Annals of Botany **107**: 229–241, 2011
doi:10.1093/aob/mcq232, available online at www.aob.oxfordjournals.org

ANNALS OF BOTANY
Founded 1887

Phylogeography of *Pinus* subsection *Australes* in the Caribbean Basin

Lev Jardón-Barbolla[1,*], Patricia Delgado-Valerio[2], Gretel Geada-López[3], Alejandra Vázquez-Lobo[1] and Daniel Piñero[1]

[1]*Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México (UNAM), Apartado Postal 70-275, Mexico D.F. 04510, Mexico,* [2]*Facultad de Agrobiología 'Presidente Juárez', Universidad Michoacana de San Nicolás de Hidalgo, Av. Revolución esquina con Berlín S/N, CP 60091, Uruapan, Michoacán, México and* [3]*Departamento de Biología, Universidad de Pinar del Río, Martí 270 esquina con 27 de Noviembre, CP 20100, Ciudad de Pinar del Río, Pinar del Río, Cuba*

** For correspondence. E-mail levjardon@yahoo.com.mx or lojb@hp.fciencias.unam.mx*

Received: 19 April 2010 Returned for revision: 26 July 2010 Accepted: 27 October 2010 Published electronically: 29 November 2010

• *Background and Aims* Four species of *Pinus* subsection *Australes* occur in the Caribbean Basin: *P. caribaea*, *P. cubensis*, *P. maestrensis* and *P. occidentalis*. This study analyses the phylogeography of these species to assess possible colonization events from Central America to the islands and subsequent population expansions during glacial periods driven by both drier climate and larger emerged land areas.
• *Methods* Allele size data were obtained for plastid microsatellites for 314 individuals from 24 populations, covering the distribution range of subsection *Australes* in the Caribbean Basin.
• *Key Results* In total, 113 plastid haplotypes were identified. The highest genetic diversity was found in populations of *P. caribaea.* Overall, Caribbean Basin populations fit the isolation by distance model. Significant phylogeographical structure was found ($R_{ST} = 0{\cdot}671 >$ permuted $R_{ST} = 0{\cdot}101$; $P < 0{\cdot}0001$). The haplotype network and a Bayesian analysis of population structure (BAPS) indicated different Central American origins for *P. caribaea* var. *bahamensis* and *P. caribaea* var. *caribaea* plastids, with Central America populations in northern and south-eastern groups. Sudden expansion times for BAPS clusters were close to three glacial maxima.
• *Conclusions* Central America contains ancestral plastid haplotypes. Population expansion has played a major role in the distribution of genetic diversity in *P. caribaea* var. *hondurensis*. Two colonization events gave rise to the *P. caribaea* var. *bahamensis* and *P. caribaea* var. *caribaea* lineages. Plastid variation in the eastern species (*P. cubensis*, *P. maestrensis* and *P. occidentalis*) evolved independently from that in *P. caribaea* var. *caribaea*. Incomplete lineage sorting between *P. cubensis* and *P. maestrensis* is apparent. Inferred expansion times for *P. caribaea* var. *bahamensis* and for the eastern lineages correspond to glacial maxima, whereas those for *P. caribaea* var. *hondurensis* correspond to the beginning of the temperature decrease that led to Marine Isotope Stage 8.

Key words: Caribbean Basin, historical demography, phylogeography, *Pinus*, plastid microsatellites, population genetics, subsection *Australes*.

INTRODUCTION

Pinus subsection *Australes* (Pinaceae) is a monophyletic group of pines (Gernandt *et al.*, 2005, 2008) occurring in North America, Mexico, Central America and the western Caribbean islands. Four taxa of subsection *Australes* are found from low savannah to mountain regions in the Caribbean Basin (Farjon and Styles, 1997): *P. caribaea*, *P. cubensis, P. maestrensis* and *P. occidentalis*.

The savannah lowlands in the west of the Caribbean Basin are dominated by the three varieties of *Pinus caribaea*: var. *hondurensis* (Central America), var. *caribaea* (western Cuba and Isla de la Juventud) and var. *bahamensis* (Bahamas and Turks & Caicos Islands). Among the three varieties, mature trees exhibit only slight morphological differences (Barrett and Golfari, 1962; Farjon and Styles, 1997). Although the population genetics and distribution of *P. caribaea* have been well studied (Dvorak *et al.*, 2000, 2005), there are no studies using molecular markers that include all three varieties throughout their natural range.

Two allopatric taxa, *P. cubensis* and *P. maestrensis*, are found in the eastern mountains of Cuba. Some authors (Farjon and Styles, 1997) have treated *P. maestrensis* as a synonym of *P. cubensis*, whereas others (López-Almirall, 1982) have considered them to be separate species. An analysis of ovulate cone morphology characters by Adams and Jackson (1997) retrieved *P. cubensis* and *P. caribaea* as sister groups but a posterior phylogenetic analysis of plastid DNA (Geada-López *et al.*, 2002) recovered *P. cubensis* and *P. maestrensis* as sister groups. According to Gernandt *et al.* (2005), *P. cubensis* is closely related to *P. occidentalis*, an endemic species from the mountains of Hispaniola. In Cuba, *P. cubensis* is distributed throughout the Sierra Nipe-Cristal and *P. maestrensis* inhabits the Sierra Maestra. *P. occidentalis* is found in the mountains of Hispaniola.

Slight morphological differences among the three varieties of *P. caribaea* suggest that they diverged recently (estimated divergence time for subsection *Australes* is 6 Myr BP; Gernandt *et al.*, 2008) and/or suggest gene flow between the continent and the islands. In contrast, morphological

differences have permitted the distinction of three species (*P. cubensis*, *P. maestrensis* and *P. occidentalis*) in eastern Cuba and Hispaniola. These eastern species share ecological affinities, such as higher altitude distributions, more humid habitats and less invasive capability, differing in these characteristics from *P. caribaea* (López-Almirall, 1982). This suggests that eastern species have evolved separately from *P. caribaea* and that gene flow between the eastern and western areas of the basin has been restricted.

The present distribution of genetic variation in natural populations of a species or a set of recently diverged species is the result of current and past processes, including gene flow and introgression, population expansion, fragmentation, bottlenecks, changes in distribution ranges and common ancestry relationships (Cruzan and Templeton, 2000; Templeton, 2004; Avise, 2009). Plastids are paternally inherited in pines (Chen *et al.*, 2002) and are more sensitive to genetic drift due to lower effective population size. The high degree of polymorphism of plastid simple sequence repeats (plastid SSRs; Vendramin *et al.*, 1996) has proved useful in elucidating the genetic structure and recent phylogeographical history of populations (Cuenca *et al.*, 2003; Afzal-Rafii and Dodd, 2007; Bucci *et al.*, 2007; Delgado *et al.*, 2007; Rodríguez-Banderas *et al.*, 2009). These markers are also suitable for studies of past demographic expansions, such as those occurring after colonization events (Navascués *et al.*, 2006). High mutation rates like those found in plastid SSRs (Provan *et al.*, 1999) also permit the analysis of more recent population processes such as recent divergence and gene flow (Kuhner, 2009). Plastid markers have also been used to reveal cryptic hybridization in pines (Jasińska *et al.*, 2010).

The interaction of four tectonic plates formed the Caribbean Basin during the Tertiary (Pindell *et al.*, 2006). At approximately 35 Mya the Greater Antilles and the Bahamas became completely separated from Central and North America (Iturralde-Vinent, 2006; Pindell *et al.*, 2006). This geological event probably preceded the arrival of the genus *Pinus* in southern Mexico (21 Mya; Graham, 1999). In the recent palaeogeography of the region, major changes in sea level have been important processes, particularly during glacial periods when a larger area of the main islands was above sea level, but there was no connection among them or with the continent (Iturralde-Vinent, 2006).

The position of Cuba and Hispaniola with respect to the nearby continental masses has remained almost unchanged for the past 10 Myr (Iturralde-Vinent, 2006). The Bahamas emerged more recently, during the Pleistocene (1·9 Mya) (Pindell *et al.*, 2006) and the low level of endemism there has been explained as a result of the recent arrival of most of its flora and fauna (Trejo-Torres and Ackerman, 2001; Ricklefs and Bermingham, 2008).

Meanwhile there has been continuous land between Honduras and North America since the Miocene (23 Mya; Iturralde-Vinent, 2006; Pindell *et al.*, 2006). The floristic composition of the Caribbean Basin, particularly the Great Antilles, suggests an important contribution from Central America (Lavin *et al.*, 2003; Negrón-Ortiz and Watson, 2003; Santiago-Valentín and Olmstead, 2003; Ricklefs and Bermingham, 2008). This is also true for some animal species, for example the vertebrate genus *Icterus* (Omland *et al.*, 1999).

Given that: (1) no land connection has existed between the islands and the continent for the last 30 Myr (Iturralde-Vinent, 2006), (2) the Greater Antilles and the Bahamas are geologically younger than Central America, (3) Central America has contributed to the Greater Antilles biota and (4) the origin of subsection *Australes* is relatively recent (6 Mya; Gernandt *et al.*, 2008), we would expect to find a pattern of colonization of islands by pines from Central America via long-distance dispersal. Thus, we predict ancestral haplotypes and higher levels of genetic diversity within populations of *P. caribaea* var. *hondurensis* and that ancestors of the present Caribbean pines arrived to the islands from Central America.

Here we analyse the distribution of genetic variation to infer the main processes that have affected the distribution of plastid DNA lineages of pines of subsection *Australes* in the Caribbean Basin. We use information from seven polymorphic plastid SSRs to study the genetic structure of populations of pines in subsection *Australes* in the Caribbean Basin. The aims of our study were: (1) to infer patterns of colonization of the Caribbean islands by pines of subsection *Australes* and to investigate whether genetic evidence suggests Central America as the origin of present island lineages; (2) to determine the role of past fragmentation among populations; and (3) to characterize historical demographic patterns based on genetic evidence collected across the entire distributional range of the different pine species in the Caribbean. We use this information to test the hypothesis that Caribbean pines arrived in the islands from Central America.

MATERIALS AND METHODS

Population sampling

Leaf needles of the four different species of pines were sampled from Central America and the Caribbean islands. Needles were collected from six to 24 trees per population in each of 24 sampled locations (shown in Fig. 1 and Table 1). Total sample size was 314 individuals. Trees sampled within populations were separated by >50 m. National forest inventories were used to avoid sampling plantations. All locations corresponded to monospecific populations, with the exception of that of Viñales, Cuba, where *P. caribaea* var. *caribaea* and *P. tropicalis*, two distantly related species, are sympatric. Geoposition and altitudes at each collection site were measured using an eTrex Vista global positioning system (Garmin, Taiwan).

DNA isolation, microsatellite amplification and size estimation

DNA was isolated using a modified cetyl trimethyl ammonium bromide (CTAB) protocol (Doyle and Doyle, 1987) from 0·2–0·4 g of tissue. A preliminary screen for polymorphic loci was carried out for 35 individuals using ten microsatellite primers (Vendramin *et al.*, 1996), widely used in phylogeographical and population genetics studies (Cuenca *et al.*, 2003; Afzal-Rafii and Dodd, 2007; Bucci *et al.*, 2007; Delgado *et al.*, 2007; Rodríguez-Banderas *et al.*, 2009). Loci tested were Pt1254, Pt9383, Pt15169, Pt26081, Pt30204, Pt45002, Pt63718, Pt71936, Pt87268 and

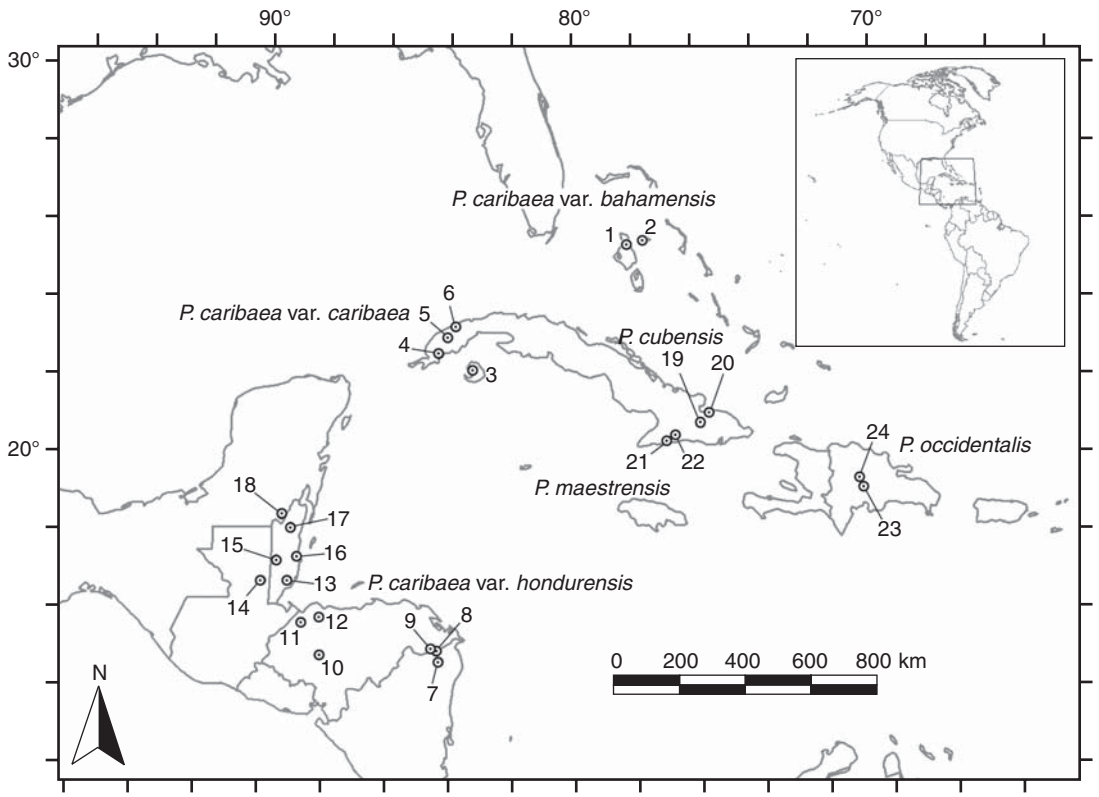

FIG. 1. Map showing the sampled populations and their species assignments: (1) Andros, (2) New Providence, (3) Isla de la Juventud, (4) Arenas Blancas, (5) Viñales, (6) Marbajita, (7) Moss, (8) Waspam, (9) Leimus, (10) San Jerónimo, (11) Cofradía, (12) Mezapa, (13) Deep River, (14) Dolores, (15) Mountain Pine Ridge, (16) Mullins River, (17) Crooked Tree, (18) Jaguactal, (19) Mayarí, (20) Cabonico, (21) Los Rodrigos, (22) Los Números, (23) Agua Blanca, (24) Manabao.

Pt110048. The polymorphic loci detected in the preliminary screen were amplified for the entire set of individuals.

PCRs were carried out in a 25-mL volume, containing 25 ng of template DNA, 0·8 mM dNTP mixture (each 0·2 μM), 2·5 mM $MgCl_2$, 1× PCR reaction buffer (Invitrogen, Carlsbad, CA, USA), 0·4 μM of each primer and 0·5 U of *Taq* polymerase (Invitrogen). PCR reactions were performed on a Mastercycler gradient (Eppendorf Scientific Inc., Hamburg, Germany) or on a 2770 Thermal Cycler (Applied Biosystems, Foster City, CA, USA), with the following profile: 7 min denaturing at 95 °C, followed by 35 cycles of 1 min at 94 °C, 1 min at 55 °C and 1 min at 72 °C, with a final hold of 7 min at 72 °C.

Amplification products of Pt71936, Pt87268 and Pt110048 were analysed by denaturing electrophoresis in 6 % polyacrylamide gels stained with silver nitrate. The size and homology of the alleles was corroborated by direct sequencing of PCR products previously purified using a QIAquick Purification Kit according to the manufacturer's instructions (QIAGEN, Valencia, CA, USA) and both strands were sequenced (High-Throughput Genomics Unit, Department of Genome Sciences, University of Washington; GenBank accession numbers: FJ705614–FJ705623, FJ750441–FJ750448 and FJ769208–FJ769217). Primers Pt1259, Pt15169 and Pt45002 were multiplexed in PCRs using forward primers labelled with fluorescent dyes (VIC, 6-FAM and VIC, respectively; Applied Biosystems) and fragment sizes were determined using an ABI 3100 Avant Genetic Analyzer (Applied Biosystems) and Gene Mapper v3·5 (Applied Biosystems).

Of the six loci retained, five were simple mononucleotides (Vendramin *et al.*, 1996) and one (Pt15169) was a compound microsatellite with three different repeat blocks [$(C)_8(T)_8A(T)_8$; Vendramin *et al.*, 1996]. To eliminate the molecularly accessible size homoplasy (MASH; Estoup *et al.*, 2002) for this locus, we purified and sequenced 42 PCR products as described above. This subsample of 42 PCR products represented the four size classes that were found for this locus and included individuals from all the populations studied.

Data analyses

Plastid haplotypes were identified by allele combinations of polymorphic SSRs. An estimation of haplotypic richness (H_r) was obtained with ADZE-v1·0 software (Szpiech *et al.*, 2008), using a rarefaction procedure to control for different sample sizes within each population. Additionally, a rarefaction curve from ADZE-1·0 was used to approximate a regression between ln(sample size) and the expected number of new plastid haplotypes, following Gillespie (2004), to validate our sample size.

The programme Arlequin v3·1 (Excoffier *et al.*, 2005) was used to obtain measures of genetic diversity for each population. For our dataset corrected for MASH through sequencing, unbiased haplotypic diversity H_E SEQ (equivalent to

TABLE 1. *Name, geographical location and genetic diversity estimates across Caribbean populations of pines of subsection* Australes

Number and population name	Latitude (°N)	Longitude (°W)	Altitude (m a.s.l.)	n	n_a	H_r	H_p	H_E SEQ (± s.d.)	θ_{IAM}	θ_{SMM}
Pinus caribaea				252	90	46·037	46·037	0·977 (0·003)	40·270	944·679
P. caribaea var. *bahamensis*				26	9	9·000	9·000	0·865 (0·035)	5·2856	26·935
1. Andros, Bahamas	24 °55′13·10″	78 ° 0′44·80″	5	14	6	3·880	3·161	0·813 (0·074)	3·476	13·829
2. New Providence, Bahamas	25 °00′31·20″	77 °30′06·90″	7	12	4	3·181	2·462	0·712 (0·105)	1·898	5·532
P. caribaea var. *caribaea*				44	32	20·380	20·380	0·9745 (0·011)	36·400	768·43
3. Isla de la Juventud, Cuba	21 °43′34·30″	82 °55′10·40″	70	6	4	4·000	4·000	0·867 (0·129)	5·389	27·639
4. Arenas Blancas, Cuba	22 °08′26·00″	83 °58′41·30″	18	11	10	5·727	5·564	0·982 (0·046)	52·144	1508·980
5. Viñales, Cuba	22 °32′48·40″	83 °42′29·50″	239	20	10	4·522	4·359	0·879 (0·048)	6·087	33·594
6. Marbajita, Cuba	22 °49′13·30″	83 °27′03·60″	114	7	6	5·285	5·286	0·954 (0·095)	18·365	238·280
P. caribaea var. *hondurensis*				182	52	20·465	20·465	0·951 (0·005)	22·192	207·746
7. Moss, Nicaragua	14 °27′13·80″	83 °54′14·40″	128	15	5	3·015	1·608	0·629 (0·125)	1·277	3·125
8. Waspam, Nicaragua	14 °43′09·80″	83 °58′47·10″	87	20	6	2·942	1·376	0·579 (0·124)	1·032	2·320
9. Leimus, Honduras	14 °45′57·00″	84 °08′08·70″	90	12	6	4·038	2·322	0·818 (0·096)	3·603	14·628
10. San Jerónimo, Honduras	14 °37′21·60″	87 °35′45·30″	530	15	9	5·044	3·206	0·933 (0·040)	12·498	111·888
11. Cofradía, Honduras	15 °26′31·10″	88 °10′03·50″	347	14	10	5·231	3·426	0·945 (0·045)	15·618	165·392
12. Mezapa, Honduras	15 °34′18·60″	87 °36′38·20″	306	19	10	4·959	2·452	0·924 (0·035)	10·714	86·065
13. Deep River, Belize	16 °29′16·00″	88 °37′46·93″	31	12	5	3·712	1·668	0·803 (0·078)	3·237	12·384
14. Dolores, Guatemala	16 °29′07·70″	89 °25′34·90″	438	14	5	3·132	1·668	0·659 (0·123)	1·468	3·808
15. Mountain Pine Ridge, Belize	16 °59′35·00″	88 °57′50·10″	501	14	4	2·286	0·947	0·396 (0·159)	0·487	0·869
16. Mullins River, Belize	17 °04′42·00″	88 °19′43·07″	18	9	5	3·667	2·398	0·722 (0·159)	2·000	5·979
17. Crooked Tree, Belize	17 °48′12·20″	88 °32′55·00″	1	24	5	2·657	2·407	0·540 (0·109)	0·878	1·862
18. Jaguactal, México	18 °08′23·17″	88 °49′37·85″	35	14	4	2·703	1·751	0·571 (0·132)	1·000	2·222
P. cubensis				18	9	7·455	4·004	0·752 (0·103)	2·350	7·629
19. Mayarí, Cuba	20 °25′13·50″	75 °45′12·60″	379	10	5	3·667	1·225	0·756 (0·129)	2·403	7·871
20. Cabonico, Cuba	20 °39′36·50″	75 °28′12·60″	114	8	5	4·000	1·524	0·786 (0·151)	2·887	10·387
P. maestrensis				18	7	7·925	4·010	0·889 (0·042)	6·772	40·081
21. Los Rodrigos, Cuba	19 °58′48·00″	76 °47′58·00″	1461	7	3	2·857	2·107	0·667 (0·160)	1·519	4·001
22. Los Números, Cuba	20 °08′40·48″	76 °31′28·86″	974	11	5	3·833	2·040	0·818 (0·083)	3·603	14·628
P. occidentalis				26	11	9·057	8·041	0·874 (0·038)	5·780	30·994
23. Agua Blanca, Dominican Republic	18 °51′1·50″	70 °41′7·80″	1563	9	7	4·988	3·111	0·889 (0·091)	6·772	40·008
24. Manabao, Dominican Republic	19 °03′18·80″	70 °47′29·10″	911	17	8	4·242	5·280	0·846 (0·066)	4·465	20·474
Eastern lineages				62	23	21·136	21·136	0·926 (0·017)	11·148	90·807
Basin total				314	113	101·010		0·982 (0·0002)	53·4303	1542·710

n, sample size; n_a, observed number of plastid haplotypes; H_r, rarefaction standardized haplotype richness (Szpiech *et al.*, 2008); H_p, rarefaction standardized private allelic richness; H_E SEQ, the unbiased haplotype diversity of data corrected for MASH; SMM and IAM correspond to the mutation models under which θ was estimated.

H_E for diploid data; Nei, 1987) was calculated. In addition, θ was estimated under the infinite allele model (IAM) and stepwise mutation model (SMM; Ohta and Kimura, 1973; Xu and Fu, 2004) using Arlequin v3·1 (Excoffier *et al.*, 2005) for our data corrected for MASH.

To discern whether data fitted the isolation by distance model (IBD) among populations, a Mantel test (10 000 permutations) was computed between the geographical distance matrix and the genetic matrix using Goldstein's distance, $\delta\mu^2$ (Goldstein *et al.*, 1995) with XLSTAT.

A minimum spanning network for plastid haplotypes was computed using TCS v.1·21 (Clement *et al.*, 2000) allowing up to four steps of missing haplotypes. Ambiguities in the network were resolved using criteria and procedures delineated by Crandall and Templeton (1993).

The R_{ST} statistic (Estoup *et al.*, 1998), which takes into account frequency and similarity in plastid haplotype size, was computed to contrast the effect of ordered versus unordered alleles on genetic differentiation. An R_{ST} value larger than expected by chance pR_{ST} (equivalent to F_{ST} and estimated based on 100 000 permutations of allele sizes; Hardy *et al.*, 2003) indicates a phylogeographical structure in which closely related plastid haplotypes are located in nearby populations (Pons and Petit, 1996; Lapègue and Petit, 1997; Afzal-Rafi and Dodd, 2007). Both Goldstein's distance and R_{ST} were computed using SPAGeDi v1·2 (Hardy and Vekemans, 2002).

The geographical structure of the genetic variation was examined through Bayesian analysis of population structure (BAPS) using BAPS 5·0 software based on linked loci (Corander *et al.*, 2008). This analysis considered only identity information of the plastid haplotypes because BAPS options for spatial clustering do not consider a stepwise mutation model. Although there could be information loss through codification in a single locus, Corander *et al.* (2008) have reported that the inclusion of spatial data is useful to avoid overestimating the number of genetic classes. This analysis was used to partition the populations in a number of K groups using spatial information to detect the most likely genetic structure among the 24 populations. The search was conducted using from two to 25 groups (K), with ten replicates for each K value. Another search with 100 initial partitions for the K

with the best likelihood was performed as delineated by Godbout *et al.* (2008).

To examine evidence of demographic expansion, mismatch distribution tests were used (Rogers and Harpending, 1992; Schneider and Excoffier, 1999) for different clades (based on BAPS) using Arlequin v3·1 (Excoffier *et al.*, 2005), and with microsatellite data binary coded (Navascués *et al.*, 2006). Estimation of τ (time to expansion in mutational units) was made by a maximum-pseudolikelihood method developed by Navascués *et al.* (2009), which uses a homoplasy correction model and has been demonstrated to provide a τ estimate closer to the true value than moment-based estimates. Time to expansion was calculated using a per-locus mutation rate estimate of $5{\cdot}5 \times 10^{-5}$ per generation (Provan *et al.*, 1999) and 42·5 years as the generation time; this generation time is the mean between the age at which maximum fecundity is reached in these species (15 years) and the approximate life span (70 years; A. López-Almirall, Museo Nacional de Historia Natural, Cuba, pers. comm.).

RESULTS

Genetic variation

Of the ten loci analysed, six were polymorphic in our data set (Pt1254, Pt15169, Pt 45002, Pt71936, Pt87268 and Pt110048). Sequencing of the locus Pt15169 (GenBank accession numbers: FJ687622–FJ687664) revealed that the first two repeat blocks were polymorphic. Each species had a distinct range of nucleotide composition in these repeat blocks within the subsample. We therefore treated them as separate loci, designated as Pt15169C (ranging from nine to ten C's) and Pt15169T (ranging from nine to 11 T's), respectively, assigning to the individuals of each species the nucleotide composition found in the subsample for its population. This eliminated the MASH in our data set. A total of 113 plastid haplotypes was identified with different combinations of alleles for the seven loci. Only haplotypes 60 and 73 had frequencies $>0{\cdot}05$ (Supplementary Data Table S1, available online).

Rarefaction analysis showed that observed differences in haplotypic richness among populations and lineages were consistent with the standardized estimators obtained (Table 1). Moreover, regression analysis demonstrated that even if we double the total sample size (to 628 individuals), we should expect to find only nine new plastid haplotypes (Supplementary Data Table S2). This indicates that our sample size was adequate to provide a reasonable estimate of genetic diversity.

The genetic diversity for *P. caribaea* populations was H_E SEQ $= 0{\cdot}977$ [confidence interval (CI) $\pm$ 0·006] ($\theta_{IAM} = 40{\cdot}270$, $\theta_{SMM} = 944{\cdot}679$). For the eastern species this estimate was significantly smaller [$H_E = 0{\cdot}926$ (CI $\pm$ 0·033); see Table 1]. As shown, the highest values for the indirect estimator of the effective population size revealed larger values for the species with a wider distribution. This finding is supported by rarefaction analysis, which shows that when controlling for sample size, the haplotypic richness of *P. caribaea* is still larger than that of the eastern lineages (Table 1).

Within *P. caribaea* var. *hondurensis* the most variable populations were in Honduras in the middle portion of the distribution range in Central America (San Jerónimo, Cofradía and Mezapa, populations 10, 11 and 12, Fig. 1, Table 1). The presence in the Arenas Blancas and Viñales populations of *P. caribaea* var. *caribaea* plastid haplotypes 28, 94 and 109, probably of Central American origin (see below), increased estimates of theta considerably.

Isolation by distance

We found a general pattern of IBD for the Caribbean Basin ($r = 0{\cdot}375$, $P < 0{\cdot}0001$; Fig. 2A). On the other hand, comparisons among populations of *P. cubensis*, *P. maestrensis* and *P. occidentalis* showed smaller $\delta\mu^2$ values on average than those obtained for pairs of populations of *P. caribaea* ($t = 4{\cdot}227$, $P = 0{\cdot}0001$). Also, pairwise comparisons among populations of *P. caribaea* and eastern species (*P. cubensis*, *P. maestrensis* and *P. occidentalis*) showed the highest genetic distance values (Fig. 2A).

At the species level, IBD was significant for *P. caribaea* ($P < 0{\cdot}05$; Fig. 2B) and the eastern species ($r = 0{\cdot}815$, $P < 0{\cdot}0001$; Fig. 2E). In contrast, IBD can be rejected within varieties *hondurensis* and *caribaea* ($P = 0{\cdot}683$ and 0·283, respectively; Fig. 2C, D).

Within *P. caribaea* var. *hondurensis*, the geographically extreme populations, Moss and Waspam (south), and Dolores, Mountain Pine Ridge, Crooked Tree and Jaguactal (north), show small θ estimates (Table 1) and strong differentiation even at small distances (mean $\delta\mu^2 = 0{\cdot}491$).

Tests of the IBD model for the three *P. caribaea* groups resulting from BAPS were non-significant (data not shown). Only the eastern group (Group IV), which corresponds to data shown in Fig. 2E, had a significant IBD structure.

Genetic relationships among taxa

The plastid haplotype network revealed a structure tightly associated with geography (Fig. 3). There were no shared plastid haplotypes among the three varieties of *P. caribaea* or between *P. caribaea* and the other three species. Rarefaction analysis showed that within each *P. caribaea* variety, sampling twice as many individuals would only result in finding one more plastid haplotype (see Supplementary Data Table S2 and Supplementary Data Figures S1 and S2). Thus, it is unlikely that increasing sample size would result in shared plastid haplotypes between varieties. In contrast, *P. cubensis* and *P. occidentalis* shared one plastid haplotype (6) and *P. cubensis* and *P. maestrensis* share three plastid haplotypes (42, 43 and 44). Shared plastid haplotypes between *P. cubensis* and *P. maestrensis* represented 44 and 61 % of sampled individuals, respectively.

Plastid haplotypes in *P. caribaea* var. *bahamensis* were more closely related to those of *P. caribaea* var. *hondurensis* (Fig. 3). Plastid haplotypes of *P. caribaea* var. *bahamensis* are only indirectly connected to plastid haplotypes of var. *caribaea* through var. *hondurensis*. Two of the most connected plastid haplotypes (60 and 66, Fig. 3) are distributed in Central America. Also, some abundant plastid DNA variants are close to the tips of the network, corresponding to the south-eastern (55, 76 and 84) and northern (65 and 73) extremes of the distribution of *P. caribaea* var. *hondurensis*.

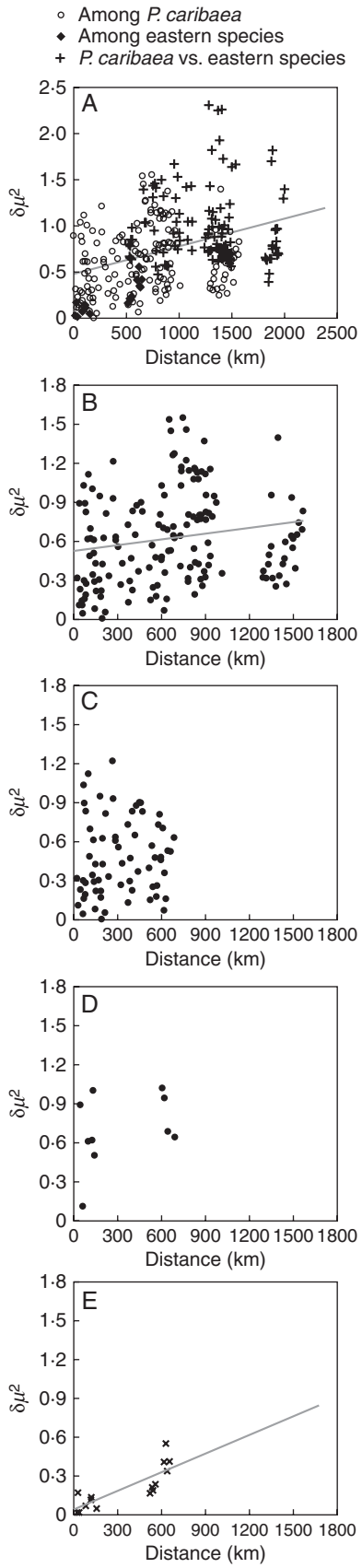

FIG. 2. Mantel test between geographical and Goldstein's genetic distance ($\delta\mu^2$). A regression line is drawn when significant correlation was found after 10 000 permutations. (A) Total data set ($r = 0{\cdot}375$, $P < 0{\cdot}0001$), showing comparisons among *P. caribaea* populations, among eastern species, and between *P. caribaea* and the eastern species, as indicated. (B) All *P. caribaea* populations ($r = 0{\cdot}178$, $P < 0{\cdot}05$). (C) Central America populations of *P. caribaea*. (D) Cuban populations of *P. caribaea*. (E) Eastern species populations ($r = 0{\cdot}815$, $P < 0{\cdot}0001$).

Phylogeographical structure

Total genetic diversity ($H_T = 0{\cdot}982$) was higher than average within populations ($H_S = 0{\cdot}770$). A permutation test revealed that R_{ST} (0·671) is significantly higher than mean permuted R_{ST} (0·101; $P < 0{\cdot}0001$), indicating geographical association of closely related plastid haplotypes. This pattern was consistent for *P. caribaea* populations for which $R_{ST} = 0{\cdot}769$ (mean permuted $R_{ST} = 0{\cdot}230$; $P < 0{\cdot}005$) and eastern species populations for which $R_{ST} = 0{\cdot}607$ (mean permuted $R_{ST} = 0{\cdot}112$; $P < 0{\cdot}005$).

Bayesian analysis of population structure gave an optimal partition of the populations into four groups ($P = 0{\cdot}999$; Fig. 4A). This structure had the largest log maximum-likelihood value among partitions (−1794·660). This analysis maximizes the posterior probability of a structure (given a maximum number of groups, *K*), assigning individuals (not populations) to each cluster. So, it was possible to identify some plastid haplotypes that can be assigned to geographically distant clusters.

P. caribaea populations formed three clusters: group I, comprising the northern range of Central America (*P. caribaea* var. *hondurensis*) and the populations from Bahamas (*P. caribaea* var. *bahamensis*); group II, the south-eastern part of the *P. caribaea* var. *hondurensis* distribution in Central America, except for nine individuals of the Mezapa population (*P. caribaea* var. *hondurensis*), which were clustered with group III; and group III, a single cluster comprising all *P. caribaea* var. *caribaea* populations except for six individuals of the Marbajita population (*P. caribaea* var. *caribaea*) which were clustered with group II. The populations of the eastern species were clustered in a fourth group (group IV).

Historical demography

The mismatch distribution analysis did not reject the null hypothesis of sudden growth in three of the clusters identified with BAPS. The only case in which this hypothesis was rejected was for *P. caribaea* var. *caribaea* group III (Fig. 4B). The time for population expansion could be estimated for the remaining groups. For example, group I of BAPS clustering corresponding to *P. caribaea* from northern Central America and Bahamas showed a distribution congruent with the sudden expansion model: $\tau = 6{\cdot}054$; estimated time since expansion, 331 100 yr BP (CI: 181 500–457 600; Navascués *et al.*, 2009). Even so, it is reasonable to suppose that *P. caribaea* populations in New Providence and Andros have a common origin because they belong to the same coral formation, and thus we computed the mismatch distribution for this and obtained a $\tau = 1{\cdot}790$ (CI: 0·833–2·637), which, considering a generation time of 42·5 years, provided

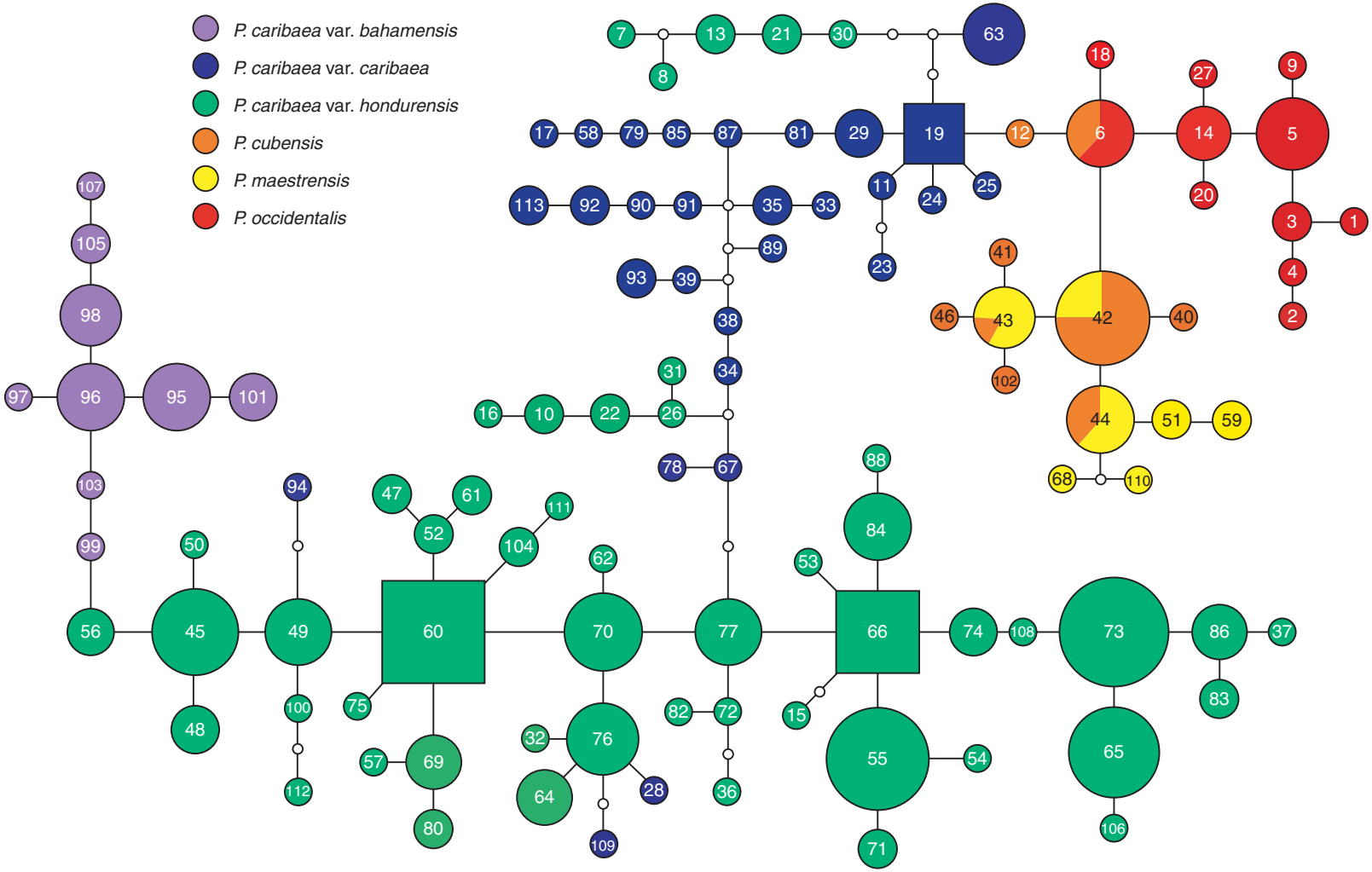

FIG. 3. Minimum spanning network for the plastid haplotypes in the *Australes* species of Caribbean pines. The area of each plastid haplotype is proportional to its abundance. Colour describes the species in which a particular plastid haplotype is found; when it is found in more than one species, colour division of the symbol is proportional to the contribution of each species to the total abundance. Each line represents a mutational step; missing plastid haplotypes are represented with circles. The squares correspond to the three most connected plastid haplotypes (19, 60 and 66).

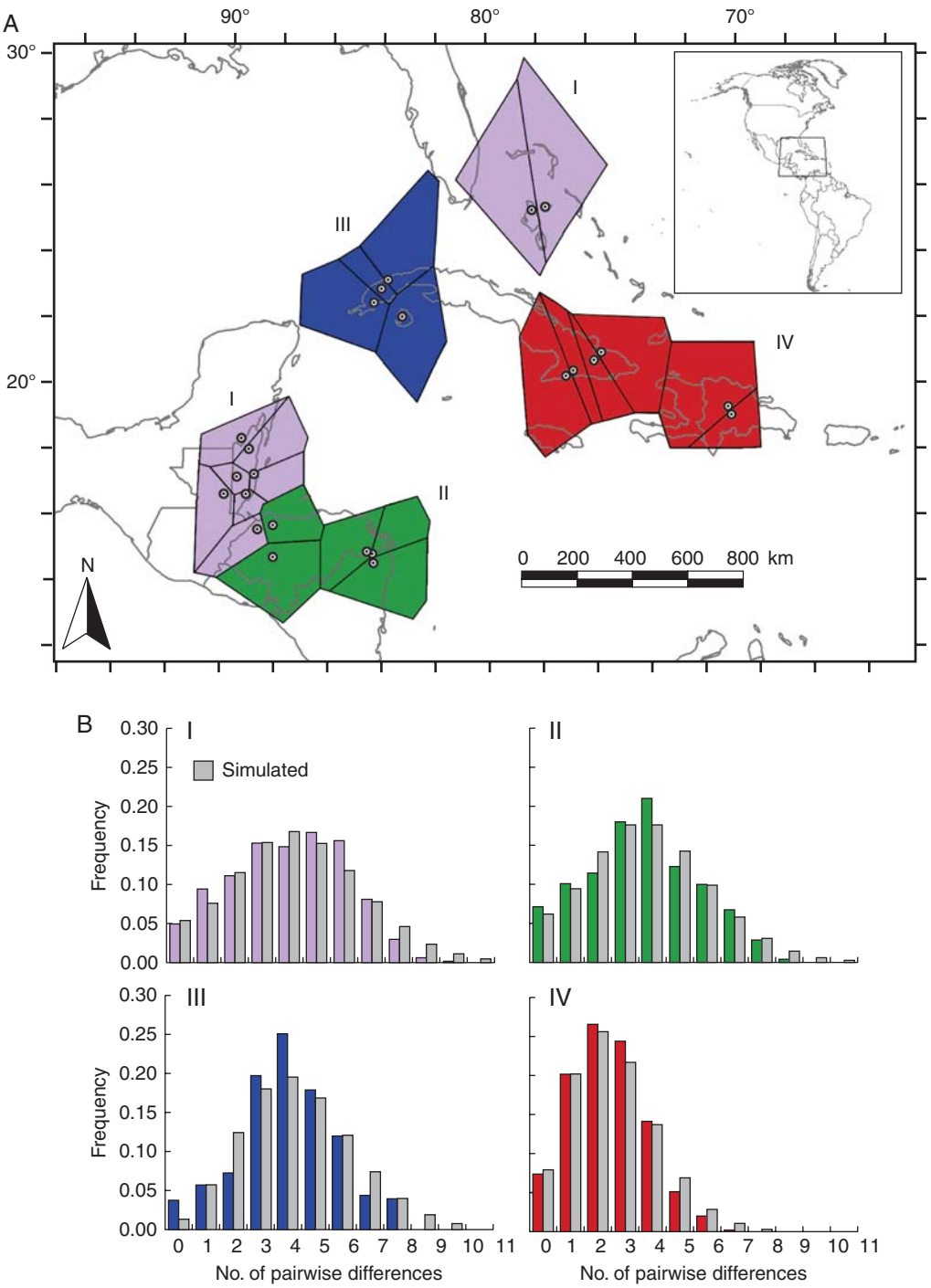

FIG. 4. (A) Bayesian analysis of population structure (BAPS) results. Clusters correspond to the best partition of the data (log maximum likelihood = −1794·6602; $P = 0{\cdot}999$) and are indicated by different colour Voronoi polygons. Collection sites are indicated by circles. (B) Mismatch distribution for each group, Roman numerals refer to the BAPS clusters; observed differences are shown in colour bars, and simulated distributions in grey bars. Parameters estimated the under sudden expansion model were: (I) $\tau = 6{\cdot}054$ (CI 3·618–8·366), (II) $\tau = 5{\cdot}965$ (CI 1·020–7·176), (IV) $\tau = 2{\cdot}921$ (CI 1·850–3·509); cluster (III) deviated significantly from the null hypothesis of sudden expansion ($P = 0{\cdot}0477$).

an estimate of time to expansion of 97 900 yr BP (CI: 45 600–144 200). The estimate of the time to expansion for group II of *P. caribaea* var. *hondurensis* was similar to that for group I ($\tau = 5{\cdot}965$ and 6·054, respectively). This is congruent with the high abundance in relatively derived plastid haplotypes (Fig. 3). Time to expansion calculated for Group II is 326 300 yr BP (CI: 150 300–511 200). Finally, group IV from BAPS (*P. cubensis*, *P maestrensis* and *P. occidentalis*) showed a lower τ estimate (2·921), which gave an estimated time to expansion of 159 800 yr BP (CI: 101 200–192 000).

DISCUSSION

Genetic variation

Genetic diversity in terms of plastid haplotype diversities (H_E, Table 1) for the taxa included was comparable with other pine species that have been studied with plastid SSRs, including *P. pinaster* (0·870, Ribeiro *et al.*, 2001; 0·927, Gómez *et al.*, 2005), *P. canariensis* (0·73, Gómez *et al.*, 2003), *P. nigra* (0·954, Afzal-Rafii and Dodd, 2007) and *P. strobiformis* (0·856, Moreno-Letelier and Piñero, 2009).

The indirect estimators of effective population size (θ) for *P. caribaea* were high in comparison with species from the east of the Caribbean Basin, namely *P. cubensis*, *P. maestrensis* and *P. occidentalis*. This pattern is congruent with the size of the species range, with the most widely distributed taxon (*P. caribaea*) having the highest theta value ($\theta_{IAM} = 40{\cdot}270$).

High values of θ found in San Jerónimo, Cofradía and Mezapa in Central America and in Arenas Blancas and Viñales in Cuba appear to be related to gene flow among these populations, as deduced from the origin of plastid haplotypes 7, 8, 13, 21, 28, 30, 94 and 109. As can be seen in Fig. 3, plastid haplotypes 7, 8, 13, 21 and 30 present in Central America populations are in a network branch derived from an ancestral plastid haplotype, 19, present only in Cuba. Similarly, plastid haplotypes 28, 94 and 109 were found in Cuban populations, but are derived from ancestral plastid haplotypes found only in Central America. Given the otherwise consistent geographical association of the network branches and the absence of shared plastid haplotypes among *P. caribaea* varieties, it is plausible that gene flow among certain populations in Central America and Cuba explains both the distribution of these plastid haplotypes and their contribution to local genetic diversity, particularly in the case of Cuban populations.

Contrasting patterns of IBD and metapopulation dynamics

Within *P. caribaea* and the eastern species, the genetic distances are correlated with geographical distance among populations, indicating that populations within each region are less isolated from each other than populations from different regions.

When analysed separately, Central American populations of *P. caribaea* did not fit the IBD model, and this result was corroborated when Mantel tests were conducted for the BAPS clusters of this region (data not shown). The IBD model assumes that populations are in equilibrium, such that genetic structure is due to geographical restriction of gene flow (Epperson, 1995; Hardy and Vekemans, 1999). Absence of IBD could be explained by the presence of highly differentiated populations with low θ_{IAM} and θ_{SMM} estimates (in the south: Moss and Waspam; in the north: Deep River, Mullins River, Mountain Pine Ridge and Jaguactal). This suggests a metapopulation dynamic (Slatkin, 1977) in which some populations have been recently founded and others have gone extinct. In this scenario, and due to the action of genetic drift, plastid haplotypes common in some populations are rare in the rest of the distribution area (Whitlock, 2004). This contributes to higher $\delta\mu^2$ values because flow has not diluted the differences among populations (Slatkin, 1977; Pannel and Charlesworth, 2000; Whitlock, 2004). This metapopulation dynamic could explain the distribution of genetic variation in *P. caribaea* from Central America and is congruent with the susceptibility of this species to the balance of humidity and periodic fire (Borhidi, 1988; Dvorak *et al.*, 2005). The precise testing of this type of dynamic would require further detailed studies about the demographic and ecological processes in these populations.

Central American origin for P. caribaea *and eastern species as a separate lineage*

Ancestral plastid haplotypes are likely to have given rise to more derived haplotypes than younger haplotypes, because in ancestral haplotypes mutation has occurred over a longer period of time (Posada and Crandall, 2001). As a consequence, older plastid haplotypes tend have more connections in a network. Although homoplasy and high mutation rates (reflected in θ) can bias this pattern, highly connected plastid haplotypes tend to be closely related to ancestral plastid haplotypes (Alavez, 2008). Thus, the presence of highly connected plastid haplotypes (Fig. 3) together with the distribution of genetic diversity could indicate that Central America is the ancestral area in the plastid haplotype distribution.

The plastid haplotype network also suggests that the distribution of genetic diversity in *P. caribaea* is the result of two independent colonization events to the islands, one to Cuba and another to Bahamas. The plastid haplotypes of *P. caribaea* var. *bahamensis* (Bahamas) were more closely related to Central American plastid haplotypes than to their Caribbean counterparts. Meanwhile, plastid haplotypes of *P. caribaea* var. *caribaea* (Cuba) had different Central American relatives. The BAPS clustering further confirmed such a close relationship between populations from the Bahamas and Central America (Fig. 4B).

Some authors have suggested a North American ancestor for *P. caribaea* (Farjon, 1996; Adams and Jackson, 1997). To date, phylogenetic analyses of *Pinus* based on plastid DNA have been unresolved with respect to the relationships between North American and Caribbean species of subsection *Australes* (Geada-López *et al.*, 2002; Gernandt *et al.*, 2005). Furthermore, it is important to remember that the lineage history of a single locus (such as the plastid genome) does not necessarily coincide with the species history (Rosenberg and Nordborg, 2002; Hey and Machado, 2003; Avise, 2009) and discordance is known to exist between nuclear and plastid phylogenetic analyses in other pines (Willyard *et al.*, 2009). Thus, even if the hypothesis of a North American origin of *P. caribaea* were true, the results of this study still reflect the historical distribution of the plastid lineages within the species. Moreover, both this study and that of Adams and Jackson (1997) agree that there have been long-distance colonization events in the evolution of subsection *Australes* pines in the Caribbean Basin, although the current study detected a different direction for such events.

The close relationship between plastid haplotypes from *P. cubensis*, *P. maestrensis* and *P. occidentalis* and those of *P. caribaea* var. *caribaea* suggests that eastern species

diverged from ancestors in western Cuba rather than from Central America. This can be explained by long-distance colonization or by an ancient fragmentation event (Cruzan and Templeton, 2000). The colonization pathway to Hispaniola from Central America through Cuba has also been inferred for other plant lineages: *Ernodea* (Negrón-Ortiz and Watson, 2003), *Exostema* (McDowell *et al.*, 2003), *Ginoria* (Graham, 2003) and Solanaceae (Santiago-Valentín and Olmstead, 2003).

Unlike the *P. caribaea* varieties, the eastern species show many morphological differences and have been recognized as different species (López-Almirall, 1982; Farjon and Styles, 1997). Nevertheless, there is a high proportion of shared plastid haplotypes among species in the east of the basin. This could be due to recurrent gene flow or shared ancestral polymorphism (Avise, 2009). Shared variation could be better explained by incomplete lineage sorting, or we should expect a smaller degree of morphological differentiation or abundant intermediate phenotypes, as a result of sustained gene flow.

The descendants of a mutation event are likely to remain in the population where they originated unless migration occurs (Posada and Crandall, 2001). What is observed in the plastid haplotype network is precisely that some descendants of such a mutation event are distributed away from where their ancestors occur. The presence in Central America of two clades of plastid haplotypes closely related to Cuban ones and, similarly, the presence of three Cuban plastid haplotypes associated with continental ones (Fig. 2) suggest gene flow between these two regions in both directions. Shared ancestral polymorphism can be discarded because there are no shared plastid haplotypes.

Stochasticity of gene genealogies may mislead the inferred processes from haplotype networks (Hey and Machado, 2003). Nevertheless, the existence of gene flow between Central America and Cuba is congruent with present biogeographical evidence for the region, showing dispersion in both directions, from Central America to the Antilles (Negrón-Ortiz and Watson, 2003; Santiago-Valentín and Olmstead, 2004; Perdices *et al.*, 2005; Ricklefs and Bermingham, 2008) and from the Antilles to Central America (Dávalos, 2007). The absence of shared plastid haplotypes among *P. caribaea* varieties could be explained if island populations have gone through bottlenecks after colonizations from Central American populations.

Two colonization events to the Caribbean islands

The R_{ST} analysis reveals a phylogeographical structure for the pines of the Caribbean Basin. The higher structure detected using ordered alleles indicates that pairs of plastid haplotypes from the same population are more similar than those of separate populations (Pons and Petit, 1996). Under a stepwise mutation model, the number of differences between plastid haplotypes reflects their genealogical relationship. Thus, high R_{ST} values reflect a spatial clustering of the majority of the closely related plastid haplotypes, a result that is congruent with the structure of the plastid haplotype network.

Results from BAPS confirm that populations of the Caribbean Basin are subdivided. The assignment of *P. caribaea* var. *bahamensis* populations to group I (mostly *P. caribaea* var. *hondurensis* populations) is striking because Bahamian populations are closer to Cuba than to Central America. Nevertheless, this finding is in agreement with the close relationship between Bahamian and Central American plastid haplotypes in the network. The plastid haplotype network relies on the number of steps that connect the haplotypes using statistical parsimony, but BAPS relies on a different statistical approach and considers only haplotype identity. Taken together, these results suggest an origin of the *P. caribaea* var. *bahamensis* plastid haplotype from Central America. Considering the recent origin of the Bahamas and that they have been separated from the continent, over-sea dispersion apparently explains these results.

Group II (the south-eastern range of *P. caribaea* var. *hondurensis*) contained some highly diverse populations, e.g. San Jerónimo and Mezapa, and also some of the most connected plastid haplotypes, (60, 66, 76 and 77). Such evidence enables higher accuracy in identifying this as a possible ancestral area.

The majority of individuals of *P. caribaea* var. *caribaea* were grouped in a separate cluster (group III), but some plastid haplotypes of this variety were assigned to the Central American group (II), which is suggestive of gene flow between different regions. Such flow must have taken place across the sea, given that Cuba has not been connected to the continent in the last 30 Myr (Iturralde-Vinent, 2006). *P. caribaea* is one of the few pines of subgenus *Pinus* with adnate seed wings, which has been assumed as an adaptation to long-distance dispersal (Farjon and Styles, 1997). Although there is variation in seed wings among *P. caribaea* varieties, ranging from almost completely articulate (var. *bahamensis*) to adnate (var. *caribaea*), in var. *hondurensis* a wide variation with both types of seeds has been described (Farjon and Styles, 1997; Eckenwalder, 2009). Seed-mediated gene flow could explain both the long distance over which it took place and the relative scarcity of plastid interchange between varieties, not reinforced by pollen gene flow, and this is reflected in the absence of plastid haplotype sharing among varieties.

Due to shorter coalescence times in linked haploid genes compared with unlinked diploid nuclear genes (Rosenberg and Nordborg, 2002; Avise, 2009), the genetic variation in plastid haplotypes reflects recent history. The absence of shared haplotypes among *P. caribaea* varieties may not occur when nuclear markers are analysed, and caution should be exercised when comparing the species and plastid histories as plastid data reflect a more recent history than nuclear data.

Our data suggest that the plastids of *P. cubensis*, *P. maestrensis* and *P. occidentalis* evolved separately from those in the rest of the basin, based on the pattern of plastid haplotype sharing and in their assignment to a single BAPS cluster. Within this group it is possible to distinguish *P. occidentalis* as closely related to *P. cubensis* and *P. maestrensis*, but sharing a small (9 %) proportion of variation with them. This is also congruent with the geological history of the ancient separation between eastern Cuba and northern Hispaniola (Iturralde-Vinent, 2006).

P. cubensis and *P. maestrensis* share a significant proportion of plastid haplotypes (33 and 43 %, respectively). It is difficult

to distinguish between recent divergence and recurrent gene flow from our data, but, as explained previously, incomplete lineage sorting is compatible with the morphological differentiation of these species.

Demographic expansions during glacial periods

The expansion time in group I (North Central America and Bahamas) of 331 100 yr BP is on the border between an interglacial period corresponding to Marine Isotope Stage (MIS) 9 and the beginning of the temperature decrease that led to MIS 8 (EPICA Community Members, 2004). The lower elevations of Andros and New Providence relative to Cuba or Hispaniola meant that a considerable proportion of the Bahamas was submerged due to higher temperatures and higher sea level of the past interglacial (Duplessy *et al.*, 2008). In this scenario some plastid lineages could have been fixed by chance after land re-emerged. According to our data, such lineages were those closely related to the Central American ones. In support of this, the expansion time estimation of 97 900 yr BP for Bahamian populations roughly corresponds to the beginning of the last glacial period (110 000 yr BP; Petit *et al.*, 1999; EPICA Community Members, 2004).

Group II, corresponding to the south-eastern range of *P. caribaea* var. *hondurensis*, also shows evidence of a sudden expansion. The time estimate of 326 300 yr BP is at the beginning of the temperature decrease that led to the MIS 8 cold period (Petit *et al.*, 1999; EPICA Community Members, 2004). Population expansion inference in groups I and II is also reinforced by the presence of abundant plastid haplotypes close to the tips of the network. Sudden expansion of group IV (eastern Cuba and Hispaniola) would have taken place around 159 800 yr BP, close to the cold period of MIS 6.

Estimates of expansion times corresponding to glacial periods are coherent with the biology of these pine species: survival of *P. caribaea* populations depends on periodic fire, otherwise they are replaced by tropical rain forest (Borhidi, 1988; Dvorak *et al.*, 2005). Present knowledge of glacial climate indicates drier environments at latitudes resembling Central America and the Caribbean (Ray and Adams, 2001). In addition, both continent and islands had more emerged area (i.e. more available habitat) during glacial periods. There is evidence that the range of *P. caribaea* var. *hondurensis* has been contracting during the Holocene interglacial due to progressive increases in humidity (Dvorak *et al.*, 2005), which supports the inference of larger populations during colder periods. Recent findings in Lake Petén-Itzá cores in Guatemala show a correlation between colder periods and an increase in pollen of *Pinus* and Poaceae concomitant with increasing fire incidence in Central America (Hodell *et al.*, 2008). This pattern reinforces the idea of expansion of *P. caribaea* savannah during glacial periods.

The wider altitudinal ranges of *P. cubensis*, *P. maestrensis* and *P. occidentalis* probably allowed them to occupy the greater available lowlands as the mountain crests became cooler and drier. In addition, the smaller reduction of monsoons in the eastern part of the basin than in Central America should have provided humid land for this group of species, adding support to population expansion during glaciations.

Our study reveals close correspondence between time to expansion and glacial periods in *P. caribaea* var. *bahamensis* (small CI in τ estimates). Previous palaeoclimate studies have shown that during the cooler periods, populations of *P. caribaea* var. *hondurensis* were larger (Dvorak *et al.*, 2005; Hodell *et al.*, 2008). In the context of current global warming, historical demography data for *P. caribaea* alert us to the possible consequences for its populations. Bahamian populations are particularly sensitive to a global warming scenario given the low elevation predominant in these islands and could be drastically reduced because the detected demographic expansion occurred when the climate was colder. Our data suggest that the cool periods were also important in the expansions of Central American *P. caribaea* var. *hondurensis*-dominated savannahs. If so, global warming may lead to *P. caribaea* population contraction or, in the case of Bahamas, an increase in its probability of extinction.

The estimated times to expansion come from only one genome (plastid) and thus may not reflect the complete history of pines in the region. Also, there is stochastic variance both in gene trees and in mutations even in genes that have undergone a similar demographic history (Hey and Machado, 2003). Nevertheless, although the variance in τ estimate is also large (Fig. 4.), it has been demonstrated through simulations that the estimate of τ obtained by the maximum-pseudolikelihood method is close to the true τ value, providing an accurate estimator of time to expansion (Navascués *et al.*, 2009). Moreover, correspondence between estimated expansion times and glacial periods is biologically meaningful and therefore its discussion in this paper was relevant.

In summary: (1) possible ancestral plastid haplotypes and highest gene diversity and θ estimates are found today in Central America; (2) population expansion has played a major role in the distribution of genetic diversity in *P. caribaea* var. *hondurensis*; (3) independent colonization events gave rise to the lineages of *P. caribaea* var. *bahamensis* and of *P. caribaea* var. *caribaea*; (4) plastid lineages of *P. cubensis*, *P. maestrensis* and *P. occidentalis* evolved independently and are separate from *P. caribaea*; (5) there is incomplete lineage sorting between *P. cubensis* and *P. maestrensis* plastids, suggesting recent divergence; and (6) all estimated expansion times correspond to glacial periods.

SUPPLEMENTARY DATA

Supplementary data are available online at www.aob.oxfordjournals.org and consist of the following files. Table S1: Plastid haplotype alellic composition and frequency of each plastid haplotype, and overall relative frequency. Table S2: Expected change in plastid haplotype diversity when doubling the sample size for each lineage. Figure S1: Expected accumulation of plastid haplotypes for the *Pinus* species studied depending on sample size. Figure S2: Expected accumulation of plastid haplotypes for *P. caribaea* varieties depending on sample size.

ACKNOWLEDGMENTS

We thank L. Rodriguez, E. García, R. Sotolongo, Y. García, A. Palmarola, M. Mejía, C. Espadas, M. Ferrer, V. Rebolledo,

F. Chi and D. Escudero, Leon University Herbarium, Pinar del Río University, Integrated Forest Enterprise of Mayarí, Guisa Forestry Station, Belize Wildlife Service, Santo Domingo Botanical Garden for work in the field. Thanks to L. Espinoza-Azuar who helped using the ABI 3100 and to O. Rodriguez who helped with lab work. R. Salas, A. Moreno-Lettelier, V. Alavez and D. Ortega provided insightful discussions about plastid haplotype networks and phylogeography. C. Espadas and E. Scheinvar are thanked for help with Fig. 4. Thanks to D.S. Gernandt and J. Jaramillo for their comments to improve this manuscript. We thank M. Navascués and C. Burgarella for providing their source code to implement pseudo-likelihood analyses. This work was supported by CONACYT, (grant number 185586 to LJB), by SEP-CONACYT, (grant number 44373 to PD), SEP-CONACYT (grant number 46925 to DP) and SEMARNAT-CONACYT (grant number 2002-C01-201 to DP). This work is part of the PhD in Biomedical Sciences of Lev Jardón-Barbolla at the Universidad Nacional Autónoma de México.

LITERATURE CITED

Adams DC, Jackson JF. 1997. A phylogenetic analysis of the southern pines (*Pinus* subsect. *Australes* Loudon): biogeographical and ecological implications. *Proceedings of the Biological Society of Washington* **110**: 681–692.

Afzal-Rafii Z, Dodd R. 2007. Chloroplast DNA supports a hypothesis of glacial refugia over postglacial recolonization in disjunct populations of black pine (*Pinus nigra*) in western Europe. *Molecular Ecology* **16**: 723–736.

Alavez V. 2008. *Evolución de microsatélites de cloroplasto en* Pinus pseudostrobus *y* Pinus montezumae*: efecto del ligamiento en la estimación de parámetros poblacionales y aplicaciones en genética de poblaciones*. BS Thesis, Universidad Nacional Autónoma de México, México.

Avise JC. 2009. Phylogeography, retrospect and prospect. *Journal of Biogeography* **36**: 3–15.

Barrett WHG, Golfari L. 1962. Descripción de dos nuevas variedades del pino de Caribe. *Caribbean Forestry* (in Spanish) **23**: 59–71.

Borhidi A. 1988. Vegetation dynamics of the savannization process on Cuba. *Plant Ecology* **77**: 177–183.

Bucci G, Gonzalez-Martinez SC, Le Provost G, *et al*. 2007. Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Molecular Ecology* **16**: 2137–2153.

Chen J, Tauer CG, Huang Y. 2002. Paternal inheritance patterns in pine hybrids detected with *trnL-trnF* intergenic region polymorphism. *Theoretical and Applied Genetics* **104**: 1307–1311.

Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**: 1657–1660.

Corander J, Sirén J, Arjas E. 2008. Bayesian spatial modelling of genetic population structure. *Computational Statistics* **23**: 111–129.

Crandall KA, Templeton AR. 1993. Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* **134**: 959–969.

Cruzan MB, Templeton AR. 2000. Palaeoecology and coalescence: phylogeographic analysis of hypotheses from the fossil record. *Trends in Ecology and Evolution* **15**: 491–496.

Cuenca A, Escalante AE, Piñero D. 2003. Long-distance colonization, isolation by distance, and historical demography in a relictual Mexican pinyon pine (*Pinus nelsonii* Shaw) as revealed by paternally inherited genetic markers (cpSSR). *Molecular Ecology* **12**: 2087–2097.

Dávalos LM. 2007. Short-faced bats (Phyllostomidae: Stenodermatina): a Caribbean radiation of strict frugivores. *Journal of Biogeography* **34**: 364–375.

Delgado P, Salas-Lizana R, Vázquez-Lobo A, *et al*. 2007. Introgressive hybridization in *Pinus montezumae* Lamb. and *Pinus pseudostrobus* Lindl. (Pinaceae): morphological and molecular (cpSSR) evidence. *International Journal of Plant Sciences* **168**: 861–875.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf material. *Phytochemical Bulletin, Botanical Society of America* **19**: 11–15.

Duplessy JC, Charbit S, Kageyama M, Masson-Delmotte V. 2008. Insolation and sea level variations during Quaternary interglacial periods: a review. *Geoscience* **340**: 701–710.

Dvorak WS, Jordon AP, Hodge GP, Romero JL. 2000. Assessing the evolutionary relationships of the pines in the *Oocarpae* and *Australes* subsections using RAPD markers. *New Forests* **20**: 163–192.

Dvorak WS, Hamrick JL, Gutierrez EA. 2005. The origin of the Caribbean pine in the seasonal swamps of the Yucatán. *International Journal of Plant Sciences* **166**: 985–994.

Eckenwalder JF. 2009. *Conifers of the world: the complete reference.* Portland, OR: Timber Press.

EPICA Community Members. 2004. Eight glacial cycles from an Antarctic ice core. *Nature* **429**: 623–628.

Epperson BK. 1995. Spatial distribution of genotypes under isolation by distance. *Genetics* **140**: 1431–1440.

Estoup A, Rousset F, Michalakis Y, Cornuet JM, Adriamanga M, Guyomard R. 1998. Comparative analysis of microsatellite and allozyme markers: a case study investigating microgeographic differentiation in brown trout (*Salmo trutta*). *Molecular Ecology* **7**: 339–353.

Estoup A, Jarne P, Cornuet JM. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetic analysis. *Molecular Ecology* **11**: 1591–1604.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3·1: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**: 47–50.

Farjon AF. 1996. Biodiversity of *Pinus* (Pinaceae) in Mexico: speciation and paleo-endemism. *Botanical Journal of the Linnean Society* **121**: 365–384.

Farjon AF, Styles BT. 1997. *Flora Neotropica,* Pinus *(Pinaceae)*. New York: The New York Botanical Garden.

Geada-López G, Kamiya K, Harada K. 2002. Phylogenetic relationships of *Diploxylon* pines (subgenus *Pinus*) based on plastid sequence data. *International Journal of Plant Sciences* **163**: 737–747.

Gernandt D, Geada-López G, Ortiz-García S, Liston A. 2005. Phylogeny and classification of *Pinus*. *Taxon* **54**: 29–42.

Gernandt D, Magallón S, Geada-López G, Zenón-Flores O, Willyard AS, Liston A. 2008. Use of simultaneous analyses to guide fossil-based calibrations of Pinaceae phylogeny. *International Journal of Plant Sciences* **169**: 1086–1099.

Gillespie JH. 2004. *Population genetics: a concise guide*, 2nd edn. Baltimore, MD: The Johns Hopkins University Press.

Godbout J, Fazekas A, Newton CH, Yehm FC, Bousquet J. 2008. Glacial vicariance in the Pacific Northwest; evidence from a lodgepole pine mitochondrial DNA minisatellite for mutiple genetically distinct and widely separated refugia. *Molecular Ecology* **17**: 2463–2475.

Goldstein DB, Ruíz-Linares A, Feldman M, Cavalli-Sforza LL. 1995. An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**: 463–471.

Gómez A, González-Martínez SC, Collada C, Climent J, Gill L. 2003. Complex population genetic structure in the endemic Canary Island pine revealed using chloroplast microsatellite markers. *Theoretical and Applied Genetics* **107**: 1123–1131.

Gómez A, Vendramin GG, González-Martínez SC, Alía R. 2005. Genetic diversity and differentiation of two Mediterranean pines (*Pinus halepensis* Mill. and *Pinus pinaster* Ait.) along a latitudinal cline using chloroplast microsatellite markers. *Diversity and Distributions* **11**: 257–263.

Graham A. 1999. Studies in neotropical palaeobotany. XIII. An Oligo-Miocene palynoflora from Simojovel (Chiapas, Mexico). *American Journal of Botany* **86**: 17–31.

Graham SA. 2003. Biogeographical patterns of Antillean Lythraceae. *Systematic Botany* **28**: 410–420.

Hardy OJ, Vekemans X. 1999. Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity* **83**: 145–154.

Hardy OJ, 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 OJ, Charbonnel N, Fréville H, Heuertz M. 2003. Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics* **163**: 1463–1482.

Hey J, Machado CA. 2003. The study of structured populations – new hope for a difficult and divided science. *Nature Reviews Genetics* **4**: 535–543.

Hodell DA, Anselmetti FS, Aiztegui D, *et al*. 2008. An 85-ka record of climate change in lowland Central America. *Quaternary Science Reviews* **27**: 1152–1165.

Iturralde-Vinent M. 2006. Meso-Cenozoic Caribbean paleogeography: implications for historical biogeography of the region. *International Geology Review* **48**: 791–827.

Jasińska AK, Wachowiak W, Muchewicz E, Boratyńska K, Montserrat JM, Boratyński A. 2010. Cryptic hybrids between *Pinus uncinata* and *P. sylvestris*. *Botanical Journal of the Linnean Society* **163**: 473–485.

Kuhner MK. 2009. Coalescent genealogy samplers: windows into population history. *Trends in Ecology and Evolution* **24**: 86–93.

Lapègue D, Petit RJ. 1997. Phylogeographic structure of white oaks throughout the European continent. *Genetics* **146**: 1475–1487.

Lavin M, Wojchiechowski MF, Gasson P, Hughes C, Wheeler E. 2003. Phylogeny of robinioid legumes (Fabaceae) revisited: *Coursetia* and *Gliricida* recircumscribed, and biogeographical appraisal of the Caribbean endemics. *Systematic Botany* **28**: 387–409.

López-Almirall A. 1982. Variabilidad del género *Pinus* (Coniferales: Pinaceae) en Cuba. *Acta Botánica Cubana* **12**: 1–32.

McDowell T, Volovosek M, Manos P. 2003. Biogeography of *Exostema* (Rubiaceae) in the Caribbean region in the light of molecular phylogenetic analyses. *Systematic Botany* **28**: 431–441.

Moreno-Letelier A, Piñero D. 2009. Phylogeographic structure of *Pinus strobiformis* Engelm. across the Chihuahuan Desert filter-barrier. *Journal of Biogeography* **36**: 121–131.

Navascués M, Vaxevanidou Z, González-Martínez SC, Climent J, Gil L, Emerson BC. 2006. Chloroplast microsatellites reveal colonization and metapopulation dynamics in the Canary Island pine. *Molecular Ecology* **15**: 2691–2698.

Navascués M, Hardy OJ, Burgarella C. 2009. Characterization of demographic expansions from the pairwise comparisons of linked microsatellite haplotypes. *Genetics* **181**: 1013–1019.

Negrón-Ortiz V, Watson LE. 2003. Hypotheses for the colonization of the Caribbean Basin by two genera of the Rubiaceae: *Erithalis* and *Ernodea*. *Systematic Botany* **28**: 442–451.

Nei M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.

Ohta T, Kimura M. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research* **22**: 201–204.

Omland KE, Lanyon SM, Fritz SJ. 1999. A molecular phylogeny of the New World orioles (*Icterus*): the importance of dense taxon sampling. *Molecular Phylogenetics and Evolution* **12**: 224–239.

Pannel JR, Charlesworth B. 2000. Effects of metapopulation processes on measures of genetic diversity. *Philosophical Transactions of the Royal Society of London* **355**: 1851–1864.

Perdices A, Doadrio I, Bermingham E. 2005. Evolutionary history of the synbranchid eels (Teleostei: Synbranchidae) in Central America an the Caribbean islands inferred from their molecular phylogeny. *Molecular Phylogenetics and Evolution* **37**: 460–473.

Petit JR, Jouzel J, Raynaud D, *et al*. 1999. Climate atmospheric history of the past 420,000 years form the Vostok ice core, Antarctica. *Nature* **399**: 429–439.

Pindell J, Kennan L, Stanek KP, Maresch WV, Draper G. 2006. Foundations of Gulf of Mexico and Caribbean evolution: eight controversies resolved. *Geologica Acta* **4**: 303–341.

Pons O, Petit RJ. 1996. Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* **144**: 1237–1245.

Posada A, Crandall KA. 2001. Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution* **16**: 37–45.

Provan J, Soranzo N, Wilson NJ, Goldstein D, Powell W. 1999. A low mutation rate for chloroplast microsatellites. *Genetics* **153**: 943–947.

Ray N, Adams JM. 2001. A GIS-based vegetation map of the World at the Last Glacial Maximum (25,000–15,000 BP). *Internet Archaeology* **11**. http://intarch.ac.uk/journal/issue11/rayadams_toc.html

Ribeiro MM, Plomion C, Petit R, Vendramin GG, Szmidt AE. 2001. Variation in chloroplast single-sequence repeats in Portuguese maritime pine (*Pinus pinaster* Ait.). *Theoretical and Applied Genetics* **102**: 97–103.

Ricklefs R, Bermingham E. 2008. The West Indies as a laboratory of biogeography and evolution. *Philosophical Transactions of the Royal Society* (B) **363**: 2393–2413.

Rodríguez-Banderas A, Vargas-Mendoza CF, Buonamici A, Vendramin GG. 2009. Genetic diversity and phylogeographic analysis of *Pinus leiophylla*: a post-glacial range expansion. *Journal of Biogeography* **36**: 1807–1820.

Rogers AR, Harpending HC. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**: 552–569.

Rosenberg NA, Nordborg M. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics* **3**: 380–390.

Santiago-Valentín E, Olmstead RG. 2003. Phylogenetics of the Antillean Goetzoideae (Solanaceae) and their relationships within the Solanaceae based on chloroplast and ITS DNA sequence data. *Systematic Botany* **28**: 452–460.

Santiago-Valentín E, Olmstead RG. 2004. Historical biogeography of Caribbean plants: introduction to current knowledge and possibilities from a phylogenetic perspective. *Taxon* **53**: 299–319.

Schneider S, Excoffier L. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites. Application to human mitochondrial DNA. *Genetics* **152**: 1079–1089.

Slatkin M. 1977. Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical Population Biology* **12**: 253–262.

Szpiech ZA, Jackobson NA, Rosenberg NA. 2008. ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* **24**: 2498–2504.

Templeton AR. 2004. Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology* **13**: 789–809.

Trejo-Torres C, Ackerman JD. 2001. Biogeography of the Antilles based on a parsimony analysis of orchid distributions. *Journal of Biogeography* **28**: 775–794.

Vendramin GG, Lelli G, Rossi P, Morgante M. 1996. A set of primers for the amplification of 20 chloroplast microsatellites in Pinaceae. *Molecular Ecology* **5**: 595–598.

Whitlock MC. 2004. Selection and drift in metapopulations. In: Hanski I, Gaggiotti O. eds. *Ecology, genetics, and evolution of metapopulations*. Amsterdam: Elsevier Academic Press, 153–173.

Willyard A, Cronn R, Liston A. 2009. Reticulate evolution and incomplete lineage sorting in the ponderosa pines. *Molecular Phylogenetics and Evolution* **52**: 498–511.

Xu H, Fu YX. 2004. Estimating effective populations size or mutation rate with microsatellites. *Genetics* **166**: 555–563.