Journal of Biogeography (J. Biogeogr.) (2013) 40, 676–692

SPECIAL ISSUE

Phylogeography of a species complex of lowland Neotropical rain forest trees (*Carapa*, Meliaceae)

Caroline Scotti-Saintagne[1]*, Christopher W. Dick[2,3], Henri Caron[4], Giovanni G. Vendramin[5], Erwan Guichoux[4], Anna Buonamici[5], Caroline Duret[1], Pierre Sire[1], Renato Valencia[6], Maristerra R. Lemes[7†], Rogério Gribel[7†] and Ivan Scotti[1]

[1]*INRA UMR 0745 EcoFoG ('Ecologie des forêts de Guyane'), Campus Agronomique, BP709 – 97387 Kourou, French Guiana, France,* [2]*Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA,* [3]*Smithsonian Tropical Research Institute, Gamboa, Republic of Panama,* [4]*INRA, Université de Bordeaux, UMR1202 BIOGECO, F-33610 Cestas, France,* [5]*Istituto di Genetica Vegetale, Sezione di Firenze, Consiglio Nazionale delle Ricerche, 50019 Sesto Fiorentino, Firenze, Italy,* [6]*Laboratorio de Ecología de Plantas y Herbario QCA, Escuela de Ciencias Biológicas, P. Universidad Católica del Ecuador, Apartado 17-01-2184, Quito, Ecuador,* [7]*Laboratório de Genética e Biologia Reprodutiva de Plantas, Instituto Nacional de Pesquisas da Amazônia, Av. André Araújo, 2936, 69083-000 Manaus, AM, Brazil*

*Correspondence: Caroline Scotti-Saintagne, INRA UMR 0745 EcoFoG ('Ecologie des forêts de Guyane'), Campus Agronomique, BP709 – 97387 Kourou, French Guiana, France.
E-mail: caroline.scotti@ecofog.gf
†Present address: Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, Rua Pacheco Leão 915, Jardim Botânico, 22460-030, Rio de Janeiro, RJ, Brazil.

ABSTRACT

Aim Many tropical tree species have poorly delimited taxonomic boundaries and contain undescribed or cryptic species. We examined the genetic structure of a species complex in the tree genus *Carapa* in the Neotropics in order to evaluate age, geographic patterns of diversity and evolutionary relationships, and to quantify levels of introgression among currently recognized species.

Location Lowland moist forests in the Guiana Shield, the Central and Western Amazon Basin, Chocó and Central America.

Methods Genetic structure was analysed using seven nuclear simple sequence repeats (nuSSR), five chloroplast SSRs (cpSSR), and two chloroplast DNA (cpDNA) intergenic sequences (*trn*H–*psb*A and *trn*C–*ycf*6). Bayesian clustering analysis of the SSR data was used to infer population genetic structure and to assign 324 samples to their most likely genetic cluster. Bayesian coalescence analyses were performed on the two cpDNA markers to estimate evolutionary relationships and divergence times.

Results Two genetic clusters (*nu_guianensis* and *nu_surinamensis*) were detected, which correspond to the Neotropical species *C. guianensis* (*sensu latu*) and *C. surinamensis*. Fourteen cpDNA haplotypes clustered into six haplogroups distributed between the two nuclear genetic clusters. Divergence between the haplogroups was initiated in the Miocene, with some haplotype structure evolving as recently as the Pleistocene. The absence of complete lineage sorting between the nuclear and chloroplast genomes and the presence of hybrid individuals suggest that interspecific reproductive barriers are incomplete. NuSSR diversity was highest in *C. guianensis* and, within *C. guianensis*, cpDNA diversity was highest in the Central and Western Amazon Basin. Regional genetic differentiation was strong but did not conform to an isolation-by-distance process or exhibit a phylogeographical signal.

Main conclusions The biogeographical history of Neotropical *Carapa* appears to have been influenced by events that took place during the Neogene. Our results point to an Amazonian centre of origin and diversification of Neotropical *Carapa*, with subsequent migration to the Pacific coast of South America and Central America. Gene flow apparently occurs among species, and introgression events are supported by inconsistencies between chloroplast and nuclear lineage sorting. The absence of phylogeographical structure may be a result of the ineffectiveness of geographical barriers among populations and of reproductive isolation mechanisms among incipient and cryptic species in this species complex.

Keywords
Amazonia, *Carapa*, historical biogeography, Mahogany family, molecular markers, Neogene diversification, phylogeography, Quaternary diversification, species complexes.

Journal of Biogeography

http://wileyonlinelibrary.com/journal/jbi
doi:10.1111/j.1365-2699.2011.02678.x
© 2012 Blackwell Publishing Ltd

INTRODUCTION

Historical biogeography has profited immensely from the use of genetic diversity patterns to infer past and present migratory and demographic patterns of species (e.g. Hewitt, 2000; Petit *et al.*, 2008). By reconstructing evolutionary relationships of lineages and estimating their divergence times, modern phylogeographical methods make it possible to infer the role of past events in shaping the current patterns of biodiversity (Excoffier, 2004). Moreover, the combination of multiple types of markers with variable mutation rates and modes of inheritance (e.g. nuclear and organellar markers) provides a means of separating the contribution of different events spanning broad time-scales (Petit *et al.*, 2005). At a shorter time-scale, this approach shows all its power in the analysis of demographic transitions in populations (Grivet *et al.*, 2009) and when investigating instances of introgression (Petit & Excoffier, 2009). More recently, the availability of analytical methods based on coalescent models (Kingman, 1982) and on the computational power of Bayesian approaches (Beaumont *et al.*, 2002) has enabled historical demography to be inferred with limited sampling of a species' genetic diversity. Such techniques are especially important for non-model species, such as tropical forest trees.

The nature of the ecological changes undergone by Neotropical forests during the glacial phases of the Pleistocene is controversial (Colinvaux *et al.*, 2000; Bush & de Oliveira, 2006; Rull, 2011), although these changes are thought to have been less severe than those in temperate forest ecosystems (Stebbins, 1974). Therefore, widespread tropical tree species may still contain genetic imprints of pre-Pleistocene events, such as the uplift of the northern Andes, the emergence of the Isthmus of Panama, tectonic arches, or the formation of major rivers such as the Amazon (e.g. Cavers *et al.*, 2003; Dick *et al.*, 2003; Hardesty *et al.*, 2010; Poelchau & Hamrick, 2013). For example, in central Panama alone over 400 tree species have geographical ranges that extend into the Amazon Basin (Dick *et al.*, 2005). If these species obtained their full geographical range prior to the Pliocene uplift of the northern Andean cordilleras, they may be useful for comparative studies of pre-Pleistocene phylogeography.

While the taxonomic diversity in tropical forests provides opportunities for pioneering research in phylogeography, it also presents some obstacles. Because many tropical taxa have been poorly studied, taxonomic confusion shrouds our understanding of many widespread taxa, which may contain cryptic species not yet described by systematists (e.g. Rosell *et al.*, 2010). Phylogeographical studies have to factor in the possibility that some genetic discontinuities may be caused by reproductive isolating mechanisms, rather than by geographical barriers or other causes. One approach to dealing with taxonomic uncertainty is to perform a 'blind population genetics' approach (Duminil *et al.*, 2006) in which Bayesian clustering and assignment methods are used to estimate the numbers of distinct genetic demes without prior taxonomic classification. Duminil *et al.* (2006) used this approach to assign trees of the genus *Carapa* Aubl. (Meliaceae) in French Guiana to two distinct genetic demes corresponding to the two *Carapa* species found in that region (*C. procera* and *C. guianensis*). The study also found evidence of introgression between the two recognized species.

We have expanded the study of *Carapa* to a broad Neotropical range, including Central America, the Amazon Basin and the Pacific coast (Chocó region) of Ecuador. At the time of our field collections, there were three recognized Neotropical species, one of which (*C. procera*) has an amphi-atlantic range that presumably originated from oceanic dispersal of the large, floating seeds. However, a recent phylogenetic analysis uncovered 27 species based on morphological data (11 in America and 16 in Africa), which will be described as new species (Kenfack, 2011; Kenfack & Perez, 2011; D. Kenfack, Harvard University, Cambridge, MA, pers. comm.). In the present study we have evaluated variation in nuclear microsatellite (nuSSR) markers, chloroplast sequences (cpDNA) and chloroplast microsatellites (cpSSR) from a continent-wide sample of Neotropical *Carapa* populations. Without prior taxonomic classifications, we aimed: (1) to identify the most likely number of distinct genetic clusters in our broad Neotropical survey; (2) to identify phylogeographic patterns within Neotropical *Carapa* sister species; (3) to determine if the historical biogeography of Neotropical *Carapa* species extends back earlier than the Quaternary; and (4) to evaluate evidence of introgression or incomplete lineage sorting among the putative species.

MATERIALS AND METHODS

Species description

The transatlantic tree genus *Carapa* belongs to the mahogany family (Meliaceae). Biogeographical studies indicate a West Gondwanan origin of the family in the Late Cretaceous [85–76 million years ago (Ma)], with a boreotropical and oceanic dispersal leading to the colonization of the Neotropics during the Palaeogene (Muellner *et al.*, 2006). The three recognized species of *Carapa* at the outset of this study were: *C. procera* DC. (range in the Guiana Shield and equatorial Africa); *C. guianensis* Aublet (Guiana Shield, Amazon Basin, Chocó and Central America); and *C. megistocarpa* A.H. Gentry & Dodson (endemic to the Chocó region) (Styles, 1981; Gentry & Dodson, 1988; Styles & White, 1991). As part of a recent global taxonomic revision of the genus, Kenfack (2011), in agreement with Noamesi (1958), elevated the Neotropical *C. procera* to species status. This taxon is now called *C. surinamensis*, while the African taxon retains the name *C. procera*. All the samples collected for the present study were initially identified as *C. guianensis* or *C. surinamensis*.

Carapa surinamensis and *C. guianensis* occur in lowland tropical forests with an annual rainfall > 3000 mm. They are canopy trees [up to 60 m in height and 2 m diameter at breast height (d.b.h.)], exploited (under the same common species

© 2012 Blackwell Publishing Ltd

name) for timber used in construction, veneer, flooring, high-quality furniture and shipbuilding. They can be a market substitute for mahogany. In addition, the oil extracted from the seeds is highly valued for its medicinal and cosmetic properties. All species are monoecious and pollinated by insects such as stingless bees, beetles and moths (Cloutier *et al.*, 2007a). Genetic studies of *C. guianensis* have demonstrated high levels of outcrossing (Hall *et al.*, 1994; Cloutier *et al.*, 2007b), and in a Brazilian population pollen dispersal distances ranging from 70 to 250 m (Cloutier *et al.*, 2007b). The seeds of all *Carapa* typically drop below the mother tree and are secondarily dispersed by scatter-hoarding rodents (Forget *et al.*, 1999). Previous population genetic studies of *Carapa* species revealed a strong regional structure in the Amazon Basin in the chloroplast genome (cpDNA), also suggesting limited gene flow by seeds (Cloutier *et al.*, 2005). However, seeds are adapted to float in water and may occasionally be dispersed over long distances along rivers and in flooded forests (Scarano *et al.*, 2003).

Sampling

This study is based on the sampling of 482 adult *Carapa* trees from Central America, coastal Ecuador, the Guiana Shield and the Amazon Basin (see Appendix S1 in Supporting Information). Of these 482 samples, 422 consisted of fresh or silica-dried leaf and cambium tissue obtained from trees in the field. A subset of the trees collected in French Guiana were not flowering at the time of collection and could not be identified to species level, because the two species, which are sympatric in the Guiana Shield, do not exhibit strong phenotypic divergence except for floral traits (Styles, 1981); the unidentified trees were treated as *Carapa* spp. For the other sites (Brazil, Costa Rica, Ecuador and Suriname), the species were assigned using the taxonomic key of the *Flora Neotropica* (Styles, 1981). In addition, we obtained DNA from 60 herbarium specimens of *Carapa* (53 of *C. guianensis* and seven of *Carapa* spp.) from Cayenne (IRD, French Guiana), Manaus (EMBRAPA and INPA, Brazil), Oxford (University Herbaria, UK) and Utrecht (National Herbarium, The Netherlands). All *C. guianensis* herbarium samples except one (from Suriname) were from sites where only *C. guianensis* has been collected by botanists (*Flora Neotropica*), and therefore no ambiguities were expected in voucher identification; in these cases the given identification was accepted.

For the purposes of this study, sampling sites were grouped into five regions: (1) the Central Amazon Basin was represented by three sites in Brazil (Tefe, Manaus, Tapajos); (2) Central America was sampled at one site in Costa Rica (Ladrillera); (3) the Guiana Shield was represented by 30 sites in French Guiana, one site in Venezuela and one site in Suriname; (4) the Pacific coast of South America (Chocó Region) was sampled at three sites in Ecuador (Esmeraldas, Pichincha, Manabí); and (5) the Western Amazon was represented by one site in Brazil (Boca do Acre). For the genetic diversity analysis at nuSSRs (324 individuals), on average 62 samples per region were collected (between 16 and 192). For cpDNA analysis (329 individuals), an average coverage of 4.5 individuals per site (between one and 28 individuals) was obtained from 74 sites across the five regions. For cpSSR analyses (166 individuals), an average of 13.4 individuals per site (between one and 48) were obtained from 19 sites.

Laboratory methods

DNA extraction

Dried cambium discs and leaf samples were flash-frozen in liquid nitrogen and later ground to a powder using a pestle and mortar. DNA extractions followed Colpaert *et al.* (2005), except for herbarium specimens, which used the Invisorb Spin Plant Mini Kit (Invitek, Berlin, Germany).

Genetic markers

Details on the markers used are given in Table 1. Nuclear genome genetic diversity was analysed at seven nuSSR loci (Dayanandan *et al.*, 1999; Vinson *et al.*, 2005). Amplification conditions are as reported in Duminil *et al.* (2006). In the cp genome, two intergenic regions, *trn*H–*psb*A and *trn*C–*ycf*6, were analysed, as described in Shaw *et al.* (2005). Sequences were aligned and edited using CodonCode Aligner 3.5.7 (Codoncode Corporation, Dedham, MA, USA). We developed an R script (strand_merger_v0.4, available from the authors) to merge the sequences obtained from the same individual and to highlight ambiguities if differences occurred between sequences. Ambiguities were finally checked and corrected by hand with CodonCode Aligner 3.5.7. Sequences of individuals used to define haplotypes have been deposited in GenBank under accession numbers JN122293–JN122320.

In addition to cpDNA sequences, seven cpSSRs were genotyped to estimate the genetic diversity at fast-evolving cp regions. Amplification conditions were as reported in Weising & Gardner (1999). Amplified products were separated on a 96-capillary automatic sequencer MegaBACE 1000 (GE Healthcare, Madison, WI, USA). Analysis of fragment lengths was performed using the software Fragment Profiler 1.2 (GE Healthcare).

Cluster determination

Nuclear marker-based population structure was inferred using the genetic clustering algorithm implemented in structure 2.3.1 (Pritchard *et al.*, 2000). We selected individuals with genotype information for at least five out of the seven loci (324 individuals) and assumed a population admixture model and correlation between allele frequencies among populations owing to migration or shared ancestry. We ran the model either with or without site-of-origin information as prior

© 2012 Blackwell Publishing Ltd

Table 1 List of the chloroplast intergenic sequences (cpDNA), chloroplast simple sequence repeats (cpSSR) and nuclear simple sequence repeats (nuSSR) analysed in the tree genus *Carapa* in the Neotropics. Amplification conditions are described in the cited references. Size (bp) is the fragment size range in base pairs.

	Name	References*	Size (bp)	Number of different alleles†
CpDNA	*trn*H–*psb*A	a	537	13 haplotypes
sequences	*trn*C–*ycf*6	a	698	7 haplotypes
CpSSR	*Ccmp2*	b	205–209	3
	Ccmp3	b	85–94	2
	Ccmp4	b	122–194	3
	Ccmp5	b	93–94	2
	Ccmp6	b	75–143	12
	Ccmp7	b	126–144	7
	Ccmp10	b	105–111	3
nuSSR	*Cg01*	c	184–212	7
	Cg5	d	166–196	15
	Cg6	d	103–123	15
	Cg7	d	215–263	20
	Cg1l	c	156–172	8
	Cg16	c	97–140	14
	Cg17	c	102–132	16

*a, Shaw *et al.* (2005); b, Weising & Gardner (1999); c, Vinson *et al.* (2005); d, Dayanandan *et al.* (1999).

†Number of haplotypes (cp DNA sequences) or number of alleles (cpSSRs and nuSSRs).

(Hubisz *et al.*, 2009). To estimate the number of ancestral populations (K), we ran a Markov chain Monte Carlo (MCMC) (burn-in length = 10,000, post-burn-in length = 100,000) for K varying between 1 and 10. Five independent runs were performed to estimate the mean and standard deviation of data probability conditional on each K. The most likely number of structured populations ($\widehat{K}$) was then inferred from the estimation of ΔK (Evanno *et al.*, 2005). We used the clumpp algorithm (Jakobsson & Rosenberg, 2007) to summarize the five simulations and to estimate cluster membership coefficients for each individual, and we used distruct 1.1 (Rosenberg, 2004) to plot them.

Although linkage disequilibrium between single nucleotide polymorphisms (SNPs) in the chloroplast region precludes inference of population clustering as a consequence of past demographic events, structure can be used to detect the structure of genetic lineages in the cpDNA haplotype tree (Pritchard *et al.*, 2000). We used the USEPOPINFO option by setting the haplogroup's membership at samples with complete SNP information. We then ran structure assuming the same ancestry and allele frequency models as for the nuSSR loci and using the same run length. Only the individuals having a probability threshold of haplogroup membership higher than 0.9 were kept for further analyses. Among those samples, we finally assigned haplotype identification to the subset of individuals without missing data to discriminate haplotypes within haplogroups.

Phylogeny and divergence time dating

The concatenated sequence of *trn*H–*psb*A and *trn*C–*ycf*6 was used, with insertions–deletions (INDELS) recoded as SNPs. Unrooted phylogenetic relationships of haplotypes were reconstructed using statistical parsimony (Templeton *et al.*, 1992) implemented in tcs 1.21 (Clement *et al.*, 2000), by setting the probability of parsimony to 0.95. In addition, we built a phylogenetic tree using beast 1.5.4 (Drummond & Rambaut, 2007) to reconstruct haplotype phylogeny and estimate divergence times. The two outgroup taxa used to root the phylogenetic tree were *Swietenia macrophylla* and *Cedrela odorata* (Meliaceae). The best-fit model of nucleotide substitution was inferred using the Akaike information criterion (AIC) implemented in the program Modeltest 3.8 (Posada & Buckley, 2004), with the starting tree obtained using the neighbour-joining algorithm in paup* 4.10b (Swofford, 2003). After determining the best evolutionary model (TVM + I) we tested the null hypothesis that the data evolved under a molecular clock by performing a likelihood-ratio test (LRT) for the set of 16 sequences (14 haplotypes and 2 outgroups). The LRT did not reject the null hypothesis of rate constancy. The likelihood score of the model using the molecular clock was significantly higher than that for the model without the molecular clock ($-2 \times \Delta LOD = -37.8$, d.f. = 14). Results remained unchanged ($-2 \times \Delta LOD = -33.6$, d.f. = 14) when INDELS were removed. Despite a higher likelihood for the model of rate constancy, we compared the phylogenetic reconstruction using both a strict and a relaxed molecular clock model assuming a Yule speciation process as tree prior and the general time-reversible model (GTR + I), which was the substitution model closest to the one selected by Modeltest. We used the relaxed molecular clock model in beast, with an exponential prior distribution for the molecular rate; this implies that changes occurred at the nodes, with the size of the change being independent of the branch length (Drummond *et al.*, 2006). The tree root prior was set as a uniform distribution varying between 67.5 and 72 Ma, which corresponds to the estimated age of the ancestor of the Cedreloideae clade (grouping the *Carapa*, *Cedrela* and *Swietenia* genera) (Muellner *et al.*, 2006). The MCMC was run for 10,000,000 and 30,000,000 generations for the strict and relaxed molecular clock models, respectively, and sampled every 1000th generation with the first 10% removed as burn-in. The program tracer 1.5 (Drummond & Rambaut, 2007) was used to check the convergence of the MCMC searches and to check whether the effective samples sizes (ESS) of the posterior probability were higher than 100 for each estimated parameter. The information contained in sampled trees (9000 and 27,000 for the strict and the relaxed model, respectively) was summarized using the TreeAnnotator program distributed with beast 1.5.4 (Drummond & Rambaut, 2007). This algorithm selects the maximum clade credibility tree and annotates it with the posterior summaries calculated for the nodes that have a posterior probability > 0.95. Finally, we

© 2012 Blackwell Publishing Ltd

used FigTree (available at http://tree.bio.ed.ac.uk/software) to view, edit and export the annotated tree.

Population genetic structure

Arlequin 3.1 (Excoffier *et al.*, 2005) was used to analyse both nuclear and chloroplast genetic diversity. Haplotype richness and haplotype diversity (h) were estimated for chloroplast markers; number of different alleles (N_A) and expected heterozygosity (H_E) were estimated for nuclear markers. To compare indices of genetic diversity between populations despite different sample sizes, we computed, for each marker type, the confidence interval of h and H_E as described in Grundmann *et al.* (2001), and estimated the allelic richness (N_A) after a rarefaction procedure using the rarefac program (Petit *et al.*, 1998). Deviation from Hardy–Weinberg (H–W) equilibrium was first tested at the regional level within putative species. However, because of the presence of sites sampled over long geographical distances in a given region, we could not exclude a deviation from H–W equilibrium resulting from population structure (Wahlund effect). Thus, we also performed H–W tests in a limited number of locations with a sufficiently large sample size. The presence of null alleles was tested using micro-checker 2.2.3 (van Oosterhout *et al.*, 2004). For loci with null alleles, the including null alleles (INA) method (Chapuis & Estoup, 2007), implemented in freena, was applied to provide a genotype data set corrected for null alleles to re-estimate genetic differentiation (F_{ST}). Pairwise F_{ST} and R_{ST} (genetic differentiation based on molecular distance between alleles assuming a stepwise mutation model; Slatkin, 1995) between populations within species and between populations belonging to different species were estimated based on 1000 permutations using Arlequin 3.1. A hierarchical analysis of molecular variance (AMOVA) was performed for each of the two measures of differentiation to estimate the partition of genetic variance among species, among populations within species, and within populations. In the same way, for cpDNA, genetic structure was inferred considering both haplotype frequencies (F_{ST}) and molecular distances between haplotypes (N_{ST}). For this purpose, a matrix of genetic distances was generated based on the number of mutational steps between haplotypes. The significance of the variance components was tested by a permutation procedure (Excoffier *et al.*, 1992). To test whether the distribution of cpDNA genetic diversity reveals a phylogeographical pattern within species, we used the procedures proposed by Pons & Petit (1996) based on the expectation that N_{ST} must not be different from the coefficient of population differentiation G_{ST} (analogue to F_{ST}) if there is no relationship between the phylogenetic and geographical proximity of alleles. The significance of the test was obtained based on 10,000 permutations using SPAGeDi 1.4 (Hardy & Vekemans, 2002). Finally, to test for the hypothesis of isolation-by-distance (IBD) to explain genetic differentiation between populations of *Carapa* we performed a Mantel test (Mantel, 1967) with genetix 4.05.2 (Belkhir *et al.*, 2004) between the matrix of genetic distances calculated with Arlequin 3.1 (F_{ST} and N_{ST}) and the matrix of geographical distances calculated by the Geographic Distance Matrix Generator 1.2.3 (Ersts, 2010).

RESULTS

Cluster determination

The cluster analysis indicated the highest probability for $K = 2$ genetic populations ($\Delta K = 1142.45$ and $\Delta K = 121.82$ respectively without and with site information as prior). The two populations were composed respectively of 118 and 155 individuals assigned with a coefficient of membership higher than or equal to 0.80 (Table 2). Among those individuals, a subset of samples from French Guiana was also used by Duminil *et al.* (2006) and showed a clear concordance between genetic clusters and diagnostic morphological characters for *Carapa* species. Samples belonging to the *C. surinamensis* genetic cluster in Duminil *et al.* (2006) grouped in the present study with the cluster containing 118 individuals, whereas samples belonging to the *C. guianensis* genetic cluster in Duminil *et al.* (2006) grouped with the second cluster. Hereafter the two clusters are respectively named '*pure_surinamensis*' and '*pure_guianensis*' to recall that the delimitation of the two 'species groups' was based on assignment probabilities close to unity at the nuSSR loci (Fig. 1a). Among the remaining 51 admixed individuals, 16 had the same probability ($0.4 \le P < 0.6$) of being assigned to either of the two clusters. These probable first-generation hybrids were discarded from intraspecific analyses. The remaining 36 hybrid individuals, having a probability ranging from 0.6 to 0.8 of being assigned to one or the other of the two clusters, may be the result of older hybridization events or recurrent backcrosses (Lepais *et al.*, 2009). These samples (called '*hybrid_BC_surinamensis*' and '*hybrid_BC_guianensis*', respectively) were added to their closest genetic clusters (respectively 9 samples were added to the cluster of *pure_surinamensis* and 27 samples to the cluster *pure_guianensis*) for subsequent intraspecific population-genetic analyses; the merger of the 'pure' groups with the 'hybrid_BC' samples will be respectively defined as '*nu_surinamensis*' and '*nu_guianensis*' (the '*nu*' prefix recalls that the classification is based on nuclear genetic markers) (Fig. 1a). However, the distinction was kept to characterize the hybridization events and to analyse introgression events (see below). The geographical distribution of the two putative species is given in Fig. 1(b). Trees belonging to the *nu_guianensis* cluster were observed in all five sampled Neotropical regions, whereas trees belonging to *nu_surinamensis* displayed a narrow distribution in the Guiana Shield and at the Manaus site in the Amazon Basin. In the same way, trees having an intermediate nuclear genome between the two defined clusters, here termed 'putative hybrid_F1 samples', were located only in French Guiana (11) and Ecuador (4).

To minimize confusion between interspecific and intraspecific evolutionary processes and patterns, subsequent analyses were performed by grouping individuals based on Bayesian clustering, and genetic diversity was analysed within each

© 2012 Blackwell Publishing Ltd

Table 2 Results of the blind population genetic survey using seven nuclear simple sequence repeat (nuSSR) loci in the tree genus *Carapa* in the Neotropics. The model with two clusters ($K = 2$) gave the best probability ($\Delta K = 1142.45$). For each sampled site (40 in total), the number of individuals assigned to one of the two clusters is given (*pure_surinamensis*, *pure_guianensis*), as well as the number of unassigned individuals. The levels of hybrid ancestry (first generation, *Hybrid_F1* or backcrosses, *BC*) are given according to the value of the membership coefficient (Q) to each group. In total, five classes are reported.

Region	Country	Site ID	Site	Q1 > 0.8	0.6 > Q1 > 0.8	0.4 > Q1 > 0.6	0.6 > Q2 > 0.8	Q2 > 0.8
				Pure_ surinamensis	*Hybrid_BC_ surinamensis*	*Hybrid_ F1*	*Hybrid_BC_ guianensis*	*Pure_ guianensis*
Central America	Costa Rica	1	Ladrillera				1	15
Pacific coast	Ecuador	3	Esmeraldas			1	3	43
		4	Manabí				2	4
		6	Pichincha	1		3	4	3
Western Amazon	Brazil	7	Boca do Acre					23
Central Amazon	Brazil	8	Tefé					4
		9	Manaus	2	3			
		10	Tapajos					9
Guiana Shield	Venezuela	11	Upata				1	4
		12	Imataca					1
	Suriname	13	Brownsberg	5				
	French Guiana	14	Chutes Voltaire	3				
		15	Bafog	9	2			
		17	Crique Valentin	4	1			
		18	Mt de Fer	5				
		19	Degrad Florian	4				
		20	Counami	6				
		21	Trou poisson	5				
		22	Piste de St Elie	5				
		23	Paracou	10				
		24	Maman Lezard	2				
		25	Mt Plomb	3				
		26	Petit saut	8				
		27	Crique Plomb	1				
		28	Mt des Singes	7		2		
		29	Risquetout	6				1
		31	Rorota		1	1	3	6
		32	Mirande				1	6
		33	Fourgassié					9
		34	Kaw					10
		35	Cacao				1	2
		36	Tibourou	6			2	
		37	Mt Tortue	9	1	5	8	1
		38	Regina			2		3
		39	Pont de Regina	8	1	1		1
		40	Parc de Regina					5
		41	Saut Maripa	1				5
		42	Nouragues	3				
		43	Saul	3				
		45	Mt Tumuc-Humac	2		1		
			Neotropics	118	9	16	26	155

cluster (*nu_guianensis/nu_surinamensis*, see above). Owing to the limited geographical distribution of the *nu_surinamensis* cluster, interspecific comparisons could only be carried out for the Guiana Shield. For the same reason, comparisons of genetic diversity between regions at the intraspecific level could only be carried out in the *nu_guianensis* cluster.

Genetic relationships between haplotypes – geographic distribution of cp lineages

The polymorphism detected in 329 samples of *Carapa* at two cpDNA regions (equivalent to 26 polymorphic sites over 1198 bp) is described in Appendix S2. A total of 17 SNPs and

© 2012 Blackwell Publishing Ltd

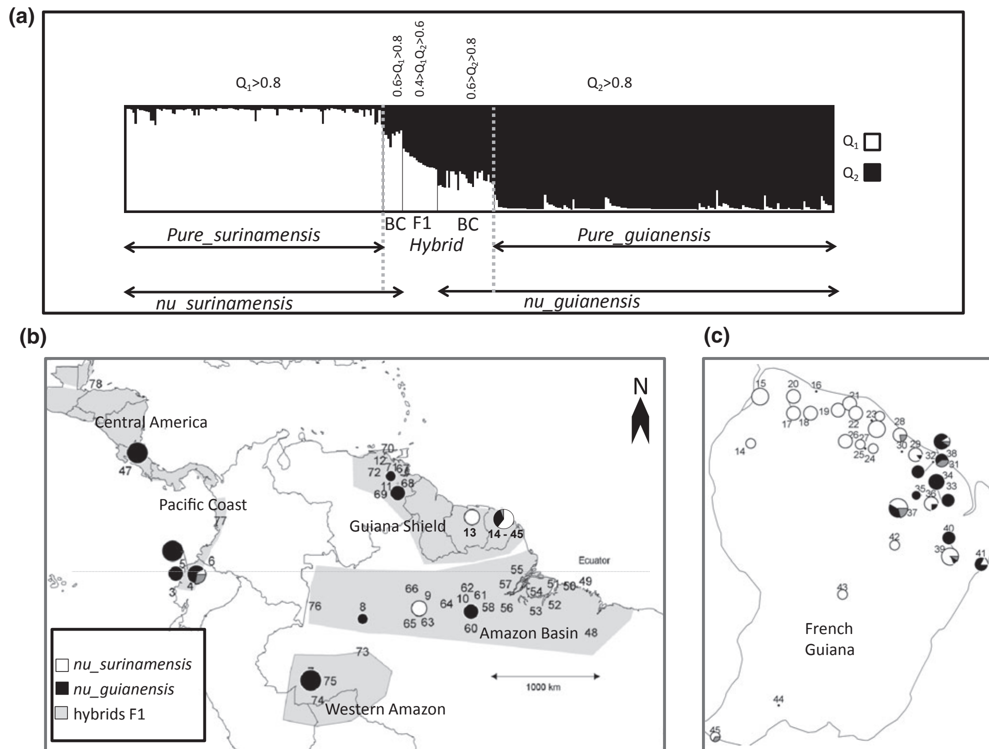

Figure 1 Bayesian clustering analysis for the tree genus *Carapa* in the Neotropics. (a) Summary plot of membership coefficient estimates (Q) derived from Bayesian clustering analysis. The 324 individuals are represented by a single vertical line broken into two shaded segments, with lengths proportional to the probability of assignment to inferred clusters Q1 and Q2. Individuals have been sorted by decreasing Q1 assignment probability value. The classification of the individuals as a function of their degree of admixture between the two clusters is reported [pure genetic cluster, putative hybrids of first generation (F1), and putative hybrids resulting from backcrosses (BC)]. This classification has been used to compare nuclear ancestry and chloroplast lineage membership. In addition, we have reported the classification (*nu_guianensis, nu_surinamensis*) used for the analysis of genetic diversity at the intraspecific level (hybrids resulting from backcrosses have been merged with their closest genetic cluster). Panels (b) and (c) show the geographical distribution of the *nu_surinamensis* and *nu_guianensis* clusters in the Neotropics (b) and French Guiana (c). The numbers (1–40) correspond to the site IDs (see Appendix S1 in Supporting Information).

two INDELS were observed in *trn*H–*psb*A (507 bp), and five SNPs and two INDELS in *trn*C–*ycf*6 (691 bp). In total, 14 haplotypes were identified; sequences have been deposited in GenBank and accession numbers are given in Appendix S2.

Six haplogroups emerged from the cladogram (Fig. 2a): G1 (five haplotypes: H1, H2, H3, H4, H5); G2 (two haplotypes: H6, H7); G3 (three haplotypes: H9, H10, H11); G4 (one haplotype, H14); G5 (two haplotypes: H12, H13); and G6 (one haplotype, H8). A total of 162 individuals without missing data for the 26 SNPs were used in STRUCTURE as reference samples to assist ancestry estimation for individuals with incomplete sequence information. The haplogroup membership of 167 additional samples was inferred with a probability equal to or higher than 0.9. Haplogroup and haplotype geographical distributions are displayed in Fig. 2(b). All herbarium samples clustered within the same haplogroups as the remaining *C. guianensis* samples. This confirmed the botanical identification of the herbarium samples selected for this study. The distribution of the cp lineages is geographically uneven: two haplogroups (G1 and G6) were observed in only one region (Guiana Shield and Central Amazon for G1 and G6, respectively); two haplogroups (G2 and G4) were observed in two regions (Guiana Shield and Central America for G2 and Central Amazon and Western Amazon for G4); one haplogroup (G5) was observed in three regions (Central and Western Amazon and the Pacific coast); finally, haplogroup G3 was present in four regions (absent from Western Amazon). The Central Amazon harbours the highest number of haplogroups, with G3, G4, G5 and G6 observed over 2600 km from the Amazonas State to the Maranhão state in Brazil. Interestingly, the two sites at the eastern (Amazon mouth) and western edge of the Central Amazon region harbour the same haplogroup (G5). In contrast to G5, the other haplogroups observed in the Central Amazon were not shared between sites (G3 in Tapajos, G4 in Tefe and G6 in Manaus). At the haplotype level, the highest diversity was observed in the Guiana Shield, with eight observed haplotypes, of which five (H1 to H5) belong to the haplogroup G1.

Dating the divergence of cp lineages

Figure 3 gives the phylogenetic tree (among 9000 posterior samples) that has the maximum sum of probabilities on its

© 2012 Blackwell Publishing Ltd

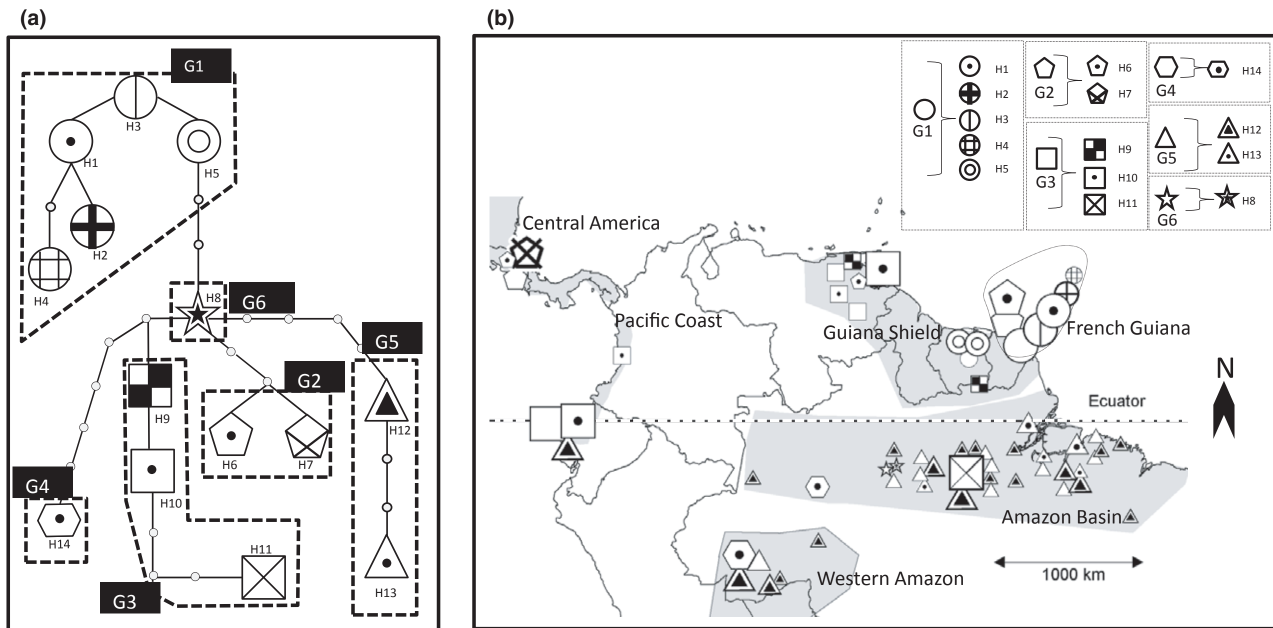

Figure 2 Geno-geographical relationships among chloroplast haplotypes of the tree genus *Carapa* in the Neotropics. Haplotypes belonging to the same haplogroup have the same shape, with different haplotypes within a haplogroup differentiated by shading. (a) Gene genealogy of 14 chloroplast (cp)DNA haplotypes (based on *trn*H–*psb*A and *trn*C–*ycf*6 sequences) observed in *Carapa*. The cladogram has been built using TCS 1.21 with parsimony probability set to 0.95. Insertion–deletions have been replaced by a single mutation event. In addition to the relationship between haplotypes, we have reported their haplogroup membership (from G1 to G6). (b) Geographical distribution in 78 sites of the Neotropics of 14 haplotypes belonging to six haplogroups observed at two cp sequences (*trn*H–*psb*A and *trn*C–*ycf*6). A total of 224 *Carapa* samples collected in South and Central America are reported. For 105 additional samples of *Carapa* for which we did not obtain haplotype identification (owing to missing data at some SNPs), we only report haplogroup membership (G1 to G6). Symbol size is proportional to the frequency of the haplotype (or haplogroup): small, 1 individual; medium, between 2 and 5 individuals; large, more than 5 individuals.

internal nodes for the *trn*H–*psb*A and *trn*C–*ycf6* sequences in *Carapa* using a strict molecular clock model and by recoding INDELS as SNPs. The phylogenetic trees obtained when INDELs were removed and when the relaxed molecular clock model was used are given in Appendix S2. The mean and the 95% credible interval for node age estimates are provided in Table 3. A comparison of the two models (fixed and relaxed molecular clock) showed conservatism in the topology of the trees but significant differences in the posterior summaries of node ages. Using the strict molecular clock, node ages displayed a lower mean (on average 2 times lower) and a lower standard error (on average 2.6 times lower) than using the relaxed molecular clock. The complete deletion of INDELS had no significant effect on the posterior summaries when the relaxed molecular clock was used but gave significantly higher mean node ages (on average 1.2 times higher) and significantly higher standard deviations (on average 1.5 times higher) when the strict molecular clock was used. The results suggest that diversification within these *Carapa* species began during the Miocene and has continued during the Quaternary. The first event was a split between the haplogroup G1 and the remaining five haplogroups [mean time varied between 13.6 and 36.6 Ma for the four tested models; extreme values of the 95% credible interval (CI) = 10.4–59.5 Ma; Table 3]. Diversification then continued during the Miocene (mean time between 9.41 and 26.2 Ma; extreme values of the 95% CI = 7.98–44.1 Ma) to give rise to the genetic lineages G2 to G6. Our analysis suggests that haplotypes began to diverge during the Miocene for G1 (mean time between 7.82 and 20 Ma; extreme values of the 95% CI = 3.26–38.7 Ma) and continued during the Pliocene and Pleistocene in haplogroups G2 (mean time varied between 1.70 and 7.73 Ma; extreme values of the 95% CI = 0.45–19.3 Ma) and G5 (mean time varied between 1.36 and 8.26 Ma; extreme values of the 95% CI = 0.67–19.3 Ma). The branches of the remaining haplogroups (G3, G4 and G6) did not have enough statistical support to estimate divergence times.

Hybridization events

The Western Amazon was the only region composed of trees belonging solely to one pure genetic cluster (all individuals had a probability $P > 0.8$ of being assigned to the *pure_guianensis* cluster) (Fig. 4a). For all the other regions we observed intermediate genotypes that probably represented first-generation hybrids and backcrossed individuals.

Chloroplast haplogroups only partially matched nuclear clusters (Fig. 4b). Two haplogroups (G4, G5) were exclusively associated with the *pure_guianensis* cluster. For the remaining haplogroups (G1, G2, G3, G6) the pattern was more complex.

© 2012 Blackwell Publishing Ltd

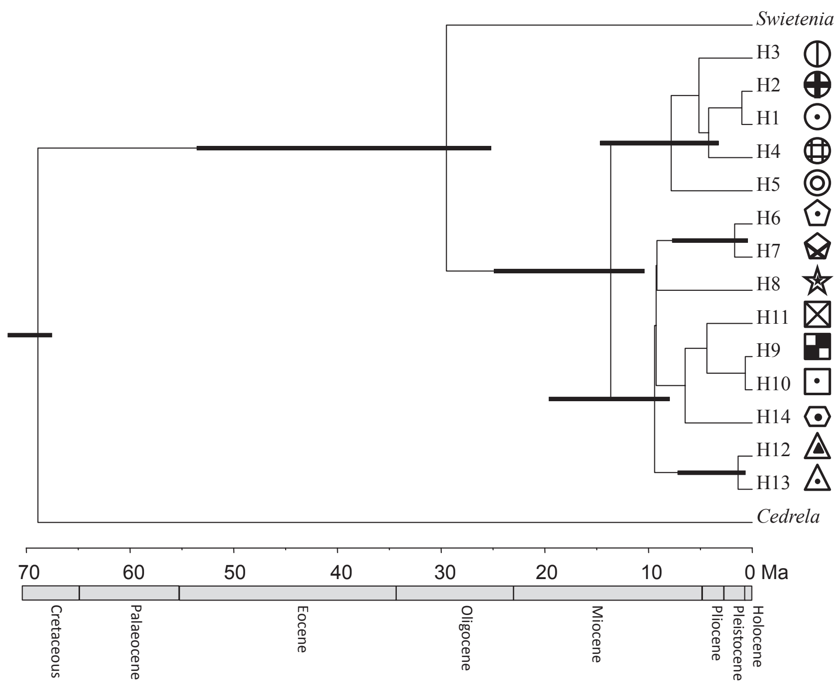

Figure 3 Phylogenetic tree for 14 haplotypes of the tree genus *Carapa* in the Neotropics and two outgroups (among 9000 posterior samples from 10,000,000 Markov chain Monte Carlo steps) that has the maximum sum of probabilities on its internal nodes for 1154 bp of the chloroplast (cp)DNA *trn*H–*psb*A and *trn*C–*ycf*6 intergenic regions using a strict molecular clock model and by recoding insertions–deletions (INDELS) as single nucleotide polymorphisms (SNPs). The phylogenetic trees obtained when INDELs were removed and when the relaxed molecular clock model was used are given in Appendix S2. Node bars display the lower and upper bounds of the 95% highest posterior density (HPD) interval of the node age in units of years. The posterior summaries have been calculated only for the nodes that have a posterior probability > 95%.

Table 3 Age estimates for key events in the genus *Carapa* in the Neotropics. The concatenated sequence of *trn*H–*psb*A and *trn*C–*ycf*6 was used, with insertions–deletions (INDELS) recoded as single nucleotide polymorphisms (with INDELS) or INDELS removed from the analysis (without INDELS). Divergence times [mean node ages in millions of years ago (Ma)] correspond to the mean of the highest posterior densities (HPD, clades that have a posterior probability ≥ 95%). The lower and upper 95% HPD of node ages are also given. Dates for the calibrated node are in bold.

	Strict molecular model					Relaxed molecular model				
	With INDELS		Without INDELS			With INDELS		Without INDELS		
Node	Mean node age (Ma)	95% HPD node age (Ma)	Mean node age (Ma)	95% HPD node age (Ma)	Geological period	Mean node age (Ma)	95% HPD node age (Ma)	Mean node age (Ma)	95% HPD node age (Ma)	Geological period
Cedreloideae	**68.9**	**67.5–71.8**	**69.7**	**67.5–71.8**	Late Cretaceous	**69.7**	**67.5–71.8**	**69.7**	**67.5–71.7**	Late Cretaceous
Carapa/*Swietenia*	29.5	25.2–53.5	n.s.	n.s.	Eocene–Oligocene	n.s.	n.s.	n.s.	n.s.	
Carapa	13.6	10.4–24.9	19.1	10.9–27.8	Oligocene–Miocene	36.6	15.9–59.5	34.9	14.3–56.8	Palaeocene–Miocene
G2,G3,G4,G5,G6	9.4	8–19.6	n.s.	n.s.	Miocene	26.2	10.6–44.1	n.s.	n.s.	Eocene–Miocene
G1 (H1 to H5)	7.8	3.3–14.6	n.s.	n.s.	Miocene–Pliocene	20	4.5–38.7	n.s.	n.s.	Eocene–Pliocene
G2 (H6, H7)	1.7	0.45–7.7	4.6	0.51–9.9	Miocene–Pleistocene	7	0.56–17.3	7.7	0.36–19.3	Miocene–Pleistocene
G3 (H9, H10, H11)	n.s.	n.s.	n.s.	n.s.		n.s.	n.s.	n.s.	n.s.	
G4 (H14)	n.s.	n.s.	n.s.	n.s.		n.s.	n.s.	n.s.	n.s.	
G5 (H12, H13)	1.4	0.7–7.2	5.1	0.78–10	Miocene–Pleistocene	7.3	0.74–16.9	8.3	0.76–19.3	Miocene–Pleistocene
G6 (H8)	n.s.	n.s.	n.s.	n.s.		n.s.	n.s.	n.s.	n.s.	

n.s., statistically not supported (posterior probability < 0.95).

Haplogroup G1 was associated at 85% with the cluster *pure_surinamensis*, and the remaining 15% were hybrids. Haplogroup G2 had a probability of 49% of being assigned to the cluster *pure_guianensis*. The number of hybrids was higher than for G1 (37%) and, more interestingly, the remaining individuals (14%, all located in French Guiana) were associated with the alternative cluster *pure_surinamensis*. Haplogroup G3 was associated at 70% with the cluster *pure_guianensis*, but a non-negligible fraction of individuals were admixed (26%, all located in Ecuador) and one tree belonged to the opposite cluster *pure_surinamensis*. Finally, one of the two individuals belonging to the haplogroup G6 was linked to the *pure_surinamensis* cluster and the other to the *hybrid_BC_surinamensis* cluster. In summary, G1 (linked to *nu_surinamensis*) was never observed in the nuclear genetic background of *nu-guianensis*, whereas G2 and G3 (linked to

© 2012 Blackwell Publishing Ltd

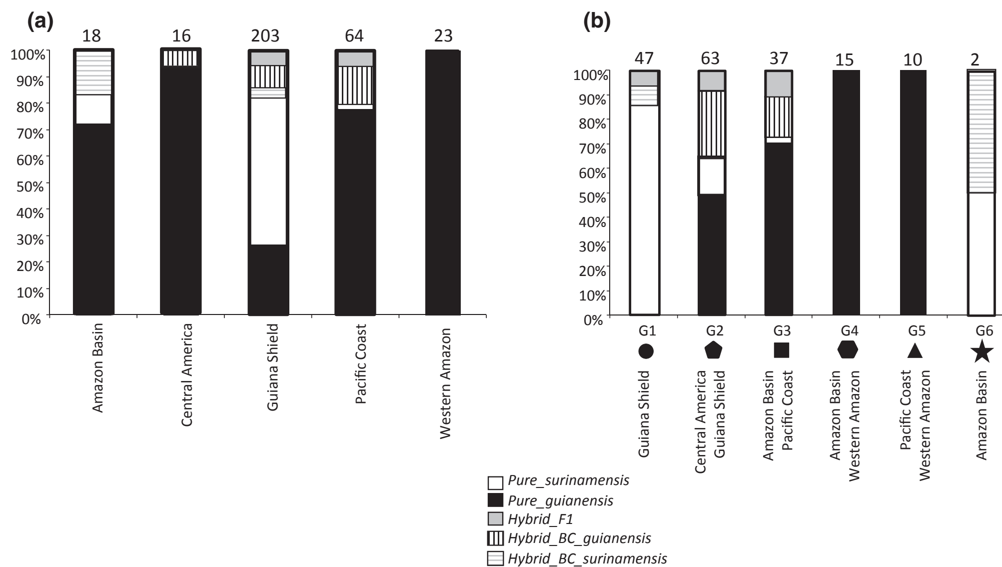

Figure 4 Comparison of nuclear and chloroplast patterns in the tree genus *Carapa* in the Neotropics. (a) Frequency by region of pure genetic clusters and hybrids as assigned by STRUCTURE (7 nuclear single sequence repeats; 324 individuals) in the tree genus *Carapa* in the Neotropics. (b) Frequency by nuclear genetic cluster of chloroplast genetic lineages defined as haplogroups (from G1 to G6) (n = 174). '*Pure_guianensis*' and '*Pure_surinamensis*' when Q2 > 0.8 and Q1 > 0.8, respectively; '*Hybrid_F1*' when 0.4 < Q1 < 0.6; '*Hybrid_BC_guianensis*' and '*Hybrid_BC_surinamensis*' when 0.6 < Q2 < 0.8 and 0.6 < Q1 < 0.8, respectively.

Table 4 Summary of genetic diversity within genetic clusters (*nu_guianensis* and *nu_surinamensis*) of the genus *Carapa* inferred by STRUCTURE for each Neotropical region averaged over seven nuclear simple sequence repeat (nuSSR) loci: number of sampled sites (S), number of gametes (n), number of different alleles (N_A), number of different alleles after rarefaction (N_{A_r}) and rarefaction size in parentheses (S_r), expected heterozygosity (H_E) and its standard error (SE), number of loci (out of seven) displaying a significant deviation (P < 0.05) from Hardy–Weinberg equilibrium (H–W).

Region	Countries	S	n	N_A	N_{A_r} (S_r)	H_E (SE)	H–W
nu_guianensis							
Neotropics	FG, VE, BR CR, EC	21	340.2	13.14	12.85 (240.4)	0.80 (0.04)	7
Central Amazon	BR	2	24.4	5.28	5.28 (24.4)	0.68 (0.05)	2
Central America	CR	1	31.7	4.71	4.39 (24.4)	0.42 (0.13)	0
Guiana Shield	FG, VE	14	130.6	9.43	6.38 (24.4)	0.67 (0.10)	5
Western Amazon	BR	1	42.3	4.57	3.82 (24.4)	0.40 (0.11)	2
Pacific coast	EC	3	105.14	8.14	5.69 (24.4)	0.64 (0.09)	5
nu_surinamensis							
Neotropics	FG, SU, BR	26	240.4	9	9 (240.4)	0.52 (0.11)	6
Guiana Shield	FG, SU	24	225.7	8.14	7.47 (130.6)	0.51 (0.12)	6

FG, French Guiana; SU, Suriname; BR, Brazil; VE, Venezuela; CR, Costa-Rica; EC, Ecuador.

nu_guianensis) were observed in the nuclear genetic background of *nu_surinamensis*.

Genetic diversity

On average, nuclear genetic diversity over seven SSR loci was higher in *nu_guianensis* than in *nu_surinamensis* (H_E = 0.8, SE = 0.037 vs. H_E = 0.52, SE = 0.11; Table 4), with five loci having significantly higher H_E (result not shown). This difference was maintained when we compared the genetic diversity estimated in the Guiana Shield, where both putative species have been sampled (H_E = 0.67, SE = 0.099 vs. H_E = 0.51, SE = 0.115; Table 4). At the intraspecific level, among the five regions analysed in the *nu_guianensis* cluster, Western Amazon and Central America harboured the lowest genetic diversity (H_E = 0.40 and 0.42, respectively; Table 4) with three and two loci, respectively (among seven), having an H_E significantly smaller than the remaining three regions (results not shown). When H_E was compared among sites no significant differences were observed (results not shown), meaning that the low genetic diversity in Central America and Western Amazon was related to the lower number of sampled

© 2012 Blackwell Publishing Ltd

Table 5 Summary of chloroplast (cp) diversity within genetic clusters (*nu_guianensis* and *nu_surinamensis*) of the Neotropical tree genus *Carapa* inferred by STRUCTURE. The following are reported for each sampling region for intergenic sequences and chloroplast simple sequence repeats (cpSSR): sample size *n*, number of different alleles (N_A), number of different alleles after rarefaction (N_{A_r}) and the rarefaction size in parentheses (*S*r), haplotypic diversity (*h*) and its confidence interval (CI) in brackets.

	Intergenic sequences*								cpSSR†			
	Haplogroup				Haplotype							
Region	*n*	N_A	N_{A_r} (Sr)	*h* [CI]	*n*	N_A	N_{A_r} (Sr)	*h* [CI]	*n*	N_A	N_{A_r} (Sr)	*h* [CI]
nu_guianensis												
Neotropics	105	4	4 (56)	0.67 [0.62–0.72]	72	6	5.62 (29)	0.77 [0.72–0.82]	91	15	5.2 (10)	0.81 [0.76–0.85]
Central Amazon	6	2	2 (6)	0.44 [0.19–0.70]	6	2	2 (6)	0.44 [0.18–0.70]	6	4	4 (6)	0.67 [0.39–0.94]
Central America	13	1	1 (6)	0	11	1	1 (6)	0	2	1	n.a.	n.a.
Guiana Shield	35	1	1 (6)	0	26	1	1 (6)	0	10	5	3.93 (6)	0.74 [0.59–0.89]
Pacific coast	31	2	1.19 (6)	0.06 [0–0.18]	12	2	1.5 (6)	0.15 [0–0.42]	52	6	2.84 (6)	0.64 [0.55–0.72]
Western Amazon	20	2	1.99 (6)	0.50 [0.45–0.54]	17	2	1.96 (6)	0.46 [0.32–0.59]	21	1	1 (6)	0
nu_surinamensis												
Neotropics	56	3	3 (56)	0.35 [0.21–0.49]	29	7	7 (56)	0.80 [0.74–0.86]	10	6	6 (10)	0.80 [0.70–0.90]
Amazon Basin	2	1	n.a.	n.a.	2	1	n.a.	n.a.	1	1	n.a.	n.a.
Guiana Shield	54	2	n.a.	0.30 [0.17–0.43]	27	6	n.a.	0.79 [0.73–0.84]	9	5	4.1 (6)	0.77 [0.66–0.87]

*DNA sequence resulting from the concatenation of *trn*H–*psb*A and *trn*C–*ycf*6.
†cpSSR haplotypes based on the polymorphism at 5 cpSSR (ccmp2, ccmp4, ccmp6, ccmp7, ccmp10, Weising & Gardner, 1999).
n.a., not applicable.

sites (only one). Genetic diversity was not significantly different among the remaining three regions. Deviation from H–W equilibrium was detected in all analysed regions except in Central America. When we performed the test on narrower geographical zones to minimize Wahlund effects, the number of loci in H–W disequilibrium was lower and probably caused by the presence of null alleles, as suggested by the MICRO-CHECKER analysis (results not shown).

Individuals were grouped based on Bayesian clustering, and genetic diversity at chloroplast markers was analysed within each cluster. Respectively, 161, 101 and 101 individuals were taken into account at the haplogroup sequence, haplotype sequence and cpSSR level. In the Guiana Shield, where both putative species had been sampled, the *nu_surinamensis* cluster harboured a significantly higher genetic diversity at cpDNA sequences than *nu_guianensis* (0 vs. 0.3 at the haplogroup level, and 0 vs. 0.82 at the haplotype level) (Table 5). At cpSSRs, *nu_surinamensis* and *nu_guianensis* had comparable genetic diversity levels (0.74 vs. 0.77, with overlapping confidence intervals). At the intraspecific level, within the *nu_guianensis* cluster, overall cpDNA genetic diversity was respectively 0.67, 0.77 and 0.81 for haplogroup, haplotypes and cpSSR (Table 5), with four haplogroups having frequencies equal to (G5) or higher than (G2, G3, G4) 10%. At the haplogroup level, the most diverse regions were the Central and Western Amazon Basin ($h = 0.44$ and 0.50, respectively), whereas the least diverse regions ($h = 0$) were Central America and the Guiana Shield, with only one haplotype observed. CpSSR genetic diversity did not differ significantly between Central Amazon, the Guiana Shield and the Pacific coast, but it was significantly lower in the Western Amazon, where only one allele was observed.

Genetic differentiation at the interspecific and intraspecific levels

Interspecific nuSSR differentiation was significant ($P < 0.001$) and nearly equal for F_{ST} (0.22) and R_{ST} (0.23). When differentiation was partitioned into between-species and between-populations components, the proportion of genetic variation among species (F_{CT}) decreased but remained significant ($F_{CT} = 0.08$, $P = 0.007$) even after correction for null alleles ($F_{CT\ (corr)} = 0.05$, $P = 0.043$) (Table 6). The proportion of genetic variation between populations and groups overall (F_{ST}) explained most of the genetic differentiation ($F_{ST} = 0.36$, $P < 0.001$; $F_{ST\ (corr)} = 0.31$, $P < 0.001$). Interestingly, the proportion of genetic variation among species when the phylogenetic distance between alleles was taken into account (R_{CT}) became not significant ($R_{CT} = 0.04$). At cpDNA, the interspecific genetic differentiation was significant for the three marker types: 0.41 ($P < 0.001$), 0.11 ($P < 0.001$) and 0.21 ($P = 0.001$) for haplogroup sequences, haplotype sequences and cpSSR, respectively. When the distance between haplotypes was taken into account, interspecific genetic differentiation (N_{ST}) remained significant ($N_{ST} = 0.32$, $P < 0.001$) and significantly higher than F_{ST} for haplotype sequences ($P = 0.035$). At cpSSR loci, in contrast, interspecific N_{ST} became not significant ($N_{ST} = 0.105$, $P = 0.008$). In common with nuSSR, interspecific genetic differentiation at cp markers decreased when the population level was included in the AMOVA and became not significant whatever the measure of genetic differentiation (haplotype, haplogroup or cpSSR) (Table 7). The population level explained most of the genetic differentiation, with F_{ST} varying between 0.51 at cpSSR and

© 2012 Blackwell Publishing Ltd

Table 6 Analysis of molecular variation (AMOVA) results at seven nuclear simple sequence repeats (nuSSRs) in the tree genus *Carapa* in the Neotropics. Degrees of freedom (d.f.) and the percentage of explained variance (PEV) are given for three hierarchical levels. Results are reported for analyses taking and not taking into account phylogenic relationships between alleles (R_{ST} and F_{ST}). In addition to F_{ST}, results are given for the genotype data set corrected for possible null alleles (F_{ST} (cor), INA method, Chapuis & Estoup, 2007).

		PEV		
	d.f.	F_{ST}	F_{ST} (cor)	R_{ST}
Among species	1	7.47	5.33	3.48
Among populations/within species	8	28.10	25.54	36.83
Within populations	604	64.43	69.13	59.69
F_{CT}		0.08*	0.05*	0.04 n.s.
F_{ST}		0.36*	0.31*	0.40*

F_{CT}, genetic variance among species/total genetic variance; F_{ST}, genetic variance among species + genetic variance among populations within species/total genetic variance.
n.s., not significant; *significant ($P < 0.001$).

0.80 at haplogroup cp sequences. Within *nu_guianensis*, the absence of significant differences between F_{ST} and N_{ST} indicated the low contribution of phylogenetic signal in the distribution of genetic diversity. In pairwise population divergence estimates (Appendix S3), differentiation was significant for all interspecific pairs, with F_{ST} at nuSSR varying between 0.25 and 0.49, and F_{ST} at haplotype cp sequences varying between 0.37 and 1. When molecular distance was taken into account, genetic differentiation varied between 0.08 and 0.79 for R_{ST} at nuSSR and between 0 and 1 for N_{ST} at cp sequences (Appendix S3). A Mantel Z-statistic performed between matrices of genetic and geographical distances for *nu_guianensis* populations rejected the presence of correlation for both F_{ST} and N_{ST} (respectively, $Z = 37179.04$, $P = 0.09$; $Z = 67278.11$, $P = 0.14$ after 120 permutations). This result does not support the isolation-by-distance hypothesis to explain the genetic differentiation observed between populations of *nu_guianensis*.

DISCUSSION

The comparison of nuclear and cpDNA patterns provided insights into the genetic delimitation of species and into hybridization processes, and permitted us to characterize the genetic structure across some selected areas in the Neotropics. Neotropical *Carapa* was grouped into six cpDNA lineages within two major clades, namely one containing haplogroup G1 and the second containing all others. The Bayesian cluster analysis of nuSSRs corroborates the existence of two distinct genetic pools that correspond to *C. surinamensis* and *C. guianensis*, as in Duminil *et al.* (2006). The two nuSSR clusters largely overlapped with haplogroup G1 and the remaining haplogroups.

Timing of speciation events

Our estimates place the divergence between *C. surinamensis* and *C. guianensis* in a pre-Pleistocene timeframe (*c.* 26 Ma). The two lineages seem to have undergone contrasting demographic trajectories. *Carapa surinamensis* has the smaller range, being restricted to the Guiana Shield and central Amazonia, and its extant haplotype diversity is of relatively recent origin (estimated 14 Ma) compared with that of *C. guianensis* (*c.* 18 Ma; Table 3). From a Neotropical perspective, it would appear that the *C. surinamensis* lineage is derived from *C. guianensis*. However, sampling of *C. procera* in Africa may change this result if the African region contains more of the ancestral cpDNA diversity. Our results indicate that ancestral

Table 7 Analysis of molecular variation (AMOVA) results at the chloroplast genome in the tree genus *Carapa* in the Neotropics. The AMOVA was perfomed at both chloroplast DNA intergenic sequences (haplotype and haplogroup levels) and chloroplast single sequence repeats (cpSSR). Degrees of freedom (d.f.) and the percentage of explained variance (PEV) are given for each hierarchical level. Results are reported for analyses taking and not taking into account phylogenic relationships between alleles (N_{ST} and F_{ST}).

	Haplotype			Haplogroup		cpSSR		
	d.f.	PEV Haplotype frequencies	PEV Molecular distance	d.f.	PEV Haplogroup frequencies	d.f.	PEV Haplotype frequencies	PEV Molecular distance
Among species	1	0	8.15	1	9.31	1	0	0
Among populations/within species	8	65.81	60.76	8	70.55	4	51.23	32.37
Within populations	91	34.2	31.08	151	20.14	94	66.4	65.93
F_{CT}		0 n.s.	0.08 n.s.		0.09 n.s.		0 n.s.	0 n.s.
F_{ST}		0.66*	0.69*		0.80*		0.51*	0.32*
Within *nu_guianensis*								
Among populations	6	84.09	74.83	6	85.43	4	65.26	46.92
Within populations	65	15.91	25.17	96	14.57	86	34.74	53.08

F_{CT}, genetic variance among species/total genetic variance; F_{ST}, genetic variance among species + genetic variance among populations within species/total genetic variance.
n.s. not significant; *significant ($P < 0.001$).

© 2012 Blackwell Publishing Ltd

populations of Neotropical *Carapa* could have been influenced by the major geographical events that define the contemporary biogeographical setting in the Neotropics, including the Pliocene uplifts of the Isthmus of Panama and the northern Andean cordilleras, and the late Miocene development of the Amazon drainage (Hoorn & Wesselingh, 2010). While our estimates of the relative timing of diversification of haplotype groups are probably accurate, the placement of divergence times into discrete geological periods comes with important caveats. First, the divergence time analysis is based on a single calibration point, which is itself an indirect estimate based on calibration points elsewhere in the Meliaceae phylogeny (Muellner *et al.*, 2006). Second, the cpDNA as a non-recombining genome does not provide information about the coalescent variance associated with lineage formation, which can include introgression events. Under the hypothesis that the nuclear genome undergoes introgression, nuSSR allele sharing is a result of gene flow rather than of incomplete lineage sorting; this generates discrepancies between the chloroplast haplotype phylogenetic tree and the coalescence of nuSSR alleles. If the coalescence of nuSSR alleles were compared with the phylogenetic tree of chloroplast sequences without taking into account gene flow, this would result in apparent incomplete lineage sorting for nuSSRs, which would mean that nuSSR coalescence underestimates chloroplast divergence. In other words, nuclear and chloroplast markers tell different evolutionary histories with respect to gene flow; reconstructing the phylogenetic history of nuclear clades requires the application of coalescent-based population-genetic methods that take into account gene flow and demography. This is beyond the scope of the present paper, but we can conservatively say that chloroplast clade divergence is an upper estimate of nuclear divergence. Factoring in all of these potential sources of error, it is best to consider the divergences in relative (i.e. older versus younger events) rather than absolute terms.

Correspondence with taxonomy

A recent morphometric and internal transcribed spacer (ITS)-based study of *Carapa* (*sensu lato*) (Kenfack, 2011) described eight new morphospecies of *Carapa* in the Neotropics. To compare our phylogeographical results with the morphospecies designations, we performed a further Bayesian clustering analysis within the two genetic clusters to test for the presence of subspecies structure (data not shown). No additional genetic structure was observed in *C. surinamensis*, but five subclusters were detected in *C. guianensis*, corresponding to the five sampling regions. Because several of the morphospecies described by Kenfack (2011) are broadly sympatric, it is not possible to link the geographic distribution of morphospecies with the genetic subclusters detected within *C. guianensis*. Our results support a taxonomic classification restricted to two species (*C. surinamensis* and *C. guianensis*) as defined by Styles (1981) in *Flora Neotropica*, at least in the sites sampled in the present study. However, the absence of common samples prevented a direct comparison of our results with those of Kenfack (2011). Given that the application of hypervariable nuclear markers allowed us to detect potential interspecific gene flow, it would be interesting to analyse genetic diversity and population divergence at nuSSRs in these new morphospecies to estimate the strength of interspecific barriers to gene flow. This would independently test the phylogenetic identity of Kenfack's (2011) new morphospecies.

Interspecific gene flow

Interestingly, the overlap of Bayesian clustering and chloroplast haplogroups was characterized by shared haplotypes resulting from either introgression via hybridization or retention of ancestral polymorphism (Linder & Rieseberg, 2004; Holland *et al.*, 2008). Nuclear–chloroplast mismatches were geographically structured, with 9 among 11 mismatches occurring in French Guiana where the two species are sympatric. This may be rather strong, albeit indirect, evidence for hybridization between the two sister species, because if multiple alleles present in a common ancestor were retained in descendant species, they should be randomly distributed in the descendant populations of both species and not concentrated geographically in the same areas.

Our results suggest asymmetric introgression (Petit *et al.*, 2003) of the *C. guianensis* chloroplast genome into *C. surinamensis* (Fig. 4b), as shown by the occasional association of *C. guianensis* chloroplast haplotypes with *C. surinamensis* nuclear genotypes. As the chloroplast is maternally inherited in this species, this implies pollination of *C. guianensis* flowers by *C. surinamensis*, followed by recurrent pollination of the resulting hybrids by *C. surinamensis*. This process will ultimately 'dilute' the nuclear *C. guianensis* genome with *C. surinamensis* background, leaving *C. guianensis* chloroplasts associated with a *C. surinamensis* nuclear background. Under a pure demographic model of population expansion (Currat *et al.*, 2008), this could infer a population expansion of *C. surinamensis* and invasion of the *C. guianensis* range. However, the association of putative first-generation hybrids with two cp lineages supports symmetrical fertilization producing viable 'F1' offspring. On the other hand, only the chloroplast lineage typical of *C. guianensis* (G2) can be found associated with a typical *C. surinamensis* nuclear background, but not vice versa. This may be attributable in part to limited sampling, or it may arise from a post-zygotic reproductive isolation mechanism that favours offspring pollinated by *C. guianensis* (Tiffin *et al.*, 2001). The morphological description of individual putative hybrids will be necessary to corroborate their hybrid status.

Historical demography

In the zone of sympatry, within-species genetic diversity is higher in *C. guianensis* than in *C. surinamensis* at nuSSR, but *C. surinamensis* displays higher cpDNA diversity. This result

© 2012 Blackwell Publishing Ltd

confirms the first observations obtained by Duminil *et al.* (2006). The high cpDNA diversity in *C. surinamensis* could be explained by the older diversification of the haplogroup G1 (associated with *C. surinamensis*) relative to haplogroup G2 (included in the *C. guianensis* clade) (see Table 3). At the continental level, the geographical distribution of haplogroups in *C. guianensis* indicates a trend towards the spatial sorting of lineages, which is in agreement with limited seed dispersal as suggested by previous ecological and genetic studies (Forget *et al.*, 1999; Cloutier *et al.*, 2005; Hardy *et al.*, 2006). However, it may be noted that the sister haplogroup G5, which is basally divergent in the *C. guianensis* clade (Fig. 3), is also central relative to the distribution range (Fig. 2b) and relatively common and widespread (from the state of Acre in Brazil in Western Amazonia to the mouth of the Amazon at the extreme east). This may indicate that the Amazon Basin is the centre of origin of the species' genetic diversity; other haplotypes may have differentiated locally from G5 after colonization. The contrast between the highest haplotypic cp diversity in the Central Amazon and the lowest in the most peripheral region (Mesoamerica) further supports this view. Lemes *et al.* (2003) found a similar pattern in mahogany (*Swietenia macrophylla*) of higher genetic diversity in South America than in Central America (Novick *et al.*, 2003). However, in this case, the lower genetic diversity in Central America was not attributed to recent colonization, but rather to more restricted habitat and more severe Pleistocene vegetation changes in Central America, leading to the loss of genetic diversity. The absence of a phylogeographical pattern between Central and South America in *Carapa* populations supports a model of colonization based on occasional long-distance dispersal events, which may have been mediated by oceanic dispersal, as has been demonstrated for some other rain forest tree species (e.g. Dick *et al.*, 2003, 2007).

CONCLUSIONS

The biogeographical history of *Carapa* appears to have been shaped by a mixture of Neogene and Quaternary events. Based on genetic data, the studied species belong to a species complex showing some broad geographic patterns of variation, while at a fine geographical scale introgression is evident. The case for introgression in *Carapa* is supported by inconsistencies between chloroplast and nuclear lineage sorting. These conclusions are different from the taxonomic definition based on morphological discontinuities. Further investigations may yet disclose the mechanisms of hybridization dynamics and natural selection that contribute to genetic exchanges between the geographically restricted *C. surinamensis* and widespread *C. guianensis*. In this sense, *Carapa* is an interesting model with which to study speciation and hybridization and their roles in the generation of diversity in tropical ecosystems.

ACKNOWLEDGEMENTS

This work was financially supported by the EU INCO 'SEEDSOURCE' project, by the EU PO-FEDER 'ENERGI-RAVI' program and by the Centre National de la Recherche Scientifique (CNRS) Programme Amazonie 'CLIPS'. The authors wish to thank Saint-Omer Cazal (UMR ECOFOG, INRA, Kourou, French-Guiana), Jean-Francois Molino [Institut de Recherche pour le Développement (IRD), UMR 'botanique et bioinformatique de l'architecture des plantes' (AMAP), Montpellier France], Daniel Sabatier (IRD, AMAP, Montpellier France), Pierre-Michel Forget (Museum National d'Histoire Naturelle, Paris), Heidy Villalobos (University of Costa Rica), Carlos Navarro (CATIE, Costa Rica), Elio Sanoja (Upata, Venezuela) and Jerôme Chave (CNRS, UMR EDB, Toulouse) for sampling, Valérie Troispoux (UMR ECOFOG, INRA, Kourou, French Guiana) for molecular biology support, and David Kenfack for discussion of unpublished findings on *Carapa* systematics. We thank two anonymous referees for their helpful suggestions and comments, which improved the manuscript.

REFERENCES

Beaumont, M.A., Zhang, W. & Balding, D.J. (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162,** 2025–2035.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. & Bonhomme, F. (2004) *GENETIX 4.05.2, logiciel sous Windows pour la génétique des populations*. Laboratoire Génome, Populations, Interactions, Université Montpellier II, Montpellier, France.

Bush, M.B. & de Oliveira, P.E. (2006) The rise and fall of the refugial hypothesis of Amazonian speciation: a paleoecological perspective. *Biota Neotropica*, **6,** 1–17.

Cavers, S., Navarro, C. & Lowe, A.J. (2003) Chloroplast DNA phylogeography reveals colonization history of a Neotropical tree, *Cedrela odorata* L., in Mesoamerica. *Molecular Ecology*, **12,** 1451–1460.

Chapuis, M.P. & Estoup, A. (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24,** 621–631.

Clement, M., Posada, D. & Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9,** 1657–1659.

Cloutier, D., Povoa, J.S.R., Procopio, L.C., Leao, N.V.M., Wadt, L.H.O., Ciampi, A.Y. & Schoen, D.J. (2005) Chloroplast DNA variation of *Carapa guianensis* in the Amazon basin. *Silvae Genetica*, **54,** 270–274.

Cloutier, D., Kanashiro, M., Ciampi, A.Y. & Schoen, D.J. (2007a) Impact of selective logging on inbreeding and gene dispersal in an Amazonian tree population of *Carapa guianensis* Aubl. *Molecular Ecology*, **16,** 797–809.

Cloutier, D., Hardy, O.J., Caron, H., Ciampi, A.Y., Degen, B., Kanashiro, M. & Schoen, D.J. (2007b) Low inbreeding and high pollen dispersal distances in populations of two Amazonian forest tree species. *Biotropica*, **39,** 406–415.

Colinvaux, P.A., De Oliveira, P.E. & Bush, M.B. (2000) Amazonian and Neotropical plant communities on glacial time-scale: the failure of the aridity and refuge hypothesis. *Quaternary Science Reviews*, **19,** 141–169.

© 2012 Blackwell Publishing Ltd

Colpaert, N., Cavers, S., Bandou, E., Caron, H., Gheysen, G. & Lowe, A.J. (2005) Sampling tissue for DNA analysis of trees: trunk cambium as an alternative to canopy leaves. *Silvae Genetica*, **54,** 265–269.

Currat, M., Ruedi, M., Petit, R.J. & Excoffier, L. (2008) The hidden side of invasions: massive introgression by local genes. *Evolution*, **62,** 1908–1920.

Dayanandan, S., Dole, J., Bawa, K. & Kesseli, R. (1999) Population structure delineated with microsatellite markers in fragmented populations of a tropical tree, *Carapa guianensis* (Meliaceae). *Molecular Ecology*, **8,** 1585–1592.

Dick, C.W., 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.

Dick, C.W., Condit, R. & Bermingham, E. (2005) Biogeographic history and the high beta diversity of rainforest trees in Panamá. *Rio Chagres: a multi-disciplinary profile of a tropical watershed* (ed. by R. Harmon), pp. 259–268. Springer Publishing Company, New York.

Dick, C.W., Bermingham, E., Lemes, M.R. & Gribel, R. (2007) Extreme long-distance dispersal of the lowland tropical rainforest tree *Ceiba pentandra* L. (Malvaceae) in Africa and the Neotropics. *Molecular Ecology*, **16,** 3039–3049.

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

Drummond, A.J., Ho, S.Y.W., Phillips, M.J. & Rambaut, A. (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology*, **4,** e88.

Duminil, J., Caron, H., Scotti, I., Cazal, S.O. & Petit, R.J. (2006) Blind population genetics survey of tropical rainforest trees. *Molecular Ecology*, **15,** 3505–3513.

Ersts, P.J. (2010) *Geographic Distance Matrix Generator (version 1.2.3)*. American Museum of Natural History, Center for Biodiversity and Conservation. Available at: http://biodiversityinformatics.amnh.org/open_source/gdmg/.

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14,** 2611–2620.

Excoffier, L. (2004) Special issue: Analytical methods in phylogeography and genetic structure. *Molecular Ecology*, **13,** 727.

Excoffier, L., Smouse, P.E. & Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131,** 479–491.

Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin version 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1,** 47–50.

Forget, P.M., Mercier, F. & Collinet, F. (1999) Spatial patterns of two rodent-dispersed rain forest trees *Carapa procera* (Meliaceae) and *Vouacapoua americana* (Caesalpiniaceae) at Paracou, French Guiana. *Journal of Tropical Ecology*, **15,** 301–313.

Gentry, A.H. & Dodson, C.H. (1988) New species and a new combination for plants from trans-Andean South America. *Annals of the Missouri Botanical Garden*, **75,** 1434–1436.

Grivet, D., Sebastiani, F., González-Martínez, S.C. & Vendramin, G.G. (2009) Patterns of polymorphism resulting from long-range colonization in the Mediterranean conifer Aleppo pine. *New Phytologist*, **184,** 1016–1028.

Grundmann, H., Hori, S. & Tanner, G. (2001) Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. *Journal of Clinical Microbiology*, **39,** 4190–4192.

Hall, P., Orrell, L.C. & Bawa, K.S. (1994) Genetic diversity and mating system in a tropical tree, *Carapa guianensis* (Meliaceae). *American Journal of Botany*, **81,** 1104–1111.

Hardesty, B., Dick, C.W., Hamrick, J., Degen, B., Hubbell, S. & Bermingham, E. (2010) Geographic influence on genetic structure in the widespread Neotropical tree *Simarouba amara* Aubl. (Simaroubaceae). *Tropical Plant Biology*, **3,** 28–39.

Hardy, O.J. & Vekemans, X. (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2,** 618–620.

Hardy, O.J., Maggia, L., Bandou, E., Breyne, P., Caron, H., Chevallier, M.-H., Doligez, A., Dutech, C., Kremer, A., Latouche-Hallé, C., Troispoux, V., Veron, V. & Degen, B. (2006) Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. *Molecular Ecology*, **15,** 559–571.

Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405,** 907–913.

Holland, B.R., Benthin, S., Lockhart, P.J., Moulton, V. & Huber, K.T. (2008) Using supernetworks to distinguish hybridization from incomplete lineage sorting. *BMC Evolutionary Biology*, **8,** 202.

Hoorn, C. & Wesselingh, F.P. (2010) *Amazonia, landscape and species evolution*. Blackwell Publishing, Oxford.

Hubisz, M.J., Falush, D., Stephens, M. & Pritchard, J.K. (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9,** 1322–1332.

Jakobsson, M. & Rosenberg, N.A. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23,** 1801–1806.

Kenfack, D. (2011) Resurrection in *Carapa* (Meliaceae): a reassessment of morphological variation and species boundaries using multivariate methods in a phylogenetic context. *Botanical Journal of the Linnaean Society*, **165,** 186–221.

Kenfack, D. & Perez, A. (2011) Two new species of *Carapa* (Meliaceae) from western Ecuador. *Systematic Botany*, **36,** 124–128.

Kingman, J.F.C. (1982) The coalescent. *Stochastic Processes and their Applications*, **13,** 235–248.

© 2012 Blackwell Publishing Ltd

Lemes, M.R., Gribel, R., Proctor, J. & Grattapaglia, D. (2003) Population genetic structure of mahogany (*Swietenia macrophylla* King, Meliaceae) across the Brazilian Amazon, based on variation at microsatellite loci: implications for conservation. *Molecular Ecology*, **12,** 2875–2883.

Lepais, O., Petit, R.J., Guichoux, E., Lavabre, J.E., Alberto, F., Kremer, A. & Gerber, S. (2009) Species relative abundance and direction of introgression in oaks. *Molecular Ecology*, **18,** 2228–2242.

Linder, C.R. & Rieseberg, L.H. (2004) Reconstructing patterns of reticulate evolution in plants. *American Journal of Botany*, **91,** 1700–1708.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27,** 209–220.

Muellner, A.N., Savolainen, V., Samuel, R. & Chase, M.W. (2006) The mahogany family "out-of-Africa": divergence time estimation, global biogeographic patterns inferred from plastid *rbcL* DNA sequences, extant, and fossil distribution of diversity. *Molecular Phylogenetics and Evolution*, **40,** 236–250.

Noamesi, G.K. (1958) *A revision of Xylocarpae (Meliaceae)*. PhD Thesis, University of Wisconsin, Madison, WI.

Novick, R.S., Dick, C.W., Lemes, M., Navarro, C., Caccone, A. & Bermingham, E. (2003) Genetic structure of Mesoamerican populations of big-leaf mahogany (*Swietenia macrophylla*) inferred by microsatellite analysis. *Molecular Ecology*, **12,** 2885–2893.

van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M. & Shipley, P. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4,** 535–538.

Petit, R.J. & Excoffier, L. (2009) Gene flow and species delimitation. *Trends in Ecology and Evolution*, **24,** 386–393.

Petit, R.J., El Mousadik, A. & Pons, O. (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12,** 844–855.

Petit, R.J., Bodenès, C., Ducousso, A., Roussel, G. & Kremer, A. (2003) Hybridization as a mechanism of invasion in oaks. *New Phytologist*, **161,** 151–164.

Petit, R.J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D. & Vendramin, G.G. (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, **14,** 689–701.

Petit, R.J., Hu, F.S. & Dick, C.W. (2008) Forests of the past: a window to future changes. *Science*, **320,** 1450–1452.

Poelchau, M.F. & Hamrick, J.L. (2013) Comparative phylogeography of three common Central American Neotropical tree species. *Journal of Biogeography*, **40,** 618–631.

Pons, O. & Petit, R.J. (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics*, **144,** 1237–1245.

Posada, D. & Buckley, T.R. (2004) Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, **53,** 793–808.

Pritchard, J., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155,** 945–959.

Rosell, J.A., Olson, M.E., Weeks, A., De-Nova, J.A., Lemos, R.M., Camacho, J.P., Feria, T.P., Gómez-Bermejo, R., Montero, J.C. & Eguiarte, L.E. (2010) Diversification in species complexes: tests of species origin and delimitation in the *Bursera simaruba* clade of tropical trees (Burseraceae). *Molecular Phylogenetics and Evolution*, **57,** 798–811.

Rosenberg, N.A. (2004) Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4,** 137–138.

Rull, V. (2011) Origins of biodiversity. *Science*, **331,** 398–399.

Scarano, F.R., Pereira, T.S. & Rocas, G. (2003) Seed germination during floatation and seedling growth of *Carapa guianensis*, a tree from flood-prone forests of the Amazon. *Plant Ecology*, **168,** 291–296.

Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W., Miller, J., Siripun, K.C., Winder, C.T., Schilling, E.E. & Small, R.L. (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany*, **92,** 142–166.

Slatkin, M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139,** 457–462.

Stebbins, G.L. (1974) *Flowering plants: evolution above the species level*. Belknap Press, Cambridge, MA.

Styles, B.T. (1981) Swietenioideae. *Flora Neotropica Monograph No. 28: Meliaceae* (ed. by T. Pennington, T. Styles and A.H. Taylor), pp. 359–418. New York Botanical Garden, New York.

Styles, B.T. & White, F. (1991) Meliaceae. *Flora of tropical East Africa* (ed. by R.M. Polhill), pp. 1–68. A.A. Balkema, Rotterdam.

Swofford, D.L. (2003) *PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4*. Sinauer Associates, Sunderland, MA.

Templeton, A.R., Crandall, K.A. & Sing, C.F. (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132,** 619–633.

Tiffin, P., Olson, S. & Moyle, L.C. (2001) Asymmetrical crossing barriers in angiosperms. *Proceedings of the Royal Society B: Biological Sciences*, **268,** 861–867.

Vinson, C.C., Azevedo, V.C.R., Sampaio, I. & Ciampi, A.Y. (2005) Development of microsatellite markers for *Carapa guianensis* (Aublet), a tree species from the Amazon forest. *Molecular Ecology Notes*, **5,** 33–34.

Weising, K. & Gardner, R. (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome*, **42,** 9–19.

© 2012 Blackwell Publishing Ltd

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 List and details of the sites sampled for fresh and herbarium material.
Appendix S2 Polymorphism at chloroplast sequences and phylogenetic reconstruction.
Appendix S3 Pairwise population genetic differentiation.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

BIOSKETCH

Caroline Scotti-Saintagne is a population geneticist interested in the demographic and environmental processes involved in the genetic divergence between closely related species of forest trees.

Author contributions: C.S-S., I.S., C.D., H.C. and G.G.V. designed the experiments; C.S-S., I.S. and H.C. performed the statistical analyses; C.S-S., H.C., I.S., R.V., M.L. and R.G. collected the samples; C.S-S., H.C., G.G.V, C.D., E.G., P.S. and A.B. contributed to the molecular analyses; and C.S-S., I.S. and C.D. wrote the paper and all authors read and approved the manuscript.

Editor: Malte Ebach

This Special Issue was motivated by research undertaken by partners in the SEEDSOURCE consortium (http://www.seedsource.net), a collaborative project funded by the European Commission under FP6. The project aimed to establish the evolutionary basis and geographic distribution of genetic variation in a range of important Neotropical tree species, for better sourcing and control of germplasm and forest conservation.

© 2012 Blackwell Publishing Ltd