The LINNEAN SOCIETY of London BOTANICAL Journal of the Linnean Society

Botanical Journal of the Linnean Society, 2013, **171**, 19–37. With 3 figures

Origin and evolution of Chrysobalanaceae: insights into the evolution of plants in the Neotropics

LÉA BARDON[1], JULIETTE CHAMAGNE[1,2], KYLE G. DEXTER[1,3], CYNTHIA A. SOTHERS[4], GHILLEAN T. PRANCE[4] and JÉRÔME CHAVE[1]*

[1]*Laboratoire Evolution et Diversité Biologique, UMR 5174, CNRS, Université Paul Sabatier, 31062 Toulouse, France*
[2]*Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland*
[3]*Royal Botanic Garden Edinburgh, 20a Inverleith Row, Edinburgh EH3 5LR, UK*
[4]*Herbarium, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB, UK*

Received 15 December 2011; revised 1 June 2012; accepted for publication 15 June 2012

Some plant families show a striking imbalance in species diversity between the Neotropics and the Palaeotropics. The woody plant family Chrysobalanaceae is a typical example of this pattern, with 80% of the 531 species in the Neotropics. In order to test alternative interpretations for this pattern, we generated a dated phylogenetic hypothesis for Chrysobalanaceae, using DNA sequence data from one nuclear and six plastid markers. Using a maximum likelihood approach, we jointly inferred ancestral areas and diversification rates in the Neotropics and Palaeotropics. We found that Chrysobalanaceae most probably originated in the Palaeotropics about 80 Mya. The family dispersed into the Neotropics at least four times beginning 40–60 Mya, with at least one back-dispersal to the Palaeotropics. Members of Chrysobalanaceae have experienced higher extinction, speciation and net diversification rates in the Neotropics. Hence, the high species diversity of Chrysobalanaceae in the Neotropics appears to be primarily caused by a higher speciation rate in this region. Several recent studies have shown high diversification rates in Neotropical plant families, but have focused on Andean-centred taxa. Ours is the first study to find a similar pattern in a family for which the centre of diversity is in eastern and central Amazonia. *Botanical Journal of the Linnean Society*, 2013, **171**, 19–37.

ADDITIONAL KEYWORDS: ancestral state reconstruction – diversification rates – fossil calibration – Neotropical plant biodiversity – phylogeny.

INTRODUCTION

Tropical forests house the majority of the 60 000–80 000 tree species in the world and, among tropical forests, those of the Neotropics stand out as the most diverse, with > 20 000 tree species (Fine & Ree, 2006). Although the diversification and dispersal of plant lineages among tropical regions have probably shaped this extant pattern, evolutionary hypotheses to explain the exceptional Neotropical plant diversity remain poorly developed and tested (Raven & Axelrod, 1974; Stebbins, 1974; Gentry, 1982). If more lineages have been present longer in the Neotropics, then higher Neotropical plant diversity could simply reflect greater time for diversification. An alternative, but not mutually exclusive, hypothesis is that diversification rates are higher in the Neotropics, which could be a result of lower extinction rates and/or higher speciation rates, as originally proposed by Gentry (1982), who speculated that the uplift of the Andes was the foremost cause of higher speciation rates in the Neotropics, although late Cretaceous plant migration between North and South America could also have played a role in the assembly of the Neotropical flora. In order to test these hypotheses, studies are needed that track the origins and intercontinental dispersals of plant lineages and compare diversification, speciation and extinction rates among

*Corresponding author. E-mail: jerome.chave@univ-tlse3.fr

tropical regions. Here, we use a molecular phylogenetic approach to address the origin, dispersal and diversification history of the pantropical plant family Chrysobalanaceae. This family is emblematic of the tropical diversity imbalance, with 80% of the 531 species being found in the Neotropics.

The number of phylogenetic hypotheses available for plant groups has grown considerably in recent years and, with fossil evidence, they have begun to shed light on the history of tropical plant lineages. For instance, numerous studies have shown that intercontinental dispersal was rampant following the breakup of Gondwana and must be accounted for when comparing diversity among tropical regions (Tiffney, 1985; Manchester, 1999; Morley, 2000, 2003; Sanmartin, Enghoff & Ronquist, 2001; Pennington & Dick, 2004). It is now known that Melastomataceae diversified in Eurasia before spreading to North and South America during the early Eocene (Renner, Clausing & Meyer, 2001). Likewise, it has been suggested that Rubiaceae (Antonelli *et al.,* 2009), *Guatteria* Ruiz & Pav. (Annonaceae) (Erkens, Maas & Couvreur, 2009) and *Begonia* L. (Begoniaceae) (Goodall-Copestake *et al.,* 2010) probably originated in the Palaeotropics and used the boreotropical route to reach South America.

Many of these pantropical plant lineages show a consistent pattern of being more diverse in the Neotropics, particularly in South America. If a lineage is more diverse in the Neotropics, but originated elsewhere, then diversification rates must have been higher in the Neotropics to explain higher Neotropical diversity. This higher net diversification could be a result of higher speciation rates (a cradle hypothesis) or lower extinction rates (a museum hypothesis). In support of the Neotropical 'cradle' hypothesis, the complex geological history of the Neotropics may have yielded more frequent opportunities for allopatric speciation. The uplift of the northern Andes since the early Neogene (*c*. 23 million years ago (Mya)) yielded a variety of new habitats, changed river flows and altered climatic conditions (Burnham & Graham, 1999; Linder, 2008; Antonelli *et al*., 2009; Hoorn *et al*., 2010). Since *c*. 3 Mya, the closure of the Isthmus of Panama and the emergence of glacial/interglacial climatic oscillations have also played a role in driving higher speciation rates and thus shaping Neotropical biodiversity patterns (Gentry, 1982; Richardson *et al.,* 2001; Bennett, 2004). In favour of the Neotropical 'museum' hypothesis, we note that the Guianan and Brazilian cratons, geological formations dating back to around 2 giga years ago (Gya), have been above sea level and were in tropical latitudes over much of the Tertiary. Thus, South America has almost always offered ample space for tropical plant species to survive and diversify, at least over the past 65 million years (Myr). If a larger block of rainforest was more continuously present through time, this may explain the high diversity of the Neotropics in comparison with other tropical regions (Fine & Ree, 2006).

In order to contribute to a broader understanding of why the Neotropics have exceptional plant diversity, we studied the biogeography and modes of diversification in the pantropical plant family Chrysobalanaceae. The pantropical distribution of Chrysobalanaceae, its abundance and diversity in Neotropical forests, and its well-understood taxonomy (Prance, 1972; Prance & White, 1988; Prance & Sothers, 2003) make it an excellent family to study the origins of high Neotropical plant diversity. It is currently unknown whether the high Neotropical diversity of the family is a result of a longer residence time in the Neotropics, greater speciation rates and/or lower extinction rates. Here, we construct a dated phylogenetic hypothesis for the family, including 17 of the 18 genera, and 74 species (14% of the total diversity), based on DNA sequencing of six plastid markers and the nuclear marker *ITS*, and several fossil calibration points. We determine the biogeographical origin of the family and subsequent intercontinental dispersal events. Further, whilst reconstructing ancestral geographical states (Neotropical or Palaeotropical), we jointly estimated the rates of speciation, extinction and migration based on a maximum likelihood technique. The results are discussed in the light of recent evidence for patterns of Neotropical plant diversification.

MATERIAL AND METHODS

Family description and species sampling

Chrysobalanaceae (*sensu* Matthews & Endress, 2008; Yakandawala, Morton & Prance, 2010) is a tropical woody plant family in Malpighiales (Davis *et al*., 2005; APG III, 2009). The sister family is the monotypic Euphroniaceae. Members of Chrysobalanaceae are trees and shrub species often ecologically dominant in the New World tropics where they occupy both moist and dry forest biomes (Prance, 1972). The most recent taxonomic treatment of the family includes 531 species and 18 genera (Prance & White, 1988; Prance & Sothers, 2003). Among these, 423 species (80%) and eight genera are found in South America. In floristic studies across the Amazon, members of Chrysobalanaceae comprise up to 10% of the trees in Amazonian forest plots and reach their highest diversity in the central and eastern Amazon (Prance & White, 1988; Hopkins, 2007). However, there is a greater generic diversification in the Palaeotropics, with 11 genera in Africa and seven in Asian tropics. A genus-level phylogenetic analysis has

been conducted for the family based on the sequencing of two DNA regions (*rbcL* and ITS), but it lacked sufficient resolution and sampling to assess the biogeographical origins of the family or to calculate diversification rates (Yakandawala *et al*., 2010).

Tissue samples were collected throughout the tropics: South America (Guyana, French Guiana, Venezuela, Brazil), Africa (Gabon, Cameroon, Benin, Central African Republic, Guinea, Senegal, Democratic Republic of Congo), South-East Asia (Indonesia, Malaysia, New Caledonia), Madagascar and La Réunion. The great majority of the collections from South America were collected for the purpose of this project in the Guianas. Collections from Africa were contributed by tissue exchange agreements with Université Libre de Bruxelles, the Royal Botanic Gardens, Kew and Université de la Réunion, and our own collections in the Central African Republic. Collections in South-East Asia were mostly obtained by exchange agreements with the Royal Botanic Gardens, Kew, and with the Botanic Garden of Sandakan, Malaysia. Appendix 1 reports the full list of samples and their collection locations. In addition, several sequences were retrieved from GenBank that were used in an earlier phylogenetic study of the family by Yakandawala *et al*. (2010).

DNA extraction, amplification and sequencing

New DNA for this study was extracted with a Biosprint 15 (Qiagen) auto-extractor, following the protocol provided by the supplier (see Appendix 2). Some of the DNA extracts used in the analysis were obtained through the DNA bank at the Royal Botanic Gardens, Kew. The extraction protocol differed in including a standard cetyltrimethylammonium bromide (CTAB)–chloroform extraction followed by ethanol precipitation and washing, and then by density gradient cleaning and dialysis.

We initially attempted to amplify 30 markers for which primers have already been published in the literature: 15 plastid intergenic spacers or introns (Hamilton, 1998; Shaw *et al*., 2007), two plastid genes (Savolainen *et al*., 2000; Dunning & Savolainen, 2010), 13 low-copy nuclear genes (Strand, Leebens-Mack & Milligan, 1997; Li *et al*., 2008) and one mitochondrial gene (Davis & Wurdack, 2004). We failed to amplify any of the low-copy nuclear genes or the mitochondrial gene. Six plastid markers were finally selected for this analysis: *psbD-trnT*, *psbA-trnH, atpI-atpH*, the *ndhA* intron, *matK* and *rbcL*a. In addition to these markers, the nuclear ITS region was amplified (see Appendix 3 for details on primers).

DNA amplification was performed using the polymerase chain reaction (PCR) protocols reported in Appendix 2. PCR products were cleaned and sequenced on an ABI3730XL Automated DNA sequencer (Genoscreen, Lille, France). The two complementary DNA strands were manually corrected and assembled using Sequencher version 4.8 (Gene Codes Corporation, 2003). The total DNA matrix was aligned for each marker with the MUSCLE alignment program (Edgar, 2004). The resulting sequence matrix was further manually edited in MEGA 4 (Tamura *et al*., 2007). The newly generated sequences are available on GenBank and EMBL (accessions JQ898692 to JQ899029, see Appendix 1). (See also Table 1 for a summary of these statistics.)

Table 1. Summary data on the sequence matrix used in phylogenetic tree reconstruction of Chrysobalanaceae. 'Length' is the amplicon length retained for the analysis in the consensus matrix, N_g is the number of sequences newly generated for the analysis, N_t is the total number of sequences, including those retrieved from public domain sequence databases, and P is the percentage of available sequences in the complete dataset

Amplicon	Length	N_g	N_t	P (%)
ITS	720	57	71	90
matK	854	55	56	73
rbcLa	650	60	72	90
psbA-trnH	415	45	45	58
ndhA	1304	47	47	61
atpI-atpH	1258	45	45	56
psbD-trnT	1696	29	29	38
Total	6897	338	362	

Phylogenetic reconstruction

Sequences were assembled into three datasets. The first dataset contained sequences obtained for the six plastid regions: *atpI-atpH*, *psbD-trnT*, *psbA-trnH*, *ndhA* intron, *matK* and *rbcLa*. The second dataset included the ribosomal intergenic spacer ITS. The aligned plastid and nuclear datasets included 6177 bp and 720 bp, respectively (Table 1). The third dataset included all sequences concatenated for the seven markers (i.e. datasets 1 plus 2). For each marker, the number of available sequences is reported in Table 1.

Maximum likelihood analyses were conducted using RaxML software version 7.0.4 (Stamatakis, Hoover & Rougemont, 2008) run on the complete dataset including the seven markers through the CIPRES supercomputer cluster (http://www.phylo.org). Our alignment was separated into four partitions: ITS, *matK*, *rbcLa* and noncoding plastid DNA regions. Branch support was assessed using a rapid bootstrap procedure (Stamatakis *et al*., 2008). Branch length in the phylogram denoted overall branch divergence. We also performed a tree reconstruction using

RaxML based on the ITS dataset only (nine of 77 ITS sequences were missing). We then compared the ITS-based tree with that based on the complete dataset to test whether our missing plastid sequences may result in tree topology issues, such as the presence of 'wildcard' taxa (Platnick, Griswold & Coddington, 1991). We finally performed a tree reconstruction based on the plastid DNA dataset only and compared the resulting tree with that based on the ITS dataset to explore potential inconsistencies among topologies. The last two trees gave similar results, and so the tree based on the plastid DNA dataset only is not shown here.

In a second step, we constructed a dated phylogenetic tree using BEAST software (Drummond & Rambaut, 2007). First, we constructed an initial phylogenetic tree obtained using the PhyML version 3.0.1 maximum likelihood phylogeny reconstruction program (Guindon *et al*., 2010), as implemented in seaview4 software (Gouy, Guindon & Gascuel, 2010). The consensus phylogram was then roughly dated using PATHd8 version 1.0 (Britton *et al*., 2007) based on four calibration points (three fossils and one temporal constraint on the root; see below). This tree then served as input for a combined analysis of divergence times and phylogenetic topology using the software BEAST version 1.6.1 (Drummond & Rambaut, 2007). The advantage of this preliminary procedure is to provide BEAST with a phylogenetic hypothesis not too far from the most likely region for the parameters, hence avoiding floating-point computing errors. Substitution and clock models were unlinked, and each marker evolved independently. For all markers, we chose the most parameterized molecular evolution model, the general time-reversible (GTR) model. This assumes site heterogeneity modelled by a Gamma distribution, and takes into account proportions of invariant sites. Divergence times were estimated under a log-normal uncorrelated relaxed clock method for each partition and using the Yule model of speciation. Several preliminary BEAST runs were performed using one Markov Chain Monte-Carlo (MCMC) for 1 000 000 generations to adjust the operators for optimal mixing of the MCMCs. Then, ten independent runs of 20 000 000 generations were conducted with sampling every 2000th generation for the combined dataset.

The burn-in part of the MCMC was discarded (10% of the total number of generations). Post burn-in trees were merged using LogCombiner version 1.6.1 (Drummond & Rambaut, 2007) and performance was evaluated using Tracer version 1.5 (Drummond & Rambaut, 2007). Mean evolutionary rates and divergence times were calculated using TreeAnnotator version 1.6.1 after the removal of 25% burn-in, keeping target heights concerning the node heights option. The effective sampling sizes (ESSs) of each parameter were checked at the end of each simulation, and were considered to be of good quality when > 200.

Age constraints

Phylogenetic analyses and clock-independent dating estimates of Davis *et al*. (2005) provided a minimum age for the stem node of Chrysobalanaceae *s.s.* of ~60 Myr (corresponding to the age found for the crown age of Chrysobalanaceae *s.s*.) and a maximum age of ~90 Myr (corresponding to the age found for the split between Chrysobalanaceae *s.s./Euphronia* Mart. and *Dichapetalum* Thouars/*Trigonia* Aubl.). The mean age found for the split between Chrysobalanaceae *s.s.* and *Euphronia guianensis* (R.H. Schomb.) Hallier f. was ~78 Myr. Here, we used this date for the Chrysobalanaceae–Euphroniaceae split which was constrained by a normal probability distribution.

In addition, two South American macrofossils and one North American palynofossil were used as minimum age constraints for three internal clades. Several endocarp fossils have been reported for the genus *Parinari* Aubl., notably from the publication of Tiffney, Fleagle & Brown (1994), who studied an Ethiopian plant assemblage dated at 16.1 Mya, and Wijninga (1996) from Colombia. Here, we used the slightly older date of the fossil endocarp also attributed to *Parinari* from the Cucaracha Formation, Panama (F. Herrera and S. Manchester, Department of Biology, Florida Museum of Natural History, University of Florida, Gainesville, FL, USA, pers. comm.), which was dated at 19.0–17.5 Mya. Second, a fossil of *Hirtella* L. from the Gandarela Basin, Minas Gerais, Brazil, published by Duarte & Mello Filha (1980), is based on leaf material only. The dating of this rock formation is also less clear, but is thought to be of the Eocene period (56–34 Mya). Third, we used the pollen type attributed to *Chrysobalanus* L. by Wodehouse (1932) found in the shales of the Green River Formation, Colorado, USA, and used by Davis *et al*. (2005). This pollen attributed to *Chrysobalanus* was dated to the early to middle Eocene (49–34 Mya), and we used the most recent age estimate (34 Myr) to constrain the tree dating.

For fossil calibrations based on a minimum age for a split, Ho (2007) suggested a log-normal prior distribution with the probability of the nodal age decreasing with time. Log-normal distributions with an offset were chosen to calibrate splits between the genera *Chrysobalanus* (BEAST parameters: mean, 2.6; standard deviation, 0.5; offset, 34), *Hirtella* (mean, 2.6; standard deviation, 0.5; offset, 34) and *Parinari* (mean, 2.9; standard deviation, 0.5; offset, 17.5) and their sister group, respectively, in BEAST analyses. The lower bounds of the age estimates

(in Myr) were chosen as offsets for the log-normal distributions.

Ancestral state reconstruction and inference of diversification rates

Here, we sought to assign extant species into Neotropical and Palaeotropical species (Appendix 1). We coded the Neotropical/Palaeotropical distinction as a binary character, and inferred the state of the internal nodes of the consensus phylogenetic tree obtained by the Bayesian reconstruction. For two species, *Parinari excelsa* Sabine and *Chrysobalanus icaco* L., sequences from both South America and Africa were available.

To carry out this analysis, we used the binary-state speciation and extinction (BiSSE) method (Maddison, Midford & Otto, 2007) in the package 'diversitree' (FitzJohn, Maddison & Otto, 2009) of the R statistical software 2.13.0 (R Development Core Team, 2009). BiSSE computes the probability of a phylogenetic tree and of the observed distribution of a binary character state (here Neotropical or Palaeotropical) among the tree tips, given a model of character evolution, speciation and extinction (Maddison *et al*., 2007). Precisely, this method jointly estimates six parameters: speciation and extinction rates $\lambda_1$and μ_1 in the Neotropical zone and λ_0 and μ_0 in the Palaeotropical zone, and two shift rates from the Neotropical to the Palaeotropical zone (q_{10}), and from the Palaeotropical to the Neotropical zone (q_{01}). BiSSE maximizes the likelihood of obtaining the observed tree, given particular values of speciation, extinction and transition rates. It is possible to fix parameters to be equal to each other by setting constraints, so that alternative models of evolution can be tested and then compared with each other by means of the assessment of likelihood or Akaike information criterion (AIC) values. We tested four alternative models. First, we fitted the 'free' model in which all six parameters were estimated. Then, we fitted a model with equal speciation rates ($\lambda_1 = \lambda_0$), one with equal extinction rates ($\mu_1 = \mu_0$), one with equal speciation and extinction rates ($\lambda_1 = \lambda_0$ and $\mu_1 = \mu_0$) and, finally, one with equal migration rates ($q_{10} = q_{01}$). Recent extensions of the BiSSE model allow parameter estimation from an incompletely sampled phylogeny and measures of parameter uncertainty (FitzJohn *et al*., 2009). The posterior probability distribution of the model parameters was approximated by MCMC, using an exponential prior for parameters whose mean was equal to twice the character-independent diversification rate (FitzJohn *et al*., 2009). To infer the model parameters in BiSSE, we ran two independent MCMCs for 10 000 steps. Ancestral states were also inferred in the diversitree package taking into account the different rates calculated from the best model.

RESULTS

Phylogenetic relationships in Chrysobalanaceae

The phylogenetic trees obtained with the complete dataset using a maximum likelihood reconstruction and a Bayesian method are displayed in Figures 1 and 2, respectively. Some genera appear to be monophyletic: *Maranthes* Blume (two species), *Dactyladenia* Welw. (two species), *Couepia* Aubl. (seven species), *Hirtella* L. (13 species) and *Parinari* Aubl. (seven species). Several clades are well supported whatever the method: clades L1, L2 and L3 include three species of *Licania* Aubl., clade L4 includes four *Licania* spp. and clade L5 includes 12 *Licania* spp. However, it is impossible to draw conclusions concerning the monophyly of *Licania* because of the weak support for basal nodes. We also define clade N which includes only Neotropical species (the genera *Hirtella*, *Couepia* and *Licania*), apart from *Afrolicania elaeosperma* Mildbr. Two of the three species of *Magnistipula* Engl. included in this analysis form a clade, but this clade excludes the accession of *Magnistipula tamenaka* (Capuron) F.White. Clade M includes only Palaeotropical species. We also define clade P, which includes *Parinari*, plus *Neocarya macrophylla* (Sabine) Prance ex F.White.

The topologies obtained with plastid markers only (result not shown) and the nuclear marker (Appendix 4) were similar, although slight inconsistencies could be noticed, in particular with respect to the position of some *Licania* spp. The phylogenetic tree obtained from the ITS dataset was consistent with the topology obtained with the full dataset, further suggesting that the poor placement of some of the species in the plastid DNA tree (*Magnistipula tamenaka*, *Acioa edulis* Prance) is the result of a wildcard taxon effect. Hereafter, we discuss analyses using the phylogenetic trees generated from the combined dataset.

Biogeography of Chrysobalanaceae

Figure 2 represents the dated phylogenetic tree with the representation of ancestral states (Neotropical or Palaeotropical) as obtained by the BiSSE analysis (the 95% highest posterior distributions, or HPDs, for ages are indicated on the phylogenetic tree presented in Appendix 5). The most recent common ancestor (MRCA) of the family appears to have been native to the Palaeotropics (probability $P = 1.00$) and, as the age of the family was constrained at 78 Myr, based on the broader analysis of Malpighiales by Davis *et al*. (2005), migration into South America after the breakup of Gondwana is highly likely.

Clade N is an almost strictly Neotropical clade, except for *Afrolicania elaeosperma*, the MRCA of

Figure 1. Phylogenetic tree of Chrysobalanaceae obtained with the full dataset (plastid plus nuclear markers) inferred from a maximum likelihood phylogeny reconstruction method (RaxML). Branches supported by bootstrap values > 80% are thick and those with bootstrap values of 60–80% are marked by an asterisk. The dating points used for Figure 2 are marked with a white star.

Figure 2. Dated phylogenetic tree from a Bayesian phylogeny reconstruction (BEAST version 1.6). Branches supported by posterior probabilities > 0.95 are thick and those with posterior probabilities of 0.85–0.95 are tagged by a star. On the same tree, an ancestral state reconstruction was inferred using a likelihood reconstruction method (binary-state speciation and extinction, BiSSE). Grey (vs. black) colour represents Neotropical (vs. Palaeotropical) species or internal nodes. States for unambiguous nodes are not represented for the sake of clarity.

which appeared to be Neotropical ($P = 0.92$). This clade may have diversified around 47 Mya (HPD, 40–54 Myr), suggesting that the arrival of Chrysobalanaceae in the Neotropics occurred before this date. Further, *Hirtella* appears to have diversified about 10 Mya (HPD, 6–14 Myr) and *Couepia* about 11 Mya (HPD, 7–14 Myr).

An inspection of clade P, including *Parinari*, shows that the MRCA of clade P, and the ancestor of all *Parinari*, was Palaeotropical ($P = 1.00$) and the split between *Parinari* and *Neocarya macrophylla* is inferred at 33 Mya (HPD, 26–45 Myr). Evidence from the genus in both Africa at 16.1 Mya and the Neotropics prior to 17 Mya suggests that the dispersal into South America is older than these dates.

Joint estimates of dispersal, speciation and extinction rates

We used the consensus phylogenetic tree produced by BEAST to test the different 'BiSSE' models. Using likelihood ratio tests and AIC scores, we determined

Table 2. Maximum likelihood value and Akaike information criterion (AIC) scores for alternative diversification models. The 'free' model represents an unconstrained model (all parameters are independently estimated). The 'Equal.λ' model assumes equal speciation rates between the Neotropics and the Palaeotropics. The 'Equal.μ' model assumes equal extinction rates. The 'Equal.q' model assumes equal migration rates across the Neotropics and the Palaeotropics. Finally, the 'Equal.$\lambda\mu$' model assumes that both the speciation and extinction rates are constrained to be equal between the two regions

Model	Log-likelihood	AIC
Free	–321.46	654.91
Equal.λ	–331.16	672.31
Equal.μ	–332.01	674.02
Equal.q	–331.20	672.41
Equal.$\lambda\mu$	–333.58	675.15

that the best model was the fully parameterized model in which speciation, extinction and transition rates were different between the Neotropics and the Palaeotropics (Table 2).

Figure 3 shows the posterior distributions of the six parameters estimated, together with diversification (speciation minus extinction) rates for each geographical area. Both speciation and extinction rates were found to be significantly higher in the Neotropics than in the Palaeotropics. The net diversification rate was higher in the Neotropics, although the difference was not significant. Likewise, migration from the Palaeotropics into the Neotropics was higher than vice versa, although this difference was not significant.

DISCUSSION

Phylogenetic relationships in Chrysobalanaceae

Overall, our analysis provides strong support for the validity of the morphological characters used to delimit genera in this family (as discussed previously in Prance & White, 1988). Chrysobalanaceae *s.s.* was divided by Prance & Sothers (2003) into 18 genera on the basis of morphological data. Among the seven genera for which at least two species have been sampled, five were confirmed as monophyletic on the basis of our sampling. Good support for genera *Hirtella* and *Couepia* was found, supporting a previously held view that these genera are well defined on a morphological basis (Martius & Zuccarini, 1832 in Prance, 1972). The monophyly of *Parinari* is also well supported in our study, as is the clade grouping of *Parinari* with *Neocarya macrophylla*. *Neocarya macrophylla* was called *Parinari senegalensis* Perr. ex DC. in the past and grouped with *Parinari excelsa* in one of the two *Parinari* sections (Prance & White, 1988).

For two genera, *Licania* and *Magnistipula*, our analysis indicates that they may not be monophyletic. Yakandawala *et al*. (2010) also suggested that they are nonmonophyletic on the basis of morphological data, although they stated that further work using molecular data and more taxa was needed to confirm these findings. Our new analysis remains inconclusive because of the weak support for deep nodes and the limited taxonomic coverage of the family in our phylogenetic analyses. Prance (1972) proposed that *Licania* should be separated into four subgenera: *Moquilea* Aubl., *Parinariopsis* Huber, *Angelesia* (Korth.) Prance & F.White and *Licania*. Clade L1 includes three species of subgenus *Moquilea*, whereas all species present in the *Licania* clades L2, L3 and L5 belong to subgenus *Licania*. Clade L4 lumps three species belonging to subgenus *Licania* and one of subgenus *Moquilea*. Thus, the subgeneric classification of *Licania* may be in need of revision. In the genus *Magnistipula*, two of the three species included in our analysis belong to subgenus *Magnistipula* and group together. The third species, *M. tamenaka*, falls out in a separate clade; it is one of the two species of *Magnistipula* from Madagascar and belongs to subgenus *Tolmiella* F.White. It cannot be excluded that the odd position of *M. tamenaka* is a result of incomplete sequencing of the markers. An alternative possibility is that *M. tamenaka* belongs to a segregate of genus *Magnistipula*. Finally, we note that the two accessions of *Acioa* Aubl. did not fall together as a single clade. However, *Acioa edulis* was represented by a single sequence and, in another phylogenetic study (C. Sothers, unpubl. data), *A. guianensis* Aubl. and *Acioa edulis* formed a well-supported clade; therefore, we believe that the results regarding the phylogenetic position of *M. tamenaka* and *A. edulis* are most probably a result of a wildcard taxon effect.

Biogeography of Chrysobalanaceae

On the basis of age estimates of Davis *et al*. (2005), Chrysobalanaceae *s.s*. originated well after the last known connection between Africa and South America (~105 Mya). The ancestral state reconstruction shows that the MRCA of Chrysobalanaceae was Palaeotropical. The stem age of clade N, which comprises the majority of South American species, estimated at 49 Mya (HPD, 44–59 Myr), and the crown age, estimated at 47 ± 6 Mya, place the temporal boundaries on the first dispersal of Chrysobalanaceae in the New World. This corresponds well with the late Palaeocene climate maximum (LPCM, 58–52 Mya), during which a boreotropical forest putatively existed in

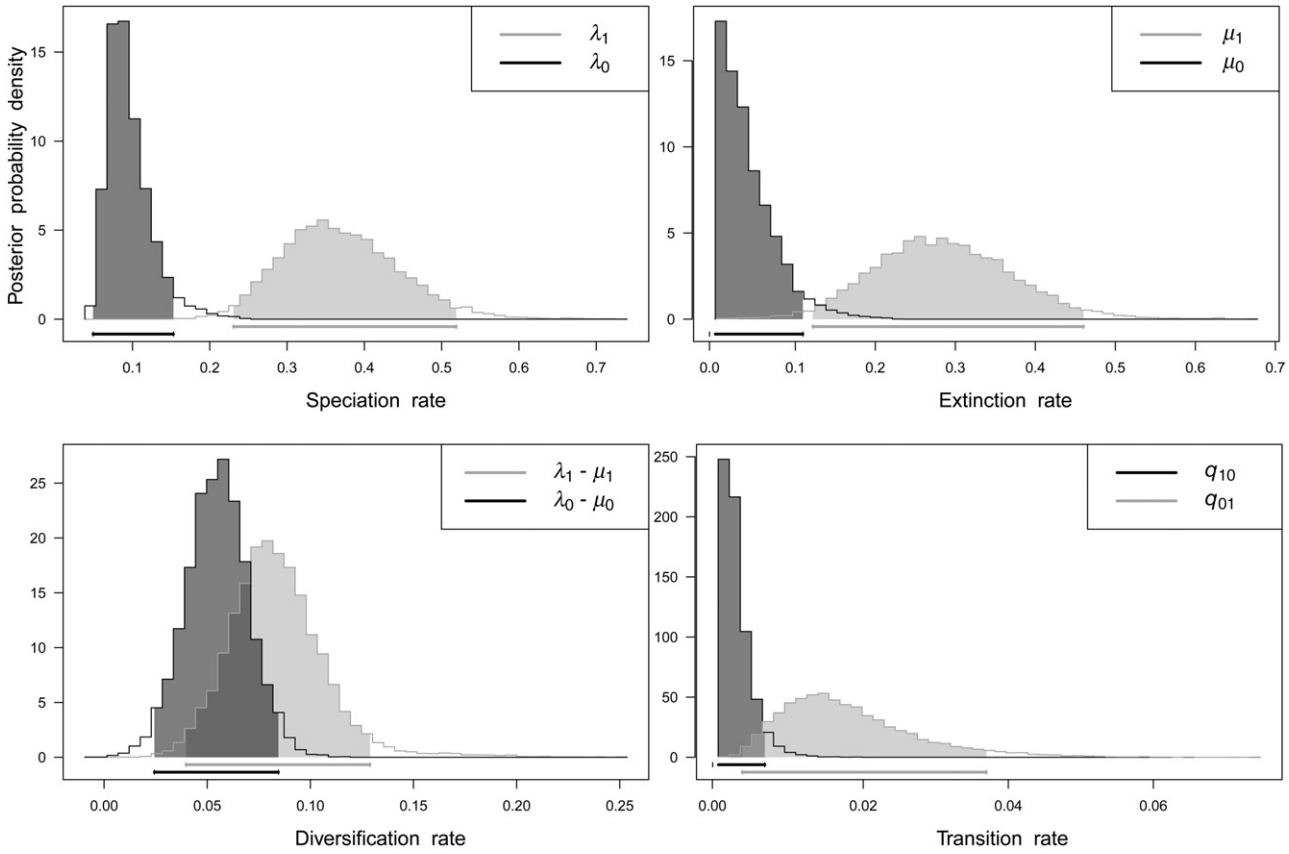


Figure 3. Posterior probability distributions of the rates of speciation, extinction, diversification and transition (species-Myr^{-1}) generated in the binary-state speciation and extinction (BiSSE) analysis. Horizontal bars indicate the 95% credibility interval. Estimates are shown for the Neotropics (in grey) and the Palaeotropics (in black).

North America and Europe and the two continents were connected by the North Atlantic land bridge (NALB; Tiffney, 1985). Wodehouse (1932) identified a middle Eocene palynofossil in the Green River Formation, Colorado, USA as *Chrysobalanus* pollen, suggesting that North America contained suitable habitat for Chrysobalanaceae 45 Mya. We therefore suggest that Chrysobalanaceae first dispersed from Africa to the Neotropics via Europe and North America during the LPCM. The family may have then subsequently dispersed from North America into South America via an island corridor of the proto-Greater Antilles, thought to have been present 45–35 Mya (Iturralde-Vinent & MacPhee, 1999). It is important to emphasize, however, that future fossil discoveries could modify or even overturn this proposed scenario.

The genus *Parinari* contains evidence of further dispersals between the Palaeo- and Neotropics. The MRCA of all *Parinari* spp. sampled was found in the Palaeotropics around 8 Mya (HPD, 8–17 Myr), whereas the stem age of the genus was estimated at *c*. 33 Mya. As *Parinari* fossils older than 16 Myr have been found in both Africa and the Neotropics, this suggests that we have underestimated the age of the *Parinari* MRCA. Further, dispersal of this genus probably occurred from Africa to the Neotropics, and the first event must have been older than 16 Myr. The presence of *P. excelsa* in both the Neotropics and Africa could suggest that there may have been one more recent dispersal event. However, our accessions of *P. excelsa* from both continents fall into two separate clades, raising the possibility that *P. excelsa* in Africa and the Neotropics are distantly related and do not represent a recent dispersal event.

The other well-known example of recent trans-oceanic dispersal is that of *Chrysobalanus icaco*. In our analysis, we have included one accession from Africa (sometimes called *Chrysobalanus ellipticus* Sol. ex Sabine, but synonymized under *Chrysobalanus icaco* by Prance & White, 1988) and one accession from South America. Our dating of their split ranges between 4 and 13 Mya. When discussing possible recent trans-oceanic dispersal events, the clade including genus *Chrysobalanus* deserves additional scrutiny. The position of *Afrolicania elaeosperma*,

which falls clearly into the Neotropical clade N, is also an interesting result, because it must be a secondary dispersal from the Neotropics into the Palaeotropics.

Another group is poorly supported (bootstrap value, 36; posterior probability, 0.74), but includes *Acioa* Aubl., *Licania* subgenus *Angelesia*, *Hunga* Prance and *Exellodendron* Prance. If confirmed, this clade would form an interesting boreotropical group from a biogeographical standpoint: *Exellodendron* (five species) and *Acioa* (four species) are found only in South America, *Hunga* (11 species) is found only in Papua New Guinea and New Caledonia, and *Licania* subgenus *Angelesia* (three species) is widespread in South-East Asia (New Guinea, Thailand, Philippines). This supports the hypothesis that *Licania splendens* (Korth.) Prance may be a different genus, and that it may bear some affinities with *Hunga*. This also suggests that a Pacific dispersal route may have been possible. A better sampling of this group should yield more results.

The genus *Maranthes* Blume was, for a long time, placed in the genus *Parinari* (Prance & White, 1988), but is here placed in clade M, also containing genera *Dactyladenia* and *Grangeria* Comm. ex Juss. Clade M and *Maranthes* are found to be Palaeotropical in our ancestral state reconstruction analysis. However, one species, *Maranthes panamensis* (Standl.) Prance & F.White, is found in Central America. It would be important to include an accession of this species in future analyses to confirm that *M. panamensis* represents a recent trans-oceanic dispersal event in the *Maranthes* clade.

Prance & White (1988) suggested that long-distance oceanic dispersal may be common in Chrysobalanaceae, particularly in the light of the fact that six genera have species that can disperse via flotation. Our results suggest that this is indeed the case, as many of the more recent dispersal events do not seem to be explicable by dispersal via continents or land bridges.

Diversification in the Neotropics

Differences in species richness between the Neotropics and Palaeotropics could be explained by differences in the time of species colonizations and/or differences in diversification rates between these regions (Moore & Donoghue, 2007). Our study shows that Chrysobalanaceae originated in the Palaeotropics, ruling out longer residence time in the Neotropics as an explanation for greater species richness there. Our diversification analyses showed that extinction rates have been higher in the Neotropics than the Palaeotropics, which rules out a 'museum' hypothesis of low extinction. This result is somewhat surprising because wide-scale aridification leading to disproportionate extinctions is often invoked to explain why the African flora is less diverse than that of South America.

The only hypothesis to explain the higher Neotropical diversity of Chrysobalanaceae that was supported by our analyses was the 'cradle' hypothesis of higher speciation rates in the Neotropics. The recent appearance and radiation of species-rich genera, such as *Hirtella* and *Couepia* (~12 Mya), lend support to the idea of a high speciation rate for Chrysobalanaceae in the Neotropics. This result suggests that the current composition of Neotropical rainforests was essentially set up during the Neogene (~23–5 Mya; Potter & Szatmari, 2009). It is tempting to explain this high speciation rate by invoking the uplift of the Andes, as described by Gentry (1982). This uplift began for the northern Andes around 23 Mya and reached peaks of intensity during the late to middle Miocene (~12 Mya) and early Pliocene (~4.5 Mya) (Richardson *et al*., 2001; Erkens *et al.,* 2009; Hoorn *et al*., 2010). However, the large majority of species of Chrysobalanaceae are in the central and eastern Amazon, and this family is predominantly a lowland rainforest group, with a few outliers in other habitats and few at high altitudes. Thus, it is likely that the rise of the Andes did not play a direct role in speciation events for Chrysobalanaceae. A similar scenario may be invoked for other Amazonian-centred Gondwanan families of woody trees, such as Caryocaraceae, Humiriaceae, Lecythidaceae, Sapotaceae and Vochysiaceae.

An alternative explanation is that the instability of the South American climate during the Pliocene and/or Pleistocene may have led to recurrent phases of retraction of species of Chrysobalanaceae into refugia, thus causing allopatry and acting as a speciation pump (Haffer, 1969; Prance, 1974). However, it is not clear that the palaeoclimate was less stable in South America than in Africa or other tropical regions over the last 10 Myr. More detailed phylogenetic hypotheses, including speciation events spanning this epoch, would be needed to adequately test this scenario.

Our best model allowed for different transition rates between regions, with the transition rate from the Palaeotropics to the Neotropics being higher than the reverse. This asymmetric dispersal may reflect the prevailing direction of past ocean currents (Renner, 2004). Some of the recent long-distance dispersal events appear to have occurred from Africa to South America (e.g. Dick, Abdul-Salim & Bermingham, 2003, for *Symphonia globulifera* L.f.). If plant lineages experience higher speciation rates in areas to which they have recently dispersed (*sensu* Moore & Donoghue, 2007), the observed asymmetric dispersal

may provide an additional explanation for the higher species diversity of Chrysobalanaceae in the Neotropics.

CONCLUDING REMARKS

The current patterns in the diversity of Chrysobalanaceae can be explained by multiple dispersal events between the Palaeotropics and the Neotropics and higher speciation rates in the Neotropics. Our biogeographical scenario for this family is comparable with that developed by Tiffney (1985) and Morley (2000, 2003), although more recent dispersal events probably represent long-distance dispersal across oceans, and we are currently unable to provide a finer grained scenario within the Palaeotropics. The causes of the high speciation rate of Chrysobalanaceae in the Neotropics remain unknown. Given the abundance of species of Chrysobalanaceae in eastern and central Amazonia and in drier habitat types, it is possible that the observed diversification is related to the complex geological and environmental history of South America and not necessarily directly to the uplift of the Andes.

ACKNOWLEDGEMENTS

We thank Alexandre Antonelli and Toby Pennington for their invitation to contribute to this issue, Fabiany Herrera and Steven Manchester for sharing unpublished dated fossils, and Bruce Tiffney for useful comments. We thank Joan Pereira, Colin Maycock (Herbarium of the Forest Reserve Centre at Sepilok, Malaysia), Olivier Hardy and Ingrid Parmentier for sharing their African collections and Maxime Réjou-Méchain and the ARF project of the Central African Republic for providing two samples from the Mbaïki experimental plot. This work has benefited from 'Investissement d'Avenir' grants managed by Agence Nationale de la Recherche (CEBA: ANR-10-LABX-0025; TULIP: ANR-10-LABX-0041). It was partly funded by Agence Nationale pour la Recherche (BRIDGE project), the Programme Interdisciplinaire de Recherche AMAZONIE (CNRS) and the Fondation pour la Recherche sur la Biodiversité (ATROPHY project).

REFERENCES

Antonelli A, Nylander JAA, Persson C, Sanmartin I. 2009. Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proceedings of the National Academy of Sciences of the United States of America* **24:** 9749–9754.

APG III. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* **161:** 105–121.

Bennett KD. 2004. Continuing the debate on the role of Quaternary environmental change for macroevolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359:** 295–303.

Britton T, Anderson C, Jacquet D, Lundqvist S, Bremer K. 2007. Estimating divergence times in large phylogenetic trees. *Systematic Biology* **56:** 741–752.

Burnham RJ, Graham A. 1999. The history of Neotropical vegetation: new developments and status. *Annals of the Missouri Botanical Garden* **86:** 546–589.

Davis CC, Webb CO, Wurdack KJ, Jaramillo CA, Donoghue MJ. 2005. Explosive radiation of Malpighiales supports a Mid-Cretaceous origin of modern tropical rain forests. *The American Naturalist* **3:** E36–E65.

Davis CC, Wurdack KJ. 2004. Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. *Science* **305:** 676–678.

Dick CW, Abdul-Salim K, Bermingham E. 2003. Molecular systematic analysis reveals cryptic tertiary diversification of a widespread tropical rain forest tree. *The American Naturalist* **162:** 691–703.

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

Duarte L, Mello Filha M. 1980. Florula Cenozoica de Gandarela, Minas Gerais. *Anais Academia Brasilera de Ciencias* **52:** 77–91.

Dunning LT, Savolainen V. 2010. Broad-scale amplification of *matK* for DNA barcoding plants, a technical note. *Botanical Journal of the Linnean Society* **164:** 1–9.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32:** 1792–1797.

Erkens RHJ, Maas JW, Couvreur TLP. 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 PVA, Ree RH. 2006. Evidence for a time-integrated species–area effect on the latitudinal gradient in tree diversity. *The American Naturalist* **168:** 796–804.

FitzJohn RG, Maddison WP, Otto SP. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic Biology* **58:** 595–611.

Gentry AH. 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.

Goodall-Copestake WP, Pérez-Espona S, Harris DJ, Hollingsworth PM. 2010. The early evolution of the mega-diverse genus *Begonia* (Begoniaceae) inferred from organelle DNA phylogenies. *Biological Journal of the Linnean Society* **101:** 243–250.

Gouy M, Guindon S, Gascuel O. 2010. Seaview version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* **27:** 221–224.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59:** 307–321.

Haffer J. 1969. Speciation in Amazonian forest birds. *Science* **165:** 131–137.

Hamilton MB. 1998. Four primers pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* **8:** 513–525.

Ho SYW. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. *Journal of Avian Biology* **38:** 409–414.

Hoorn C, Wesselingh FP, ter Steege H, Bermudez MA, Mora A, Sevink J, Sanmartín I, Sanchez-Meseguer A, Anderson CL, Figueiredo JP, Jaramillo C, Riff D, Negri FR, Hooghiemstra H, Lundberg J, Stadler T, Särkinen T, Antonelli A. 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* **330:** 927–931.

Hopkins MJG. 2007. Modelling the known and unknown plant biodiversity of the Amazon Basin. *Journal of Biogeography* **34:** 1400–1411.

Iturralde-Vinent MA, MacPhee RDE. 1999. Paleogeography of the Caribbean region: implications for Cenozoic biogeography. *Bulletin of the American Museum of Natural History* **238:** 3–90.

Li M, Wundera J, Bissolia G, Scarponia E, Gazzania S, Barbaroa E, Saedlera H, Varottoa C. 2008. Development of COS genes as universally amplifiable markers for phylogenetic reconstructions of closely related plant species. *Cladistics* **24:** 1–19.

Linder HP. 2008. Plant species radiations: where, when, why? *Philosophical Transactions of the Royal Society B: Biological Sciences* **363:** 3097–3105.

Maddison WP, Midford PE, Otto SP. 2007. Estimating a binary character's effect on speciation and extinction. *Systematic Biology* **56:** 701–710.

Manchester SR. 1999. Biogeographical relationships of North American Tertiary floras. *Annals of the Missouri Botanical Garden* **86:** 472–522.

Matthews ML, Endress PK. 2008. Comparative floral structure and systematics in Chrysobalanaceae *s.l.* (Chrysobalanaceae, Dichapetalaceae, Euphroniaceae, Trigoniaceae; Malpighiales). *Botanical Journal of the Linnean Society* **157:** 249–309.

Moore BR, Donoghue MJ. 2007. Correlates of diversification in the plant clade Dipsacales: geographic movement and evolutionary innovations. *The American Naturalist* **170** (suppl)**:** S28–S55.

Morley RJ. 2000. *Origin and evolution of tropical rain forests*. New York: Wiley.

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

Pennington RT, Dick CW. 2004. The role of immigrants in the assembly of the South American rainforest tree flora. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359:** 1611–1622.

Platnick NI, Griswold CE, Coddington JA. 1991. On missing entries in cladistic analysis. *Cladistics* **7:** 337–343.

Potter PE, Szatmari P. 2009. Global Miocene tectonics and the modern world. *Earth-Science Reviews* **96:** 279–295.

Prance GT. 1972. *Flora neotropica. Chrysobalanaceae*. New York: Hafner Publishing Company.

Prance GT. 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 GT, Sothers CA. 2003. *Chrysobalanaceae 1 & 2, species plantarum: flora of the world, parts 9 &10*. Canberra: Australian Biological Resources Study.

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

R Development Core Team. 2009. *R: a language and environment for statistical computing*. Vienna: R foundation for Statistical Computing.

Raven PH, Axelrod DI. 1974. Angiosperm biogeography and past continental movements. *Annals of the Missouri Botanical Garden* **61:** 539–673.

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

Renner SS, Clausing G, Meyer K. 2001. Historical biogeography of Melastomataceae: the role of Tertiary migration and long-distance dispersal. *American Journal of Botany* **88:** 1290–1300.

Richardson JE, Pennington RT, Pennington TD, Hollingsworth PM. 2001. Rapid diversification of a species-rich genus of Neotropical rain forest trees. *Science* **293:** 2242–2245.

Sanmartin I, Enghoff H, Ronquist F. 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. *Biological Journal of the Linnean Society* **73:** 345–390.

Savolainen V, Chase MW, Hoot SB, Morton CM, Soltis DE, Bayer C, Fay MF, De Bruijn AY, Sullivan S, Qiu Y-L. 2000. Phylogenetics of flowering plants based on combined analysis of plastid *atpB* and *rbcL* gene sequences. *Systematic Biology* **49:** 306–362.

Shaw J, Lickey EB, Schilling EE, Small RL. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* **94:** 275–288.

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

Stebbins GL. 1974. *Flowering plants: evolution above the species level*. Cambridge, MA: Harvard University Press.

Strand AE, Leebens-Mack J, Milligan BG. 1997. Nuclear DNA-based markers for plant evolutionary biology. *Molecular Ecology* **6:** 113–118.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24:** 1596–1599.

Tiffney BH. 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 BH, Fleagle JG, Brown TM. 1994. Early to Middle Miocene angiosperm fruits and seeds from Fejej, Ethiopia. *Tertiary Research* **15:** 25–42.

Wijninga VM. 1996. Neogene ecology of the Salto de Tequendama site (2475m altitude, Cordillera Oriental, Colombia): the paleobotanical record of montane and lowland forests. *Review of Palaeobotany and Palynology* **92:** 97–156.

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

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

APPENDIX 1

The full list of samples with their herbarium numbers, their collection locations and their accession numbers on GenBank and EMBL for each marker.

Name	Bar code	Sampling country	*psbD-trnT*	*atpI-atpH*	*ndhA* intron	*matK*	ITS	*rbcLa*	*trnH-psbA*
Acioa guianensis Aubl.	KD5285	French Guiana	JQ898765	JQ898947	JQ898785		JQ898973	JQ898703	JQ898918
Acioa guianensis Aubl.	Prance 30841	Brazil					GQ424453	GQ424473	
Acioa edulis Prance	cynthia34568	Brazil						JQ898704	
Atuna racemosa Raf.	2118-KEW	Indonesia	JQ898752	JQ898928	JQ898781	JQ898828			JQ898909
Atuna racemosa Raf.	Morton 89	Indonesia					GQ424454	GQ424474	
Bafodeya benna (Scott-Elliot) Prance ex F.White	Saiden1997	Guinea					GQ424455	GQ424475	
Chrysobalanus ellipticus Sol. ex Sabine	MH2313	Benin	JQ898770	JQ898958	JQ898782	JQ898857			JQ898923
Chrysobalanus ellipticus Sol. ex Sabine	MH2314	Benin					JQ898992	JQ898692	
Chrysobalanus icaco L.	KD5286	French Guiana	JQ898766	JQ898948	JQ898786	JQ898850	JQ898993	JQ898693	JQ898919
Chrysobalanus icaco L.	Morton 64	Indonesia						GQ424476	
Chrysobalanus icaco L.	Prance 30833	Dominican Republic					GQ424456		
Couepia bracteosa Benth.	P00610245	French Guiana	JQ898775		JQ898787	JQ898862			
Couepia bracteosa Benth.	NH200200	French Guiana					JQ899003	JQ898696	JQ898883
Couepia caryophylloides Benoist	B443011	French Guiana	JQ898761	JQ898938	JQ898783	JQ898839	JQ899002	JQ898700	JQ898884
Couepia excelsa Ducke	JC 152	French Guiana	JQ898764		JQ898784	JQ898848	JQ899021		JQ898917
Couepia guianensis Aubl.	B421045	French Guiana	JQ898760	JQ898936	JQ898795	JQ898837			JQ898885
Couepia guianensis Aubl.	NH200659	French Guiana					JQ899001	JQ898699	
Couepia habrantha Standl.	LV109174	French Guiana		JQ898954					
Couepia habrantha Standl.	P00610420	French Guiana				JQ898865	JQ898998	JQ898697	
Couepia habrantha Standl.	P00610747	French Guiana							JQ898886
Couepia joaquinae Prance	NH200492	French Guiana	JQ898772	JQ898970	JQ898810	JQ898859	JQ899000	JQ898701	JQ898887
Couepia magnoliifolia Benth. ex Hook.f.	P01860371	French Guiana	JQ898779		JQ898817	JQ898869			JQ898925
Couepia magnoliifolia Benth. ex Hook.f.	LV109012	French Guiana					JQ898999	JQ898698	
Dactyladenia pallescens (Baill.) Prance & F.White	McPherson 16317	Gabon					GQ424458	GQ424478	
Dactyladenia pallescens (Baill.) Prance & F.White	PM5525	Cameroon						JQ898694	
Dactyladenia staudtii (Engl.) Prance & F.White	PM5392	Cameroon			JQ898788	JQ898829			
Dactyladenia staudtii (Engl.) Prance & F.White	PM5282	Cameroon					JQ898974	JQ898695	
Exellodendron barbatum (Ducke) Prance	A110225	French Guiana	JQ898759	JQ898935	JQ898794	JQ898835	JQ898991	JQ898705	JQ898888
Grangeria borbonica Lam.	LR 84	Reunion		JQ898966	JQ898823	JQ898877			
Grangeria borbonica Lam.	Derleth 61	Madagascar					GQ424460	GQ424480	
Hirtella araguariensis Prance	JC 132	French Guiana		JQ898945	JQ898801	JQ898846			
Hirtella araguariensis Prance	NL110214	French Guiana					JQ898988	JQ898735	JQ898889
Hirtella bicornis Mart. & Zucc.	M17116087	French Guiana	JQ898769	JQ898957	JQ898808	JQ898856			
Hirtella bicornis Mart. & Zucc.	M17116415	French Guiana					JQ898984	JQ898736	JQ898890
Hirtella glandulosa Spreng.	NH200142	French Guiana					JQ898986	JQ898733	JQ898891
Hirtella ciliata Mart. & Zucc.	KD 5565	Guyana		JQ898951	JQ898804	JQ898853	JQ899024	JQ898748	
Hirtella davisii Sandwith	PE 4153	French Guiana		JQ898971	JQ898826	JQ898881	JQ898982	JQ898732	

Hirtella glandistipula Ducke	P00610010	French Guiana					JQ898985	JQ898731		
Hirtella hebeclada Moric.	JPC1	Brazil	JQ898763	JQ898942	JQ898798	JQ898843	JQ899020	JQ898746	JQ898915	
Hirtella macrosepala Sandwith	B445070	French Guiana		JQ898939		JQ898840	JQ898989	JQ898734	JQ898892	
Hirtella physophora Mart. & Zucc.	JC 076	French Guiana		JQ898943						
Hirtella physophora Mart. & Zucc.	JC113	French Guiana					JQ898981	JQ898737		
Hirtella silicea Griseb.	PE 1100	French Guiana			JQ898819	JQ898871	JQ899014	JQ898751	JQ898926	
Hirtella suffulta Prance	NH200131	French Guiana	JQ898771	JQ898959	JQ898809	JQ898858	JQ898983	JQ898738	JQ898893	
Hirtella tenuifolia Prance	JC 97	French Guiana		JQ898941	JQ898800	JQ898845			JQ898916	
Hirtella tenuifolia Prance	JC084	French Guiana					JQ898987	JQ898730		
Hirtella triandra Sw.	KD 5694	Guyana		JQ898952	JQ898805	JQ898854	GQ424461	GQ424481	JQ898922	
Hirtella triandra Sw.	S. R. Hill 29095	Dominican Republic					GQ424462	GQ424482		
Hunga gerontogea (Schltr.) Prance	McPherson 6093	New Caledonia					GQ424463	GQ424483		
Kostermanthus robustus Prance	Morton tree1099	Malaysia						JQ898729		
Afrolicania elaeosperma Mildbr.	Le11505	Cameroon							JQ898894	
Licania alba (Bernoulli) Cuatrec.	P00610185	French Guiana	JQ898774	JQ898931	JQ898812	JQ898861	JQ899004	JQ898707		
Licania alba (Bernoulli) Cuatrec.	NH200372	French Guiana					JQ899017			
Licania amapaensis Prance	CETP104	French Guiana	JQ898762	JQ898940	JQ898797	JQ898841	JQ899023	JQ898720	JQ898921	
Licania apetala Fritsch	KD5423	French Guiana	JQ898768	JQ898950	JQ898796	JQ898852			JQ898895	
Licania canescens Benoist	B422043	French Guiana		JQ898937		JQ898838	JQ899006	JQ898710		
Licania canescens Benoist	NH200019	French Guiana								
Licania densiflora Kleinhoonte	JC 140	French Guiana		JQ898946		JQ898847		JQ898715		
Licania densiflora Kleinhoonte	PE5401	French Guiana							JQ898896	
Licania densiflora Kleinhoonte	P01860172	French Guiana								
Licania glabriflora Prance	A 110083	French Guiana	JQ898757	JQ898933	JQ898792	JQ898833	JQ899016			
Licania glabriflora Prance	B445068	French Guiana						JQ898723		
Licania guianensis Kuntze	PE1075	French Guiana						JQ898722		
Licania heteromorpha Benth.	LV109473	French Guiana		JQ898955						
Licania heteromorpha Benth.	B421016	French Guiana					JQ898978	JQ898726	JQ898897	
Licania hypoleuca Benth.	NH220023	French Guiana						JQ898711		
Licania irwinii Prance	KD5287	French Guiana	JQ898767	JQ898949	JQ898803	JQ898851	JQ899010	JQ898709	JQ898920	
Licania kunthiana Hook.f.	LV113008	French Guiana		JQ898956	JQ898807		JQ899007	JQ898714		
Licania laevigata Prance	M17116871	French Guiana		JQ898967	JQ898824	JQ898879	JQ899979	JQ898728		
Licania latifolia Benth. ex Hook.f.	PE5397	French Guiana						JQ898727		
Licania latistipula Prance	CETP 119	French Guiana		JQ898960		JQ898873	JQ898995	JQ898725	JQ898914	
Licania laxiflora Fritsch	P00610253	French Guiana	JQ898776		JQ898813	JQ898863	JQ899005	JQ898708		
Licania laxiflora Fritsch	B445017	French Guiana							JQ898898	
Licania licaniiflora S.F.Blake	P01860126	French Guiana	JQ898778			JQ898867			JQ898899	
Licania licaniiflora S.F.Blake	P00610743	French Guiana					JQ898990	JQ898717		
Licania longistyla Fritsch	JC 182	French Guiana		JQ898965	JQ898802	JQ898849	JQ899022	JQ898747		
Licania macrophylla Klotzsch	KGD5284	French Guiana					JQ898994	JQ898724		
Licania majuscula Sagot	A110139	French Guiana	JQ898758	JQ898934	JQ898793	JQ898834			JQ898900	
Licania majuscula Sagot	NL110261	French Guiana					JQ899008	JQ898712		
Licania membranacea Sagot ex Laness.	NH200005	French Guiana					JQ899009	JQ898716		
Licania membranacea Sagot ex Laness.	M17116066	French Guiana				JQ898878				
Licania membranacea Sagot ex Laness.	P00610550	French Guiana							JQ898901	
Licania micrantha Miq.	NH200679	French Guiana	JQ898773		JQ898811	JQ898860				

APPENDIX 1. *Continued*

Name	Bar code	Sampling country	*psbD-trnT*	*atpI-atpH*	*ndhA* intron	*matK*	ITS	*rbcLa*	*trnH-psbA*
Licania micrantha Miq.	P00610368	French Guiana					JQ899011	JQ898706	JQ898902
Licania minutiflora Cuatrec.	NH200220	French Guiana		JQ898969	JQ898825	JQ898880	JQ898996	JQ898721	
Licania minutiflora Cuatrec.	NH200162	French Guiana							JQ898903
Licania octandra Pilg.	LV109008	French Guiana		JQ898953	JQ898806	JQ898855	JQ899025	JQ898718	
Licania ovalifolia Kleinhoonte in Pulle	B437044	French Guiana					JQ899012	JQ898713	JQ898904
Licania parvifructa Fanshawe & Maguire	P01120631	French Guiana			JQ898815	JQ898866	JQ899027	JQ898749	JQ898924
Licania pyrifolia Griseb.	AF28563	Venezuela				JQ898836			
Licania splendens (Korth.) Prance	2120-KEW	Indonesia	JQ898755	JQ898929	JQ898790	JQ898831	JQ899015	JQ898742	JQ898912
Licania sprucei Fritsch	P01860159	French Guiana			JQ898816	JQ898868			
Licania sprucei Fritsch	P00610114	French Guiana					JQ898997	JQ898719	JQ898905
Licania tomentosa Fritsch	Morton 97	Indonesia					GQ424464	GQ424484	
Magnistipula butayei De Wild.	T. B. Hart 1362	Zaire					GQ424465		
Magnistipula butayei sub sp sargosii (Pellegr.) F.White	F2P2C2 603	Central African Republic		JQ898963	JQ898822	JQ898876	JQ899018		
Magnistipula glaberrima Engl.	PM 5354	Cameroon				JQ898872			
Magnistipula glaberrima Engl.	PM4906	Cameroon					JQ898980	JQ898702	
Magnistipula tamenaka (Capuron) F.White	16117-KEW	Madagascar	JQ898756	JQ898932	JQ898791	JQ898832		JQ898744	JQ898913
Maranthes glabra (Oliv.) Prance	MH 838	Gabon		JQ898968			JQ899026		
Maranthes glabra (Oliv.) Prance	Bos7352	Cameroon					GQ424466	GQ424486	
Maranthes kerstingii (Engl.) Prance ex F.White	F1P1C4 986	Central African Republic	JQ898780	JQ898962	JQ898821	JQ898875		JQ898745	
Neocarya macrophylla (Sabine) Prance ex F.White	Goudiaby_ Sambou1998	Senegal					GQ424467	GQ424487	
Parinari brasiliensis Hook.f.		Brazil	JQ898753		JQ898818	JQ898870	JQ899028	JQ898750	
Parinari campestris Aubl.	P00610261	French Guiana	JQ898777		JQ898814	JQ898864			JQ898906
Parinari campestris Aubl.	P00610286	French Guiana					JQ898976	JQ898739	
Parinari canarioides Kosterm.	SAN 152716	Malaysia		JQ898961	JQ898820	JQ898874	JQ899029		JQ898910
Parinari excelsa Sabine	GK0056	Gabon		JQ898964		JQ898842			
Parinari excelsa Sabine	B429025	French Guiana					JQ898977	JQ898741	JQ898907
Parinari excelsa Sabine	GK0054	Gabon					JQ899019		
Parinari montana Aubl.	JC 085	French Guiana		JQ898944	JQ898799	JQ898844			
Parinari montana Aubl.	LV109130	French Guiana					JQ898975	JQ898740	
Parinari montana Aubl.	NH200145	French Guiana							JQ898908
Parinari oblongifolia Hook.f.	SAN 152336	Malaysia		JQ898972	JQ898827	JQ898882		JQ898743	JQ898927
Parinari sumatrana Kurz	2119-KEW	Indonesia	JQ898754	JQ898930	JQ898789	JQ898830	JQ899013		JQ898911
Euphronia guianensis (R.H.Schomb.) Hallier f.	Berry6562	Venezuela				EF135540	AY674597	EF135540	

APPENDIX 2

LABORATORY PROTOCOLS: DNA EXTRACTION AND AMPLIFICATION

DNA extraction

To extract DNA, leaf material was placed into 2-mL Eppendorf tubes with three autoclaved balls and was ground to dust with a Bio/Tissue Lyser. Cell lysis was performed as follows: 800 µL of cetyl trimethylammonium bromide (CTAB) and 10 µL of proteinase K were added to all tubes, which were then incubated for 2 h at 55 °C. Tubes were then centrifuged for 10 min at 20 817 g (14 000 rpm). DNA was extracted from the supernatant using the auto-extractor Biosprint 15 (Qiagen), following the protocol provided by the supplier. This protocol involves a precipitation in isopropanol, rinsing in pure ethanol and conservation in sterile water.

DNA amplification

DNA amplification was performed using classic polymerase chain reaction (PCR) protocols. A total of 1 µL of DNA of each extract was added to 10 µL of PCR buffer, 1 µL of deoxynucleoside triphosphate (dNTP), 1 µL of each primer at a concentration of 20 µM, 0.2 µL of Taq polymerase and 35.8 µL of sterile water. Then, DNA extracts were placed into a thermal cycler to realize a PCR. For markers newly amplified on Chrysobalanaceae, the first PCR protocol tested was that provided in the original publication. If needed, the second protocol tested involved template denaturation at 93 °C for 1 min, followed by 35 cycles of denaturation at 93 °C for 10 s, primer annealing at 50 °C for 1 min and primer extension at 68 °C for 1 min, followed by a final extension step of 2 min at 68 °C (henceforth called LongRange). The protocol was occasionally optimized by adding $MgCl_2$ and performing temperature gradients. A touch down amplification was programmed for the ITS region. Cycling conditions were 20 decreasing cycles (30 s at 94 °C, 30 s at the hybridization temperature and 45 s at 72 °C) for which the hybridization temperature was 62 °C at the first cycle and then decreased by 0.6 °C every cycle until it reached 50 °C. This process eliminates poor-quality strands during the amplification. This protocol was achieved by 20 normal cycles at a hybridization temperature of 50 °C to amplify all the remaining strands. For trnH-psbA, cycling conditions involved template denaturation at 80 °C for 5 min, followed by 35 cycles of denaturation at 94 °C for 30 s, primer annealing at 56 °C for 30 s and primer extension at 72 °C for 1 min, followed by a final extension step of 10 min at 72 °C. Finally, for rbcLa, a first denaturation stage at 95 °C for 1 min was followed by 35 amplification cycles (1 min of denaturation at 95 °C, 30 s of hybridization at 50 °C and 1 min of strand polymerization at 72 °C) and 7 min at 72 °C to achieve the polymerization of all the strands. In a few cases, we used more than one tissue sample to generate sequences for a species.

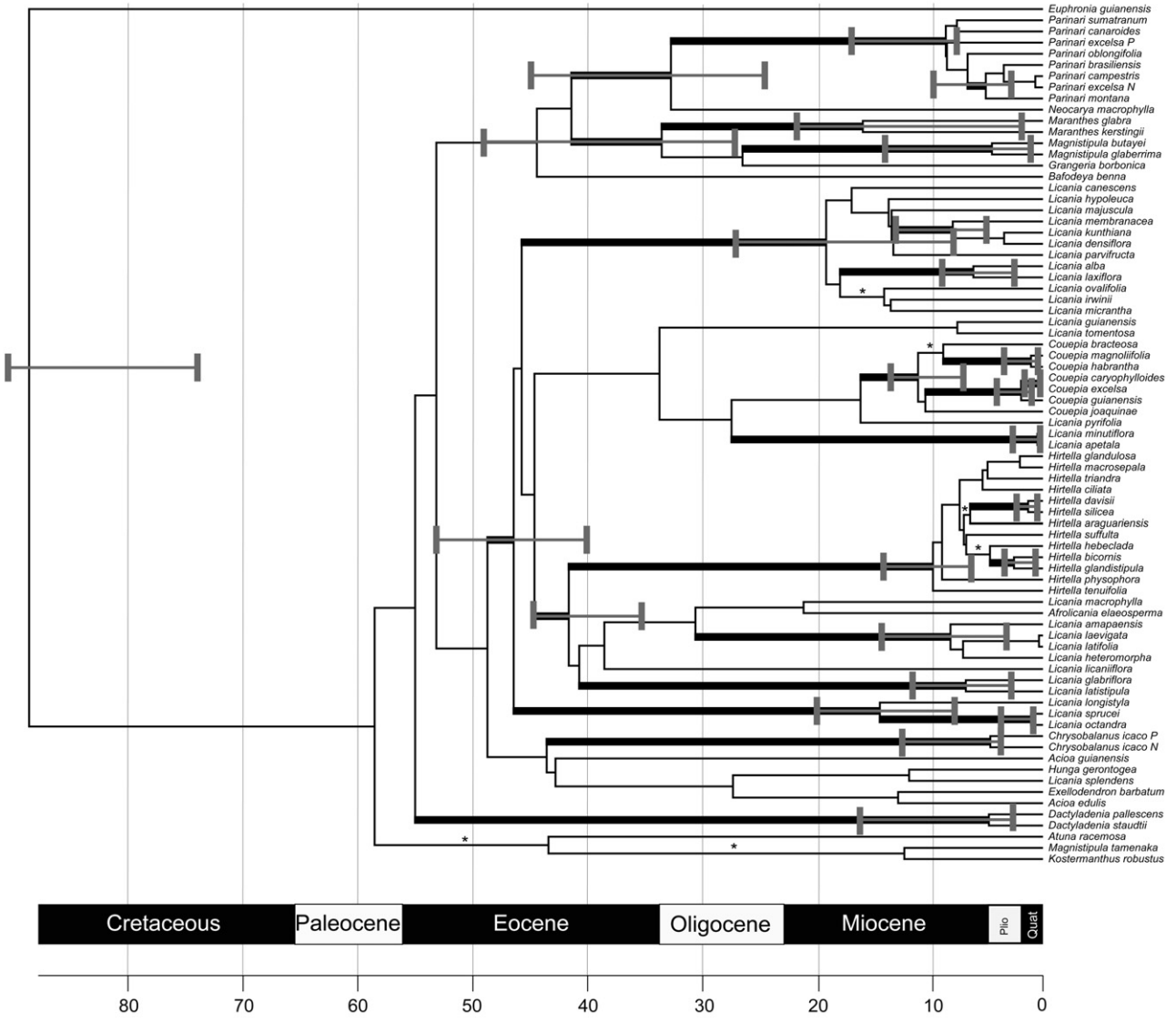
APPENDIX 3

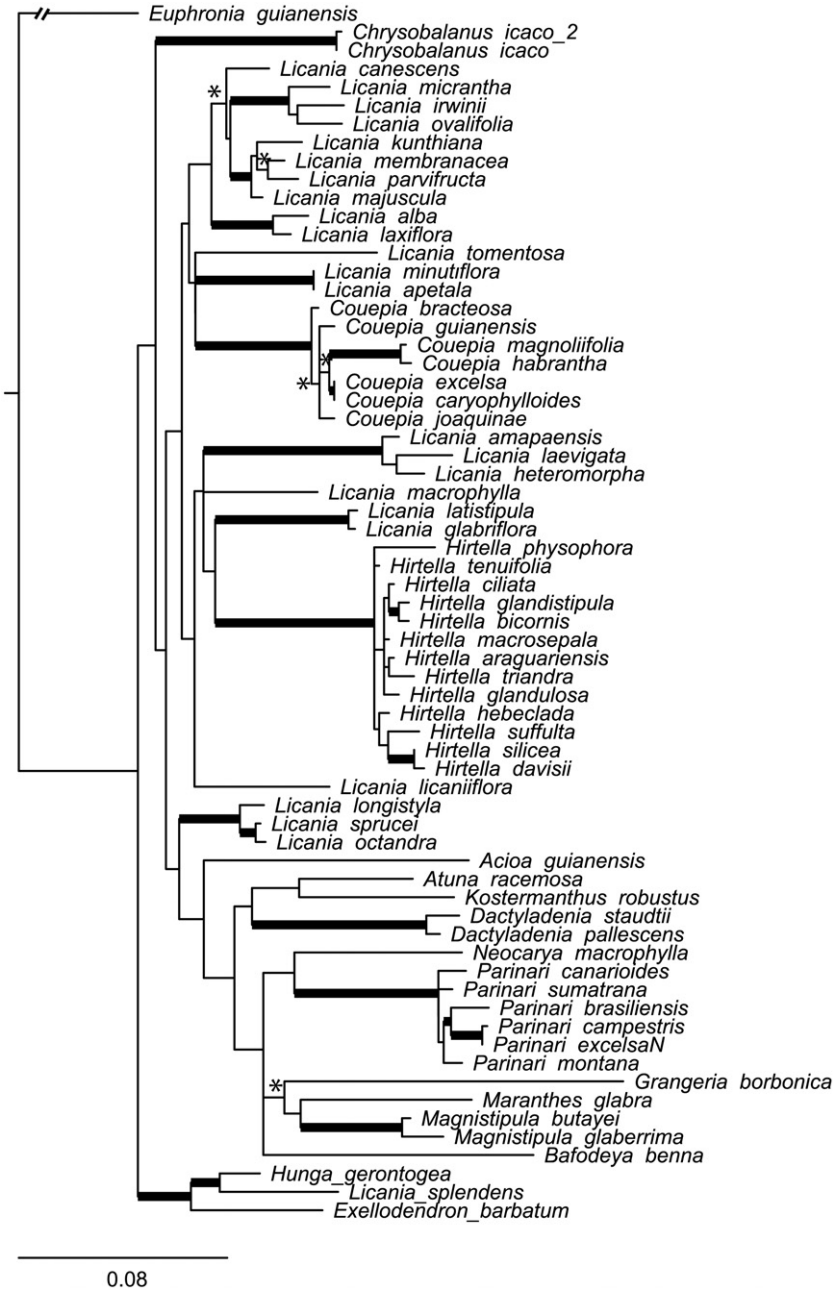
Amplified region names, associated primer names and sequences, amplification protocols used and references for primer sequences.

Region	Characteristics	Primer name and sequence (5'–3')	Amplification protocol	Reference
rpl14-rps8-infA-rpl36	Plastid intergenic spacer	**rpl14:** GGRTTGGAACAAATTACTATAATTCG **rpl36:** AAGGAAATCCAAAAGGAACTCG	LongRange	Shaw *et al*. (2007)
psbD-trnT	Plastid intergenic spacer	**psbD:** CTCCGTARCCAGTCATCCATA **trnT (GGU)-R:** CCCTTTTAACTCAGTGGTAG	LongRange	Shaw *et al*. (2007)
psaI-accD	Plastid intergenic spacer	**accD:** AATYGTACCACGTAATCYTTTAAA **psaI-75R:** AGAAGCCATTGCAATTGCCGGAAA	LongRange	Shaw *et al*. (2007)
atpI-atpH	Plastid intergenic spacer	**atpI:** TATTTACAAGYGGTATTCAAGCT **atpH:** CCAAYCCAGCAGCAATAAC	LongRange	Shaw *et al*. (2007)
rps16-trnK	Plastid intergenic spacer	**rpS16x2F2:** AAAGTGGGTTTTTATGATCC **trnK (UUU) x1:** TTAAAAGCCGAGTACTCTACC	LongRange	Shaw *et al*. (2007)
ndhA intron	Plastid intron	**ndhAx1:** GCYCAATCWATTAGTTATGAAATACC **ndhAx2:** GGTTGACGCCAMARATTCCA	LongRange	Shaw *et al*. (2007)
psbJ-petA	Plastid intergenic spacer	**psbJ:** ATAGGTACTGTARCYGGTATT **petA:** AACARTTYGARAAGGTTCAATT	LongRange	Shaw *et al*. (2007)
ndhJ-trnF	Plastid intergenic spacer	**ndhJ:** ATGCCYGAAAGTTGGATAGG **TabE:** GGTTCAAGTCCCTCTATCCC	LongRange	Shaw *et al*. (2007)
trnH-psbA	Plastid intergenic spacer	**trnH RKr:** ACTGCCTTGATCCACTTGGC **psbA FKr:** CGAAGCTCCATCTACAAATGG		Hamilton (1998)
matK	Plastid gene	**1R_KIM:** ACCCAGTCCATCTGGAAATCTTGGTTC **3F_KIM:** CGTACAGTACTTTTGTGTTTACGAG		Dunning & Savolainen (2010)
rbcLa	Plastid gene	**rbcL 1F:** ATGTCACCACAAACAGAAAC **rbcL 724R:** TCGCATATGTACCTGCAGTAGC		Savolainen *et al*. (2000)
ITS	Ribosomal intergenic spacer	**ITS juliette1f:** AGTGTTCGGATCGCGC **ITS juliette 1r:** GCCGTTACTAGGGGAATCCT		Unpublished Unpublished

APPENDIX 4

Phylogenetic tree for Chrysobalanaceae obtained with the ITS dataset inferred from a maximum likelihood method (RaxML). Branches supported by bootstrap values > 80% are thick and those with bootstrap values between 60 and 80% are indicated with a star.





APPENDIX 5

Dated phylogenetic tree obtained with the full dataset and inferred using Bayesian phylogeny reconstruction software (BEAST version 1.6), with the 95% highest posterior distributions (HPDs) for ages. Branches supported by posterior probabilities > 0.95 are thick and those with posterior probabilities between 0.85 and 0.95 are indicated with a star.