dated at 23.8 mya (Simpson, 1961; Xi et al., 2012), and this places a minimal age constraint on the stem of Balanopaceae. *Paleoclusia chevalieri* (89.3 mya; Crepet and Nixon, 1998; Ruhfel et al., 2013) constrains the minimal age of the stem of Clusiaceae. Finally, a pollen fossil assigned to *Caryocar* and dated at 55.5 mya (Germeraad et al., 1968; Xi et al., 2012) constrains the minimal age of stem of Caryocaraceae. We also constrained the stem of Malpighiales at 111 mya with a 95% confidence interval of (106–117), consistent with previous estimates of Magallón et al. (2015) and Xi et al. (2012).

Ancestral area reconstruction—Ancestral area reconstruction models are often predicated on the assumption that in addition to cladogenic events, as summarized in the user-supplied phylogenetic tree (several tree realizations or a single consensus tree), two major processes may occur: dispersal across ranges (irrespective of the extent of the range), and extinction in a region. Such models have been implemented in several statistical packages (Ronquist, 1997; Ree and Smith, 2008; Landis et al., 2013). Here we used the R package BioGeoBEARS (Matzke, 2013a,b; 2014), which allows

TABLE 1. List of the species with sequenced plastid genomes included in this study. Details on geographical origin and accession numbers are provided in Appendix S1.

Family	Species	Source
Chrysobalanaceae	*Acioa guianensis* Aubl.	This study
Chrysobalanaceae	*Afrolicania elaeosperma* Mildbr.	This study
Chrysobalanaceae	*Angelesia splendens* Korth.	This study
Chrysobalanaceae	*Atuna racemosa* Raf.	This study
Chrysobalanaceae	*Chrysobalanus icaco* L.	Malé et al. (2014)
Chrysobalanaceae	*Couepia caryophylloides* Benoist	This study
Chrysobalanaceae	*Couepia grandiflora* (Mart. & Zucc.) Benth. ex Hook.f.	This study
Chrysobalanaceae	*Couepia guianensis* Aubl.	Malé et al. (2014)
Chrysobalanaceae	*Couepia ovalifolia* (Schott) Benth. ex Hook.f.	This study
Chrysobalanaceae	*Couepia paraensis* (Mart. & Zucc.) Benth.	This study
Chrysobalanaceae	*Couepia paraensis* subsp *cerradoana* Prance	This study
Chrysobalanaceae	*Couepia polyandra* (Kunth) Rose	This study
Chrysobalanaceae	*Couepia rankiniae* Prance	This study
Chrysobalanaceae	*Couepia sandwithii* Prance	This study
Chrysobalanaceae	*Couepia subcordata* Benth. ex Hook.f.	This study
Chrysobalanaceae	*Dactyladenia bellayana* (Baill.) Prance & F.White	This study
Chrysobalanaceae	*Dactyladenia buchneri* (Engl.) Prance & Sothers	This study
Chrysobalanaceae	*Dactyladenia floretii* Breteler	This study
Chrysobalanaceae	*Exellodendron barbatum* (Ducke) Prance	This study
Chrysobalanaceae	*Gaulettia elata* (Ducke) Sothers & Prance	This study
Chrysobalanaceae	*Grangeria borbonica* Lam.	This study
Chrysobalanaceae	*Hirtella macrosepala* Sandwith	This study
Chrysobalanaceae	*Hirtella physophora* Mart. & Zucc.	Malé et al. (2014)
Chrysobalanaceae	*Hirtella racemosa* Lam.	This study
Chrysobalanaceae	*Hirtella suffulta* Prance	This study
Chrysobalanaceae	*Hirtella zanzibarica* Oliv.	This study
Chrysobalanaceae	*Kostermanthus robustus* Prance	This study
Chrysobalanaceae	*Licania alba* (Bernoulli) Cuatrec.	Malé et al. (2014)
Chrysobalanaceae	*Licania canescens* Benoist	This study
Chrysobalanaceae	*Licania glabriflora* Prance	This study
Chrysobalanaceae	*Licania heteromorpha* Benth.	Malé et al. (2014)
Chrysobalanaceae	*Licania macrophylla* Benth.	This study
Chrysobalanaceae	*Licania majuscula* Sagot	This study
Chrysobalanaceae	*Licania membranacea* Sagot ex Laness.	This study
Chrysobalanaceae	*Licania michauxii* Prance	This study
Chrysobalanaceae	*Licania micrantha* Miq.	This study
Chrysobalanaceae	*Licania minutiflora* (Sagot) Fritsch	This study
Chrysobalanaceae	*Licania ovalifolia* Kleinh.	This study
Chrysobalanaceae	*Licania sprucei* (Hook.f.) Fritsch	Malé et al. (2014)
Chrysobalanaceae	*Licania tomentosa* (Benth.) Fritsch	This study
Chrysobalanaceae	*Magnistipula butayei* De Wild.	This study
Chrysobalanaceae	*Maranthes gabunensis* (Engl.) Prance	This study
Chrysobalanaceae	*Maranthes glabra* (Oliv.) Prance	This study
Chrysobalanaceae	*Maranthes kerstingii* (Engl.) Prance ex F.White	This study
Chrysobalanaceae	*Neocarya macrophylla* (Sabine) Prance ex F.White	This study
Chrysobalanaceae	*Parastemon urophyllus* (Wall. ex A.DC.) A.DC.	This study
Chrysobalanaceae	*Parinari campestris* Aubl.	Malé et al. (2014)
Chrysobalanaceae	*Parinari capensis* Harv. subsp *incohata* F.White	This study
Chrysobalanaceae	*Parinari curatellifolia* Planch. ex Benth.	This study
Chrysobalanaceae	*Parinari oblongifolia* Hook.f.	This study
Chrysobalanaceae	*Hunga gerontogea* (Schltr.) Prance	This study
Euphroniaceae	*Euphronia guianensis* (R.H.Schomb.) Hallier f.	This study
Oxalidaceae	*Oxalis latifolia* Kunth.	Moore et al. (2010)
Centroplacaceae	*Bhesa* sp.	Xi et al. (2012)
Irvingiaceae	*Irvingia malayana* Oliv. Ex Benn.	Xi et al. (2012)
Pandaceae	*Microdesmis caesarifolia* Planch.	Xi et al. (2012)
Rhizophoraceae	*Cassipourea guianensis* Aubl.	This study
Rhizophoraceae	*Rhizophora mangle* L.	Xi et al. (2012)
Bonnetiaceae	*Ploiarium* sp.	Xi et al. (2012)
Clusiaceae	*Clusia rosea* Jacq.	Xi et al. (2012)
Humiriaceae	*Humiria balsamifera* (Aubl.) J.St.-Hil.	This study
Violaceae	*Amphirrhox longifolia* (A.St.-Hil.) Spreng.	This study
Goupiaceae	*Goupia glabra* Aubl.	This study
Phyllanthaceae	*Hieronyma alchorneoides* Allemao	This study

continued

TABLE 1, *continued*

Family	Species	Source
Phyllanthaceae	*Flueggea suffruticosa* (Pall.) Baill.	Xi et al. (2012)
Euphorbiaceae	*Conceveiba guianensis* Aubl.	This study
Peraceae	*Pera bumeliifolia* Griesb.	Xi et al. (2012)
Peraceae	*Pogonophora schomburgkiana* Miers ex Benth.	This study
Balanopaceae	*Balanops pachyphylla* Baill. Ex Guillaumin	Xi et al. (2012)
Caryocaraceae	*Anthodiscus peruanus* Baill.	Xi et al. (2012)
Dichapetalaceae	*Dichapetalum zenkeri* Engl.	Xi et al. (2012)
Trigoniaceae	*Trigonia sp.*	Xi et al. (2012)
Caryocaraceae	*Caryocar glabrum* (Aubl.) Pers.	This study

probabilistic inference of ancestral geographic ranges on a phylogeny and comparison of different models of range evolution. BioGeoBEARS also explicitly models founder-event speciation cladogenic events, wherein the new species arises in a range outside of the ancestral range (Matzke, 2014).

To infer the ancestral areas in Chrysobalanaceae, we used a subtree of the main chronogram, restricted to the chrysobalanoid clade (143 species). We coded the species as being either present or absent in three geographical areas: tropical America, tropical Africa (including the Mascarene islands and Madagascar), and Southeast Asia. We inferred the evolution of geographical ranges on a single consensus tree in a likelihood framework, without constraining the directionality or timing of dispersal. Then, we assessed the fit of each model using AIC weights and selected the best model.

Inferring diversification rates—We tested for the presence of shifts in diversification rates along the entire time-calibrated phylogenetic tree using the BAMM software (Rabosky, 2014). BAMM was run on the maximum clade credibility tree obtained from the BEAST analysis during 10 million generations, and the results were analyzed using the BAMMtools package (Rabosky et al., 2014). Speciation, extinction, and net diversification by clade were then inferred using BayesRate v1.63 (Silvestro et al., 2011). Markov chain Monte Carlo (MCMC) sampling for the birth-death model was conducted during 40,000 generations per tree on a subsample of 10 trees from the output of the BEAST tree reconstruction, and only the cold chain parameter estimates were reported. In both BAMM and BayesRate, we accounted for the incomplete taxon sampling for each clade. Taxon sampling per clade was assessed on the basis of recent monographic revisions (Prance and Sothers, 2003; Sothers et al., 2014; Sothers and Prance, 2014).

RESULTS

Phylogenetic resolution—We sampled 139 Chrysobalanaceae species (approximately 26%) and included another 23 species as outgroups. The topology inferred using plastid genome data (74 species; 17,594 distinct alignment patterns) showed strong overall support within the family (Fig. 1), irrespective of the partitioning scheme (Appendix S3 a,b). The chronogram, as inferred using the full data set, also showed a good overall support (Fig. 2). *Kostermanthus* (3 spp.) represented the earliest diverging lineage in the family. The rest of the family included five main clades (clade names are provided in Fig. 1). Clade A included the pantropical genus *Parinari* (39 spp.) and its African sister genus *Neocarya* (1 sp.). Based on a single sequence, *Bafodeya benna* was tentatively placed in clade A (Fig. 1; posterior probability: 0.77; bootstrap support: 39%). Clade B had a predominantly African affinity, with genera *Magnistipula* (14 spp.) being found in Africa and Madagascar, and *Grangeria* (2 spp.), restricted to Madagascar and the Mascarene Islands. Clade B also contained *Maranthes* (12 spp.), found mostly in Africa with the exception of two species (*M. corymbosa* in Southeast Asia, and *M. panamensis* in Central America; neither of which were included in the current study), and the two Asian genera *Parastemon* (3 spp.) and *Atuna* (8 spp.). Clade C included the African genus *Dactyladenia* (32 spp.) plus the African species *Hirtella zanzibarica*. Clade D included two South American genera (*Acioa*, 6 spp.; *Exellodendron*, 5 spp.), two Asian genera (*Angelesia*, 3 spp.; *Hunga*, 11 spp.) and one amphiatlantic genus found both in America and Africa (*Chrysobalanus*, 3 spp.). Finally, clade N (for Neotropical) contained over 75% of Chrysobalanaceae species in our sample, all Neotropical *Licania* (ca. 220 spp.), *Couepia* (58 spp.), *Gaulettia* (9 spp.) and *Hirtella* (105 spp.) species, and the African *Afrolicania elaeosperma* Mildbr. *Licania*, as currently circumscribed, could be separated into at least four lineages. First, *Licania michauxii* Prance, the only North American species, represented an early diverging subclade in clade N (clade Licania1). Second, a clade of five species (Licania2), including *L. minutiflora* (Sagot) Fritsch was sister to *Couepia*. Third, a clade of 27 species, included for instance *L. sprucei* (Hook.f.) Fritsch and *L. alba* (Bernoulli) Cuatrec. (Licania3). The fourth and last clade included *L. heteromorpha* Benth and was sister to *Hirtella* (Licania4). Two *Licania* species had an uncertain position: *L. brittoniana* Fritsch, close to clade Licania2; and *L. licaniiflora* (Sagot) S.F.Blake, which was close to clade Licania4.

Divergence time estimation and ancestral areas reconstruction—According to our dated phylogenetic reconstruction, Chrysobalanaceae diversified in the late Eocene to early Oligocene (Fig. 2; crown age: 33.4 mya; HPD: 30.2–37.3 mya; Appendix S3c). For the BioGeoBears ancestral area reconstruction, we found that the BayesArea model including the founder-speciation parameter was favored (Appendix S3d). The family originated in the Paleotropics, and the common ancestor of the family, excluding *K. robustus*, was found to be African. The first arrival of Chrysobalanaceae into the Neotropics is consistent with the stem age of the clade (D+N), dated at 29.5 mya (HPD: 26.9–32.5 mya). Secondary dispersal events from South America into Africa occurred between 11.5 and 24.2 Ma (*Afrolicania elaeosperma, Chrysobalanus icaco*, respectively) and from South America into Asia approximately 13.7 mya (HPD: 7.0–18.2 mya) as evidenced by the grouping of *Exellodendron* and (*Angelesia* + *Hunga*). Within the Neotropical clade, the largest *Licania* clade (clade Licania3) and *Couepia* diversified from around 23.6 mya (crown age; HPD: 21.9–25.4) and 19.5 mya (crown age; HPD: 17.5–21.9 mya), respectively. Diversification in

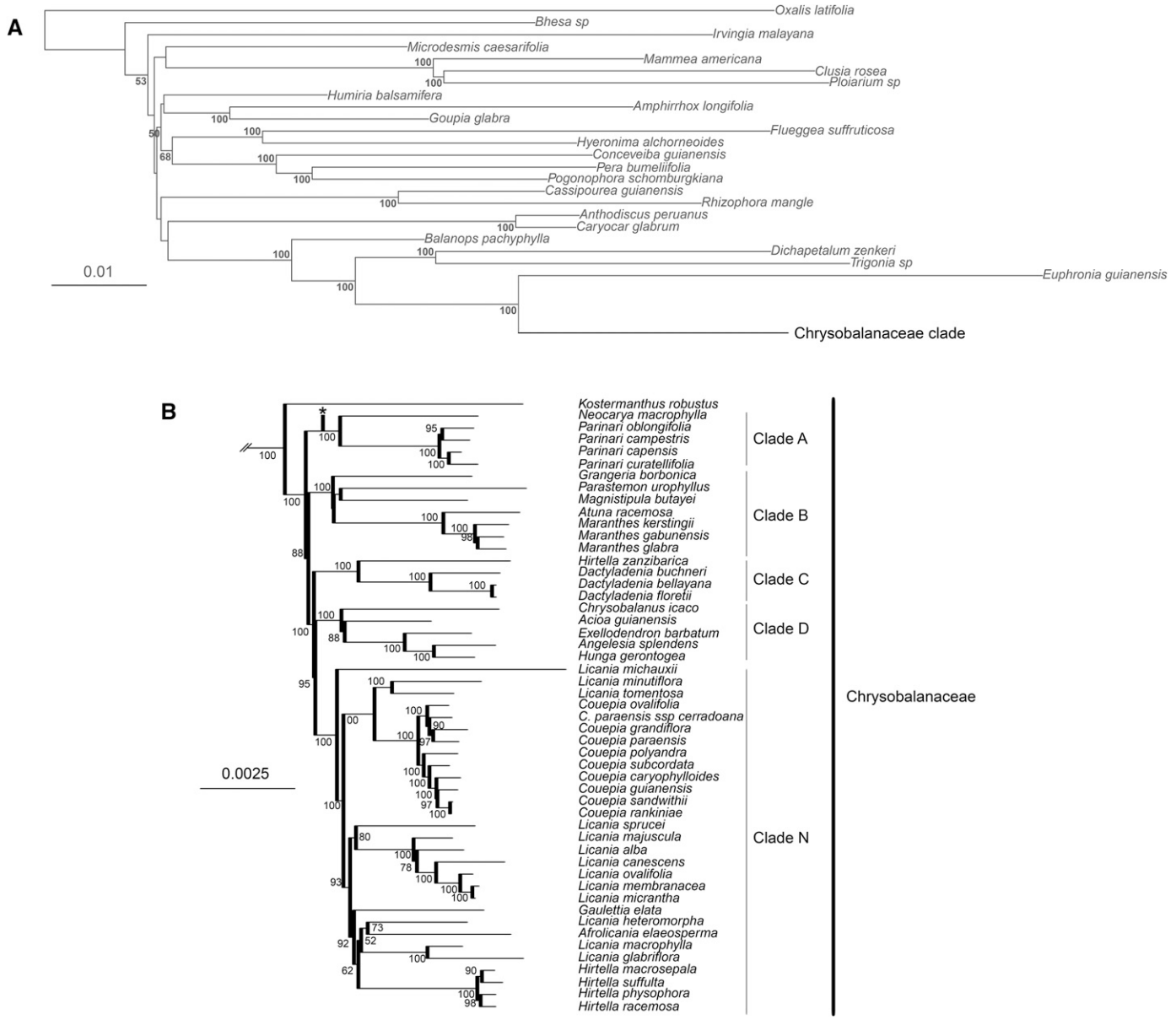


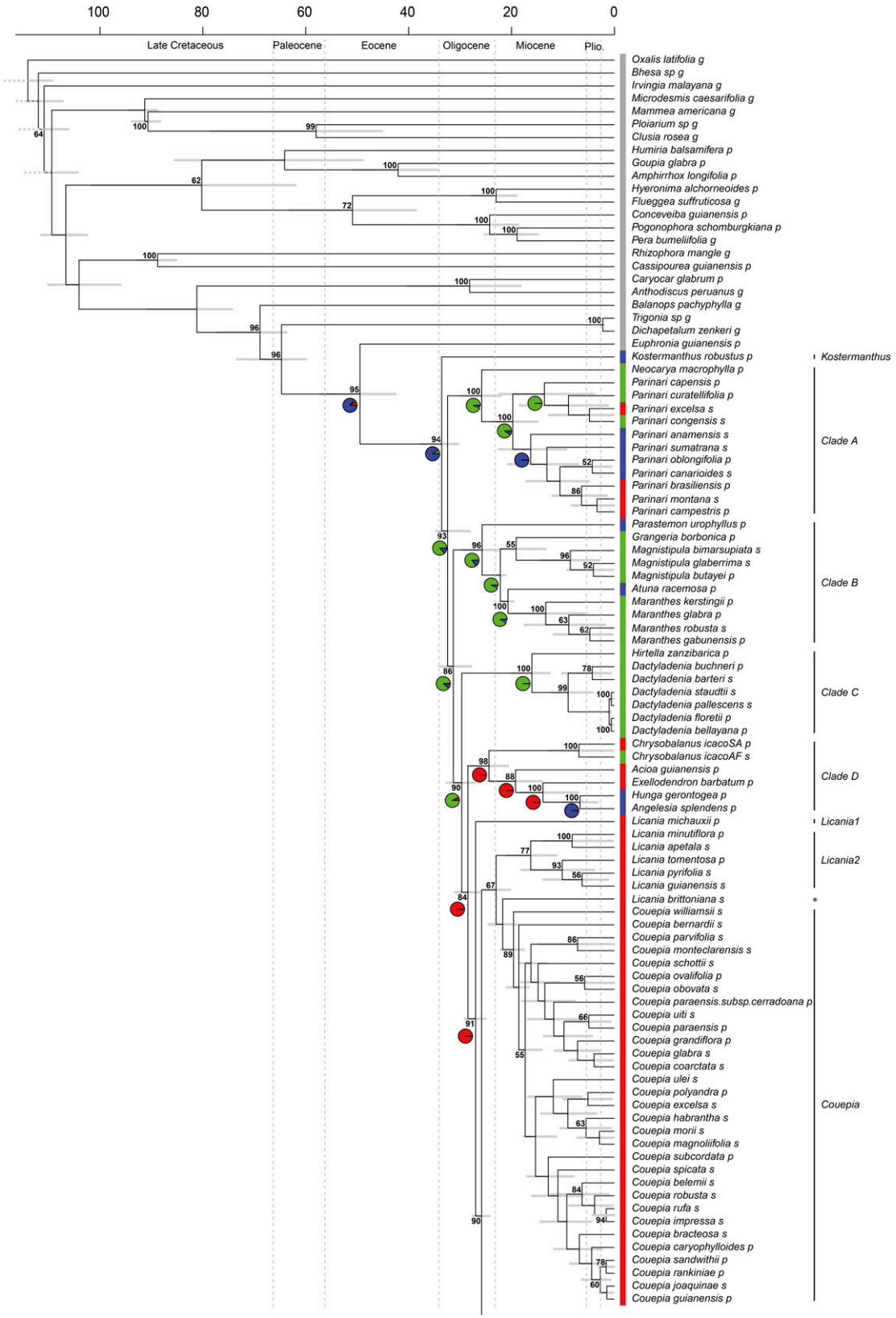
FIGURE 1 Phylogram obtained with 76 protein-coding sequences of 74 plastid genomes, and a maximum likelihood reconstruction method (RAxML). Bootstrap values (only > 50%) are indicated at each node: (A) tree including the outgroups only; (B) magnification (×4) of the Chrysobalanaceae clade. The star symbol locates the probable position of *Bafodeya benna* Prance ex. F White.

Neotropical *Hirtella* was found to be even younger, around 17.2 mya (crown age; HPD: 16.8–17.8 mya).

Inferring diversification rates—In the BAMM analysis, the most probable configuration of rate shifts included one shift in the branch leading to Chrysobalanaceae, and a secondary shift in a clade of seven species including *L. membranacea* (Appendix S3e). The posterior distributions of speciation, extinction, and net diversification rates were computed for all five clades A, B, C, D, and N with BayesRate (Fig. 3; Appendix S3f). It was higher in the Neotropical clade (λ_{net} = 0.15 species/mya) than in the other clades (λ_{net}= 0.10–0.11 species/mya), with the exception of clade C (*Dactyladenia* spp.), which presented a very high net diversification rate (0.17 species/mya), but also a much broader posterior distribution.

DISCUSSION

Phylogenetic reconstruction—Almost all nodes in the tree constructed from full plastid genomes were strongly supported (Fig. 1), irrespective of the partitioning scheme (Fig. S2). Our matrix had almost 18,000 patterns, so character sampling was greatly improved. This robust phylogenetic hypothesis represents a significant advance of this study over previously published analyses (Yakandawala et al., 2010; Bardon et al., 2013). Phylogenetic resolution and support should depend on both the number of characters and of taxa included in a study, but the optimal strategy has generated much debate (Philippe et al., 2011). In some studies, branch support was found to increase more by increasing taxon sampling than character sampling (Pollock et al., 2002; Heath


100
80
60
40
20
0
Late Cretaceous
Paleocene
Eocene
Oligocene
Miocene
Plio.
Oxalis latifolia g
Bhesa sp g
Irvingia malayana g
Microdesmis caesarifolia g
Mammea americana g
Ploiarium sp g
Clusia rosea g
Humiria balsamifera p
Goupia glabra p
Amphirrhox longifolia p
Hyeronima alchorneoides p
Flueggea suffruticosa g
Conceveiba guianensis p
Pogonophora schomburgkiana p
Pera bumeliifolia g
Rhizophora mangle g
Cassipourea guianensis p
Caryocar glabrum p
Anthodiscus peruanus g
Balanops pachyphylla g
Trigonia sp g
Dichapetalum zenkeri g
Euphronia guianensis p
Kostermanthus robustus p
Neocarya macrophylla p
Parinari capensis p
Parinari curatellifolia p
Parinari excelsa s
Parinari congensis s
Parinari anamensis s
Parinari sumatrana s
Parinari oblongifolia p
Parinari canarioides s
Parinari brasiliensis p
Parinari montana s
Parinari campestris p
Parastemon urophyllus p
Grangeria borbonica p
Magnistipula bimarsupiata s
Magnistipula glaberrima s
Magnistipula butayei p
Atuna racemosa p
Maranthes kerstingii p
Maranthes glabra p
Maranthes robusta s
Maranthes gabunensis p
Hirtella zanzibarica p
Dactyladenia buchneri p
Dactyladenia barteri s
Dactyladenia staudtii s
Dactyladenia pallescens s
Dactyladenia floretii p
Dactyladenia bellayana p
Chrysobalanus icacoSA p
Chrysobalanus icacoAF s
Acioa guianensis p
Exellodendron barbatum p
Hunga gerontogea p
Angelesia splendens p
Licania michauxii p
Licania minutiflora p
Licania apetala s
Licania tomentosa p
Licania pyrifolia s
Licania guianensis s
Licania brittoniana s
Couepia williamsii s
Couepia bernardii s
Couepia parvifolia s
Couepia monteclarensis s
Couepia schottii s
Couepia ovalifolia p
Couepia obovata s
Couepia paraensis.subsp.cerradoana p
Couepia uiti s
Couepia paraensis p
Couepia grandiflora p
Couepia glabra s
Couepia coarctata s
Couepia ulei s
Couepia polyandra p
Couepia excelsa s
Couepia habrantha s
Couepia morii s
Couepia magnoliifolia s
Couepia subcordata p
Couepia spicata s
Couepia belemii s
Couepia robusta s
Couepia rufa s
Couepia impressa s
Couepia bracteosa s
Couepia caryophylloides p
Couepia sandwithii p
Couepia rankiniae p
Couepia joaquinae s
Couepia guianensis p
Kostermanthus
Clade A
Clade B
Clade C
Clade D
Licania1
Licania2
*
Couepia

(Continued)

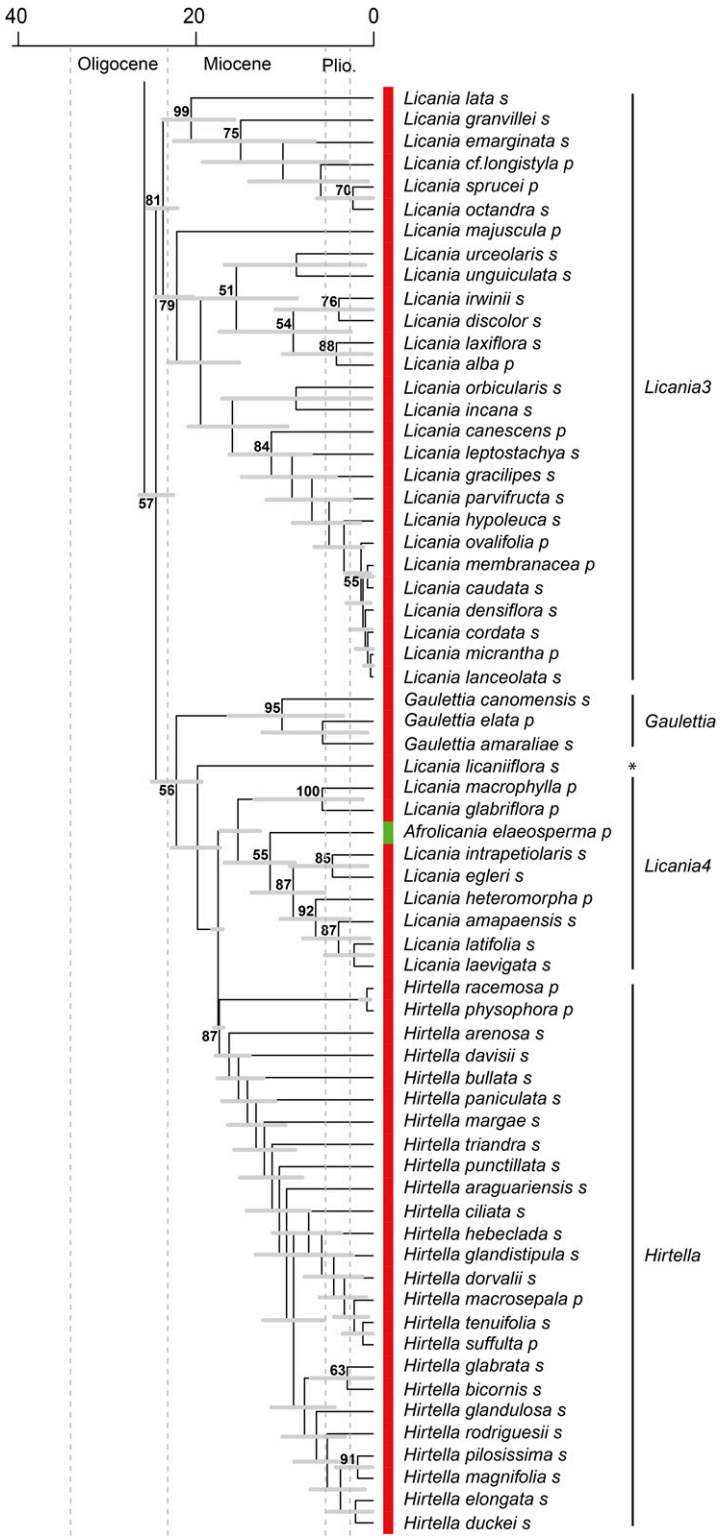


FIGURE 2 Maximum clade credibility tree of Chrysobalanaceae obtained from the 162-species data set with the geographical area displayed in color codes (red: Neotropics; green: Africa; blue: southeast Asia and Oceania). Bootstrap values (only >50%) are reported near the corresponding nodes, and the inferred age of the nodes is provided with the 95% confidence interval. Ancestral states within the chrysobalanoid clade, as inferred with the best model (BayArea+j), are also reported next to the nodes. The letter next to species name describes the sequence material used for this reconstruction (p: newly generated plastid genome; s: Sanger sequences; g: plastid genome coding sequences retrieved from GenBank). The time scale is in millions of years (Ma) and geological eras are reported near the time axis (Plio., Pliocene).

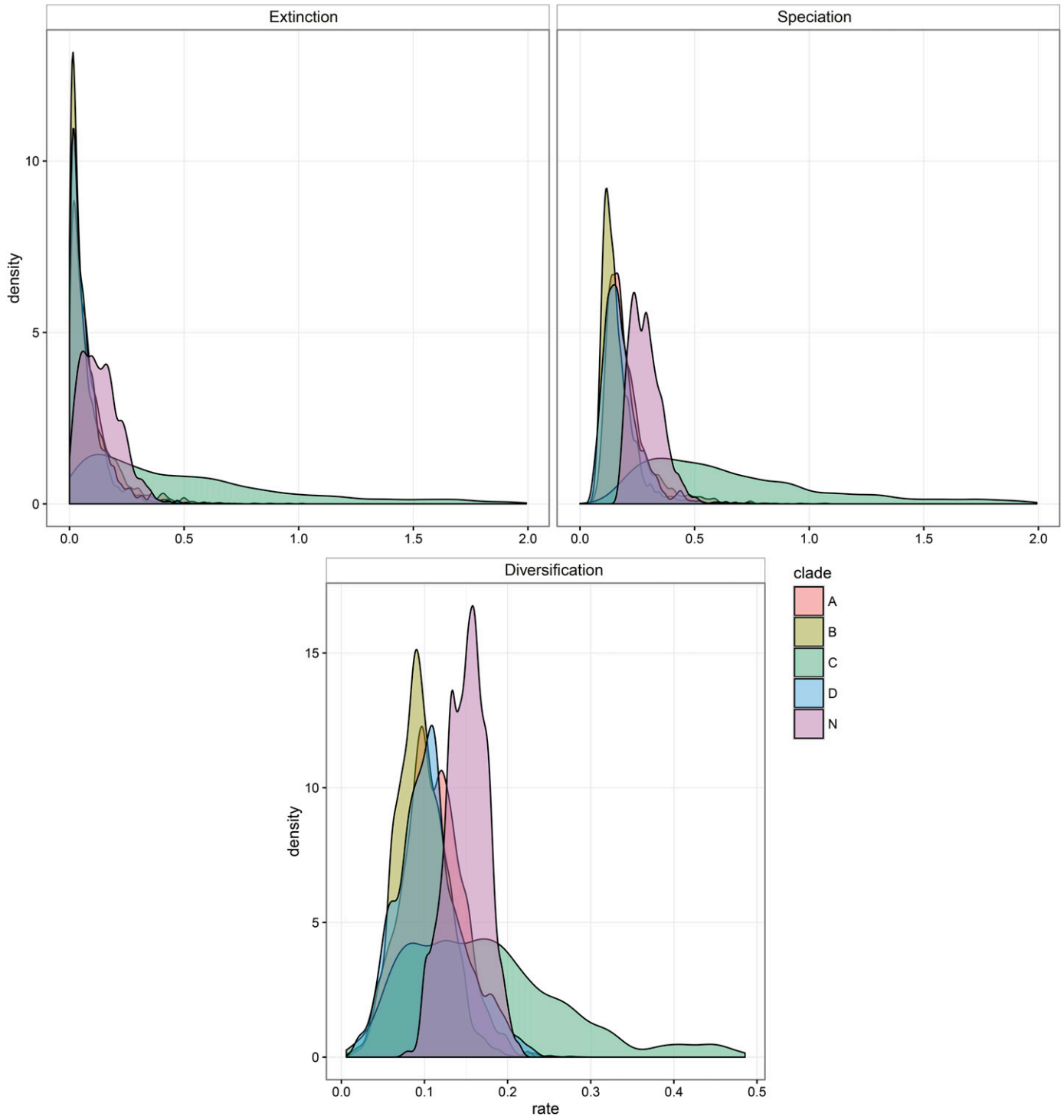


FIGURE 3 Posterior distribution of the speciation, extinction, net diversification rates for the five main clades in Chrysobalanaceae (clade names as in Fig. 2), represented by the posterior rate estimates (relative densities).

et al., 2008; Xi et al., 2012), but striking a suitable balance between missing taxa and missing characters is not easy and depends on the nature of the problem (Wiens, 2003; Wiens and Moen, 2008). Plastid genome sequencing has rejuvenated this old debate (Parks et al., 2009; Bayly et al., 2013; Williams et al., 2016), and the current study contributes to this debate. Increasing taxon sampling with amplicon sequencing, we found a slight increase in resolution, but not systematically across the clades. In clade A, the plastid-genome tree had 2/3 of the branches supported (BS > 70%), vs. 3/11 in the full phylogeny; in clade B, 4/6 vs. 6/9; in clade C, 3/3 vs. 5/5; in clade D, 3/4 vs. 5/5; and in clade N, 21/28 vs. 25/103. The *Hirtella* clade, with 1/24 branches supported, contributed disproportionately to the problem, and deeper sampling within this clade is a forthcoming challenge.

Phylogenetic relationships within Chrysobalanaceae—We were able to discuss the phylogenetic position of all 20 currently described genera in Chrysobalanaceae. The earliest-diverging lineage is represented by the little known tropical Asian genus *Kostermanthus* (Prance, 1979). This position of *Kostermanthus* is unexpected in view of the morphological features of this genus, which suggest an affinity to the phylogenetically more distantly related *Acioa* and *Atuna* (Yakandawala et al., 2010). It would be important to explore whether this hypothesis is supported using morphological characters. Clade A is the second earliest-diverging lineage and it confirms the close affinity of *Neocarya* and *Parinari*. About a fourth of the species in *Parinari* were included in this analysis, and the monophyly of *Parinari* is confirmed. We also show that based on available sequence data, the genus *Bafodeya* belongs to clade A.

Clade B was already recognized by Yakandawala et al. (2010), as grouping *Grangeria*, *Magnistipula*, and *Maranthes*. We found that *Atuna* and *Parastemon* are also included in this clade, with *Atuna* close to *Maranthes* and *Parastemon* as the early diverging branch of this clade.

Clade C includes all sampled species of *Dactyladenia* and *Hirtella zanzibarica*. In previous systematic treatments, the genus *Hirtella* was described as present predominantly in the Neotropics, but with two species in Africa (Prance and Sothers, 2003). One of these African *Hirtella* species, *H. zanzibarica* Oliv, was included for the first time in our analysis and was found to belong to clade B, sister to the African genus *Dactyladenia*. The remaining 25 Neotropical *Hirtella* species sampled formed a distinct monophyletic group within clade N. Du Petit-Thouars (1806) described the genus *Thelira*, based on a specimen from Madagascar (*H. thouarsiana* Baill. ex Laness.). Baillon (1868) later placed *Thelira* in synonymy with *Hirtella*, a circumscription accepted in more recent revisions (Prance and White, 1988; Prance and Sothers, 2003). Our analysis suggests that *Thelira* should be resurrected.

Clade D includes *Acioa, Chrysobalanus*, *Exellodendron, Angelesia*, and *Hunga* (see also Bardon et al., 2013). *Chrysobalanus* was found to be the early diverging group within this clade. *Hunga* and *Angelesia* were placed next to each other. Members of this clade exhibit large floral diversity, including *Chrysobalanus* with actinomorphic flowers versus the other related genera, which have weakly-to-strongly zygomorphic flowers (Prance and Sothers, 2003).

Finally, the Neotropical species are almost all included in a single clade (clade N). Recently, Sothers et al. (2014) used five markers to focus on the circumscription of species in the genus *Couepia*. Previously described *Couepia* species were found to belong to four separate clades: *Acioa*, *Hirtella*, *Couepia* sensu stricto, and a clade representing a novel genus, which they named *Gaulettia*. The three species included here belonging to *Gaulettia* formed a clade. Our analysis confirms these recent findings although the phylogenetic position of *Gaulettia* within clade N remains unclear (Fig. S2).

In *Licania*, we found four different clades, entirely restricted to clade N (Fig. 2). The first lineage (Licania1), sister to the rest of clade N, consists of a single species *L. michauxii*—the only North American species in the genus—that is included here for the first time in a molecular phylogeny of the family. A second *Licania* clade, called Licania2 in Fig. 2, was sister to *Couepia* sensu stricto. All species in this clade are representatives of subgenus *Moquilea* (Aubl.) section *Moquilea* sensu Prance and Sothers (2003), with the exception of *L. apetala* (E.Mey.) Fritsch. A third *Licania* clade, called Licania3 in Fig. 2, grouped most *Licania* species. Within this clade, we identified two well-supported subclades. The first one includes species belonging to subgenus *Moquilea* section *Moquilea*, and the second with species belonging to subgenus *Licania* section *Licania*, apart from *L. unguiculata* Prance (subgenus *Moquilea* section *Leptobalanus*) (Prance and Sothers, 2003). Finally, the last distinctive *Licania* clade, denoted clade Licania4 in Fig. 2, was close to the Neotropical *Hirtella*, and was overall less strongly supported. It had an affinity with subgenus *Licania* section *Hymenopus*, as defined in Prance and Sothers (2003). A different analysis of this group of species would be that clade Licania4 contains only six species (including *L. heteromorpha*), and that the four remaining species: *L. licaniiflora*, (*L. macrophylla*+*L. glabriflora*) and *Afrolicania elaeosperma* Mildbr. do not belong in this clade. *Afrolicania elaeosperma* is the only African species previously assigned to *Licania* (Prance and White, 1988), and segregated into *Afrolicania* by Prance and Sothers (2003), after Dissanayake (1999); this assignment to *Licania*, however, has always had a debated position. The present analysis does not completely resolve this problem, but based on plastid genome sequences only, there is strong evidence that it belongs to clade Licania4 (Fig. S2).

A revised biogeographical history for Chrysobalanaceae—An important insight of this study is the profound revision of the temporal scale over which Chrysobalanaceae have evolved and dispersed across the tropics. The revised stem age for Chrysobalanaceae is 49.2 mya (highest posterior density, HPD: 42.3–56.8 mya), and the crown age is dated at 33.4 mya (30.2–37.3 mya), around the Eocene-Oligocene transition. This is considerably younger than the ages inferred by Bardon et al. (2013; 86 mya) and Xi et al., (2012; 66.2 mya HPD: 60.3–74.9 mya), but is in line with the estimates of Magallón et al. (2015; stem age, 54.5 mya; HPD: 40.2–72.8 mya). This recent age is consistent with the revised fossil record of the family, which rejected all fossils older than the Oligocene-Miocene transition (Jud et al., 2016).

Even if the Paleotropical origin of Chrysobalanaceae is confirmed here, as was hypothesized by Bardon et al. (2013), the first arrival of Chrysobalanaceae into the Neotropics is now predicted to have occurred in the Oligocene, with clade D + clade N (Fig. 1) dating back to 29.5 mya. Different paths and land bridges have been suggested to explain such a dispersal pathway from Africa into the Neotropics. Given our dating of the arrival of the family into the Neotropics, the opportunity for a short-distance overseas dispersal via island chains in the southern Atlantic (Morley, 2003; Pennington and Dick, 2004) had passed. Alternatively, a migration through Laurasia, and through the North Atlantic land bridge to reach North America, has been proposed to explain the pantropical distribution of some plant families (Davis et al., 2002; Antonelli et al., 2009; Weeks et al., 2014; Fritsch et al., 2015). Based on our revised dating of the migration event, we can exclude that Chrysobalanaceae dispersed into South America around the Eocene thermal maximum through the North Atlantic Land Bridge (Tiffney, 1985; Lavin and Luckow, 1993). The presence of North American *Licania michauxii* as an early diverging lineage of clade N, plus the presence of *Chrysobalanus icaco* in North America (clade D), are both indicative of a dispersal route through North America, but the precise biogeographic scenario remains uncertain. The phylogenetic position of *L. michauxii* bears interesting similarities with that of *Protium fragrans* (Rose) Urb., which is endemic to Cuba, and belongs to the earliest diverging lineage of the Neotropical *Protium* clade, also dating to the Oligocene (Fine et al., 2014).

Clades N and D display three examples of probable, long, transoceanic migration during the Miocene. One is the dispersal of the *Angelesia+Hunga* clade in Southeast Asia and Oceana around 13.8 mya (HDP: 7.0–18.2 mya), which split from the South American genus *Exellodendron*. In addition, the African *Afrolicania elaeosperma* originated from a migration event ca. 11.5 mya (HPD: 8.8–16.6 mya) from the Neotropics. Finally, *Chrysobalanus icaco* is found both in South America and in Africa, but our data are insufficient to precisely date the dispersal event (probably from South America to Africa) that led to this disjunction.

Clade B originated 31.1 Ma ago (HPD: 27.7–33.9 mya), and the ancestral area reconstruction predicted an African origin with over 80% confidence over a Southeast Asian origin. Yet, there does not seem to be a simple dispersal scenario between these two continents. The Southeast Asian *Parastemon* is at the base of this clade, but the phylogenetic position of *Grangeria*, endemic to the Mascarenes Islands and Madagascar remains uncertain, so an Asian origin cannot be ruled out. Importantly in this discussion, *Atuna* has two endemic species in the western Ghats of India. This reinforces the hypothesis that this clade has dispersed across the Indian Ocean, but the exact timing of these dispersal events needs to be evaluated in the light of the late Eocene geography of this region (Ali and Aitchison, 2008).

Finally, both *Maranthes* (12 species, 20.5 Ma, HPD: 19.5–21.8 Ma for the stem age) and *Parinari* (39 species, 25.6 Ma, HPD: 22.0–29.9 Ma for the stem age) have achieved pantropical distribution. *Maranthes* probably has an African origin, with a single species in Central America [*M. panamensis* (Standl.) Prance & F.White], and another species widespread in Southeast Asia (*M. corymbosa* Blume). These last two species were not included in our analysis, but Sothers et al. (2014) confirmed that they belong in the genus *Maranthes*. For *Parinari*, our scenario suggests an African origin with an early dispersal event in Southeast Asia followed by a secondary dispersal into the Neotropics from Southeast Asia (Fig. 2). Note also that *Parinari excelsa* Sabine occurs both in Africa and South America. However, our taxon sampling remains limited for this group, and the phylogenetic tree is still poorly supported, so this scenario will require reexamination.

Rates of diversification—Almost 75% of total species diversity in Chrysobalanaceae is in the Neotropical clade N. We compared the rates of speciation, extinction, and net diversification among clades. We found that clade N had both higher speciation rate and higher extinction rate than the other clades in the family, with an inferred net diversification rate of 0.15 species/mya, compared to approximately 0.11 species/mya in the other clades. The Miocene appears to have been a major epoch for diversification in South America (Hoorn et al., 2010). Several groups have given rise to diverse plant genera within a short timeframe. The genus *Guatteria* (Annonaceae; 177 spp.; Maas et al., 2015) was predicted to have diversified during the Miocene (crown age 11–18 mya, Erkens et al., 2007, 2009), but have an old stem age (49–52 mya), and clarifying the time calibration and the phylogenetic position within the family has thus proven difficult (Maas et al., 2015). Other examples of large Neotropical genera having diversified around the Oligocene-Miocene include the Neotropical Protieae (Burseraceae; Fine et al., 2014), the *Dacryodes-Trattinnikia* clade (Burseraceae; Weeks et al., 2014), genus *Mimosa* (Fabaceae; 530 spp., Simon et al., 2009), genus *Astrocaryum* (Arecaceae; Roncal, et al., 2013), and two clades within the Meliaceae: the Neotropical *Trichilia*, and *Ruagea+Guarea* (Koenen et al., 2015). In our study, genera *Hirtella*, and *Licania* sensu stricto (clade NL3 of Fig. 2) both have over 100 species, all of which have appeared since the Miocene in South America. In addition, *Couepia* (58 spp.) also dates back to the Oligocene-Miocene period, according to our analysis (Fig. 2). Together, these results reinforce the idea that the diversification of the Neotropical flora has largely occurred within the past 23 mya.

Possible causes for the higher diversification rate for Chrysobalanaceae in the Neotropics involve the complex geological history of this continent. For instance, Clade D is about the same age as clade N, and it is essentially Neotropical, but contains only 28 described species, yet it did not diversify as rapidly (Sothers et al., 2014). Biological attributes, such as generation time, specialization to abiotic conditions, or pollination mechanisms, also influence diversification rates (Barraclough et al., 1998; Baker et al., 2014). *Hirtella*, which represents about a fourth of the Neotropical species in the family, is predominantly constituted of small tree species that are found in the understory of Neotropical forests, and this ecology could relate to a different mode of diversification (Smith, 2001; Baker et al., 2014). More generally, Neotropical Chrysobalanaceae are found in contrasted habitats including the arid caatinga, poor-soil South American campinas, and Southeast Asian kerangas, or the margin of mangrove swamps (Prance and White, 1988). Diversification in clade N may have thus been driven by niche specialization for fire resistance, drought tolerance, or other biological attributes associated with the fluctuating climatic conditions since the Miocene (Simon et al., 2009; Hernandez-Hernandez et al., 2014; Fine et al., 2014).

CONCLUSIONS

We confirmed the Paleotropical origin of Chrysobalanaceae, suggested *Kostermanthus* as the early diverging lineage, and found evidence for two distinct clades in *Hirtella*, and at least four in *Licania*, as currently circumscribed. Our results demonstrate that Neotropical diversity mostly originated from a single dispersal event. We also found evidence for the propensity of transoceanic dispersal for this family. Given that Neotropical Chrysobalanaceae began to diversify during the Eocene, their extensive diversity is not the result of rapid radiations such as those described for Andean taxa (which have mostly taken place from the Pliocene). Nor is it the result of constant diversification rates coupled with low extinction rates in response to a stable tropical ecosystem. In contrast, the diversification pattern in Chrysobalanaceae suggests that Amazonian rainforest diversity is best explained by the joint influence of large in situ speciation and extinction rates.

ACKNOWLEDGEMENTS

The authors thank J. Chamagne, G. Dauby, B. David, K. G. Dexter, J. Engel, O. Hardy, V. Hequet, N. J. Matzke, O. Maurin, C. Maycock, I. Parmentier, J. Pereira, M. Réjou-Méchain, D. Sabatier, D. Strasberg, P. Taberlet, and A. Vicentini, for advice and/or for sharing plant material, and L. Chatrou, P.V.A. Fine, and A. Willyard for their insightful comments at the review stage. We also thank the following institutions for facilitating material sharing: Institut de Recherche Pierre Fabre, Herbarium of the Forest Reserve Centre at Sepilok, University of Johannesburg, the ARF project of the Central African Republic and the Mbaïki experimental plot, Association Solidarité France-Sahel, and Centre IRD de Nouméa. It is a pleasure to

acknowledge funding by the ANR-funded project METABAR, "Investissement d'Avenir" grants managed by Agence Nationale de la Recherche (CEBA: ref. ANR-10-LABX-25-01, TULIP: ANR-10-LABX-0041) and by a Fondation pour la Recherche sur la Biodiversité grant (Atrophy). This work is dedicated to the memory of our colleague Jean-Christophe Pintaud.

LITERATURE CITED

Ali, J. R., and J. C. Aitchison. 2008. Gondwana to Asia: plate tectonics, paleogeography and the biological connectivity of the Indian sub-continent from the Middle Jurassic through latest Eocene (166–35 Ma). *Earth-Science Reviews* 88: 145–166.

Antonelli, A., J. A. A. Nylander, C. Persson, and I. Sanmartín. 2009. Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proceedings of the National Academy of Sciences, USA* 106: 9749–9754.

Bacon, C. D., D. Silvestro, C. Jaramillo, B. T. Smith, P. Chakrabarty, and A. Antonelli. 2015. Biological evidence supports an early and complex emergence of the Isthmus of Panama. *Proceedings of the National Academy of Sciences, USA* 112: 6110–6115.

Baillon, H. E. 1868. Sur le genre *Thelyra* de Dupetit Thouars. *Adansonia* 8: 159–161.

Baker, T. R., R. T. Pennington, S. Magallon, E. Gloor, W. F. Laurance, M. Alexiades, E. Alvarez, et al. 2014. Fast demographic traits promote high diversification rates of Amazonian trees. *Ecology Letters* 17: 527–536.

Bardon, L., J. Chamagne, K. G. Dexter, C. A. Sothers, G. T. Prance, and J. Chave. 2013. Origin and evolution of Chrysobalanaceae: insights into the evolution of plants in the Neotropics. *Botanical Journal of the Linnean Society* 171: 19–37.

Barraclough, T. G., A. P. Vogler, and P. H. Harvey. 1998. Revealing the factors that promote speciation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 353: 241–249.

Bartish, I. V., A. Antonelli, J. E. Richardson, and U. Swenson. 2011. Vicariance or long-distance dispersal: historical biogeography of the pantropical subfamily Chrysophylloideae (Sapotaceae). *Journal of Biogeography* 38: 177–190.

Bayly, M. J., P. Rigault, A. Spokevicius, P. Y. Ladiges, P. K. Ades, C. Anderson, et al. 2013. Chloroplast genome analysis of Australian eucalypts–*Eucalyptus, Corymbia, Angophora, Allosyncarpia* and *Stockwellia* (Myrtaceae). *Molecular Phylogenetics and Evolution* 69: 704–716.

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, et al.. 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10: e1003537.

Bush, M. B. 1994. Amazonian speciation: a necessarily complex model. *Journal of Biogeography* 21: 5–17.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17: 540–552.

Chambers, K. L., and G. O. Poinar Jr. 2010. The Dominican amber fossil *Lasiambix* (Fabaceae, Caesalpinioideae?) is a *Licania* (Chrysobalanaceae). *Journal of the Botanical Research Institute of Texas* 4: 217–218.

Cody, S., J. E. Richardson, V. Rull, C. Ellis, and R. T. Pennington. 2010. The Great American Biotic Interchange revisited. *Ecography* 33: 326–332.

Couvreur, T. L. P., M. D. Pirie, L. W. Chatrou, R. M. K. Saunders, Y. C. F. Su, J. E. Richardson, and R. H. J. Erkens. 2011. Early evolutionary history of the flowering plant family Annonaceae: steady diversification and boreotropical geodispersal. *Journal of Biogeography* 38: 664–680.

Cracraft, J. 1973. Continental drift, paleoclimatology, and the evolution and biogeography of birds. *Journal of Zoology* 169: 455–543.

Cracraft, J. 1985. Historical biogeography and patterns of differentiation within the South American avifauna: areas of endemism. *Ornithological Monographs* 36: 49–84.

Crepet, W., and K. Nixon. 1998. Fossil Clusiaceae from the Late Cretaceous (Turonian) of New Jersey and implications regarding the history of bee pollination. *American Journal of Botany* 85: 1122–1133.

Davis, C. C., C. D. Bell, S. Mathews, and M. J. Donoghue. 2002. Laurasian migration explains Gondwanan disjunctions: evidence from Malpighiaceae. *Proceedings of the National Academy of Sciences, USA* 99: 6833–6837.

Dissanayake, D. 1999. Phylogenetic research on the family Chrysobalanaceae. *PhD thesis, Department of Botany, University of Reading.*

Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 214.

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

Du Petit-Thouars, L.-M. A. A. 1806. *Genera Nova Madagascariensia*. Paris.

Erkens, R. H. J., L. W. Chatrou, J. W. Maas, T. van der Niet, and V. Savolainen. 2007. A rapid diversification of rain forest trees (*Guatteria*, Annonaceae) following dispersal from Central into South America. *Molecular Phylogenetics and Evolution* 44: 399–411.

Erkens, R. H., J. Maas, and T. L. P. Couvreur. 2009. From Africa via Europe to South America: migrational route of a species-rich genus of Neotropical lowland rain forest trees (*Guatteria*, Annonaceae). *Journal of Biogeography* 36: 2338–2352.

Fine, P. V. A., F. Zapata, and D. C. Daly. 2014. Investigating processes of Neotropical rain forest tree diversification by examining the evolution and historical biogeography of the Protieae (Burseraceae). *Evolution* 68: 1988–2004.

Fritsch, P. W., S. R. Manchester, R. D. Stone, B. C. Cruz, and F. Almeda. 2015. Northern Hemisphere origins of the amphi-Pacific tropical plant family Symplocaceae. *Journal of Biogeography* 42: 891–901.

Gentry, A. H. 1982. Neotropical floristic diversity: phytogeographical connections between Central and South America, Pleistocene climatic fluctuations, or an accident of the Andean orogeny? *Annals of the Missouri Botanical Garden* 69: 557–593.

Germeraad, J. H., C. A. Hopping, and J. Muller. 1968. Palynology of Tertiary sediments from tropical areas. *Review of Palaeobotany and Palynology* 6: 189–348.

Gregory-Wodzicki, K. M. 2000. Uplift history of the Central and Northern Andes: a review. *Geological Society of America Bulletin* 112: 1091–1105.

Heath, T. A., S. M. Hedtke, and D. M. Hillis. 2008. Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution* 46: 239–257.

Hernández-Hernández, T., J. W. Brown, B. O. Schlumpberger, L. E. Eguiarte, and S. Magallon. 2014. Beyond aridification: multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *The New Phytologist* 202: 1382–1397.

Hoorn, C., F. P. Wesselingh, H. ter Steege, M. A. Bermudez, A. Mora, J. Sevink, I. Sanmartín, et al. 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* 330: 927–931.

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

Insel, N., C. J. Poulsen, and T. A. Ehlers. 2009. Influence of the Andes Mountains on South American moisture transport, convection, and precipitation. *Climate Dynamics* 35: 1477–1492.

Jud, N. A., C. W. Nelson, and F. Herrera. 2016. Fruits and wood of *Parinari* from the early Miocene of Panama and the fossil record of Chrysobalanaceae. *American Journal of Botany* 103: 277–289.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.

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

Koenen, E. J. M., J. J. Clarkson, T. D. Pennington, and L. W. Chatrou. 2015. Recently evolved diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on the origins of rainforest hyperdiversity. *The New Phytologist* 207: 327–339.

Landis, M. J., N. J. Matzke, B. R. Moore, and J. P. Huelsenbeck. 2013. Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology* 62: 789–804.

Lavin, M., and M. Luckow. 1993. Origins and relationships of tropical North America in the context of the boreotropics hypothesis. *American Journal of Botany* 80: 1–14.

Litt, A., and M. W. Chase. 1999. The systematic position of *Euphronia*, with comments on the position of *Balanops*: an analysis based on *rbc*L sequence data. *Systematic Botany* 23: 401–409.

Lohmann, L. G., C. D. Bell, M. F. Calió, and R. C. Winkworth. 2013. Pattern and timing of biogeographical history in the Neotropical tribe Bignonieae (Bignoniaceae). *Botanical Journal of the Linnean Society* 171: 154–170.

López-Giráldez, F., and J. P. Townsend. 2011. PhyDesign: an online application for profiling phylogenetic informativeness. *BMC Evolutionary Biology* 11: 152.

Luebert, F., and M. Weigend. 2014. Phylogenetic insights into Andean plant diversification. *Frontiers in Ecology and Evolution* 2: 27.

Maas, P. J. M., L. Y. T. Westra, S. A. Guerrero, A. Q. Lobão, U. Scharf, N. A. Zamora, and R. H. J. Erkens. 2015. Confronting a morphological nightmare: revision of the Neotropical genus *Guatteria* (Annonaceae). *Blumea* 60: 1–219.

Magallón, S., S. Gómez-Acevedo, L. L. Sánchez-Reyes, and T. Hernández-Hernández. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *The New Phytologist* 207: 437–453.

Malé, P.-J. G., L. Bardon, G. Besnard, E. Coissac, F. Delsuc, J. Engel, E. Lhuillier, et al. 2014. Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. *Molecular Ecology Resources* 14: 966–975.

Manns, U., N. Wikström, C. M. Taylor, and B. Bremer. 2012. Historical biogeography of the predominantly Neotropical subfamily Cinchonoideae (Rubiaceae): Into or out of America? *International Journal of Plant Sciences* 173: 261–286.

Matzke, N. J. 2013a. BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. *PhD Thesis, University of California, Berkeley*.

Matzke, N. J. 2013b. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography* 5: 242–248.

Matzke, N. J. 2014. Model Selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology* 63: 951–970.

Moore, M. J., P. S. Soltis, C. D. Bell, J. G. Burleigh, and D. E. Soltis. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences, USA* 107: 4623–4628.

Morley, R. J. 2003. Interplate dispersal paths for megathermal angiosperms. *Perspectives in Plant Ecology, Evolution and Systematics* 6: 5–20.

Nürk, N. M., C. Scheriau, and S. Madriñán. 2013. Explosive radiation in high Andean *Hypericum*—rates of diversification among New World lineages. *Frontiers in Genetics* 4: 17.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics (Oxford, England)* 20: 289–290.

Parks, M., R. Cronn, and A. Liston. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology* 7: 84.

Pennington, R. T., and C. W. Dick. 2004. The role of immigrants in the assembly of the South American rainforest tree flora. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 359: 1611–1622.

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.

Poinar, G. O. Jr., K. L. Chambers, and A. E. Brown. 2008. *Lasiambix dominicensis* gen. and sp. nov., a eudicot flower in Dominican amber showing affinities with Fabaceae subfamily Caesalpinioideae. *Journal of the Botanical Research Institute of Texas* 2: 463–471.

Pollock, D. D., D. J. Zwickl, J. A. McGuire, and D. M. Hillis. 2002. Increased taxon sampling is advantageous for phylogenetic inference. *Systematic Biology* 51: 664–671.

Prance, G. T. 1974. Phytogeographic support for the theory of Pleistocene forest refuges in the Amazon basin, based on evidence from distribution patterns in Caryocaraceae, Chrysobalanaceae, Dichapetalaceae and Lecythidaceae. *Acta Amazonica* 3: 5–28.

Prance, G. T. 1979. New genera and species of Chrysobalanaceae from Malesia and Oceania. *Brittonia* 31: 79–95.

Prance, G. T., and C. A. Sothers. 2003. *Chrysobalanaceae 1&2, Species Plantarum: Flora of the World, parts 9 &10*. Canberra: Australian Biological Resources Study.

Prance, G. T., and F. White. 1988. The genera of Chrysobalanaceae: a study in practical and theoretical taxonomy and its relevance to evolutionary biology. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 320: 1–184.

Rabosky, D. L. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS One* 9: e89543.

Rabosky, D. L., M. Grundler, C. Anderson, P. Title, J. J. Shi, J. W. Brown, H. Huang, and J. G. Larson. 2014. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution* 5: 701–707.

Raven, P. H., and D. I. Axelrod. 1974. Angiosperm biogeography and past continental movements. *Annals of the Missouri Botanical Garden* 61: 539–673.

Ree, R. H., and S. A. Smith. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* 57: 4–14.

Renner, S. S. 2004. Plant dispersal across the tropical Atlantic by wind and sea currents. *International Journal of Plant Sciences* 165: S23–S33.

Roncal, J., F. Kahn, B. Millan, T. L. P. Couvreur, and J.-C. Pintaud. 2013. Cenozoic colonization and diversification patterns of tropical American palms: evidence from *Astrocaryum* (Arecaceae). *Botanical Journal of the Linnean Society* 171: 120–139.

Ronquist, F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology* 46: 195–203.

Ruhfel, B. R., M. A. Gitzendanner, P. S. Soltis, D. E. Soltis, and J. G. Burleigh. 2014. From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evolutionary Biology* 14: 23.

Ruhfel, B. R., P. F. Stevens, and C. C. Davis. 2013. Combined morphological and molecular phylogeny of the clusioid clade (Malpighiales) and the placement of the ancient rosid macrofossil Paleoclusia. *International Journal of Plant Sciences* 174: 910–936.

Sepulchre, P., T. Arsouze, Y. Donnadieu, J. C. Dutay, C. Jaramillo, J. Le Bras, E. Martin, et al. 2014. Consequences of shoaling of the Central American Seaway determined from modeling Nd isotopes. *Paleoceanography* 29: 176–189.

Silvestro, D., J. Schnitzler, and G. Zizka. 2011. A Bayesian framework to estimate diversification rates and their variation through time and space. *BMC Evolutionary Biology* 11: 311.

Simon, M. F., R. Grether, L. P. de Queiroz, C. Skema, R. T. Pennington, and C. E. Hughes. 2009. Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proceedings of the National Academy of Sciences, USA* 106: 20359–20364.

Simpson, G. G. 1980. Splendid isolation: the curious history of South American mammals. Yale University Press, New Haven.

Simpson, J. B. 1961. The Tertiary pollen-flora of Mull and Ardnamurchan. *Transactions of the Royal Society of Edinburgh* 64: 421–468.

Smith, J. F. 2001. High species diversity in fleshy-fruited tropical understory plants. *American Naturalist* 157: 646–653.

Sothers, C., G. T. Prance, S. Buerki, R. de Kok, and M. W. Chase. 2014. Taxonomic novelties in Neotropical Chrysobalanaceae: towards a monophyletic *Couepia*. *Phytotaxa* 172: 176–200.

Sothers, C. A., and G. T. Prance. 2014. Ressurrection of *Angelesia*, a Southeast Asian genus of Chrysobalanaceae. *Blumea* 59: 103–105.

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

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

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Systematic Biology* 57: 758–771.

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

Stull, G. W., R. D. de Stefano, D. E. Soltis, and P. S. Soltis. 2015. Resolving basal lamiid phylogeny and the circumscription of Icacinaceae with a plastome-scale data set. *American Journal of Botany* 102: 1794–1813.

Talavera, G., and J. Castresana. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* 56: 564–577.

Tiffney, B. H. 1985. The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. *Journal of the Arnold Arboretum* 66: 243–273.

Tiffney, B. H., J. G. Fleagle, and T. M. Brown. 1994. Early to Middle Miocene angiosperm fruits and seeds from Fejej, Ethiopia. *Tertiary Research* 15: 25–42.

Weeks, A., F. Zapata, S. K. Pell, D. C. Daly, J. D. Mitchell, and P. V. A. Fine. 2014. To move or to evolve: Contrasting patterns of intercontinental connectivity and climatic niche evolution in "Terebinthaceae" (Anacardiaceae and Burseraceae). *Frontiers in Genetics* 5: 1–20.

Wesselingh, F. P., M. E. Räsänen, G. Irion, H. B. Vonhof, R. Kaandorp, W. Renema, L. R. Pittman, and M. Gingras. 2002. Lake Pebas: a palaeoecological reconstruction of a Miocene, long-lived lake complex in western Amazonia. *Cainozoic Research* 1: 35–81.

Wiens, J. J. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. *Systematic Biology* 52: 528–538.

Wiens, J. J., and D. S. Moen. 2008. Missing data and the accuracy of Bayesian phylogenetics. *Journal of Systematics and Evolution* 46: 307–314.

Williams, A. V., J. T. Miller, I. Small, P. G. Nevill, and L. M. Boykin. 2016. Integration of complete chloroplast genome sequences with small amplicon datasets improves phylogenetic resolution in Acacia. *Molecular Phylogenetics and Evolution* 96: 1–8.

Willis, C. G., B. F. Franzone, Z. Xi, and C. C. Davis. 2014. The establishment of Central American migratory corridors and the biogeographic origins of seasonally dry tropical forests in Mexico. *Frontiers in Genetics* 5: 433.

Wodehouse, R. P. 1932. Tertiary pollen. I. Pollen of the living representatives of the Green River Flora. *Bulletin of the Torrey Botanical Club* 59: 313–340.

Wyman, S. K., R. K. Jansen, and J. L. Boore. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics (Oxford, England)* 20: 3252–3255.

Xi, Z., B. R. Ruhfel, H. Schaefer, A. M. Amorim, M. Sugumaran, K. J. Wurdack, P. K. Endress, et al. 2012. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proceedings of the National Academy of Sciences, USA* 109: 17519–17524.

Yakandawala, D., C. M. Morton, and G. T. Prance. 2010. Phylogenetic relationships of the Chrysobalanaceae inferred from chloroplast, nuclear, and morphological data. *Annals of the Missouri Botanical Garden* 97: 259–281.

Zerbino, D. R., and E. Birney. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* 18: 821–829.