MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 3668–3687 doi: 10.1111/mec.13269

Phylogeny and biogeography of the American live oaks (*Quercus* subsection *Virentes*): a genomic and population genetics approach

JEANNINE CAVENDER-BARES,* ANTONIO GONZÁLEZ-RODRÍGUEZ,† DEREN A. R. EATON,‡ ANDREW A. L. HIPP,§¶ ANNE BEULKE* and PAUL S. MANOS**

*Department of Ecology, Evolution and Behavior, University of Minnesota, Saint Paul, MN 55108, USA, †Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Nacional Autónoma de México, Morelia, 58190 Michoacán, México, ‡Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520-8106, USA, §The Morton Arboretum, Lisle, IL 60532, USA, ¶The Field Museum, Chicago, IL 60605, USA, **Department of Biology, Duke University, Raleigh, NC 27708, USA

Abstract

The nature and timing of evolution of niche differentiation among closely related species remains an important question in ecology and evolution. The American live oak clade, *Virentes*, which spans the unglaciated temperate and tropical regions of North America and Mesoamerica, provides an instructive system in which to examine speciation and niche evolution. We generated a fossil-calibrated phylogeny of *Virentes* using RADseq data to estimate divergence times and used nuclear microsatellites, chloroplast sequences and an intron region of nitrate reductase (NIA-*i*3) to examine genetic diversity within species, rates of gene flow among species and ancestral population size of disjunct sister species. Transitions in functional and morphological traits associated with ecological and climatic niche axes were examined across the phylogeny. We found the *Virentes* to be monophyletic with three subclades, including a southwest clade, a southeastern US clade and a Central American/Cuban clade. Despite high leaf morphological variation within species and transpecific chloroplast haplotypes, RADseq and nuclear SSR data showed genetic coherence of species. We estimated a crown date for *Virentes* of 11 Ma and implicated the formation of the Sea of Cortés in a speciation event ~5 Ma. Tree height at maturity, associated with fire tolerance, differs among the sympatric species, while freezing tolerance appears to have diverged repeatedly across the tropical–temperate divide. Sympatric species thus show evidence of ecological niche differentiation but share climatic niches, while allopatric and parapatric species conserve ecological niches, but diverge in climatic niches. The mode of speciation and/or degree of co-occurrence may thus influence which niche axis plants diverge along.

Keywords: conservation, ecological and climatic niches, fossil calibration, genomic data, introgression, phylogeography, RADseq, Sea of Cortés, *Virentes*

Received 3 January 2015; revision received 15 May 2015; accepted 1 June 2015

Introduction

Understanding drivers of speciation and adaptive shifts along multiple dimensions of species niches is a long-standing concern in ecology and evolution. Studies of species complexes that span tropical and temperate regions help elucidate how historical and environmental factors influence speciation and adaptive evolution. Here, we undertook a synthetic examination of the phylogeny, functional ecology and biogeographic history of the live oak clade (*Quercus* subsection *Virentes)* that spans the temperate and dry tropical zones of unglaciated North America, Central America and Cuba (Nixon

Correspondence: Jeannine Cavender-Bares, Fax: +1 612 624 6777; E-mail: cavender@umn.edu

& Muller 1997; Manos *et al.* 1999; Cavender-Bares *et al.* 2011), to gain insight into the factors that drive speciation and shifts in species ecological niches.

The *Virentes* fall within a diverse and ecologically important woody genus in this geographic region but also a group notorious for introgressive gene flow (Whittemore & Schaal 1991; Howard *et al.* 1997; Dumolin-Lapègue *et al.* 1999; Belahbib *et al.* 2001; Dodd & Kashani 2003; Valbuena-Carabana *et al.* 2005; de Dios *et al.* 2006; Curtu *et al.* 2007a) and are sister to the more diverse and widespread white oaks of section *Quercus*. Hybridization between *Virentes* and other white oaks is possible but uncommon (Muller 1961a,b; Nixon 1985). The clade of seven named species (*Quercus virginiana* Miller, *Quercus geminata* Small, *Quercus minima* Small, *Quercus brandegeei* Goldm., *Quercus fusiformis* Small, *Quercus oleoides* S. C. and *Q. sagraeana* Nutt.) is strikingly distinct phylogenetically and morphologically from the other white oaks (Nixon 1985; Manos *et al.* 1999; Cavender-Bares *et al.* 2004a; Pearse & Hipp 2009; Hubert *et al.* 2014) and includes widespread and narrow endemic species that collectively cover the southeastern US, eastern Mexico, southern Baja California, Central America and Cuba (Fig. 1) (Muller 1961a; Nixon 1985; Nixon & Muller 1997).

Species of *Virentes* vary widely in range size and in the degree of contact with other species in the lineage; three have broad distributions (*Q. fusiformis*, *Q. virginiana* and *Q. oleoides*), and two are geographically isolated and narrowly distributed (*Q. brandegeei* and *Q. sagraeana*) (Fig. 1). The three southeastern US species are sympatric (*Q. virginiana*, *Q. geminata* and *Q. minima*), while the remaining species are parapatric or allopatric. Previous studies reveal contrasting mechanisms that limit gene flow between sympatric and parapatric species. For example, differences in flowering time are implicated in limiting gene flow between sympatric species (*Q. geminata* and *Q. minima*) (Cavender-Bares & Pahlich 2009), while differences in freezing tolerance are implicated in asymmetrical gene flow between tropical *Q. oleoides* and temperate *Q. virginiana* (Cavender-Bares *et al.* 2011). Geographic barriers to gene flow were associated with the isolation and formation of the Cuban oak, *Q. sagraeana* (Gugger & Cavender-Bares 2013), as well as the fixation of a single chloroplast haplotype at the southern range limit of *Q. oleoides* in Costa Rica (Cavender-Bares *et al.* 2011). No attempt has yet been made to link phylogeographic patterns of populations within species to the phylogenetic relationships among them to address how biogeographic processes and limits to gene flow influence macroevolution, mechanisms of speciation and niche evolution across species.

The lineage is unusual within the oaks in being restricted to low-elevation habitats and occurring largely on well-drained sandy soils or volcanic tuff (Muller 1961a; Boucher 1983; Nixon 1985; Cavender-Bares *et al.* 2004a). *Virentes* are distinguished by the synapomorphies of fused cotyledons and fused stellate trichomes (Candolle 1862; Engelmann 1876-1877; Lewis 1911; Coker 1912; Camus 1936-1938). All species are wind-pollinated and interfertile (Nixon 1985; Nixon & Muller 1997) and have unusually high wood density. *Virentes* maintain a green or mostly green canopy through the winter (southeastern US and Texas) or during the dry season (Central America) with a leaf lifespan of approximately 1 year (Nixon 1985; Cavender-Bares & Holbrook 2001).

Climatic distributions and exposure to freezing temperatures vary among species (Fig. 1), but *Virentes* are restricted to temperate climates with mild winters or seasonally dry tropical climates. Variation in vulnerability to freezing and tolerance of drought likely influences distribution and migration patterns (Cavender-Bares 2007; Cavender-Bares & Pahlich 2009; Koehler *et al.* 2012). Species also have contrasting life history traits and growth forms, varying from fire-dependent shrubs with underground rhizomes (*Q. minima*) to drought-adapted, fire-tolerant short trees with pubescent leaves resistant to water loss (*Q. geminata*) and large fire-intolerant trees (to 20 m) that are less drought tolerant (*Q. virginiana*) (Cavender-Bares *et al.* 2004a).

The IUCN-red-listed *Q. brandegeei* in the Cape of southern Baja California and its geographically most proximate relative *Quercus fusiformis* in eastern central Texas and northern Mexico (Fig. 1) represent one of the broadest disjunctions known in American *Quercus* and are hypothesized to have split from a once-widespread taxon (Muller 1967; Nixon 1985). Estimating the divergence time between these two species would provide insight into the causes of vicariance and speciation, including the potential role of the formation of the Sea of Cortés, which separated Baja California from continental Sonora, and was complete by about 5 million years ago (Riddle *et al.* 2000; Garrick *et al.* 2009). A second disjunction occurs within *Q. oleoides* between the isolated population at the southern range limit in Costa Rica and Honduras across the Nicaraguan Depression, which has increasingly been implicated as the vicariant event associated with a common phylogeographic break in many different kinds of taxa (Arrivillaga *et al.* 2002; Gutiérrez-García & Vázquez-Domínguez 2013; Poelchau & Hamrick 2013; Rodríguez-Correa *et al.* 2015). A previous study hypothesized that local climatic change associated with the rise of mountain chains in northwestern Costa Rica was a primary factor in the initiation and persistence of a disjunct *Q. oleoides* population there (Cavender-Bares *et al.* 2011). Synthesizing prior and new data, this study addresses four outstanding questions:

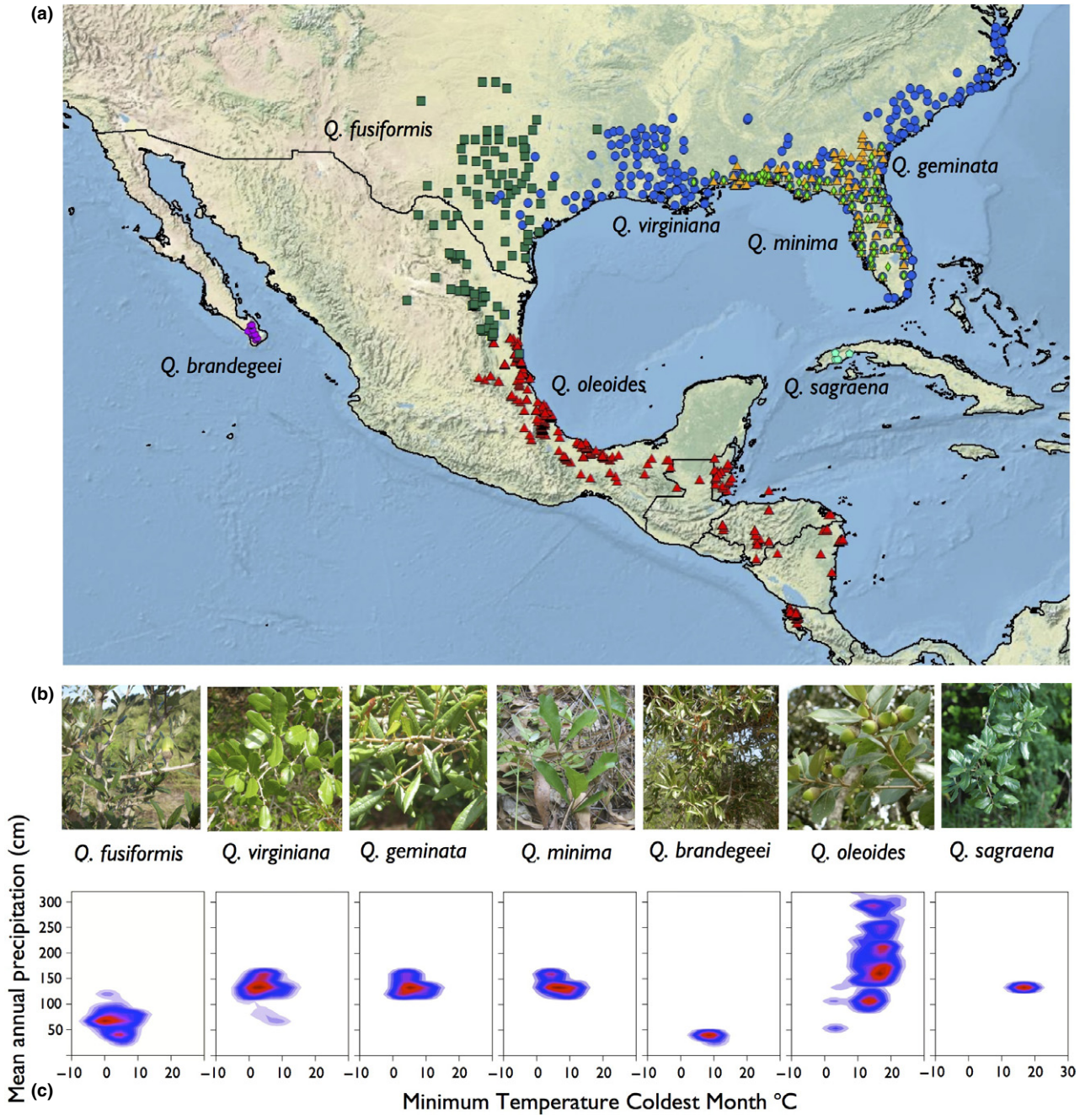

Fig. 1 (a) Distribution map of *Virentes* based on species occurrences. Legend: purple circles = *Quercus brandegeei*, dark green squares = *Quercus fusiformis*, orange triangles = *Quercus geminata*, green diamonds = *Quercus minima*, red triangles = *Quercus oleoides*, cyan hexagons = *Quercus sagraeana*, and blue circles = *Quercus virginiana*. (b) Leaf-level photos of each *Virentes* species. (c) Species climatic distributions are shown for mean annual precipitation (cm) and mean minimum temperature of the coldest month (°C) for each species based on occurrence data and WorldClim bioclimatic variables 6 and 12 (Hijmans *et al*. 2005) using two-dimensional kernel density estimation. Red colours indicate climatic regions with highest density of occurrence. Photos were taken by J.C.B., except *Q. fusiformis*, taken by F. Hoerner.

1 What are the major clades and sister-species relationships within the *Virentes*?
2 Given the phylogenetic relationships of the *Virentes*, do population-level markers reveal concordant structure and species coherence across the entire range of *Virentes*?
3 Can major geographic events, such as the formation of the Sea of Cortés separating Baja California from

continental Mexico or the Nicaraguan Depression in Central America, be implicated in genetic isolation and/or speciation events within the *Virentes*?

4 How do functional traits, ecological habitats and climatic distributions shift among species throughout the distribution of the clade among sympatric and allopatric species?

To address these questions, we generated a phylogenetic hypothesis for the *Virentes* using restriction site associated DNA sequences (RADseq) and examined population genetic diversity, structure and isolation, using simple sequence repeats (SSRs), chloroplast sequences and a low-copy intron region of nitrate reductase (NIA-*i3*). We also generated a fossil-calibrated phylogeny using RAD sequence data to infer divergence times and biogeographic historical events involved in speciation. Finally, we examined leaf morphological traits, growth form (tree height) and vulnerability to freezing to examine shifts in functional traits that are linked to ecological and climatic niches. We tested biogeographic hypotheses using both a phylogenomic approach with many loci but limited individuals and population-level analyses with many individuals but fewer loci. We applied a previously developed RADseq data processing pipeline (Eaton 2014) to generate a concatenated data matrix and used two phylogenetic methods for reconstructing the phylogeny. The first approach allowed us to test for species coherence, while the second allowed us to calibrate the phylogeny using minimum fossil ages.

Methods

Taxon sampling

Individual trees of each *Virentes* species were sampled throughout their ranges. Identification of species was based on leaf, bark and stem height characters following Muller (1961a), Nixon & Muller (1997) and Kurz & Godfrey (1962). A list of the total samples and their geographic localities is provided on Dryad (doi:10.5061/dryad.855 pg). Of these, 27 from across the major geographic regions were selected for RAD sequencing (Table S1, Supporting information) but chosen randomly within regions, excluding individuals from the putative hybrid zone between *Quercus fusiformis* and *Quercus oleoides*. Voucher specimens are housed in the University of Minnesota Bell Museum of Natural History. Permit and collection authorizations are provided in Appendix S1 (Supporting information). DNA extraction and sequencing methods followed Hipp *et al.* 2014 (see Appendix S1, Supporting information for details).

RADseq data

Illumina sequencing. A RAD sequencing library was prepared at Floragenex Inc. (Eugene, Oregon) as described in Hipp *et al.* (2014) using *Pst*I restriction enzyme. Samples were pooled into a multiplexed library and sequenced on an Illumina HiSeq 2000 to generate 100-bp single-end reads. Data from eight additional *Quercus* species (*Q. acutissima* Carruth., *Q. chrysolepis* Liebm., *Q. durata* Jepson, *Q. douglasii* Hook. & Arn., *Q. arizonica* Sarg., *Q. engelmannii* Greene, *Q. hemisphaerica* Bartram ex Willd. and *Q. nigra* L.), generated by Hipp *et al.* (2014), were included as outgroup taxa. The libraries for these data were generated by the same technique. Seven are 100-bp single-end reads from a lane Illumina HiSeq 2000, and one is 60-bp single-end read run on an Illumina GAIIx. All sequencing was done at Floragenex Inc. (Eugene, OR, USA).

Data filtering. Raw sequence data were analysed in the software pipeline PYRAD v.1.4 (Eaton & Ree 2013), which filters and clusters RAD sequences to identify putatively orthologous loci. This pipeline is suited to the phylogenetic scale of our study because of its use of global alignment clustering which can cluster highly divergent sequence while taking into account indel variation. Filtering parameters were set to replace base calls of $Q < 20$ with an ambiguous base (N) and discard sequences containing more than three Ns. Reads clustered at 85% and 92% similarity yielded similar results; therefore, we reported only analyses run at 85% clustering similarity. Consensus base calls were made for clusters with a minimum depth of coverage >5. After correcting for errors, loci containing more than two alleles were excluded as potential paralogs since all taxa in the study are diploid. Consensus loci were then clustered across samples at Floragenex Inc. and aligned. A final filtering step excluded loci that contain any site that is heterozygous across more than three samples, as this is more likely to represent a fixed difference among clustered paralogs than a true polymorphism at the scale of this study.

RADseq data sets. The samples sequenced for this study had an average of 895 ± 544 K reads that passed quality filtering. These clustered into an average of 45 ± 15 K clusters per sample, with a mean depth of 15.4, giving rise to 41 ± 15K consensus sequences per sample (Table S2, Supporting information). When clustered across samples, the largest data set 'All_min4' contains 74K RADseq loci with ~63% missing data. The other data sets contain fewer loci, but with less missing sequence.

Phylogenetic analyses

Maximum-likelihood trees were inferred for each concatenated supermatrix, with missing data coded as 'N's,

using RAXML 7.2.8 (Stamatakis 2006), with bootstrap support estimated from 200 replicate searches from random starting trees using the GTR + GAMMA nucleotide substitution model.

Divergence times. To estimate divergence dates, we inferred fossil-calibrated time trees using BEAST v.1.75 enabled by parallel processing with BEAGLE (2009–2013 Phylogenetic Likelihood Working Group) at the University of Minnesota Supercomputing Institute (MSI) facilities. A subset of the total taxa was used, including one to three individuals from each species within the *Virentes*, and eight outgroup taxa, to reduce run time for convergence of the Markov Chain Monte Carlo (MCMC). A total of 25 taxa were included, of which 17 were *Virentes* individuals, with 817 555 bp of concatenated sequences (sub_c85d6m20p3, dx.doi.org/10.5061/dryad.524mf). A lognormal relaxed molecular clock was enforced with a GTR substitution model with gamma site heterogeneity, four rate categories, and a Yule process tree prior. It was not possible to partition the concatenated sequences by individual loci (>8175); thus, we assumed a common mutation rate across the genome. The MCMC length was 100 000 000 thinned every 10 000 for analysis. Analyses were run three times with different starting seeds. BEAST log files were analysed with Tracer for convergence, the phylogenetic results during the burn-in period were removed, and the combined tree files were used to generate a maximum clade credibility tree with median heights in TREEANNOTATOR v. 1.7.5.

Fossil calibration. To calibrate the tree, fossil dates were imposed as priors at three nodes: (i) the white oak clade, *Quercus* section *Quercus*, which includes the *Virentes*; (ii) the American oak clade, which includes the red (*Quercus* section *Lobatae*), white (*Quercus* section *Quercus*) and golden oaks (*Quercus* section *Protobalanus*); and (iii) the root node for the genus *Quercus.* No definitive fossils are available within the *Virentes*. Fossil leaves from Oligocene deposits in Colorado (MacGinitie 1953) have been attributed to *Virentes* but never substantiated and were not used. These fossils led to the hypothesis of an ancient origin of the *Virentes* (MacGinitie 1953) discussed in Nixon (1985). The white oak clade was assigned an informative lognormal prior with a median age of 31.3 million years before the present with 95% of the distribution between 28 and 38.2 million years BP. The median date matched the mid-point of the minimum fossil age based on Elsik & Yancey (2000) of 28.4–33.9 Ma for oaks in the Americas, but we allowed a longer tail. This date is a more accurate estimate for the stratum that contains fossil cupules named *Quercus huntsvillensis* Daghlian et Crepet (Daghlian & Crepet 1983). An informative lognormal prior with a median of 40 million years ago and 95% of the distribution within 37.25 and 44.0 Ma is based on dates for the earliest documented *Quercus* macrofossils in the Americas (see Borgardt & Pigg 1999). The root node for the genus *Quercus* was assigned an informative exponential prior with a median of 50.4 Ma and 95% of the distribution within 40.8 and 85 Ma. The exponential prior is appropriate for the root given we are using minimum fossil ages estimated at 50 million years BP, but the clade could be older. Dates for the root node were based on pollen fossil evidence (Denk & Grimm 2009; Denk *et al.* 2012).

Genetic diversity and structure

SSRs. We extracted DNA and amplified eleven previously published microsatellite loci located on seven chromosomes: *QpZAG 1/2*, *QpZAG 1/5*, *QpZAG 9*, *QpZA 15*, *QpZAG 16*, *QpZAG 36*, *QpZAG 46*, *QpZAG 102*, *QpZAG 110* (Steinkeller *et al.*, 1997), *QrZAG 11* and *QrZAG 30* (Kampfer *et al.*, 1998), for individuals of *Quercus minima* ($N = 38$), *Q. fusiformis* ($N = 92$) and *Q. brandegeei* ($N = 35$), following methods described in Cavender-Bares & Pahlich (2009). These data were combined with previously published data for *Quercus geminata*, *Q. virginiana*, *Q. oleoides* and *Q. sagraeana* (Cavender-Bares & Pahlich 2009; Cavender-Bares *et al.* 2011; Gugger & Cavender-Bares 2013) for a total of 672 individuals across all species.

Samples were assigned to eight population groups, which included the seven morphological/geographic species and an unnamed but genetically distinct Costa Rican population of *Q. oleoides* previously identified (Cavender-Bares *et al.* 2011). For each of these groups, we estimated the number of alleles, N_A; the effective number of alleles, N_{Ae}; the allelic richness expressed as the expected number of alleles among two gene copies, A_R ($k = 2$); and the gene diversity corrected for sample size, H_E (Nei, 1978). The SPAGEDI software (Hardy & Vekemans, 2002) was used for these calculations. Pairwise and overall genetic differentiation (F_{ST}) among the groups was determined with the method of Weir (1996) implemented in the FREENA program (Chapuis & Estoup, 2007). A bootstrapping procedure over loci with 1000 replicates was performed to obtain mean F_{ST} values and their 95% confidence intervals. This software was also used to estimate the frequency of null alleles for each locus and group with the expectation-maximization algorithm (Dempster *et al.* 1977).

To test for underlying genetic structure and admixture among the populations and to determine how well the genetic structure in the molecular data corresponds to species designations, we used a Bayesian clustering algorithm implemented in the program STRUCTURE v.

2.3.1 (Pritchard *et al.* 2000). We used the admixture model and the correlated model with a burn-in length of 1 000 000 iterations with 100 000 MCMC replicates. We allowed K to range from 1 to 14. For each K, we ran 10 iterations and averaged the log probabilities (Ln *P*) of the data (D) [Ln *P(D)*]. We followed the method of Evanno *et al.* (2005) to examine the most probable value of K as determined by the maximum value of ΔK, which represents a large magnitude second derivative of the log likelihood.

Chloroplast DNA. Of the individuals in this data set, 391 sequences were available for a region within *trn*D–*trn*T (newly sequenced for *Q. minima, Q. fusiformis* and *Q. brandegeei* or previously published, Cavender-Bares *et al.* 2011), and 327 were available for the *rpl*32–*trn*L^{UAG} chloroplast region (Shaw *et al.*, 2007) (newly sequenced for *Q. minima, Q. fusiformis* and *Q. brandegeei* or previously published, Gugger & Cavender-Bares 2013). Sequences for both regions were available for 215 individuals and were concatenated for a total of 1450 bp. Parsimony networks with insertion–deletions coded as a fifth state and ignoring poly-A repeats were constructed for each chloroplast region separately as well as for the concatenated sequences using the *haploNet* function in the PEGAS package (Paradis *et al.* 2010) in R. The networks were constructed using an infinite site model. All three haplotype networks were qualitatively very similar, but the concatenated data set resolved the highest number of haplotypes, and only the concatenated network is reported. For each of the eight populations, we determined the total number of haplotypes, N_H; the rarefied haplotype richness, H_R; and the haplotype diversity with unordered alleles, *h* (Pons & Petit 1996), with SPAGEDI (Hardy & Vekemans, 2002). This program was also used to estimate pairwise and overall cpDNA haplotype differentiation (G_{ST}) among groups according to Pons & Petit (1996). Significance of the G_{ST} values was determined by 10 000 random permutations of individuals among groups (Hardy & Vekemans, 2002). Genetic diversity (H) was also calculated based on RADseq data as the proportion of heterozygous base calls across all sites with sufficient coverage across all loci that passed paralog filtering.

Ancestral population size

To examine whether there was evidence of range retraction in *Q. fusiformis* and *Q. brandegeei* from a common ancestor of these two sister taxa, we used the isolation-with-migration model in IMa (February 2008 version; Hey & Nielsen, 2007) to estimate the effective population sizes before and after the split between the two species. The input data included the two chloroplast regions described above (*trn*D-*trn*T: 616 bp, $N = 56$; *rpl*32: 815 bp, $N = 54$); an intron region of the low-copy nuclear gene nitrate reductase (NIA-*i3*: 945 bp, $N = 23$); and nine nuclear microsatellites described above (Zag110, Zag16, Zag46, Zag15, Zag9, Zag102, Zag1x5, Zag11 and Zag36, N range = 170–214). We sequenced NIA-*i3* using primers published by Howarth & Baum (2002) following the methods described in Cavender-Bares *et al.* 2011. Despite introgression of *Q. fusiformis* with other taxa, IMa estimates have been found to be quite robust to moderate violations of the model assumptions (Strasburg & Rieseberg 2010). A burn-in period of 10 000 000 was used (following Cavender-Bares *et al.* 2011), and Metropolis coupling was implemented using 40 chains. The analysis was repeated three times at MSI facilities; each showed convergence and yielded very similar results. We were most interested in comparing current effective population sizes with the ancestral effective population size rather than predicting actual numbers of individuals. Results are thus reported as Θ values, where $\Theta = 4N\mu$, N = the effective population size, and μ = the mutation rate. As such, it is not necessary to estimate a mutation rate.

Leaf morphometric analysis, tree height, freezing vulnerability and climatic distributions

Leaf morphology. For a subsample of individuals from each site, ten dried, pressed leaves were scanned and analysed per individual for laminar leaf area and leaf shape using the leaf imaging software SHAPE 1.2 (Iwata & Ukai, 2002) for a total of 5762 leaves from 580 individuals. This program employs a geometric morphometrics approach based on a quantitative evaluation of the contour shape of each leaf with elliptic Fourier descriptors (EFDs) (Viscosi *et al.* 2009). This software also performs a principal components analysis to summarize the information contained in the EFDs, so that the scores of principal components can be used as observed values of morphological traits in subsequent analysis (Iwata & Ukai 2002).

Tree height. Height values for a total of 110 trees from all species were available from their full-range extents, estimated using a clinometer or extendible pole from reproductive individuals.

Vulnerability to freezing. Predicted vulnerability to freezing temperatures (VF) at −15°C was calculated from mean minimum temperature of the coldest month (Bioclim variable BIO6) at the location of each collected specimen. The relationship is based on an empirical regression ($R^2 = 0.67$) between experimentally determined vulnerability to freezing and mean minimum

temperature of the coldest month of source populations of saplings grown from seed for four species (*Q. geminata*, *Q. virginiana*, *Q. fusiformis* and *Q. oleoides*) using the electrolyte leakage method of freezing injury in a previous study (Koehler *et al.* 2012). The study showed genetically based variation in freezing tolerance both within and across species in controlled environments that was strongly associated with climate of origin. Specifically, FV = 100 −(81.202 − 1.4075X), where X is the mean minimum temperature of the coldest month in the source location. Any negative values were assumed to indicate zero vulnerability to freezing at −15°C.

Climatic distributions. Species climatic distributions are described based on bioclimatic variables from the WorldClim data (Hijmans *et al.* 2005), including mean minimum temperature of the coldest month (Bioclim 6) and mean annual precipitation (Bioclim 12) using two-dimensional kernel density estimation with the kde2d function in the MASS version 7.3–34 R package (Ripley *et al.* 1998). The approach estimates the density of each species in two-dimensional climatic niche space based on locality data from collections reported in this study and cleaned GBIF data (http://www.gbif.org/, 7 July 2008) reported in Cavender-Bares *et al.* 2011.

Results

Phylogenetic reconstruction

The RAXML analyses showed that *Virentes* are monophyletic, with strong support for three subclades: a southwest clade (*Quercus fusiformis, Q. brandegeei*), which is sister to all other *Virentes*; a Florida clade (*Q. virginiana*, *Q. minima, Quercus geminata*); and a Central American clade that groups *Q. oleoides* and the Cuban oak *Quercus sagraeana* (Fig. 2a). The same topology was recovered with the reduced taxon matrix in BEAST. Accessions of *Q. fusiformis* were inferred to be paraphyletic, with sampled populations from Mexico appearing more closely related to accessions of *Q. brandegeei* than to accessions of *Q. fusiformis* from Texas. Similarly, accessions of the Cuban oak *Q. sagraeana* appear paraphyletic with Central American samples of *Q. oleoides* nested within it.

Comparison of population-wide markers and RADseq data

STRUCTURE analyses of the nuclear SSR data assign almost all individuals from each of the named species in the *Virentes* to a distinct ancestral population, with some evidence of mixed ancestry or misclassifications (Fig. 2b). This degree of species coherence aligns closely with previous work in the genus (e.g. Bacilieri *et al.* 1995; Craft *et al.* 2002; González-Rodríguez *et al.* 2004a,b; Curtu *et al.* 2007b; Hipp & Weber 2008; Cavender-Bares & Pahlich 2009; Aldrich & Cavender-Bares 2011), which suggests that morphologically defined oak species are largely genetically coherent despite introgressive hybridization. The RADseq data similarly suggest that species in the *Virentes* are predominantly monophyletic with the exclusion of a few probable hybrids (Fig. 2a), as found in previous phylogenetic analyses of the genus (Nixon 1985; Pearse & Hipp 2009).

STRUCTURE results. The largest ΔK values with the full data set indicated most probable K values of 2 and 3. These ancestral groups delineate clades rather than individual species. The next highest ΔK was for $K = 6$. Results for $K = 6$ and $K = 7$ are highly similar (Appendix S2, Supporting information), although the latter distinguishes the Cuban group, *Q. sagraeana*, and provides a level of structure best associated with species boundaries. Partitioning the data into the two or three major groups delimited by the analysis when $K = 2$ or $K = 3$ and running those data sets individually yield the same seven ancestral groups shown in Fig. 2b (see Appendix S2, Supporting information).

Quercus minima and *Q. geminata* could not be distinguished on the basis of the SSR data although the RADseq data did separate them. Some admixture was evident between *Q. virginiana* and the *Q. minima* + *Q. geminata* clade, particularly at the Big Shoals site in northern central Florida. Considerable admixture was found between *Q. fusiformis* and *Q. oleoides*, particularly in the mixed zone where their ranges overlap in northeastern Mexico. *Quercus sagraeana* was found to be a genetically distinct group, despite admixture with *Q. oleoides* and *Q. virginiana*, as previously reported (Gugger & Cavender-Bares 2013).

Quercus brandegeei was genetically distinct based on SSR data. The species harbours a unique chloroplast haplotype that is one mutation away from an ancestral haplotype that is widespread within the *Virentes* and also found within *Q. fusiformis* (Fig. 3). The combined data support *Q. brandegeei* as a distinct species. The *Quercus oleoides* population in Costa Rica also forms a distinct group, as was reported previously (Cavender-Bares *et al.* 2011). In the RADseq phylogenetic analysis, the two Costa Rican samples were clustered, but were nested within the *Q. oleoides* + *Q. sagraeana* clade.

Genetic diversity and differentiation of species

The 11 nuclear SSRs revealed that the eight morphological/geographic groups differ in their genetic diversity (Table 1). In general, the three groups with a restricted

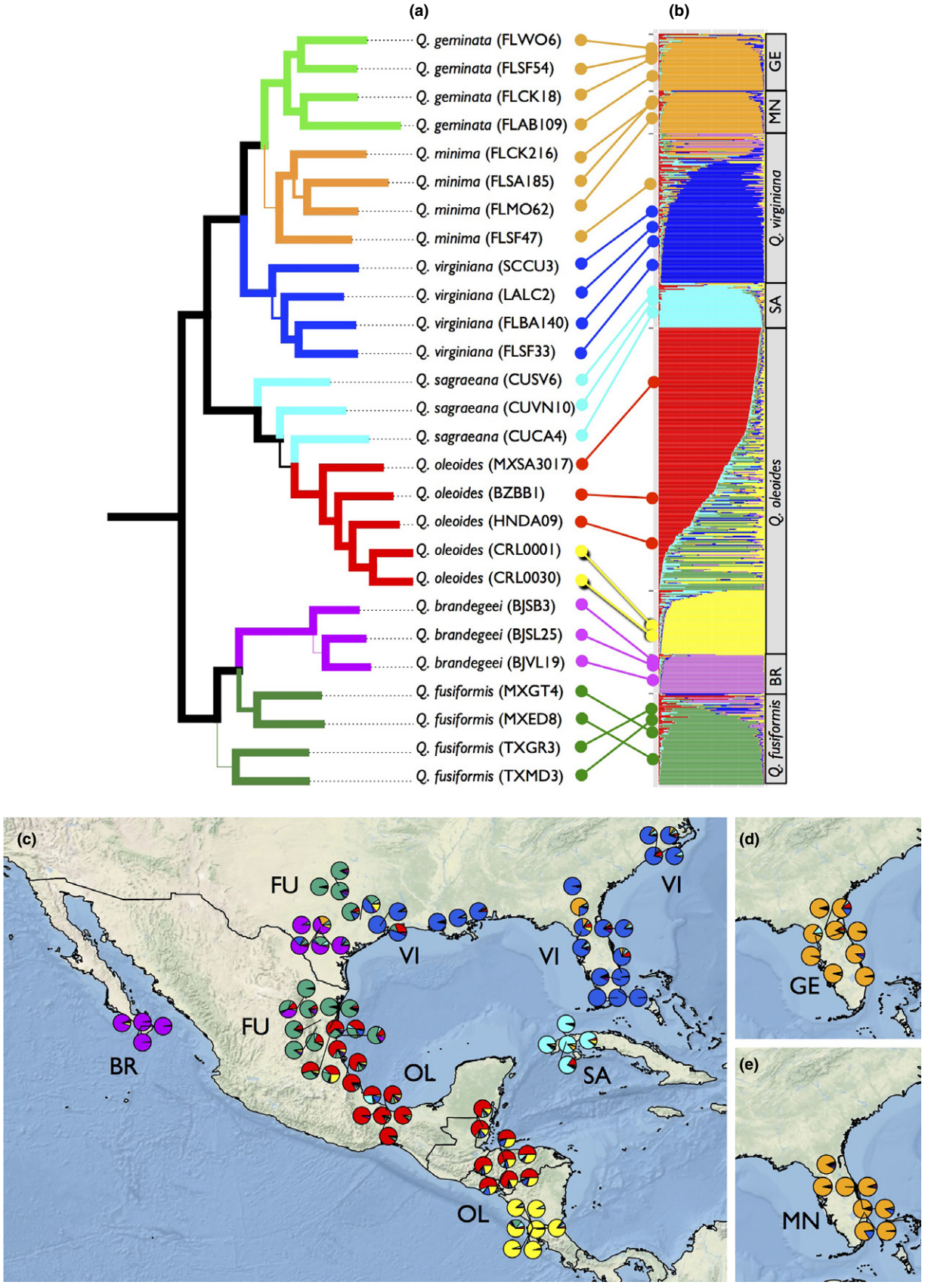

Fig. 2 (a) Phylogenetic tree inferred from RADseq (min20) data for 27 *Virentes* individuals (8 outgroup taxa not shown) using RAXML. (b) Results of STRUCTURE analysis showing proportion of ancestry ($K = 7$) for 672 individuals across all species based on 11 nuclear SSRs. Ancestral groups corresponding to species groups are given the same colours as in (a), except that *Quercus minima* and *Quercus geminata* were not differentiated using SSRs and are both shown with orange. Lines connect the individuals in the phylogenetic tree to the same individuals in the STRUCTURE analysis. Species names or abbreviations are shown to the right of the groups. (c) Geographic distribution of ancestral groups ($K = 7$). Pie charts show ancestral group proportions, averaged by site for each collection site and morphological species. (d) *Q. geminata* and (e) *Q. minima* are shown separately for clarity. Abbreviations are as follows: VI = *Quercus virginiana*; OL = *Quercus oleoides*; FU = *Quercus fusiformis*; BR = *Quercus brandegeei*; SA = *Quercus sagraeana*; MN = *Quercus minima*; GE = *Quercus geminata*.

distribution range (*Q. brandegeei*, *Q. sagraeana* and the isolated Costa Rican population of *Q. oleoides*) have lower genetic variation than the other five groups. Metrics of genetic diversity from chloroplast and SSR markers were significantly predicted by estimates of current range size (mean R^2 values were 0.51, range: 0.43–0.61; mean $P = 0.035$, range: 0.019–0.051) with variation depending on the metric of diversity and the range size

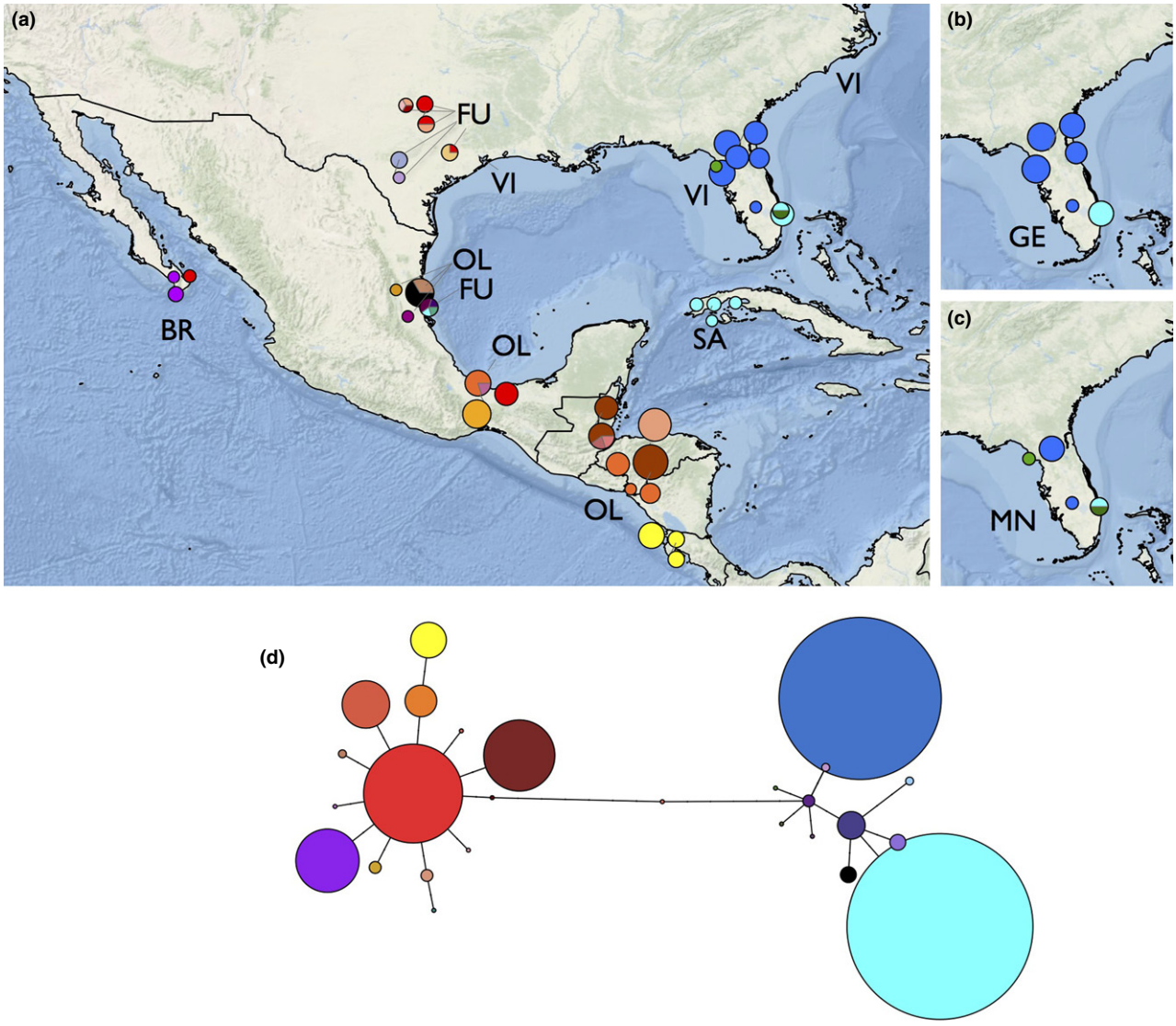

Fig. 3 (a) Chloroplast haplotypes within the *Virentes*. Circle size is proportional to sample size of the haplotype. Pie charts show haplotype proportions, averaged for site locations. Morphological species are indicated with abbreviations and are the same as in the Fig. 2 legend. (b) *Quercus minima* and (c) *Q. geminata* are shown separately for clarity. (d) The minimum-spanning network for 26 haplotypes constructed using an infinite-sites model.

estimation method. In contrast, the heterozygocity of diversity of nonconserved loci in the RADseq data (H-RAD), measured as the proportion of sites within an individual that were called heterozygous, did not covary with range size (Table 1; methods in Appendix S1, Supporting information). All interspecific genetic differentiation values (F_{ST}) were significantly different from zero.

Two major chloroplast lineages are apparent, one of which is restricted to north and east of northern Mexico; one major chloroplast haplotype is shared among four species, including all of the widespread species (Fig. 3). The number of chloroplast haplotypes in each of the morphological/geographic groups varied considerably (Table 1). The groups with restricted geographic range had one (Costa Rican population of *Q. oleoides* and *Q. sagraeana*) or two haplotypes (*Q. brandegeei*), while *Q. virginiana* had seven, *Q. fusiformis* nine and *Q. oleoides* eleven haplotypes. However, after rarefaction, haplotype richness was highest in *Q. fusiformis*, followed by *Q. oleoides* and *Q. virginiana*. From the 26 haplotypes identified, six were shared between two or more groups and the rest were exclusively found in one group. Genetic differentiation (G_{ST}) was significant for all pairwise comparisons except between *Q. geminata* and *Q. minima* and between *Q. virginiana* and *Q. minima*, but values were fairly low (Table 2). Significant values ranged from 0.10 (*Q. geminata* and *Q. virginiana*) to 1.0 (between *Q. sagraeana* and the Costa Rican population of *Q. oleoides*). Chloroplast haplotypes thus show geographic patterns but do not separate along species boundaries, given that genetic differentiation among species groups is low.

Range retraction of Q. fusiformis and Q. brandegeei ancestor

Estimates from IMa of ancestral population size of the common ancestor of *Q. fusiformis* and *Q. brandegeei* indicate at least a 30-fold larger effective population size than current estimates for both species combined (Fig. 4), indicating a once-broader distribution.

Node dates and divergence times

The estimated divergence of *Virentes* from the rest of the white oaks is on the order of 28 million years ago (27–31 Ma) with the crown age of the *Virentes* estimated at 11 Ma (8.4–14.1 Ma 95% HPD; Fig. 5a). The split between the southeastern US clade and the *Q. sagraeana* + *Q. oleoides* clade is estimated at 9.3 Ma (6.9–11.7 Ma 95% HPD). The divergence of *Q. brandegeei* from *Q. fusiformis* is estimated at 5.2 Ma (2.6–8.1 Ma 95% HPD). At the southern range limit of *Virentes*, the divergence of the geographically disjunct Costa Rican population from the Honduran population of *Q. oleoides* is estimated at 1.9 Ma (1.0–3.1 Ma 95% HPD), earlier than previously estimated (Cavender-Bares *et al.* 2011).

Functional and morphological traits

Differentiation in freezing vulnerability and leaf morphology was associated with divergence in climatic distributions among allopatric species, while growth form and tree height (Myers 1990; Cavender-Bares *et al.* 2004b), associated with fire tolerance, diverged among sympatric species in the southeastern US clade (Fig. 5c–e).

Table 1 Indices of genetic diversity from 11 nuclear microsatellite simple sequence repeats (SSR) for a total of 672 individuals and chloroplast haplotypes for a total of 215 individuals across the seven *Virentes* species and the geographically isolated Costa Rican population of *Quercus oleoides*. Geographic range sizes are estimated by a minimum convex polygon (MCP) of occurrence points. Abbreviations are as follows: NA_SSR, number of SSR alleles; NAe, effective number of SSR alleles (Nielsen *et al.* 2003); AR(k = 2)_SSR, allelic richness (expected number of alleles among 2 gene copies); He, gene diversity corrected for sample size (Nei, 1978). H-RAD is the heterozygocity (proportion of sites within an individual that were heterozygous) averaged across all sampled individuals in the taxon; SD is the standard deviation. Indices of genetic diversity from chloroplast DNA sequences are abbreviated as follows: Nh, number of haplotypes; Hr, haplotype richness (rarefacted); h, gene diversity with unordered alleles (Pons & Petit 1996).

Species	NA_SSR	NAe_SSR	AR(k = 2)_SSR	He_SSR	H-RAD	SD	Nh_cp	Hr_cp	h_cp	Range (MCP) km^2
Quercus brandegeei	6.91	2.93	1.59	0.59	0.35	0.01	2	1.97	0.44	1660
Quercus fusiformis	11.91	6.03	1.71	0.71	0.45	0.01	9	5.26	0.85	676 070
Quercus geminata	15	7.2	1.73	0.73	0.38	0.04	2	1.94	0.39	217 876
Quercus minima	11.82	6.2	1.76	0.76	0.41	0.13	4	4	0.64	274 344
Quercus oleoides	17.27	6.66	1.77	0.77	0.34	0.04	11	5.24	0.87	624 469
Q. oleoides_Costa Rica	9.45	4.05	1.65	0.65	0.30	0.01	1	1	0.00	
Quercus sagraeana	6.91	3.94	1.61	0.61	0.39	0.06	1	1	0.00	1431
Quercus virginiana	18.73	7.22	1.8	0.80	0.37	0.05	7	4.18	0.74	1 117 033

Table 2 Above diagonal: genetic differentiation (F_{ST}) values among *Virentes* species and the geographically isolated Costa Rican population of *Quercus oleoides* based on 11 nuclear microsatellite loci not correcting for null alleles; the correction gives nearly identical results; global F_{ST} not using ENA: 0.116, F_{ST} using ENA: 0.113. Below diagonal: chloroplast DNA haplotype differentiation (G_{ST}) among *Virentes* species and the geographically isolated Costa Rican population of *Q. oleoides*; values in bold are significant.

	Quercus brandegeei	*Quercus fusiformis*	*Quercus geminata*	*Quercus minima*	*Quercus oleoides*	*Quercus oleoides*_ Costa Rica	*Quercus sagraeana*	*Quercus virginiana*
Quercus brandegeei	—	**0.16**	**0.25**	**0.24**	**0.18**	**0.24**	**0.21**	**0.16**
Quercus fusiformis	**0.28**	—	**0.20**	**0.18**	**0.06**	**0.17**	**0.14**	**0.10**
Quercus geminata	**0.58**	**0.37**	—	**0.02**	**0.15**	**0.23**	**0.16**	**0.12**
Quercus minima	**0.44**	**0.22**	0.03	—	**0.14**	**0.23**	**0.16**	**0.11**
Quercus oleoides	**0.33**	**0.11**	**0.37**	**0.24**	—	**0.08**	**0.07**	**0.06**
Quercus oleoides _Costa Rica	**0.78**	**0.57**	**0.80**	**0.68**	**0.57**	—	**0.09**	**0.10**
Quercus sagraeana	**0.78**	**0.55**	**0.74**	**0.68**	**0.57**	**1.00**	—	**0.08**
Quercus virginiana	**0.37**	**0.14**	**0.10**	−0.01	**0.17**	**0.63**	**0.58**	—

Leaf traits. Leaf size and shape are associated with both climate and resource acquisition (Wright *et al.* 2004). The tropical species with high mean annual precipitation (MAP) (Fig. 1c), *Q. oleoides* and *Q. sagraeana*, had significantly larger leaves than all other species (Fig. 5b). *Quercus oleoides* had the largest leaf size (1628.7 mm^2) and was significantly differentiated from *Q. sagraeana* (1272.5 mm^2), which had the next largest leaf size. *Quercus brandegeei* had the smallest leaf size (461.4 mm^2), followed by *Q. geminata* and *Q. fusiformis*, and its leaves were significantly smaller than all species except *Q. geminata*. These three species occur either in areas with low MAP (*Q. brandegeei* and *Q. fusiformis*) or in xeric soils (*Q. geminata*). *Quercus minima* and *Q. virginiana* had intermediate leaf sizes (Fig. 5b).

The first principal component of the Fourier descriptors of leaf shape (Appendix S3, Supporting information) also showed significant differentiation among all

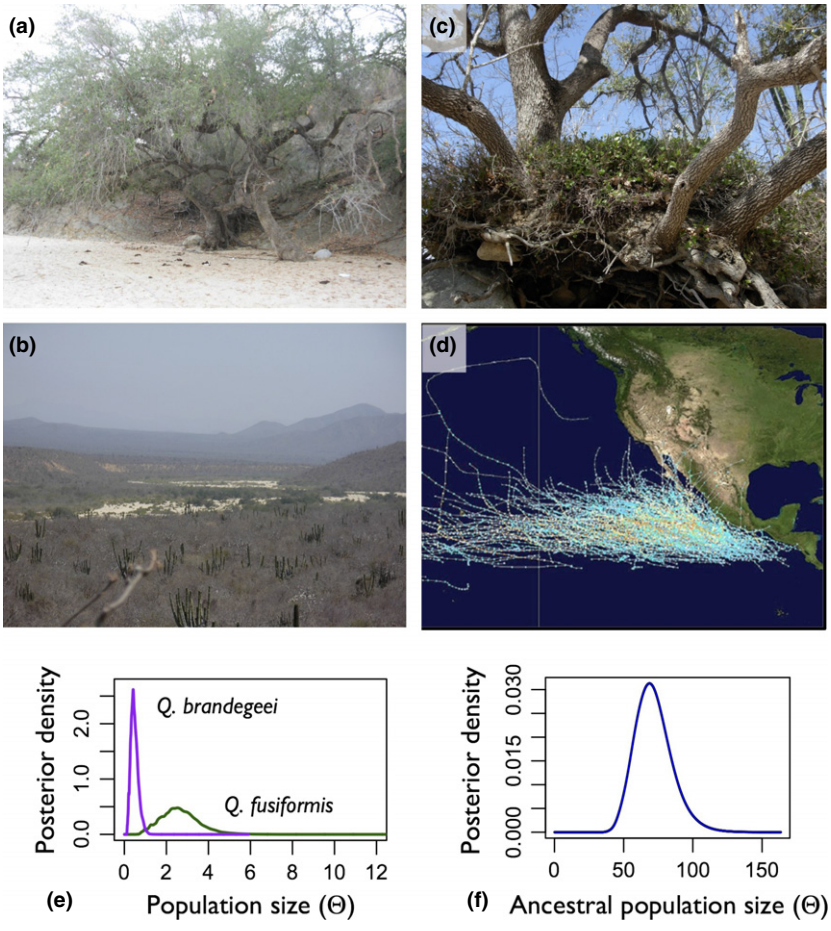

Fig. 4 (a) *Quercus brandegeei,* an IUCN-red-listed endangered species, occurs only along ephemeral riverbeds embedded (b) within a narrow range in the desert of southern Baja California. (c) Vegetative resprouts from the mother tree. Photos (a–c) were taken by J.C.B. (d) Dried riverbeds likely are temporarily inundated during hurricane season given the regular occurrence of hurricanes. Shown are Pacific hurricane tracks from 1980 to 2005 (image licensed under public domain; commons.wikimedia.org/wiki/File:Pacific_hurricane_tracks_1980–2005). (e) Current effective population sizes of *Q. brandegeei* and its sister species *Q. fusiformis* are given in units of Θ, equivalent to 4 Nμ, where N is effective population size and μ is the mutation rate. (f) Effective population size of the common ancestor is given in the same units, revealing that it was much larger. Values of Θ were predicted using the isolation-with-migration model (Hey and Nielson 2007) based on chloroplast sequences, NIA-i3 sequences and nuclear SSR data.

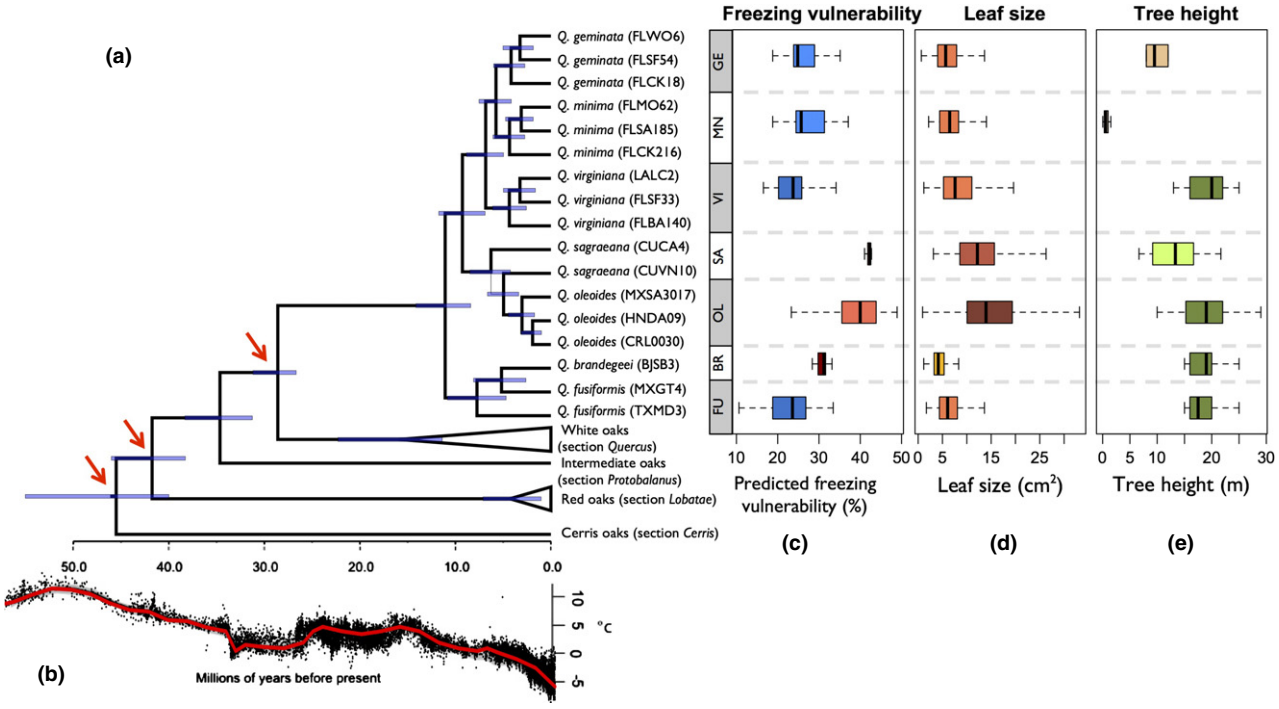

Fig. 5 (a) Time-calibrated phylogeny inferred from RADseq data for 20 *Virentes* individuals and three outgroup taxa in BEAST v.1.7.5 using three priors for node ages based on fossil data. Line widths indicate support for nodes. Red arrows designate calibrated nodes. Blue bars show the 95% highest posterior density values around each age estimate. (b) Reconstructed mean ocean temperature based on oxygen isotope records from deep-sea sediments (Zachos *et al.* 2001) is shown below the time axis, millions of years before the present. Means and variance are shown with box and whisker plots for each species for predicted freezing vulnerability (c), leaf size (d) and tree height (e). Colours are associated with species means. In the left panel, blue hues indicate lower predicted vulnerability to freezing, and orange hues indicate higher vulnerability to freezing and lack of freezing tolerance. In the middle panel, darker hues indicate larger leaves; in the right panel, light colours indicate shorter height.

species, with the exception that *Q. virginiana* was not differentiated from *Q. brandegeei* or *Q. fusiformis*, and *Q. minima* and *Q. geminata* were not differentiated. Within the sympatric species in the southeastern US clade, the three coexisting species were significantly differentiated with respect to leaf morphology and size. However, morphology was highly variable within species and also varied significantly among sites (Appendix S2, Supporting information).

Tree height was also differentiated among the coexisting species (Fig. 5e, rightmost panel). *Quercus minima* is a short, fire-dependent shrub, while *Q. geminata* is a fire-tolerant short-to-intermediate tree and *Q. virginiana* is a fire-intolerant large tree. All of the other species in the *Virentes* have similar tree heights to *Q. virginiana*. Within the *Virentes* species in the United States, the species that do not get tall place their trunks and the majority of their biomass belowground (Kurz & Godfrey 1962). The only ones to do this are in Florida, primarily *Q. minima*, and to a lesser extent *Q. geminata*. The relationship between tree height and fire tolerance strategies was reported in previous work showing that tree height has a negative relationship with rhizome resprouting and belowground investment in biomass that is protected from aboveground fire in oaks in the southeastern US (Myers 1990; Cavender-Bares *et al.* 2004b).

Predicted freezing vulnerability separates species across the tropical–temperate divide. The tropical *Q. oleoides*, *Q. sagraeana* and *Q. brandegeei* have higher freezing vulnerability than their temperate counterparts. *Quercus fusiformis* has the lowest freezing vulnerability of any of the *Virentes* species, and the southeastern US clade also has low freezing vulnerability (Koehler *et al.* 2012). In contrast, *Q. oleoides* has high vulnerability and lacks any freezing tolerance. Data from such a small number of tips are unlikely to support any ancestral character state reconstructions conclusively. However, our data are compatible with freezing tolerance as the ancestral state in the *Virentes*, as additionally supported by the fact that the root of the American oak clade appears to be North American, not Mexican (Pearse & Hipp 2009; Hipp *et al.* 2014). If this inference is correct, then freezing tolerance was

lost twice within the *Virentes*, once in the ancestor of *Q. oleoides* and *Q. sagraeana* and once in *Q. brandegeei* (Fig. 5c).

Discussion

Phylogenetic hypothesis: major clades and sister-species relationships within the Virentes

We present a robust phylogenetic reconstruction of the *Virentes* based on RADseq data. While there have been previous efforts to understand the phylogenetic and biogeographic history of particular *Virentes* species (Nixon 1985; Manos *et al.* 1999; Cavender-Bares *et al.* 2004a, 2011; Cavender-Bares & Pahlich 2009; Pearse & Hipp 2009; Gugger & Cavender-Bares 2013), no study to date has included all seven species in a molecular analysis. Here, we show evidence for two main clades. The first is the Mexican–Texas clade comprising *Quercus fusiformis* and *Quercus brandegeei*, spanning the Atlantic Coast to central Mexico and Texas and including the endemic species of *Q. brandegeei* in southern Baja California. The second clade comprises species of the southeastern US, including the widespread *Quercus virginiana*, *Quercus geminata* and *Quercus minima*, and its sister clade that includes the Cuban oak, *Quercus sagraeana*, and the widespread *Quercus oleoides* that extends from northern Mexico to Costa Rica. There is strong support for a sister relationship between *Q. geminata* and *Q. minima* that form a clade sister to *Q. virginiana*. The crown age of *Virentes* based on fossil calibration is estimated at 11.1 Ma (8.4–14.1 95% HPD), indicating a fairly recent diversification of the clade and providing evidence against a hypothesized ancient origin (MacGinitie 1953) although the stem lineage of *Virentes* is the same age as its sister group (section *Quercus*, the white oaks) dated to be at least 30 Ma. We acknowledge that it is unknown how the inability to model rate heterogeneity among loci in the RADseq data influences branch length estimates.

Nixon (1985) proposed three possible clades within the group based on phenetic characters: (i) *Q. oleoides*, *Q. geminata* and *Q. sagraeana* (which he called *Q. oleoides* var. *sagraeana)*; (ii) *Q. virginiana* and *Q. minima*, with the latter possibly a derivative of a *Q. virginiana* -like ancestor; and (iii) *Q. brandegeei* and *Q. fusiformis*. Muller (1961b), in contrast, interpreted *Q. fusiformis* as a reticulate derivative of *Q. virginiana* and *Q. brandegeei*. Nixon (1985) further suggested that *Q. geminata* was more closely related to *Q. minima* than to *Q. virginiana* given that *Q. geminata* and *Q. minima* share characters of the leaves and pistillate flowers, including 'reflexed styles with a pronounced stigmatic groove' extending from the surface of the stigma that contrast the straight styles of all other live oak species. The placement of the Cuban oak, *Q. sagraeana*, as paraphyletic to *Q. oleoides* is consistent with Nixon's (1985) hypothesis, although in contrast to that view, *Q. oleoides* does not fall out in the same clade as *Q. geminata*.

Phylogeographic patterns, genetic diversity and population structure

The largely monophyletic or paraphyletic relationships of populations within species based on the RADseq data provide evidence for coherence at the species level, despite introgression. Significant but low differentiation (F_{ST}) between species using nuclear SSR and chloroplast sequence data for many individuals, sampled widely across species ranges, tells a similar story of species coherence with porous boundaries. Morphologically, leaf traits reveal high variation within species and high overlap in leaf shape and size across all species. Nevertheless, significant differences were found in leaf traits among species (Appendix S3, Supporting information) indicating that species show morphological cohesion.

Introgression among sympatric species in the southeastern US clade. Earlier flowering time in *Q. virginiana* than in *Q. geminata* has been consistently observed (Sargent 1918; Nixon 1985; Cavender-Bares & Pahlich 2009). However, *Q. minima* and *Q. geminata* have similar flowering times (Nixon 1985; Cavender-Bares *et al.* 2004b) and probably do not have phenological isolating mechanisms. *Quercus minima* and *Q. virginiana* occur in different ecological habitats, whereas *Q. minima* and *Q. geminata* can co-occur, although they differ ecologically in being fire dependent vs fire tolerant (Kurz & Godfrey 1962; Cavender-Bares *et al.* 2004b). Both the ecological overlap and lack of an isolating mechanism may explain why the two species cannot be separated with SSR (Fig. 2) or chloroplast data (Fig. 3).

Introgression between parapatric species. Significant introgression between *Q. fusiformis* and *Q. virginiana* at the range boundary in Texas and at the range boundary between *Q. fusiformis* and *Q. oleoides* in northern Mexico is apparent (Figs 2 and 3). Nixon (1985) noted introgression between the species where they come into close proximity but suggested that climatic and water availability differences associated with elevation might limit gene flow. In Texas, *Q. fusiformis* occurs in more xeric, higher elevation sites than *Q. virginiana*, which is found more in wetter coastal environments; similarly, in Mexico, *Q. fusiformis* occurs on higher elevation piedmont and *Q. oleoides* on lower elevations near the coast. Transpecific chloroplast haplotypes in

the southeast US clade and *Q. sagraeana* clearly indicate gene flow from Florida to Cuba, as detected previously (Gugger & Cavender-Bares 2013). Transpecific patterns of haplotype variation reflect a combination of introgression and ancestral polymorphism. This is a long-standing issue in *Quercus* biology and one that has been addressed multiple times in multispecies studies of oaks (see Muir & Schlotterer 2005). Previous multispecies oak surveys have revealed that haplotype variation is more associated with geography than species delineations (Whittemore & Schaal 1991; Dumolin-Lapègue *et al.* 1997, 1999; Manos *et al.* 1999; Petit *et al.* 2002). Within the *Virentes*, only *Q. brandegeei* seems completely isolated from gene flow, as indicated by the SSR data, although it shares a putatively ancestral chloroplast haplotype that is widespread within the *Virentes*.

Genetic diversity patterns. We found a significant association between range size and genetic diversity, similar to classic patterns theorized and observed for terrestrial plant populations (Stebbins 1942; Hamrick & Godt 1989, 1996; Ellstrand & Elam 1993; Frankham 1997; Gitzendanner & Soltis 2000). A previous study of the population history of the most widespread species, *Q. virginiana* and *Q. oleoides*, indicated that historical differences in climatic stability from the tropics to the temperate zone were likely influential in driving higher genetic differentiation among populations in *Q. virginiana* relative to *Q. oleoides* and higher genetic diversity within populations in the tropics (Cavender-Bares *et al.* 2011). Despite these trends, the range size of *Q. virginiana* is nearly double that of *Q. oleoides*, which may explain its higher genetic diversity.

The low genetic diversity of the endemic species in Baja California and Cuba is of particular concern. *Quercus brandegeei* occurs in a very narrow geographic range only in sites adjacent to ephemeral river beds that we believe fill up after hurricanes, given the high number that reach the Cape region of Baja California (Fig. 2a–c). Hurricane systems develop over the warm waters off the west coast of southern and central Mexico from July through November, but there is very little precipitation in other parts of the year (Turner & Brown 1982). As a consequence, recruitment is likely highly episodic or very limited; we (JCB and AGR) saw no evidence of seedling recruitment or juvenile regeneration other than vegetative root sprouts directly connected to the mother trees (Fig. 2d).

Biogeographic inferences and vicariance

Inception of the Sea of Cortés and formation of the Baja California Peninsula. Separation between the Baja California peninsula and adjacent continental Mexico has been repeatedly implicated in animal phylogeographic studies as a prominent vicariance event critical to biotic diversification in the region (e.g. Case & Cody 1983; Riddle *et al.* 2000; Grismer 2002; Crews & Hedin 2006; Douglas *et al.* 2006; Riddle & Hafner 2006; Ross & Markow 2006; Pfeiler *et al.* 2007). Between 8 and 13 million years ago, most of Baja California was submerged beneath the Pacific Ocean against the northwest coast of mainland Mexico. By 6 Ma, a shallow epicontinental seaway had formed (Grismer 2002), and by about 5.5 Ma, the Baja California peninsula began to separate from the Mexican mainland as a result of plate-boundary expansion between the North American and Pacific plates, leading to inception of the Sea of Cortés and permanent separation of the peninsula by 5 Ma. At the same time, peninsula ranges were uplifted, causing rain shadows and severe localized drying trends in Baja California. The peninsula was also likely fragmented by transpeninsular seaways connecting the Pacific Ocean and Sea of Cortés. Marine transgressions and seaways have provided a basis for explaining late Neogene (5.5–1 Ma) biogeographic disjunctions in vertebrates (Grismer 2000; Riddle *et al.* 2000).

Overall drying trends in North America accompanied global cooling trends (Fig. 5b) beginning during the Eocene and continued through the Neogene. Intermittent glacial periods of the Pleistocene and rain shadows caused by the uplift of the Peninsular Ranges brought severe localized drying trends to Baja California. Miocene climates were probably favourable to the dispersal of *Virentes* populations across the continent; however, repeated drying and cooling trends may have caused intervening populations to disappear, leaving the Baja California population isolated from the rest of the *Virentes*. Muller (1967) provided fossil evidence that the tree flora of Baja California likely consists of relictual populations that were once more broadly distributed; he cited putative fossil equivalents of *Q. brandegeei* in Miocene deposits at Tehachapi in southern California, USA (as *Quercus mohavensis* Axelrod), 1500 km northward of the current range limit, as evidence for the range retraction hypothesis. Vicariance caused by the drying of the interior of North America coupled with the formation of the Sea of Cortés would explain our estimated divergence time between *Q. brandegeei* and *Q. fusiformis* of 5.17 Ma with a 95% probability density interval of 2.6–8.1 Ma. The presence of a previously widespread ancestor of *Q. fusiformis* and *Q. brandegeei* that experienced severe range retraction due to climate change is supported by the IMa results indicating a much larger ancestral effective population size than current effective population size estimates of both species combined (Fig. 4e and f).

Changes in environmental conditions that accompanied Pleistocene and Holocene climatic fluctuations are known to have influenced other plant species distributions (Grismer 2000, 2002), and range shifts of Sonoran Desert floral communities have been inferred from plant macrofossils in packrat middens (Vandevender *et al.* 1994; Holmgren *et al.* 2011). In the plant genus *Guaiacum* (Zygophyllaceae), vicariance from a common ancestor was demonstrated using nuclear SSRs, likely due to range retraction as a consequence of climatic drying. The Baja California shrub *Guaiacum unijugum* shows similar ecological and reproductive patterns to *Q. brandegeei* occurring only along occasional waterways and showing very limited seedling recruitment (McCauley *et al.* 2010). Other distribution shifts consistent with climatic drying include the northward expansion of columnar cacti along the Baja California peninsula (Nason *et al.* 2002; Clark-Tapia & Molina-Freaner 2003) and the expansion of the desert plant *Euphorbia lomelii* along the north–south axis of the peninsula (Garrick *et al.* 2009).

Nicaraguan depression. Divergence of the Costa Rican population of *Q. oleoides* from the Central American widespread population, estimated at 1.9 Ma with a 95% probability density interval of 0.99 to 3.11 Ma, appears to be older than previously estimated (Cavender-Bares *et al.* 2011). This timing implicates the formation of the Nicaraguan Depression and associated volcanic activity as the cause of vicariance, rather than the rise of the Guanacaste cordillera associated with the decline of wet tropical forest and spread of dry tropical forest species. Three main tectonic phases affected the Nicaraguan Depression including Miocene convergence, Pliocene extension, and Pleistocene-to-present transtensional deformation (Funk *et al.* 2009). Lake Nicaragua may have divided an already dispersed population or prevented dispersion, with one or more long-distance dispersal events giving rise to the current Costa Rican population whose single chloroplast haplotype reveals long-term isolation.

Long-distance dispersal and origins of the Cuban oak. Phylogenetic results indicate a Central American origin of *Q. sagraeana*. The origin of the Cuban oak has long been of interest to biogeographers (Muller 1955, 1961a; Lopez-Almirall 1979; Nixon 1985; Gugger & Cavender-Bares 2013). It was originally described as a distinct species by Nuttall (1842) based on a specimen collected by R. de la Sagra. Later taxonomists described it as a subspecies of *Q. oleoides* and suggested a hybrid origin of the Cuban oak from *Q. oleoides* in the Yucatán region of Mexico and *Q. geminata* (Muller 1955, 1961a; Nixon 1985) or *Q. virginiana* (Lopez-Almirall 1979) from Florida based on morphological evidence. Vicariance caused by the separation of Cuba from Central America can be ruled out (Muller 1955; Nixon 1985; Gugger & Cavender-Bares 2013), given Cuba has been isolated since the early Cretaceous >35 Ma (Pindell & Dewey 1982; Iturralde-Vinenta 2006), making long-distance dispersal the only plausible scenario. Nixon (1985) hypothesized that long-distance dispersal by passenger pigeons (*Ectopistes migratorius* L.) could have transported propagules from Florida and Yucatán during periods of low sea level. Gugger & Cavender-Bares (2013) found molecular support for a Florida origin from *Q. virginiana* during the Pleistocene based on nuclear SSR and chloroplast data, despite introgression with *Q. oleoides* in Central America. While the single chloroplast haplotype in *Q. sagraeana* is shared by the Florida species (Fig. 3), the RADseq data contradict the earlier interpretation and indicate a Central American origin (Figs 2 and 5). Introgression with the southeastern US clade could distort relationships, however (see Eaton *et al.* 2015).

One possible scenario is that ancestral *Virentes* was structured phylogenetically into western and eastern groups spread across North America. At a time of low sea level, the eastern population was distributed across what is now the southeast US, Cuba, Yucatán and other parts of Central America; with sea level rise, this population became disjunct with isolated populations in Central America, Cuba and the southeast US. Note that Yucatán does not currently support *Virentes*, nor any other oak species, perhaps due to saline edaphic conditions, but it may have in the past. Thus *Q. oleoides* and *Q. sagraeana* were likely formed from a widespread ancestral population that also gave rise to the southeastern US clade.

Evolution of life history traits and morphology

A readily apparent pattern in the ecological data reveals that the three sympatric species that form the southeastern US clade show much greater ecological niche differentiation than the allopatric and parapatric species. This pattern is consistent with some degree of adaptive radiation involving interspecific competition or the development of ecologically based reproductive isolating mechanisms. Niche differences within this clade are apparent given their sharply contrasting growth forms (shrub, short tree, tall tree, Fig. 5), reflecting differentiation in fire responses, that is, fire dependence (*Q. minima*), fire tolerance (*Q. geminata*) and fire intolerance (*Q. virginiana*) (Kurz & Godfrey 1962). Only *Q. minima* forms a rhizomatous shrub in large clonal patches rarely taller than 0.5 m (Fig. 5), although other *Virentes* can produce rhizomatous stems (see Fig. 4). The rhizomatous shrub habit has been considered an adaptation to fire (Myers 1990). These niche differences are paral-

leled by differences in soil moisture and nutrient preferences (Cavender-Bares *et al.* 2004a,b) and offset flowering phenology (between *Q. geminata* and *Q. virginiana*) (Sargent 1918; Nixon 1985; Cavender-Bares & Pahlich 2009). Ecological niche differentiation among the sympatric species does not necessarily implicate sympatric speciation, although this remains a distinct possibility, particularly given flowering time differences that provide a barrier to gene flow. Nevertheless, repeated changes in sea level throughout the last 8 Ma, which led to periodic formation of barrier islands where the Florida Peninsula currently exists, would have provided ample opportunities for isolating barriers allowing allopatric speciation. In contrast, the allopatric and parapatric species whose ranges overlap only at the range margins show differentiation in predicted freezing tolerance and adaptation to climate (Koehler *et al.* 2012) but not in growth form (Fig. 5). Allopatric speciation is implicated in the split between *Q. fusiformis* and *Q. brandegeei*, between the southeastern US clade and the Central America + Cuba clade, and between *Q. oleoides* and *Q. sagraeana* within that clade.

Conclusions

The biogeographic history of *Virentes* has been shaped primarily by geologic and climatic events, including the formation of the Sea of Cortés, and an increasingly drier climate in coastal and inland areas of North and Mesoamerica. Range retraction, population migration and long-distance dispersal have been important processes in the diversification of the *Virentes*. Close examination of the evolutionary history of a small but widespread clade allowed us to gain insight into the contrasting factors that drive shifts along ecological vs climatic niche axes (cf. Emery *et al.* 2012; Ackerly *et al.* 2006). In this system, we observed that sympatric species have evolved traits that allow habitat differentiation as would be expected in adaptive radiations to avoid resource competition (Schluter 2000), limit gene flow (Levin 2006) and reduce density-dependent mortality due to phylogenetically conserved pests and pathogens (Webb *et al.* 2006; Gilbert & Webb 2007), while allopatric and parapatric species diverge in their climatic niches but maintain very similar ecological niches. Thus, while allopatric and parapatric species evolve niches to adapt to local climates, which may subsequently limit gene flow between them, sympatric species show divergence in traits that allow ecological niche partitioning within a given climatic region. These patterns suggest that the nature of speciation and degree of coexistence are critical in determining which niche axis shows most divergence. If allopatric speciation allows for climatic niche evolution, sympatry fosters ecological differentiation. Future investigations can test whether these patterns hold for oaks, generally, and in other biological systems.

Acknowledgements

We thank Hernando Rodríguez-Correa, John McVay, Matthew Kaproth, Elizabeth Fallon, Kari Koehler, Ben Lowe, Nicholas Deacon, Will Pearse, José Meireles, Shan Kothari and Anna Schweiger for comments and/or technical support. We thank the National Science Foundation IOS: 0843665 and DEB: 1146380 for funding and the Minnesota Supercomputing Institute.

References

Ackerly DD, Schwilk DW, Webb CO (2006) Niche evolution and adaptive radiation: testing the order of trait divergence. *Ecology*, **87**, S50–S61.

Aldrich PR, Cavender-Bares J (2011) Genomics and breeding of oaks and their slightly less-domesticated wild oak relatives. In: *Wealth of Wild Species: Genetic, Genomic and Breeding Resources* (ed. Kole C), pp. 89–130. Springer-Verlag, New York city, New York.

Arrivillaga J, Norris D, Feliciangeli M, Lanzaro G (2002) Phylogeography of the neotropical sand fly *Lutzomyia longipalpis* inferred from mitochondrial DNA sequences. *Infection, Genetics and Evolution*, **2**, 83–95.

Bacilieri R, Ducousso A, Kremer A (1995) Genetic, morphological, ecological and phenological differentiation between *Quercus petraea* (Matt) Liebl and *Quercus robur* L in a Mixed Stand of Northwest of France. *Silvae Genetica*, **44**, 1–10.

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

Borgardt SJ, Pigg KB (1999) Anatomical and developmental study of petrified *Quercus* (Fagaceae) fruits from the middle Miocene, Yakima Canyon, Washington, USA. *American Journal of Botany*, **86**, 307–325.

Boucher DH (1983) *Quercus oleoides* (Roble Encino, Oak). In: *Costa Rican Natural History* (ed. Janzen DH), pp. 319–322. The University of Chicago Press, Chicago, Illinois.

Camus A (1936-1938) *Monographie du Genre Quercus. Vol. 1. Sous-Genre Cyclobalanopsis. Sous-Genre Euquercus, Sections Cerris et Mesobalanus*. Lechevalier, Paris.

Candolle AD (1862) Note sur un nouveau caracter observé dans le fruit des chênes. *Archives des Sciences Physiques et Naturelles, Genéve*, **15**, 89–101.

Case TJ, Cody ML (1983) *Island Biogeography in the Sea of Cortez*. University of California Press, Berkeley.

Cavender-Bares J (2007) Chilling and freezing stress in live oaks (*Quercus* section *Virentes*): Intra- and interspecific variation in PS II sensitivity corresponds to latitude of origin. *Photosynthesis Research*, **94**, 437–453.

Cavender-Bares J, Holbrook NM (2001) Hydraulic properties and freezing-induced cavitation in sympatric evergreen and deciduous oaks with contrasting habitats. *Plant Cell and Environment*, **24**, 1243–1256.

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 DD, Baum DA, Bazzaz FA (2004a) Phylogenetic overdispersion in Floridian oak communities. *American Naturalist*, **163**, 823–843.

Cavender-Bares J, Kitajima K, Bazzaz FA (2004b) Multiple trait associations in relation to habitat differentiation among 17 Floridian oak species. *Ecological Monographs*, **74**, 635–662.

Cavender-Bares J, Gonzalez-Rodriguez 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.

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

Clark-Tapia R, Molina-Freaner F (2003) The genetic structure of a columnar cactus with a disjunct distribution: *Stenocereus gummosus* in the Sonoran desert. *Heredity*, **90**, 443–450.

Coker WC (1912) The seedlings of the live oak and white oak. *Journal of the Elisha Mitchell Scientific Society*, **28**, 34–41.

Craft KJ, Ashley MV, Koenig WD (2002) Limited hybridization between *Quercus lobata* and *Quercus douglasii* (Fagaceae) in a mixed stand in central coastal California. *American Journal of Botany*, **89**, 1792–1798.

Crews SC, Hedin M (2006) Studies of morphological and molecular phylogenetic divergence in spiders (Araneae: Homalonychus) from the American southwest, including divergence along the Baja California Peninsula. *Molecular Phylogenetics and Evolution*, **38**, 470–487.

Curtu AL, Gailing O, Finkeldey R (2007a) Evidence for hybridization and introgression within a species-rich oak (*Quercus* spp.) community. *BMC Evolutionary Biology*, **7**, 218.

Curtu AL, Gailing O, Leinemann L, Finkeldey R (2007b) Genetic variation and differentiation within a natural community of five oak species (*Quercus spp.*). *Plant Biology*, **9**, 116–126.

Daghlian CP, Crepet WL (1983) Oak catkins, leaves and fruits from the Oligocene Catahoula formation and their evolutionary significance. *American Journal of Botany*, **70**, 639–649.

Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.

Denk T, Grimm GW (2009) Significance of pollen characteristics for infrageneric classification and phylogeny in Quercus (Fagaceae). *International Journal of Plant Sciences*, **170**, 926–940.

Denk T, Grímsson F, Zetter R (2012) Fagaceae from the early Oligocene of Central Europe: persisting new world and emerging old world biogeographic links. *Review of Palaeobotany and Palynology*, **169**, 7–20.

de Dios RS, Benito-Garzon M, Sainz-Ollero H (2006) Hybrid zones between two european oaks: a plant community approach. *Plant Ecology*, **187**, 109–125.

Dodd RS, Kashani N (2003) Molecular differentiation and diversity among the California red oaks (Fagaceae; *Quercus* section Lobatae). *TAG Theoretical and Applied Genetics*, **107**, 884–892.

Douglas ME, Douglas MR, Schuett GW, Porras LW (2006) Evolution of rattlesnakes (Viperidae; Crotalus) in the warm deserts of western North America shaped by Neogene vicariance and Quaternary climate change. *Molecular Ecology*, **15**, 3353–3374.

Dumolin-Lapègue S, Demesure B, Fineschi S, Corre VL, Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. *Genetics*, **146**, 1475–1487.

Dumolin-Lapègue S, Kremer A, Petit RJ (1999) Are chloroplast and mitochondrial DNA variation species independent in oaks? *Evolution*, **53**, 1406–1413.

Eaton DAR (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.

Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using genomic RADseq data: An example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology*, **62**, 689–706.

Eaton DAR, Gonzalez-Rodriguez A, Hipp A, Cavender-Bares J (2015) Introgression obscures and reveals historical relationships among the American live oaks. *bioRxiv*, 016238.

Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size - implications for plant conservation. *Annual Review of Ecology and Systematics*, **24**, 217–242.

Elsik WC, Yancey TE (2000) Palynomorph biozones in the context of changing paleoclimate, middle Eocene to lower Oligocene of the northwest Gulf of Mexico. *Palynology*, **24**, 177–186.

Emery NC, Forrestel EJ, Jui G, Park MS, Baldwin BG, Ackerly DD (2012) Niche evolution across spatial scales: climate and habitat specialization in California Lasthenia (Asteraceae). *Ecology*, **93**, S151–S166.

Engelmann G (1876-1877) About the oaks of the United States. *Transaction of the Academy of Science St. Louis*, **3**, 372–400.

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

Frankham R (1997) Do island populations have less genetic variation than mainland populations? *Heredity*, **78**, 311–327.

Funk J, Mann P, McIntosh K, Stephens J (2009) Cenozoic tectonics of the Nicaraguan depression, Nicaragua, and Median trough, El Salvador, based on seismic-reflection profiling and remote-sensing data. *Geological Society of America Bulletin*, **121**, 1491–1521.

Garrick R, Nason J, Meadows C, Dyer R (2009) Not just vicariance: phylogeography of a Sonoran Desert euphorb indicates a major role of range expansion along the Baja peninsula. *Molecular Ecology*, **18**, 1916–1931.

Gilbert GS, Webb CO (2007) Phylogenetic signal in plant pathogen-host range. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 4979–4983.

Gitzendanner MA, Soltis PS (2000) Patterns of genetic variation in rare and widespread plant congeners. *American Journal of Botany*, **87**, 783–792.

González-Rodríguez A, Arias DM, Valencia S, Oyama K (2004a) Morphological and RAPD analysis of hybridization between *Quercus affinis* and *Q. laurina* (Fagaceae), two Mexican red oaks. *American Journal of Botany*, **91**, 401–409.

González-Rodríguez A, Bain JF, Golden JL, Oyama K (2004b) Chloroplast DNA variation in the *Quercus affinis-Q. laurina* complex in Mexico: geographical structure and associations with nuclear and morphological variation. *Molecular Ecology*, **13**, 3467–3476.

Grismer LL (2000) Evolutionary biogeography on Mexico's Baja California peninsula: a synthesis of molecules and historical

geology. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 14017–14018.

Grismer LL (2002) *Amphibians and Reptiles of Baja California Including its Pacific Islands and the Islands in the Sea of Cortés*. University of California Press, Berkeley, California.

Gugger PF, Cavender-Bares J (2013) Molecular and morphological support for a Florida origin of the Cuban oak. *Journal of Biogeography*, **40**, 632–645.

Gutiérrez-García TA, Vázquez-Domínguez E (2013) Consensus between genes and stones in the biogeographic and evolutionary history of Central America. *Quaternary Research*, **79**, 311–324.

Hamrick J, Godt M (1989) Allozyme diversity in plant species. In: *Plant Population Genetics, Breeding, and Genetic Resources* (eds Brown AHDCM, Kahler AL, Weir BS), pp. 43–63. Sinauer Associates, Sunderland, Massachusetts.

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, **351**, 1291–1298.

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.

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

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hipp AL, Weber JA (2008) Taxonomy of Hill's Oak (*Quercus ellipsoidalis*: Fagaceae): evidence from AFLP Data. *Systematic Botany*, **33**, 148–158.

Hipp AL, Eaton DAR, Cavender-Bares J, Fitzek E, Nipper R, Manos PS (2014) A framework phylogeny of the American Oak Clade based on sequenced RAD Data. *PLoS ONE*, **9**, e102272.

Holmgren CA, Betancourt JL, Rylander KA (2011) Vegetation history along the eastern, desert escarpment of the Sierra San Pedro Mártir, Baja California, Mexico. *Quaternary Research*, **75**, 647–657.

Howard DJ, Preszler RW, Williams J, Fenchel S, Boecklen WJ (1997) How discrete are oak species? Insights from a hybrid zone between *Quercus grisea* and *Quercus gambelii*. *Evolution*, **51**, 747–755.

Howarth DG, Baum DA (2002) Phylogenetic utility of a nuclear intron from nitrate reductase for the study of closely related plant species. *Molecular Phylogenetics and Evolution*, **23**, 525–528.

Hubert F, Grimm GW, Jousselin E, Berry V, Franc A, Kremer A (2014) Multiple nuclear genes stabilize the phylogenetic backbone of the genus *Quercus*. *Systematics and Biodiversity*, **12**, 405–423.

Iturralde-Vinenta MA (2006) Meso-Cenozoic Caribbean Paleogeography: implications for the Historical Biogeography of the Region. *International Geology Review*, **48**, 791–827.

Iwata H, Ukai Y (2002) SHAPE: a computer program package for quantitative evaluation of biological shapes based on elliptic fourier descriptors. *Journal of Heredity*, **93**, 384–385.

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

Koehler K, Center A, Cavender-Bares J (2012) Evidence for a freezing tolerance - growth rate trade-off in the live oaks (*Quercus* series *Virentes*) across the tropical-temperate divide. *New Phytologist*, **193**, 730–744.

Kurz H, Godfrey RK (1962) *Trees of Northern Florida*. University of Florida, Gainesville.

Levin DA (2006) The spatial sorting of ecological species: ghost of competition or of hybridization past? *Systematic Botany*, **31**, 8–12.

Lewis IM (1911) The seedling of *Quercus virginiana*. *Plant World*, **14**, 119–123.

Lopez-Almirall A (1979) Variabilidad de las hojas de *Quercus oleoides sagraena* (Fagales: Fagaceae). *Ciencias Biologicias*, **3**, 59–64.

MacGinitie HD (1953) Fossil plants of the Florissant Beds, Colorado. *Carnegie Institute of Washington Publication*, **599**, 1–98.

Manos PS, Doyle JJ, Nixon KC (1999) Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). *Molecular Phylogenetics and Evolution*, **12**, 333–349.

McCauley RA, Cortés-Palomec AC, Oyama K (2010) Distribution, genetic structure, and conservation status of the rare microendemic species, *Guaiacum unijugum* (Zygophyllaceae) in the Cape Region of Baja California, Mexico. *Revista Mexicana de Biodiversidad*, **81**, 745–758.

Muir G, Schlotterer C (2005) Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). *Molecular Ecology*, **14**, 549–561.

Muller CH (1955) The origin of *Quercus* on Cuba. *Revista de la Sociedad Cuban de Botanica*, **7**, 41.

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

Muller CH (1961b) The origin of *Quercus fusiformis* Small. *Journal of the Linnean Society*, **58**, 186–192.

Muller CH (1967) Relictual origins of insular endemics in *Quercus*. In: *Proceedings of the Symposium on the Biology of the Channel Islands* (ed. Philbrick RN), pp. 73–77. Santa Barbara Botanic Garden, Santa Barbara, California.

Myers RL (1990) Scrub and High Pine. In: *Ecosystems of Florida* (eds Myers RL, Ewel JJ), pp. 150–193. University of Central Florida Press, Orlando.

Nason JD, Hamrick J, Fleming TH (2002) Historical vicariance and postglacial colonization effects on the evolution of genetic structure in *Lophocereus*, a Sonoran Desert columnar cactus. *Evolution*, **56**, 2214–2226.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

Nielsen R, Tarpy DR, Reeve HK (2003) Estimating the effective paternity numbdetecter in social insects and the effective number of alleles in a population. *Molecular Ecology*, **12**, 3157–3164.

Nixon KC (1985) *A Biosystematic Study of Quercus Series Virentes (the live oaks) with Phylogenetic Analyses of Fagales, Fagaceae and Quercus*, Ph.D. Thesis. University of Texas, Austin.

Nixon KC, Muller CH (1997) *Quercus* Linnaeus sect. *Quercus* White oaks. In: *Flora of North America North of Mexico* (eds Committee FoNAE), pp. 436–506. Oxford University Press, New York City, New York.

Nuttall T (1842) *North American Sylva*. vol. **1**, pp. 1–56. J. Dobson, Printed by William Amphlett, New Harmony, Indiana.

Paradis E, Schliep K, Potts A, Winter D (2010) PEGAS: an R package for population genetics with an integrated-modular approach. *Bioinformatics*, **26**, 419–420.

Pearse IS, Hipp AL (2009) Phylogenetic and trait similarity to a native species predict herbivory on non-native oaks. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 18097–18102.

Petit RJ, Csaikl UM, 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.

Pfeiler E, Erez T, Hurtado LA, Markow TA (2007) Genetic differentiation and demographic history in *Drosophila pachea* from the Sonoran Desert. *Hereditas*, **144**, 63–74.

Pindell J, Dewey JF (1982) Permo-Triassic reconstruction of western Pangea and the evolution of the Gulf of Mexico/Caribbean region. *Tectonics*, **1**, 179–211.

Poelchau MF, Hamrick JL (2013) Comparative phylogeography of three common Neotropical tree species. *Journal of Biogeography*, **40**, 618–631.

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

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

Riddle B, Hafner D (2006) A step-wise approach to integrating phylogeographic and phylogenetic biogeographic perspectives on the history of a core North American warm deserts biota. *Journal of Arid Environments*, **66**, 435–461.

Riddle BR, Hafner DJ, Alexander LF, Jaeger JR (2000) Cryptic vicariance in the historical assembly of a Baja California Peninsular Desert biota. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 14438–14443.

Ripley B, Venables B, Bates DM, Hornik K, Gebhardt A, Firth D (1998) MASS R Package 7.3-34, Support functions and datasets for Venables and Ripley's MASS. CRAN.

Rodríguez-Correa H, Oyama K, MacGregor-Fors I, González-Rodríguez A (2015) How are oaks distributed in the Neotropics? A perspective from species turnover, areas of endemism, and climatic niches. *International Journal of Plant Sciences*, **176**, 222–231.

Ross CL, Markow TA (2006) Microsatellite variation among diverging populations of *Drosophila mojavensis*. *Journal of Evolutionary Biology*, **19**, 1691–1700.

Sargent CS (1918) Notes on North American trees. 1. Quercus. *Botanical Gazette*, **65**, 423–459.

Schluter D (2000) Introduction to the symposium: species interactions and adaptive radiation. *The American Naturalist*, **156** (Supplement), S1–S3.

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

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

Stebbins G (1942) The genetic approach to problems of rare and endemic species. *Madroño*, **6**, 241–258.

Steinkeller CL, Turetschek E, Gloessl J (1997) Conservation of (GA)n microsatellite loci between Quercus species. *Molecular Ecology*, **6**, 1189–1194.

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

Turner RM, Brown DE (1982) Tropical-subtropical desertlands. In: *Biotic Communities of the American Southwest–United States and Mexico* (ed. Brown DE), pp. 180–222. University of Utah Press, Salt Lake City, Utah.

Valbuena-Carabana M, Gonzalez-Martinez SC, Sork VL *et al.* (2005) Gene flow and hybridisation in a mixed oak forest (*Quercus pyrenaica* Willd. and *Quercus petraea* (Matts.) Liebl.) in central Spain. *Heredity*, **95**, 457–465.

Vandevender TR, Burgess TL, Piper JC, Turner RM (1994) Paleoclimatic implications of Holocene plant remains from the Sierra-Bach, Sonora, Mexico. *Quaternary Research*, **41**, 99–108.

Viscosi V, Lepais O, Gerber S, Fortini P (2009) Leaf morphological analyses in four European oak species (*Quercus*) and their hybrids: a comparison of traditional and geometric morphometric methods. *Plant Biosystems*, **143**, 564–574.

Webb CO, Gilbert GS, Donoghue MJ (2006) Phylodiversity-dependent seedling mortality, size structure, and disease in a Bornean rain forest. *Ecology*, **87**, S123–S131.

Weir BS (1996) *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*, 2nd edn. Sinauer Associates, Sunderland, Massachusetts.

Whittemore AT, Schaal BA (1991) Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences of the United States of America*, **88**, 2540–2544.

Wright IJ, Reich PB, Westoby M, et al. (2004) The worldwide leaf economics spectrum. *Nature*, **428**, 821–827.

Zachos J, Pagani M, Sloan L, Thomas E, Billups K (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, **292**, 686–693.

All authors contributed intellectually. J.C.B. conceived of and designed the study; collected the specimens; managed the laboratory work and analyses; ran BEAST, Structure and HaploNet analyses; collected the trait data; and wrote the manuscript. A.G.R. was involved in all sample collection in Mexico, ran genetic diversity analyses, contributed to leaf scanning and height measurements and writing of the manuscript. D.E. conducted RADseq pipeline, ran RAXML analyses and contributed to writing of the manuscript. A.H. contributed to RADseq pipeline, various analyses and writing of the manuscript. A.B. contributed to morphological analyses, data management and IMa analyses. P.M. contributed to fossil dating in BEAST analysis and writing of the manuscript.

Data accessibility

DNA sequences for chloroplast and NIA-i3: GenBank accession numbers: BankIt1824105: KR923000–KR923022 (cp *trn*D–*trn*T); BankIt1824526: KR923023–KR923114

(cp *rpl*32–*trn*L^{UAG}); BankIt1824762: KR923115–KR923185 (NIA-i3).

The RADseq data are available on the NCBI SRA in demultiplexed form. All new data generated for this study are listed under project number PRJNA277574.

phy files for RADseq-concatenated alignments used for RAXML phylogenetic inferences: Dryad Digital Repository. doi:10.5061/dryad.855 pg

xml files with data assumptions, fossil-based priors and RADseq matrices used in BEAST analyses: Dryad Digital Repository. doi:10.5061/dryad.855 pg

Nuclear SSRs for STRUCTURE analysis including sampling locations and geographic coordinates; concatenated and separate chloroplast sequences for haplotype network analysis; and IMa input files: Dryad Digital Repository. doi:10.5061/dryad.855 pg

Freezing vulnerability, leaf size and shape and tree height data with sampling locations: Dryad Digital Repository. doi:10.5061/dryad.855 pg

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 List of species, collection locations and geographic coordinates of 27 live oak individuals used for RAD sequencing.

Table S2 RADseq data showing species, collection location, identification code, raw and filtered sequence lengths, average depth, conserved loci, estimated H, and H.

Appendix S1 Additional details on data collection, data processing, and analysis.

Appendix S2 STRUCTURE results.

Appendix S3 Leaf morphological analyses.