Journal of Biogeography (J. Biogeogr.) (2011)

Molecular and morphological support for a Florida origin of the Cuban oak

Paul F. Gugger* and Jeannine Cavender-Bares

Department of Ecology, Evolution and Behavior, University of Minnesota, 100 Ecology Building, 1987 Upper Buford Circle, Saint Paul, MN, USA

ABSTRACT

Aim The origins of the Cuban biota are of long-standing interest in biogeography, and the source of a small live oak (*Quercus* series *Virentes*) population on Cuba remains unresolved. Based on morphological evidence, previous authors have hypothesized a Florida origin from either *Q. geminata* or *Q. virginiana* or both; a Mexican origin from *Q. oleoides*; or a hybrid origin from both sources. We use molecular data and taxonomically informative leaf morphology to identify the source species and timing of colonization.

Location Cuba, Central America, Mexico and the south-eastern United States.

Methods We collected representative samples of Cuban oaks and each putative source species and genotyped each sample at 12 nuclear microsatellites and two chloroplast DNA sequences. We estimated population structure using a Bayesian clustering analysis and *F*-statistics, pairwise migration rates among taxa, and divergence time using an isolation-with-migration model. We measured seven leaf traits and conducted an analysis of similarity (ANOSIM) to determine which putative source species was most similar to Cuban oaks.

Results Cuban oak contains one chloroplast DNA haplotype, which is common in southern Florida. Bayesian clustering analysis of microsatellites revealed that the Cuban oak forms a distinct and pure population cluster, and *F*-statistics showed that Cuban oaks are differentiated least from *Q. virginiana* and most from *Q. geminata*. Migration rates were highest out of Cuba to *Q. oleoides*. Molecular diversity, the ratio of allelic richness to allele size range, and effective population size of the Cuban oak were relatively low, suggesting a founder effect. Divergence time estimates fell entirely within the Pleistocene (628–6 ka), considering a range of mutation rates and generation times. Cuban oaks were morphologically most similar to *Q. virginiana* and least similar to *Q. geminata*.

Main conclusions Molecular and morphological data support a Pleistocene dispersal of *Q. virginiana* from Florida to Cuba, followed by isolation and divergence, then limited dispersal and introgression from Cuba to *Q. oleoides* in Central America. Birds could have dispersed acorns to Cuba during a glacial period when sea levels were low. These results highlight the varied origin of the Cuban biota and the possible role of Pleistocene glaciations in the establishment of temperate taxa in the tropics.

Keywords

cpDNA, Cuba, founder effect, morphology, nuclear microsatellites, over-water dispersal, phylogeography, *Quercus oleoides*, *Quercus sagraeana*, *Quercus virginiana*.

*Correspondence: Paul F. Gugger, Department of Ecology and Evolutionary Biology, University of California, 4145 Terasaki Life Sciences Building, 610 Charles E. Young Drive E, Los Angeles, CA 90095-7239, USA.
E-mail: pfg@ucla.edu

INTRODUCTION

The origins and migration history of the Cuban biota are of long-standing interest in biogeography (Wallace, 1881; Darlington, 1938; Simpson, 1956). The Cuban flora is highly diverse (> 6000 species), with *c.* 50% of species endemic to Cuba (Borhidi, 1996). A high proportion of the flora in the Cuban Archipelago (*c.* 45%) is shared with the Yucatán Peninsula in Mexico (Chiappy-Jhones *et al.*, 2001), suggesting direct exchange of populations between the two areas during geological periods when they may have been briefly connected or by over-water long-distance dispersal (Coney, 1982; Iturralde-Vinent, 1982, 1988; Briggs, 1984; Hedges *et al.*, 1992). An intercontinental Yucatán–Cuba land bridge during the late Miocene and early Pliocene (*c.* 11.6–3.6 Ma) has been hypothesized to explain the migration of freshwater fish to Cuba (Rivas, 1958; Doadrio *et al.*, 2009), although geological evidence for this connection is lacking (Iturralde-Vinent, 2006). Even in the absence of a land bridge, over-water dispersal may have been facilitated by past and present ocean currents proceeding northwards through the Yucatán Channel around western Cuba, and then eastwards through the Straits of Florida. Western Cuba also has floristic affinities with the Florida Peninsula (Chiappy-Jhones *et al.*, 2001), indicating exchanges between these regions, although these are less prevalent. No land bridge has existed between Cuba and Florida for at least the last 35 Myr (Iturralde-Vinent & MacPhee, 1999), and thus dispersal by birds or hurricanes could explain floristic similarities. Finally, a strict vicariance model for the development of the Antillean biota would have been possible through the movement of a proto-Antillean Cretaceous land mass in Central America that subsequently fragmented and dispersed eastwards in the early Cenozoic (Rosen, 1975; Savage, 1982).

Phylogenetic analyses of Cuban plant taxa suggest a variety of origins, including vicariance, dispersal from eastern North America, dispersal from Mexico or Central America, and dispersal from South America (Santiago-Valentín & Olmstead, 2004; Graham, 2010). Overall, over-water dispersal has played a prominent role in speciation and community assembly on the Caribbean islands (Calsbeek & Smith, 2003; Graham, 2003; Negrón-Ortiz & Watson, 2003; Liu *et al.*, 2004; Glor *et al.*, 2005).

The source, identity and coherence of the single oak (*Quercus*) population on Cuba remain unresolved areas of controversy and speculation, addressed only by morphological evidence to date (Fig. 1). The entire Cuban oak population is confined to the western province of Pinar del Río. Phylogenetically, it falls within a small monophyletic lineage of evergreen or brevideciduous live oaks, section *Quercus* series *Virentes* (Muller, 1961; Nixon, 1985; Manos *et al.*, 1999; Cavender-Bares *et al.*, 2004; Pearse & Hipp, 2009). The live oaks span the tropical and temperate zones from the south-eastern USA, where three species are found (*Q. virginiana, Q. geminata* and *Q. minima*), throughout Mexico and Central America to north-western Costa Rica, where *Q. oleoides* occurs. The Cuban oak was first described by Nuttall (1842) as *Quercus sagraeana* based on the specimen collected by R. de la Sagra. Using the same specimen, but apparently unaware of the original name, Richard (1853) named the Cuban oak *Q. cubana*. Trelease (1924) did not view *Q. sagraeana* as a species in its own right but as a variety of *Q. virginiana.* He proposed the name *Q. virginiana* var. *sagraeana*, implying a Florida origin.

Figure 1 *Quercus sagraeana*, the lone oak species on Cuba, is found only in the western province of Pinar del Río, where these photos were taken. (a) Large mature tree, (b) close-up of leaves, and (c) leaves of a different individual, demonstrating the variability in leaf morphology. Photo credit, J. Cavender-Bares.

Muller (1961) renamed it *Q. oleoides* var. *sagraeana*, hypothesizing that the Cuban oak population was a subspecies of *Q. oleoides*, originating from the Yucatán region of Mexico as recently as the Pliocene, but with introgression from *Q. geminata* coming from Florida (Muller, 1955). He viewed the morphologically variable, highly 'heterozygous' Cuban population as a hybrid swarm that had stabilized and was distinct from the other live oaks of the series *Virentes*. Based on high leaf morphological variation in Cuban populations, López-Almirall (1979) supported Muller's hypothesis of a hybrid origin of the Cuban oaks but assumed that the source population from Florida was *Q. virginiana* rather than *Q. geminata*. Nixon (1985) generally concurred with Muller's (1955) perspective of a largely Mexican origin for the Cuban oaks with some introgression from *Q. geminata* in Florida, and hypothesized long-distance dispersal by passenger pigeons (*Ectopistes migratorius*) because viable acorns do not float in sea water. Although *Quercus* fossils are known from as far back as the late Miocene on the Gulf Coast of Mexico (Graham, 1975) and the Eocene in the south-eastern USA (Graham, 1964), none is known in Cuba (Graham, 2003; Peros *et al.*, 2007).

Origins of the North American oaks (Manos *et al.*, 1999; Manos & Stanford, 2001) post-date the timing of a proto-Antillean land mass connection with Central America (Pindell & Barrett, 1990; Iturralde-Vinent, 2006). The possibility of a vicariant origin of the Cuban oaks can thus be ruled out, as concluded by previous authors (Muller, 1955; Nixon, 1985). Three distinct scenarios emerge for the origins, timing of colonization, and migration of the live oaks to and from Cuba; these are variously supported by previous authors based on the distributions and morphological similarities of species within the group.

1. A Mexican origin with migration of *Q. oleoides* to Cuba from the Yucatán.

2. A Florida origin with migration of either *Q. virginiana* or *Q. geminata*, or both, to Cuba.

3. Colonization from both Mexico and Florida with introgression in Cuba.

In the absence of land bridges, over-water dispersal of acorns by birds, or more recently humans, could have occurred, especially during periods of low sea level during Pleistocene glaciations (Nixon, 1985). The earliest people are thought to have inhabited the region 6 ka (Rouse & Allaire, 1978; MacPhee *et al.*, 2007). Like many Native Americans (e.g. Jacknis, 2007), indigenous Cubans probably used acorns as a food source, suggesting that oaks could have dispersed as a crop.

In the present study, we sampled all known subpopulations of the Cuban oak and representative populations from each of its three putative source species (*Q. oleoides*, *Q. virginiana*, *Q. geminata*). We present the first molecular evidence to test hypotheses about (1) the integrity of the population as a species in its own right (*Q. sagraeana*), and (2) the origins and migration history of the population. We also conducted analyses of leaf morphology of the Cuban oaks relative to putative source populations in Florida, Mexico and Central America.

MATERIALS AND METHODS

Sampling

We analysed a representative range-wide sample of Cuban oaks (n_{ind} = 40, n_{site} = 5) and each of the three putative source taxa (*Q. oleoides*, n_{ind} = 66, n_{site} = 8; *Q. virginiana*, n_{ind} = 52, n_{site} = 13; *Q. geminata*, n_{ind} = 24, n_{site} = 4; section *Quercus* series *Virentes*) (Fig. 2; see Appendix S1 in Supporting Information). As outgroups, we also sampled representative individuals of another live oak (*Q. fusiformis*), a white oak (*Q. alba*; section *Quercus*), two red oaks (*Q. buckleyi* and *Q. ellipsoidalis*; section *Lobatae*), and an intermediate oak (*Q. chrysolepis*; section *Protobalanus*). Samples were shipped or carried to the University of Minnesota and stored at −80 °C, and pressed herbarium vouchers were made for each sample.

DNA preparation

Total genomic DNA was extracted using DNeasy Plant Mini Kits (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol. We amplified 12 nuclear microsatellite (nSSR) loci from seven chromosomes: *Qp*ZAG1/2, *Qp*ZAG1/5, *Qp*ZAG9, *Qp*ZAG15, *Qp*ZAG16, *Qp*ZAG36, *Qp*ZAG46, *Qp*ZAG102, *Qp*ZAG108, *Qp*ZAG110 (Steinkellner *et al.*, 1997); *Qr*ZAG11, *Qr*ZAG30 (Kampfer *et al.*, 1998). Nuclear SSR data for the three putative source taxa came from Cavender-Bares & Pahlich (2009) and Cavender-Bares *et al.* (2011). Two chloroplast DNA (cpDNA) regions were sequenced: a portion of the *trn*Y^{GUA}–*trn*E^{UUC}–*trn*T^{GGU} (*trn*D^{GUC}–*trn*T^{GGU}) region (Cavender-Bares *et al.*, 2011) and the *rpl*32–*trn*L^{UAG} region (Shaw *et al.*, 2007). Because cpDNA is maternally inherited, haploid, and non-recombining, these sequences were concatenated for each individual. A parsimony network with insertion–deletions coded as a fifth state and ignoring poly-A repeats was constructed in TCS 1.21 (Clement *et al.*, 2000).

Genetic diversity and neutrality

All analyses were performed pooling samples within taxa. We measured genetic diversity for each species as heterozygosity (H_E) for nSSRs and as haplotype diversity (h) for cpDNA sequences in ARLEQUIN 3.5 (Excoffier *et al.*, 2005a). For each locus, we tested the assumptions that we made in our analyses below. For nSSRs, we tested selective neutrality using the Ewens–Watterson homozygosity test (Ewens, 1972; Watterson, 1978; Slatkin, 1994, 1996) implemented in PYPOP 0.7.0 (Lancaster *et al.*, 2007), linkage equilibrium using a likelihood ratio test comparing models with and without association among loci (Slatkin & Excoffier, 1996), and constant population size using M, the ratio of allelic richness to the allele size range, which is expected to be below 0.68 under a bottleneck (Garza & Williamson, 2001; Excoffier *et al.*, 2005b). M was

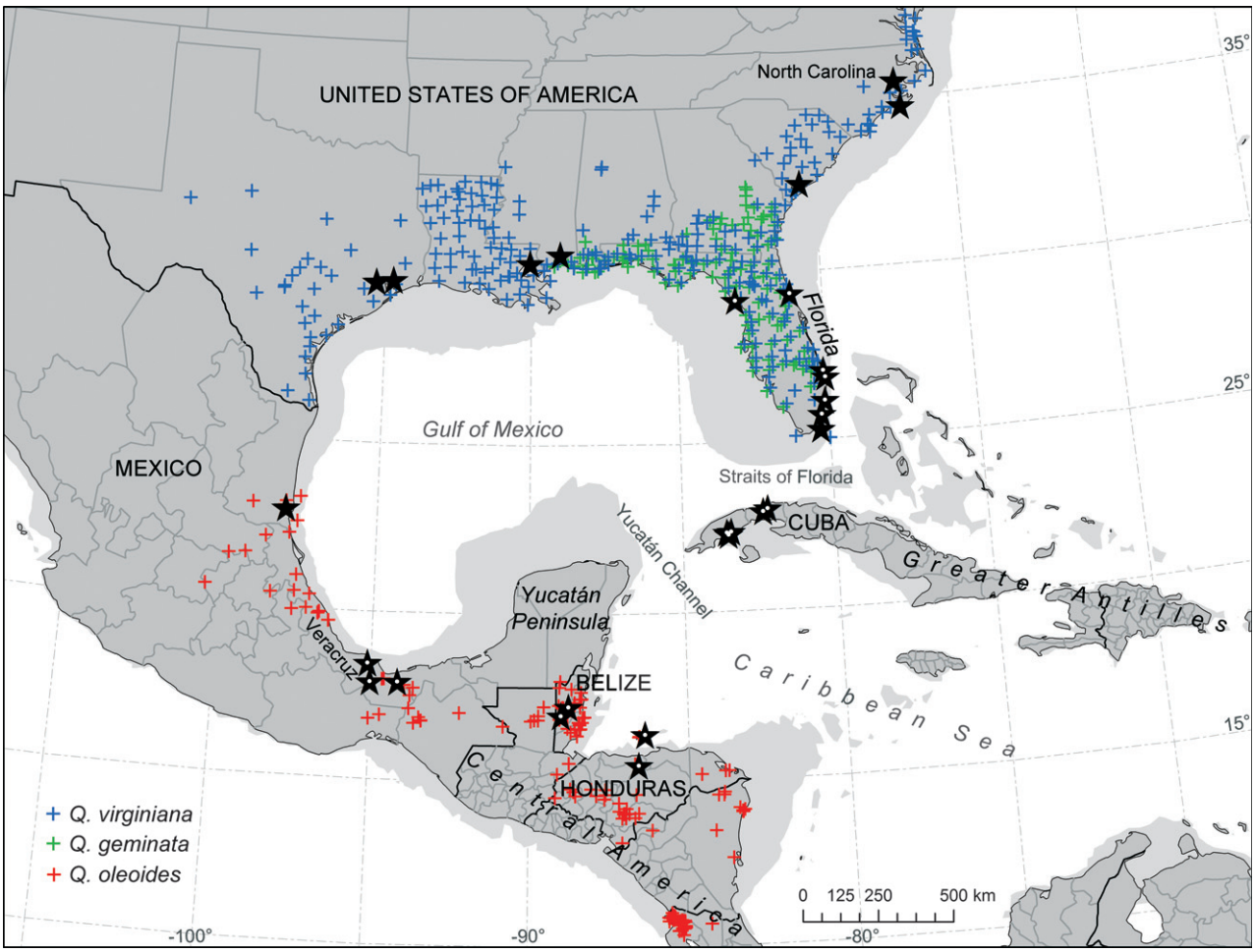

Figure 2 Map of species distributions of the live oaks (*Quercus* series *Virentes*), *Quercus virginiana*, *Q. geminata*, *Q. oleoides* and Cuban oak in south-eastern USA, Mexico, Central America and Cuba based on herbarium data extracted from the Global Biodiversity Information Facility portal (http://data.gbif.org/species). Sample sites for molecular analyses are shown as black stars; black stars with white points are sites where morphological data were also measured. The continental limits during a Pleistocene glacial are shown in light grey extending beyond modern limits.

chosen over tests of population bottlenecks based on heterozygosity (e.g. Cornuet & Luikart, 1996) because it is expected to be less sensitive to biases caused by pooling sample sites (e.g. Wahlund, 1928) or by occasional genotyping errors caused by null alleles. We tested nSSRs for the presence of null alleles using MICRO-CHECKER 2.2.3 to identify loci with homozygote excess (van Oosterhout *et al.*, 2004). For cpDNA, neutrality and constant population size were tested with Tajima's D (Tajima, 1989) and Fu's F_S (Fu, 1997), and recombination was tested with a four-gamete test (Hudson & Kaplan, 1985) in DNASP 5.0 (Librado & Rozas, 2009).

Population structure

Population structure with admixture was estimated for nSSRs using a Bayesian Markov chain Monte Carlo (MCMC) clustering analysis in INSTRUCT 1.0 (Gao *et al.*, 2007). We ran three replicates, each with different random seeds, two MCMC chains, 2,000,000 steps per chain, and 1,000,000 burn-in steps for 1–16 population clusters (K). Although INSTRUCT can estimate the 'optimal' K using the deviance information criterion (DIC), we were most interested in the uppermost hierarchical level of population structure that delineated the three putative source species. Unlike other similar clustering algorithms (Pritchard *et al.*, 2000), INSTRUCT does not assume Hardy–Weinberg equilibrium within loci. Results were visualized using DISTRUCT 1.1 (Rosenberg, 2004) and ARCGIS 9.3 (ESRI, Redlands, CA, USA).

We also estimated global and pairwise genetic differentiation among Cuban samples and putative source species (samples pooled for each taxon) using R_{ST} (Slatkin, 1995) and F_{ST} for nSSRs in GENALEX 6.1 (Peakall & Smouse, 2006) and N_{ST} and F_{ST} for cpDNA sequences using DISTON 1.0 and PERMUT-CPSSR 2.0 (Pons & Petit, 1996). R_{ST} and N_{ST} are analogous to F_{ST}, but, in addition to allele frequency differences, they consider mutational distance among alleles under a stepwise mutation model (SMM) (Ohta & Kimura, 1973) and using the uncorrected number of nucleotide differences, respectively.

Gene flow

We used a MCMC maximum likelihood approach for estimating bidirectional gene flow rates as the number of migrants per generation (N_em) among all pairs of taxa (three putative source taxa and Cuban population) using MIGRATE-N 3.2.6 (Beerli & Felsenstein, 1999, 2001). All runs used 10 short chains of 10,000 recorded genealogies and four long chains of 100,000 recorded genealogies with a sampling increment of 50–100 and a burn-in of 20,000 genealogies. An unweighted pair-group method using arithmetic averages (UPGMA) starting tree and static heating with default temperatures were selected. Nuclear SSR data were run assuming a Brownian motion approximation of the SMM, and cpDNA data were run assuming the F84 model of sequence evolution (Felsenstein & Churchill, 1996). Runs were replicated with different random seeds at least five times for each marker type to ensure convergence.

Divergence time and effective population size

Divergence time among each putative source species and the Cuban population as well as effective population sizes were estimated using an isolation-with-migration model in IMa version 12/19/2009 (Hey & Nielson, 2007). To ensure convergence, we compared three replicate runs (ESS > 50) of 40 geometrically heated chains ($g_1 = 0.8$, $g_2 = 0.9$) with a 10,000,000-step burn-in followed by > 20,000,000 steps. We estimated the maximum values for priors drawn from a truncated uniform distribution on the basis of short preliminary runs (Won & Hey, 2005). We assumed a Hasegawa–Kishino–Yano (HKY) model (Hasegawa *et al.*, 1985) of evolution for sequences and a SMM for microsatellites (Ohta & Kimura, 1973). Gene flow estimates were not trusted because simulations have shown that those estimates can be severely biased when not all interbreeding populations are included, as is the case in our pairwise approach (Strasburg & Rieseberg, 2010). The effect on divergence time estimates, however, is expected to be minimal.

To convert the model estimate of divergence time to demographic units, we considered a range of mutation rates and generation times. For cpDNA, we calculated a mutation rate for our locus by dividing the Jukes & Cantor (1969) corrected sequence divergence rate per base pair among red oaks and white oak/live oaks by two times the estimated divergence time of the sections (*c.* 40 Ma) (Daghlian & Crepet, 1983; Crepet & Nixon, 1989; Manos *et al.*, 1999). For nSSRs, we considered published rates ranging from 1.76×10^{-4} per generation in tomatoes (*Solanum lycopersicum*; Azaiez *et al.*, 2006) to 8.8×10^{-4} per generation in *Arabidopsis thaliana* (Marriage *et al.*, 2009). Per-generation nSSR mutation rates were converted to per-year rates considering mean generation times in live oaks of 100–220 years (Cavender-Bares *et al.*, 2011).

Morphological analysis

A large subset of sites was investigated for morphological variation (Fig. 2; see Appendix S1): most samples of *Q. virginiana* and *Q. geminata* from Florida, all samples of *Q. oleoides* from Belize, Honduras, and Veracruz, Mexico, and all Cuban samples. We measured taxonomically informative leaf morphological traits important in distinguishing the three putative source taxa (Muller, 1961; Kurz & Godfrey, 1962; Nixon & Muller, 1997; Cavender-Bares & Pahlich, 2009); these included the following.

1. Rugose venation (0 = no, 0.5 = moderate, 1 = extreme).
2. Revolute margins (0 = no, 0.5 = moderate, 1 = extreme).
3. Abaxial pubescence (0 = none or sparse fused-stellate hairs, 0.5 = intermediate density, 1 = densely pubescent, including stellate and additional felted or erect hairs).
4. Midvein thickness at 1/3 distance between base of petiole and tip of leaf relative to leaf area (mm mm^{-2}).
5. Leaf mass per area (g m^{-2}) as a measure of 'thick and leathery'.
6. Leaf length to width ratio.
7. Leaf area (mm^2).

Each trait was measured for three leaves per sample under a dissecting microscope or in ImageJ (Rasband, 2010) without knowledge of taxonomic identity. Average values from the three leaves for traits 4–7 were rescaled between 0 and 1, where 0 more closely matched the taxonomic description of *Q. virginiana*/*Q. oleoides* and 1 more closely matched that of *Q. geminata*. We tested for statistically significant differences among putative source species and Cuban oak using an analysis of similarity (ANOSIM) of Bray–Curtis similarities of all traits implemented in Primer5 (Primer-E Ltd., Ivybridge, UK). ANOSIM is a multivariate, nonparametric test based on the rank order of matrix values (Clarke, 1993). Similarity percentage (SIMPER) analysis was used to determine which traits contributed most to differences among taxa; traits that distinguish taxa have a high average contribution to Bray–Curtis dissimilarities among taxa and are consistent across samples (low standard deviation). To identify groups of individuals with similar morphological characteristics, we also performed a hierarchical cluster analysis using Bray–Curtis similarities in Primer5.

RESULTS

Genetic diversity and neutrality

No cpDNA variation was found in Cuban oak, and its chlorotype (C1) was also common in *Q. virginiana* and *Q. geminata* in southern Florida and present in two *Q. oleoides* individuals in Belize (Fig. 3a) (GenBank accessions JF506259–JF506632). Two deeply divergent chlorotype clades were shared among *Q. virginiana* and *Q. oleoides*, but the centre of diversity for clade C1–C8 was Florida and the centre of diversity for clade C9–C16 was Mexico and Central America. Among *Q. virginiana* and *Q. geminata*, chlorotype depends more on location than on species (Fig. 3a), with an introgression ratio, IG = 0.91, when considering all sympatric sites as a single population (Belahbib *et al.*, 2001). IG is a relative measure ranging from 0 to 1 of how often sympatric species share haplotypes locally. The 'outgroups' *Q. alba* and *Q. fusiformis* also shared chloro-

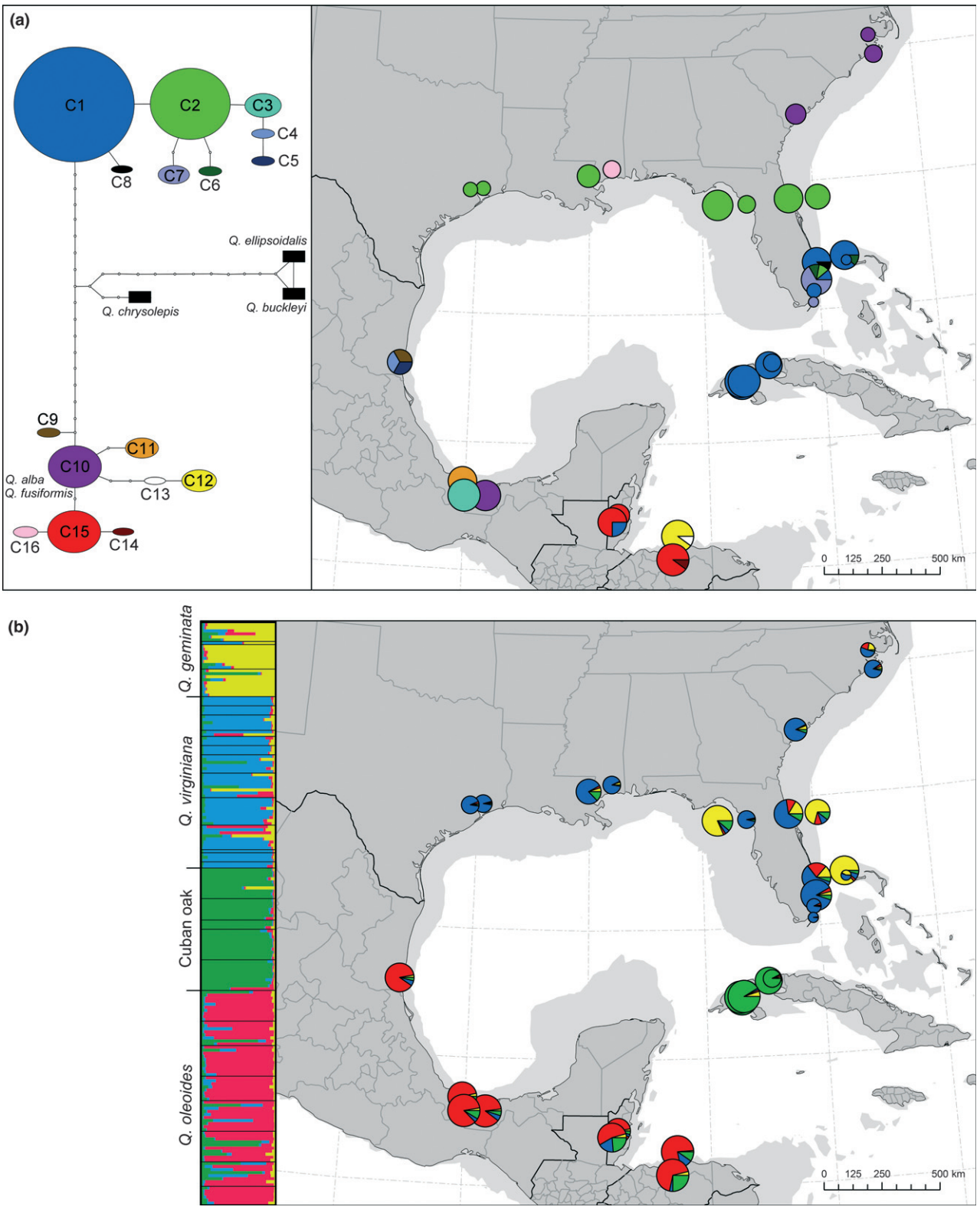

Figure 3 (a) Map of south-eastern USA, Mexico, Central America and Cuba showing the sample sites of the live oaks, *Quercus virginiana*, *Q. geminata*, *Q. oleoides* and Cuban oak, coloured according to cpDNA haplotype in proportion to frequency. A parsimony network is inset. (b) Map of sample sites of the same four live oak species coloured according to four InStruct clusters with admixture plot. *Quercus geminata* samples are shown in the ocean ± 1.3° longitude of their actual location for comparison with *Q. virginiana* samples, which are shown on land at their true location.

types with *Q. virginiana* (C10). Chloroplast DNA sequence data did not deviate from neutral expectations according to *D*, but there was evidence of a bottleneck in *Q. virginiana* according to F_S. Finally, the four-gamete test detected no evidence of intralocus recombination.

Four nSSR loci were discarded from subsequent analyses: *Qp*ZAG46 and *Qp*ZAG108 for failure to amplify in Cuban and some Belizean individuals, and *Qp*ZAG1/2 and *Qr*ZAG30 for showing significant departures from neutral expectation after Bonferroni correction for multiple tests (Ewens–Watterson, $P < 0.004$). For the remaining eight nSSR loci, diversity was high, but lowest in Cuban oak (Table 1). Mean *M* ranged from 0.19 in Cuban oak to 0.33 in *Q. virginiana*, suggesting bottlenecks in all taxa (Table 1). MICRO-CHECKER flagged three loci (*Qr*ZAG11, *Qp*ZAG15 and *Qp*ZAG102) as having homozygote excess and thus possible evidence of null alleles. However, *M* was consistently low across all eight loci, whether or not they may have had null alleles and when estimated within taxa or within individual sample sites. No pairs of loci were consistently found to be in linkage disequilibrium across taxa after Bonferroni correction.

Population structure

The three putative source species and Cuban oak samples were distinguished at $K = 4$ (mean DIC = 10,057.7) (Fig. 3b). At $K = 2$ (DIC = 10,405.5), Cuban samples belonged to a cluster more common in *Q. oleoides* but present in *Q. virginiana* and *Q. geminata* (see Appendix S2). At $K = 3$ (DIC = 10,057.7), Cuban samples were delineated from a primarily *Q. oleoides* cluster and a *Q. virginiana*/*Q. geminata* cluster (Appendix S2). The 'optimal' $K = 9$ (DIC = 9797.3) resulted in Cuban oak samples almost entirely assigned to one genetic cluster and the three putative source species assigned to various nearly equal mixtures of the other eight genetic clusters (Appendix S2).

Genetic differentiation among taxa was low, but significant for nSSRs ($F_{ST} = 0.024$, $R_{ST} = 0.09$) and moderate for cpDNA sequences ($F_{ST} = 0.374$, $N_{ST} = 0.509$) (Table 2). By most measures, Cuban oaks were least differentiated from *Q. virginiana* and most differentiated from *Q. geminata*, and Cuban oaks were more differentiated from mainland taxa than mainland taxa were from each other. Although these differences were not that large, weak differentiation is common among oak species (Petit *et al.*, 1997; Zeng *et al.*, 2010).

Gene flow

For nSSRs, migration between Cuban oak and mainland taxa was modest ($0.25 < N_em < 0.51$), and generally lower than migration estimates among mainland taxa (Table 2). Gene flow estimates were highest from *Q. virginiana* to Cuban oak than from *Q. oleoides* or *Q. geminata*. Gene flow was slightly

Table 1 Sample size of individuals (*N*), mean number of nSSR alleles (*A*), mean nSSR heterozygosity (H_E), ratio of allelic richness to allele size range (*M*), number of cpDNA haplotypes (n_h), haplotype diversity (*h*), Tajima's *D*, and Fu's F_S (the bold value indicates significance after Bonferroni correction) for Cuban oak, *Quercus virginiana* (south-eastern USA), *Q. geminata* (Florida) and *Q. oleoides* (Mexico and Central America).

Population	*N*	*A*	H_E	*M*	n_h	*h*	*D*	F_S
Cuban oak	40	7.9	0.73	0.19	1	0	–	–
Q. geminata	24	11.5	0.75	0.25	3	0.52	1.2	1.4
Q. virginiana	52	14.4	0.80	0.33	7	0.76	1.3	**9.9**
Q. oleoides	66	14.1	0.81	0.32	11	0.84	1.6	6.9

Table 2 Global and pairwise morphological dissimilarity (*R*-statistic from ANOSIM) and molecular differentiation (R_{ST}, N_{ST} and F_{ST}) among Cuban oak, *Quercus virginiana* (south-eastern USA), *Q. geminata* (Florida) and *Q. oleoides* (Mexico and Central America). Bold values were significantly different from zero after Bonferroni correction. N_em_1 and N_em_2 indicate mean population migration rates (effective number of migrants per generation) into the first and second population listed in each pairwise comparison to the left and were estimated from nSSRs across five independent runs of MIGRATE-N.

	Morphology	cpDNA		nSSR			
Groups	*R*	N_{ST}	F_{ST}	R_{ST}	F_{ST}	N_em_1	N_em_2
Global	**0.514**	**0.509**	**0.374**	**0.090**	**0.024**	–	–
Cuban oak–*Q. geminata*	**0.787**	0.579	0.612	0.087	0.157	0.25	0.30
Cuban oak–*Q. oleoides*	**0.316**	0.678	0.566	0.05	0.093	0.30	0.51
Cuban oak–*Q. virginiana*	**0.15**	0.239	0.528	0.055	0.086	0.40	0.35
Q. virginiana–*Q. geminata*	**0.994**	0.178	0.063	0.023	0.093	0.57	0.62
Q. virginiana–*Q. oleoides*	**0.144**	0.365	0.174	0	0.059	0.65	0.53
Q. oleoides–*Q. geminata*	**0.987**	0.668	0.313	0.009	0.114	0.56	0.46

higher from *Q. virginiana* to Cuban oak than vice versa ($N_em_{VI \rightarrow CU} = 0.40$; $N_em_{CU \rightarrow VI} = 0.35$) and higher from Cuban oak to *Q. oleoides* than the reverse ($N_em_{OL \rightarrow CU} = 0.30$; $N_em_{CU \rightarrow OL} = 0.51$). Preliminary simulations (P. Beerli, Florida State University, pers. comm.) suggest that estimates of the directionality of migration based on multiple loci should be robust to violating the assumption of constant population size (i.e. bottleneck; Tables 1 and 3) and to the assumption that shared genetic variation is not due to recent divergence from a common ancestor. The magnitudes, however, may be inaccurate. N_em estimates for cpDNA sequences did not converge, even across longer runs with more heated chains.

Divergence time and effective population size

All divergence time estimates among putative source species and the Cuban population fell within the Pleistocene (Table 3). The Cuban oak–*Q. virginiana* split may have occurred 628–6 ka, including the 90% highest posterior density interval of estimates based on low mutation rate/long generation time and high mutation rate/short generation time (Table 3). Divergence time estimates among other taxa and other details are given in Table 3. The scaled effective population size estimate for the Cuban population is *c*. 15% of *Q. virginiana* and 1.4% of their ancestor, supporting a strong bottleneck.

The conversions of scaled divergence time estimates to demographic units incorporated our cpDNA mutation rate estimate of 1.27×10^{-10} bp^{-1} year^{-1} ($\pm$ 0.41×10^{-10} bp^{-1} year^{-1}), which we converted to a per-locus per-year rate for IMa by multiplying by the sequence length of 1805 bp. Our mutation rate estimates are similar to those calculated for *rbc*L based on the divergence of *Quercus* and *Castanea* (0.71×10^{-10} bp^{-1} year^{-1}; Frascaria *et al.*, 1993) and those calculated based on only the *trn*Y–*trn*E–*trn*T locus (0.8×10^{-10} bp^{-1} year^{-1}; Cavender-Bares *et al.*, 2011).

Morphological analyses

Analysis of similarity showed global morphological differences among species ($R = 0.51$, $P < 0.0001$). All pairwise comparisons among taxa were significant (Table 2), with Cuban oak most different from *Q. geminata* ($R = 0.79$, $P < 0.0001$), intermediately different from *Q. oleoides* ($R = 0.32$, $P < 0.0001$), and least different from *Q. virginiana* ($R = 0.15$, $P < 0.001$). The most important traits for distinguishing taxa were as follows: for Cuban oak and *Q. geminata*, revolute margins (contribution to dissimilarity = 31.2%) and rugose venation (29.9%); for Cuban oak and *Q. oleoides*, leaf area (25.5%) and pubescence (21.6%); and for Cuban oak and *Q. virginiana*, pubescence (22.2%) and rugose venation (19.6%) (see Appendix S3). Hierarchical clustering revealed two groups of individuals: one containing all *Q. geminata* and two Cuban individuals, and one containing a mixture of *Q. oleoides*, *Q. virginiana*, and the remaining Cuban individuals (Appendix S3).

Table 3 Divergence times (t) and 90% highest posterior density intervals (HPDI) estimated from IMa among Cuban oak, *Quercus virginiana* (south-eastern USA), *Q. geminata* (Florida) and *Q. oleoides* (Mexico and Central America) were converted to demographic units (ka) assuming high mutation rates and a short generation time of 100 years (t_{low}), and assuming low mutation rates and a long generation time of 220 years (t_{high}). Scaled effective population size estimates are given for taxon 1 (q_1), taxon 2 (q_2) and their ancestor (q_A). Values for runs with highest effective sample size are shown.

Comparison	Priors (t, m_1, m_2, q_1, q_2, q_A)	q_1	q_2	q_A	t (90% HPDI)	t_{low} (90% HPDI) (ka)	t_{high} (90% HPDI) (ka)
Q. virginiana–Cuban oak	3, 2, 9, 0.5, 0.3, 5	5.49 (2.47–10.82)	0.85 (0.37–1.71)	59.99 (41.92–88.64)	0.13 (0.03–0.42)	22 (6–69)	200 (52–628)
Q. geminata–Cuban oak	20, 0.7, 6, 1, 0.5, 7	11.69 (5.77–20.52)	1.24 (0.60–2.19)	54.97 (26.97–146.63)	0.55 (0.09–9.09)	91 (15–1502)	825 (135–13,639)
Q. oleoides –Cuban oak	2, 3, 20, 1, 0.3, 6	8.55 (3.70–16.90)	0.70 (0.22–1.65)	59.05 (40.40–90.67)	0.20 (0.04–0.76)	33 (6–125)	296 (56–1139)
Q. virginiana–*Q. oleoides*	2, 3, 5, 0.5, 0.7, 5	6.57 (3.52–10.90)	6.19 (3.36–10.16)	70.73 (51.27–99.02)	0.21 (0.08–0.57)	35 (12–95)	314 (113–860)

DISCUSSION

Our combined molecular and morphological data support a Pleistocene origin of Cuban oaks from *Q. virginiana* in Florida. The Cuban population does not show any evidence of hybrid origin. Rather, the combined effect of a dispersal bottleneck from a single source and persistent isolation as a small population has yielded a relatively pure, low-diversity population. Some evidence supports subsequent dispersal out of Cuba to *Q. oleoides* populations in Belize and Honduras.

Florida origin from *Q. virginiana*

Evidence for a *Q. virginiana*, Florida origin is three-fold. First, cpDNA haplotypes in Cuban oak (C1) are identical to the most common chlorotype in southern Florida, shared by both *Q. virginiana* and *Q. geminata* (Fig. 3a); that chlorotype falls within a clade whose centre of cpDNA diversity is Florida and that is highly divergent from the clade found primarily in *Q. oleoides* in Central America and Mexico. Second, gene flow estimates based on nSSRs were higher into Cuban oak from *Q. virginiana* than from the other mainland taxa (Table 2). Third, Cuban oaks are morphologically and genetically most similar to *Q. virginiana* and least similar to *Q. geminata* (Table 2). However, we cannot rule out that a small proportion of colonizers were *Q. geminata* or that a small proportion of introgressed *Q. geminata* genes were carried in *Q. virginiana*. For example, hierarchical clustering analysis of leaf morphology grouped two Cuban individuals with an otherwise purely *Q. geminata* cluster (Appendix S3). Despite morphological variability (Fig. 1 & Appendix S3; Muller, 1955; López-Almirall, 1979), genetic diversity (A, H_E, n_h, h; Table 1) is relatively low in Cuban oak, and there is almost no signal of genetic admixture, even when assuming a wide range of ancestral population numbers (Fig. 2b & Appendix S2). This provides strong evidence against a predominantly hybrid origin for the Cuban oak and suggests continued isolation following colonization.

Pleistocene colonization of Cuba

If Cuban populations derived from *Q. virginiana*, colonization probably occurred in the Pleistocene (Table 3). The low diversity and low M-value of the Cuban oak (Table 1) and the low effective population size of Cuban oaks relative to *Q. virginiana* and their ancestor (Table 3) indicate a strong founder effect. The combination of cooler temperatures and low sea level during a glacial period may have facilitated limited dispersal to Cuba, highlighting the role that Pleistocene glaciations might have played in the establishment of temperate taxa in the tropics (Gugger *et al.*, 2011).

Passenger pigeons have been hypothesized as a potential dispersal vector (Nixon, 1985). Although they were only rarely observed in Cuba before extinction in the 20th century, they could have been more common there during cool, glacial periods. Other birds, such as crows (*Corvus*) or jays (e.g. *Aphelocoma*, *Cyanocitta*), could have also been plausible dispersing agents. However, divergence time estimates preclude humans as a vector. The only divergence time estimates that overlap with human occupation of the region (6 ka) are at the extreme tail of the 90% highest posterior density interval of estimates based on high mutation rates and short generation times (Table 3). Estimates based on the unusually high nSSR mutation rates in *Arabidopsis* (Marriage *et al.*, 2009) are less plausible in oaks than estimates based on low mutation rates (Cavender-Bares *et al.*, 2011).

The actual divergence of Cuban populations from *Q. virginiana* may be somewhat older than reported here because mutation rates may be lower in trees than in herbaceous plants (Petit & Hampe, 2006). Given that nSSRs represent eight of the nine loci used in our analysis, our estimates strongly depend on nSSR rates, which were estimated from herbaceous plants. A study using the same nSSRs but with two additional DNA sequence loci found an older divergence time between *Q. oleoid*es and *Q. virginiana* than when our data are used for that estimate (Table 3; Cavender-Bares *et al.*, 2011).

Isolation and speciation

Following colonization, the isolation of Cuban oaks in small populations permitted genetic and morphological differentiation. Leaf morphological differences among Cuban oak and mainland species are comparable to differences among mainland species, and genetic differentiation among Cuban oak and mainland species is stronger than among mainland taxa (Tables 2 & 3). Morphological and molecular differentiation, together with the isolation and genetic purity of the Cuban population, argues for the taxonomic designation of the Cuban oak at the species level as *Q. sagraeana*.

Post-colonization dispersal to Central American *Q. oleoides*

Although gene flow into Cuba following colonization appears to have been lacking, there is evidence for dispersal out of Cuba. In particular, the Cuban population cluster (Fig. 3b) is present in moderate proportion in Belize and Honduras. Moreover, gene flow estimates were higher out of Cuba to *Q. oleoides* than into Cuba (Table 2). Gene flow by pollen dispersal to Central America could have been, or could be, facilitated by prevailing north-easterly trade winds in the region. Chlorotype C1 was found in two individuals in Belize, suggesting that seed also dispersed there. Those samples were collected near modern buildings at Mountain Pine Ridge, a site that has a long history of active management for forest resources (Anonymous, 2002). Thus, people may have planted them there from a Cuban source. Alternatively, birds or other natural agents could have been the dispersal vector.

Alternative explanations

An alternative scenario less consistent with our data is an origin from *Q. oleoides* in Central America, especially Belize. For example, the Cuban nSSR cluster (Fig. 3b) and chlorotype C1 are found in Belize, two nSSR loci failed to amplify in both Cuban and some Belizean samples, and gene flow into Cuban oak from *Q. oleoides* was only a little less than that from *Q. virginiana*. A *Q. oleoides* origin would assume a strong dispersal bottleneck in which only relatively rare Central American genotypes arrived in Cuba. This would also imply that, subsequently, seeds dispersed from Cuba to Florida, where chlorotype C1 was captured (Rieseberg & Soltis, 1991; Tsitrone *et al.*, 2003) by both *Q. virginiana* and *Q. geminata*. We regard this scenario as less parsimonious because (1) the dispersal route would run counter to the diversity gradient for the chlorotype clade containing C1, (2) common Central American genotypes (e.g. C15 and *Q. oleoides* nSSR cluster) are not found in Cuba, and (3) it relies on the assumption that *Q. oleoides* previously inhabited the harsh limestone soils of the Yucatán Peninsula near Cuba, which is unclear in the fossil record (Leyden *et al.*, 1996; Islebe & Sánchez, 2002; Graham, 2003; Carrillo-Bastos *et al.*, 2010). Nonetheless, it must be cautioned that our molecular data represent a limited snapshot relative to genome-wide variation.

Finally, we note that our molecular and morphological data contain a limited signal of migration into Cuba. Bayesian clustering analysis suggests that Cuban oak forms a distinct population cluster (Fig. 2b), and migration rate estimates (Table 2) often suggest more migration out of Cuba than into it. Thus a third scenario is that the identity of the source species is obscured by an old colonization of Cuba, subsequent divergence, and recent, limited migration out of Cuba into both Florida and Central America.

Chloroplast introgression

The introgression of the chloroplast genome is common in sympatric oaks within the same taxonomic section (Whittemore & Schaal, 1991; Petit *et al.*, 2002). Our cpDNA data show a strong signal of chloroplast introgression among *Q. virginiana* and *Q. geminata* in Florida (IG = 0.91), similar to that among closely related European white oaks (Belahbib *et al.*, 2001). Chloroplast introgression across sympatric species within section *Quercus* could also explain the unusual pattern of genetic diversity observed here, where each of two deeply divergent chlorotype clades are widely distributed from Central America to North Carolina and are shared among *Q. virginiana* and *Q. oleoides*. Chloroplast capture (Rieseberg & Soltis, 1991; Tsitrone *et al.*, 2003) and pollen swamping (Potts & Reid, 1988; Bacilieri *et al.*, 1996; Petit *et al.*, 2004) have been proposed as mechanisms for shared chloroplast variation among species. However, shared ancestral variation owing to incomplete lineage sorting cannot be ruled out. A history of extensive pollen swamping (causing asymmetrical introgression), as proposed in European white oaks (Petit *et al.*, 2004), could cloud migration patterns inferred from cpDNA because species can effectively colonize new territory exclusively by pollen. Although chloroplast introgression can hamper phylogeographic approaches to understanding intraspecific history, we still found some molecular and morphological signals of colonization and divergence in the Cuban oak.

CONCLUSIONS

Molecular and morphological data support the hypothesis that oaks colonized Cuba during the Pleistocene from a single source: *Q. virginiana* in Florida. Subsequent isolation and divergence suggest that the Cuban oak should be regarded as a distinct species as per its original designation (*Q. sagraeana* Nuttall, 1842). Although Cuba shares strong floristic similarities with the Yucatán Peninsula (Chiappy-Jhones *et al.*, 2001), our data emphasize that the Cuban biota arose from diverse sources driven by a variety of geological and climatic forces at different time-scales (Graham, 2003; Santiago-Valentín & Olmstead, 2004).

ACKNOWLEDGEMENTS

We thank M. Merriman and K. Koehler for laboratory assistance; M. McQuaid (Office of International Programs, University of Minnesota) for assistance in acquiring travel authorization to Cuba; E. Logan for logistical assistance in travelling to Cuba; A. López Almirall (Museo Nacional de Historia Natural, La Habana, Cuba) for assistance with collections in Cuba; the Cornell University Computational Biology Service Unit for computing resources; H. Rodríguez Correa (Universidad Nacional Autónoma de México, Morelia) for GIS data of continental margins during Pleistocene glaciations; and C. Dick and S. Cavers for inviting us to contribute to this Special Issue. We also thank the family of P.F.G. in Cienfuegos, Cuba for their hospitality: que pronto podamos reunirnos en familia otra vez para seguir compartiendo ese cariño tan grande que siempre hemos guardado a través de los años. Funding was provided by the National Science Foundation IOS: 0843665 to J.C.-B.

REFERENCES

Anonymous (2002) *Mountain Pine Ridge Carbon Sequestration Project, Belize C.A*. Progress Report No. 1, Global Forest Nursery Development, Inc., Belize City, Belize.

Azaiez, A., Bouchard, E.F., Jean, M. & Belzile, F.J. (2006) Length, orientation, and plant host influence the mutation frequency in microsatellites. *Genome*, **49,** 1366–1373.

Bacilieri, R., Ducousso, A., Petit, R.J. & Kremer, A. (1996) Mating system and asymmetric hybridization in a mixed stand of European oaks. *Evolution*, **50,** 900–908.

Beerli, P. & Felsenstein, J. (1999) Maximum likelihood estimation of migration rates and population numbers of two populations using a coalescent approach. *Genetics*, **152,** 762–773.

Beerli, P. & Felsenstein, J. (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences USA*, **98,** 4563–4568.

Belahbib, N., Pemonge, M.H., Ouassou, A., Sbay, H., Kremer, A. & Petit, R.J. (2001) Frequent cytoplasmic exchanges between oak species that are not closely related: *Quercus suber* and *Q. ilex* in Morocco. *Molecular Ecology*, **10,** 2003–2012.

Borhidi, A. (1996) *Phytogeography and vegetation ecology of Cuba*. Akadémiai Kiadó, Budapest.

Briggs, J. (1984) Freshwater fishes and biogeography of Central America and the Antilles. *Systematic Zoology*, **33,** 428–435.

Calsbeek, R. & Smith, T.B. (2003) Ocean currents mediate evolution in island lizards. *Nature*, **426,** 552–555.

Carrillo-Bastos, A., Islebe, G.A., Torrescano-Valle, N. & González, N.E. (2010) Holocene vegetation and climate history of central Quintana Roo, Yucatán Península, Mexico. *Review of Palaeobotany and Palynology*, **160,** 189–196.

Cavender-Bares, J. & Pahlich, A. (2009) Molecular, morphological and ecological niche differentiation of sympatric sister oak species, *Quercus virginiana* and *Q. geminata* (Fagaceae). *American Journal of Botany*, **96,** 1690–1702.

Cavender-Bares, J., Ackerly, D.D., Baum, D.A. & Bazzaz, F.A. (2004) Phylogenetic overdispersion in Floridian oak communities. *The American Naturalist*, **163,** 823–843.

Cavender-Bares, J., González-Rodríguez, A., Pahlich, A., Koehler, K. & Deacon, N. (2011) Phylogeography and climatic niche evolution in live oaks (*Quercus* series *Virentes*) from the tropics to the temperate zone. *Journal of Biogeography*, **38,** 962–981.

Chiappy-Jhones, C., Rico-Gray, V., Gama, L. & Giddings, L. (2001) Floristic affinities between the Yucatán Peninsula and some karstic areas of Cuba. *Journal of Biogeography*, **28,** 535–542.

Clarke, K.R. (1993) Nonparametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, **18,** 117–143.

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

Coney, P.J. (1982) Plate tectonic constraints on the biogeography of Middle America and the Caribbean region. *Annals of the Missouri Botanical Garden*, **69,** 432–443.

Cornuet, J.M. & Luikart, G. (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144,** 2001–2014.

Crepet, W.L. & Nixon, K.C. (1989) Earliest megafossil evidence of Fagaceae: phylogenetic and biogeographic implications. *American Journal of Botany*, **76,** 842–855.

Daghlian, C.P. & Crepet, W.L. (1983) Oak catkins, leaves and fruits from the Oligocene Catahoula Formation and their evolutionary significance. *American Journal of Botany*, **70,** 639–649.

Darlington, P. (1938) The origin of the fauna of the Greater Antilles, with discussion of dispersal of animals over water and through the air. *Quarterly Review of Biology*, **13,** 274–300.

Doadrio, I., Perea, S., Alcaraz, L. & Hernández, N. (2009) Molecular phylogeny and biogeography of the Cuban genus *Girardinus* Poey, 1854 and relationships within the tribe *Girardinini* (Actinopterygii, Poeciliidae). *Molecular Phylogenetics and Evolution*, **50,** 16–30.

Ewens, W.J. (1972) Sampling theory of selectively neutral alleles. *Theoretical Population Biology*, **3,** 87–112.

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

Excoffier, L., Estoup, A. & Cornuet, J.-M. (2005b) Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics*, **169,** 1727–1738.

Felsenstein, J. & Churchill, G.A. (1996) A hidden Markov model approach to variation among sites in rate of evolution. *Molecular Biology and Evolution*, **13,** 93–104.

Frascaria, N., Maggia, L., Michaud, M. & Bousquet, J. (1993) The *rbc*L gene sequence from chestnut indicates a slow rate of evolution in the Fagaceae. *Genome*, **36,** 668–671.

Fu, Y.X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147,** 915–925.

Gao, H., Williamson, S. & Bustamante, C.D. (2007) A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics*, **176,** 1635–1651.

Garza, J.C. & Williamson, E.G. (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, **10,** 305–318.

Glor, R.E., Losos, J.B. & Larson, A. (2005) Out of Cuba: overwater dispersal and speciation among lizards in the *Anolis carolinensis* subgroup. *Molecular Ecology*, **14,** 2419–2432.

Graham, A. (1964) Origin and evolution of the biota of southeastern North America: evidence from the fossil plant record. *Evolution*, **18,** 571–585.

Graham, A. (1975) Late Cenozoic evolution of tropical lowland vegetation in Veracruz, Mexico. *Evolution*, **29,** 723–735.

Graham, A. (2003) Historical phytogeography of the Greater Antilles. *Brittonia*, **55,** 357–383.

Graham, S.A. (2010) Revision of the Caribbean genus *Ginoria* (Lythraceae), including *Haitia* from Hispañola. *Annals of the Missouri Botanical Garden*, **97,** 34–90.

Gugger, P.F., González-Rodríguez, A., Rodríguez-Correa, H., Sugita, S. & Cavender-Bares, J. (2011) Southward Pleistocene migration of Douglas-fir into Mexico: phylogeography, ecological niche modeling, and conservation of 'rear edge' populations. *New Phytologist*, **189,** 1185–1199.

Hasegawa, M.H., Kishino, H. & Yano, T. (1985) Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22,** 160–174.

Hedges, S.B., Hass, C.A. & Maxson, L.R. (1992) Caribbean biogeography: molecular evidence for dispersal in West Indian terrestrial vertebrates. *Proceedings of the National Academy of Sciences USA*, **89,** 1909–1913.

Hey, J. & Nielson, R. (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences USA*, **104,** 2785–2790.

Hudson, R.R. & Kaplan, N.L. (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111,** 147–164.

Islebe, G. & Sánchez, O. (2002) History of Late Holocene vegetation at Quintana Roo, Caribbean coast of Mexico. *Plant Ecology*, **160,** 187–192.

Iturralde-Vinent, M.A. (1982) Aspectos geológicos de la biogeografía de Cuba. *Ciencias de la Tierra y el Espacio, Academia de Ciencias, Cuba*, **5,** 84–100.

Iturralde-Vinent, M.A. (1988) *Naturaleza Geológica de Cuba*. Editorial Científico-Técnico, Havana, Cuba.

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

Iturralde-Vinent, M.A. & MacPhee, R.D.E. (1999) Paleogeography of the Caribbean region: implications for Cenozoic biogeography. *Bulletin of the American Museum of Natural History*, **238,** 1–95.

Jacknis, I. (2007) *Food in California Indian culture*. Phoebe A. Hurst Museum of Anthropology, Berkeley, CA.

Jukes, T.H. & Cantor, C. (1969) Evolution of protein molecules. *Mammalian protein metabolism* (ed. by M.N. Munro), pp. 21–132. Academic Press, New York.

Kampfer, S., Lexer, C., Glossl, J. & Steinkellner, H. (1998) Characterization of $(GA)_n$ microsatellite loci from *Quercus robur*. *Hereditas*, **129,** 183–186.

Kurz, H. & Godfrey, R.K. (1962) *Trees of northern Florida*. University of Florida, Gainesville, FL.

Lancaster, A.K., Single, R.M., Solberg, O.D., Nelson, M.P. & Thomson, G. (2007) PyPop update: a software pipeline for large-scale multilocus population genomics. *Tissue Antigens*, **69,** 192–197.

Leyden, B.W., Brenner, M., Whitmore, T., Curtis, J.H., Piperno, D.R. & Dahlin, B.H. (1996) A record of long- and short-term climatic variation from northwest Yucatán: Cenote San José Chulchacá. *The managed mosaic: Ancient Maya agriculture and resource use* (ed. by S.L. Fedick), pp. 30–50. University of Utah Press, Salt Lake City, UT.

Librado, P. & Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25,** 1451–1452.

Liu, H., Trusty, J., Oviedo, R., Anderberg, A. & Francisco-Ortega, J. (2004) Molecular phylogenetics of the Caribbean genera *Rhodogeron* and *Sachsia* (Asteraceae). *International Journal of Plant Sciences*, **165,** 209–217.

López-Almirall, A. (1979) Variabilidad de las hojas de *Quercus oleoides sagraeana* (Fagales: Fagaceae). *Ciencias Biológicas*, **3,** 59–64.

MacPhee, R.D.E., Iturralde-Vinent, M.A. & Jiménez-Vázquez, O. (2007) Prehistoric sloth extinctions in Cuba: implications of a new 'last' appearance date. *Caribbean Journal of Science*, **43,** 94–98.

Manos, P.S. & Stanford, A.M. (2001) The historical biogeography of Fagaceae: tracking the Tertiary history of temperate and subtropical forests of the Northern Hemisphere. *International Journal of Plant Sciences*, **162,** S77–S93.

Manos, P.S., Doyle, J.J. & Nixon, K.C. (1999) Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). *Molecular Phylogenetics and Evolution*, **12,** 333–349.

Marriage, T.N., Hudman, S., Mort, M.E., Orive, M.E., Shaw, R.G. & Kelly, J.K. (2009) Direct estimation of the mutation rate at dinucleotide microsatellite loci in *Arabidopsis thaliana* (Brassicaceae). *Heredity*, **103,** 310–317.

Muller, C.H. (1955) The origin of *Quercus* on Cuba. *Revista de la Sociedad Cubana de Botánica*, **7,** 41–47.

Muller, C.H. (1961) The live oaks of the series *Virentes*. *American Midland Naturalist*, **65,** 17–39.

Negrón-Ortiz, V. & Watson, L.E. (2003) Hypotheses for the colonization of the Caribbean basin by two genera of the Rubiaceae: *Erithalis* and *Ernodea*. *Systematic Botany*, **28,** 442–451.

Nixon, K.C. (1985) *A biosystematic study of* Quercus *series* Virentes *(the live oaks) with phylogenetic analyses of Fagales, Fagaceae and* Quercus. PhD Thesis, University of Texas, Austin, TX.

Nixon, K.C. & Muller, C.H. (1997) *Quercus* Linnaeus sect. Quercus white oaks. *Flora of North America north of Mexico* (ed. by F.O.N.A.E. Committee), pp. 436–506. Oxford University Press, New York.

Nuttall, T. (1842) *The North American sylva, or, a description of the forest trees of the United States, Canada, and Nova Scotia, not described in the work of F. Andrew Michaux*. J. Dobson, Philadelphia, PA.

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.

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.

Peakall, R. & Smouse, P.E. (2006) genalex 6: genetic analysis in Excel: population genetic software for teaching and research. *Molecular Ecology Notes*, **6,** 288–295.

Pearse, I.S. & Hipp, A.L. (2009) Phylogenetic and trait similarity to a native species predict herbivory on non-native oaks. *Proceedings of the National Academy of Sciences USA*, **106,** 18097–18102.

Peros, M.C., Reinhardt, E.G. & Davis, A.M. (2007) A 6000-year record of ecological and hydrological changes from Laguna de la Leche, north coastal Cuba. *Quaternary Research*, **67,** 69–82.

Petit, R.J. & Hampe, A. (2006) Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics*, **37,** 187–214.

Petit, R.J., Pineau, E., Demesure, B., Bacilieri, R., Ducousso, A. & Kremer, A. (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences USA*, **94,** 9996–10001.

Petit, R.J., Csaikl, U.M., Bordács, S. *et al.* (2002) Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management*, **156,** 5–26.

Petit, R.J., Bodenes, C., Ducousso, A., Roussel, G. & Kremer, A. (2004) Hybridization as a mechanism of invasion in oaks. *New Phytologist*, **161,** 151–164.

Pindell, J.L. & Barrett, S.F. (1990) Geological evolution of the Caribbean region: a plate-tectonic perspective. *The geology of North America, Vol. H, The Caribbean region* (ed. by G. Dengo and J.E. Case), pp. 405–432. The Geological Society of America, Boulder, CO.

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

Potts, B.M. & Reid, J.B. (1988) Hybridization as a dispersal mechanism. *Evolution*, **42,** 1245–1255.

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

Rasband, W.S. (2010) *ImageJ*. U.S. National Institutes of Health, Bethesda, MD. Available at: http://imagej.nih.gov/ij/ (accessed February 2010).

Richard, A. (1853) Fanerogamia ó plantas vasculares. *Historia física, política y natural de la Isla de Cuba* (ed. by R. de la Sagra), p. 230. Arthur Bertrand, Paris.

Rieseberg, L.H. & Soltis, D.E. (1991) Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants*, **5,** 65–84.

Rivas, L.R. (1958) The origin, evolution, dispersal, and geographical distribution of the Cuban poeciliid fishes of the tribe *Girardinini*. *Proceedings of the American Philosophical Society*, **102,** 281–320.

Rosen, D.E. (1975) A vicariance model of Caribbean biogeography. *Systematic Zoology*, **24,** 431–464.

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

Rouse, I. & Allaire, L. (1978) The Caribbean. *Chronologies in New World archeology* (ed. by R.E. Taylor and C.W. Meighan), pp. 431–481. Academic Press, New York.

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

Savage, J.M. (1982) The enigma of the Central American herpetofauna: dispersals or vicariance? *Annals of the Missouri Botanical Garden*, **69,** 464–547.

Shaw, J., Lickey, E.B., Schilling, E.E. & Small, R.L. (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany*, **94,** 275–288.

Simpson, G.G. (1956) Zoogeography of West Indian land mammals. *American Museum Novitates*, **1759,** 1–28.

Slatkin, M. (1994) An exact test for neutrality based on the Ewens sampling distribution. *Genetical Research*, **64,** 71–74.

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

Slatkin, M. (1996) A correction to the exact test based on the Ewens sampling distribution. *Genetical Research*, **68,** 259–260.

Slatkin, M. & Excoffier, L. (1996) Testing for linkage disequilibrium in genotypic data using the expectation–maximization algorithm. *Heredity*, **76,** 377–383.

Steinkellner, H., Lexer, C., Turetschek, E. & Glossl, J. (1997) Conservation of $(GA)_n$ microsatellite loci between *Quercus* species. *Molecular Ecology*, **6,** 1189–1194.

Strasburg, J.L. & Rieseberg, L.H. (2010) How robust are "isolation with migration" analyses to violations of the IM model? A simulation study. *Molecular Biology and Evolution*, **27,** 297–310.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123,** 585–595.

Trelease, W. (1924) The American oaks. *Memoirs of the National Academy of Sciences*, **20,** 1–255.

Tsitrone, A., Kirkpatrick, M. & Levin, D.A. (2003) A model for chloroplast capture. *Evolution*, **57,** 1776–1782.

Wahlund, S. (1928) Zusammensetzung von Populationen und Korrelationsercheinungen vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas*, **11,** 65–106.

Wallace, A.R. (1881) *Island life or the phenomena and causes of insular fauna and floras, including a revision and attempted solution of the problem of geological climates*. Harper & Brothers, New York.

Watterson, G.A. (1978) The homozygosity test of neutrality. *Genetics*, **88,** 405–417.

Whittemore, A.T. & Schaal, B.A. (1991) Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences USA*, **88,** 2540–2544.

Won, Y.J. & Hey, J. (2005) Divergence population genetics of chimpanzees. *Molecular Biology and Evolution*, **22,** 297–307.

Zeng, Y.-F., Liao, W.-J., Petit, R.J. & Zhang, D.-Y. (2010) Exploring species limits in two closely related Chinese oaks. *PLoS ONE*, **5,** e15529.

SUPPORTING INFORMATION

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

Appendix S1 Sample site coordinates, species composition, sample sizes, and whether or not morphological data were measured.

Appendix S2 Maps and admixture plots for two, three and nine (lowest deviance information criterion) population clusters defined by InStruct.

Appendix S3 Supplementary morphological analyses: results of similarity percentage (SIMPER) analysis and hierarchical clustering based on leaf traits.

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.

BIOSKETCHES

Paul F. Gugger is currently a postdoctoral scholar at the University of California, Los Angeles and is interested in the phylogeography, population genetics, genomics and palaeoecology of trees.

Jeannine Cavender-Bares is an associate professor in the Department of Ecology, Evolution and Behavior at the University of Minnesota. Her work examines the physiological ecology, community assembly, climate adaptation and phylogeography of plants with a focus on the New World oaks (*Quercus*).

Editor: Malte Ebach

This Special Issue was motivated by research undertaken by partners in the SEEDSOURCE consortium (http://www.seed-source.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.