

# Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity

Andrew L. Hipp<sup>1,2</sup>, Paul S. Manos<sup>3</sup>, Antonio González-Rodríguez<sup>4</sup>, Marlene Hahn<sup>1</sup>, Matthew Kaproth<sup>5,6</sup>, John D. McVay<sup>3</sup>, Susana Valencia Avalos<sup>7</sup> and Jeannine Cavender-Bares<sup>5</sup>

<sup>1</sup>The Morton Arboretum, 4100 Illinois Route 53, Lisle, IL 60532, USA; <sup>2</sup>The Field Museum, 1400 S Lake Shore Drive, Chicago, IL 60605, USA; <sup>3</sup>Duke University, Durham, NC 27708, USA; <sup>4</sup>Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Nacional Autónoma de México, Antigua Carretera a Pátzcuaro No. 8701, Col. Ex Hacienda de San José de la Huerta, Morelia, Michoacán 58190, México; <sup>5</sup>Department of Ecology, Evolution and Behavior, University of Minnesota, Saint Paul, MN 55108, USA; <sup>6</sup>Department of Biological Sciences, Minnesota State University, Mankato, MN 56001, USA; <sup>7</sup>Herbario de la Facultad de Ciencias, Departamento de Biología Comparada, Universidad Nacional Autónoma de México, Circuito Exterior, s.n., Ciudad Universitaria, Coyoacán, CP 04510, México City, México

## Summary

Authors for correspondence:

Andrew L. Hipp

Tel: +1 630 725 2094

Email: ahipp@mortonarb.org

Paul S. Manos

Tel: +1 919 660 7358

Email: pmanos@duke.edu

Jeannine Cavender-Bares

Tel: +1 612 624 6337

Email: cavender@umn.edu

Received: 2 June 2017

Accepted: 29 July 2017

New Phytologist (2017)

doi: 10.1111/nph.14773

**Key words:** American oaks, convergence, niche evolution, phylogeny, *Quercus*, restriction site-associated DNA sequencing (RAD-seq), sympatric parallel diversification, woody plants.

- Oaks (*Quercus*, Fagaceae) are the dominant tree genus of North America in species number and biomass, and Mexico is a global center of oak diversity. Understanding the origins of oak diversity is key to understanding biodiversity of northern temperate forests.
- A phylogenetic study of biogeography, niche evolution and diversification patterns in *Quercus* was performed using 300 samples, 146 species. Next-generation sequencing data were generated using the restriction-site associated DNA (RAD-seq) method. A time-calibrated maximum likelihood phylogeny was inferred and analyzed with bioclimatic, soils, and leaf habit data to reconstruct the biogeographic and evolutionary history of the American oaks.
- Our highly resolved phylogeny demonstrates sympatric parallel diversification in climatic niche, leaf habit, and diversification rates. The two major American oak clades arose in what is now the boreal zone and radiated, in parallel, from eastern North America into Mexico and Central America.
- Oaks adapted rapidly to niche transitions. The Mexican oaks are particularly numerous, not because Mexico is a center of origin, but because of high rates of lineage diversification associated with high rates of evolution along moisture gradients and between the evergreen and deciduous leaf habits. Sympatric parallel diversification in the oaks has shaped the diversity of North American forests.

## Introduction

Oaks (*Quercus* L., Fagaceae) constitute the dominant woody plant genus of North America, including Mexico, in biomass and species richness based on forest inventory data (Banda *et al.*, 2016; Cavender-Bares, 2016). Encompassing 65% of the estimated 435 species in the genus *Quercus* (Nixon, 1997; Manos *et al.*, 1999; Valencia, 2004; de Beaulieu & Lamant, 2010), the predominantly American oak clade, comprising the white oaks, red oaks, and intermediate oaks (Manos *et al.*, 1999), ranges from wetlands of the upper Midwestern US to droughty mountains of the arid southwest, and from southern Canada to the northern Andes of Colombia (Nixon, 1993, 2006). Yet despite the ecological and economic importance of oaks, key questions about oak diversification are unanswered. Why are there so many red oaks (*Quercus* sect. *Lobatae*) and white oaks (*Quercus* sect. *Quercus*) in the Americas and in Mexico in particular? What is the biogeographic and

ecological oak diversification? What ecological factors shape diversification in the American oaks? Transitions in climate, ring porosity (in the wood), and animal dispersal of seeds have been suggested to drive an increase in diversification at or near the base of the 'quercoids' (*Quercus*, *Lithocarpus*, *Castanopsis*), modulated by the acquisition of deciduous leaves (Bouchenak-Khelladi *et al.*, 2015; Larson-Johnson, 2016). Wind pollination may have played a role in increased diversification of *Quercus* relative to close relatives (Manos & Stanford, 2001). Little is known, however, about drivers of diversification within *Quercus*.

Even fine-scale relationships within the genus are poorly understood. While a phylogenetic framework for the genus was worked out two decades ago (Manos *et al.*, 1999), subsequent work has mostly been taxonomically narrow (Cavender-Bares *et al.*, 2015; Eaton *et al.*, 2015; Owusu *et al.*, 2015; Fitz-Gibbon *et al.*, 2017; Leroy *et al.*, 2017; McVay *et al.*, 2017b), sparse (Pearse & Hipp, 2009; Hipp *et al.*, 2014; Pham *et al.*, 2017), or

insufficient in locus-sampling to provide robust phylogenetic estimates (Oh & Manos, 2008; Denk & Grimm, 2010; Hubert *et al.*, 2014). A recent study of the white oaks and closely related *Virentes* and *Ponticae* clades (McVay *et al.*, 2017b) addressed the taxon- and locus-sampling issues, but it sampled only one major clade. These papers all provide important contributions to our understanding of oak diversification, but none provides the species-level phylogeny needed to understand the diversification of this ecologically and economically important genus.

In this study, we address two interrelated sets of questions. First, did the American oaks originate in Mexico, where they are most diverse, and disperse northward? Conversely, if the Mexican lineages are derived, what explains the high diversity of Mexican oaks? Second, what role have niche evolution and the evolution of the deciduous and evergreen leaf habit played in the diversification of the American oaks? We utilize a next-generation sequencing method, restriction site-associated DNA sequencing (RAD-seq) (Baird *et al.*, 2008; Ree & Hipp, 2015) to investigate the phylogeny of the American oak clade based on 293 ingroup samples representing 146 species, including nearly complete sampling of US species and the most thorough sampling to date of the Mexican and Central American species. We combine these with fossil data to calibrate divergence times, biogeographic data to investigate the timing and frequency of transitions among geographic regions, climatic and soils data to investigate rates of niche evolution, and leaf habit data to investigate the evolution of leaf traits. Our analyses of these data provide unprecedented phylogenetic resolution of the American oak clade and a clear picture of how sympatric parallel diversification has shaped oak diversity in the Americas.

## Materials and Methods

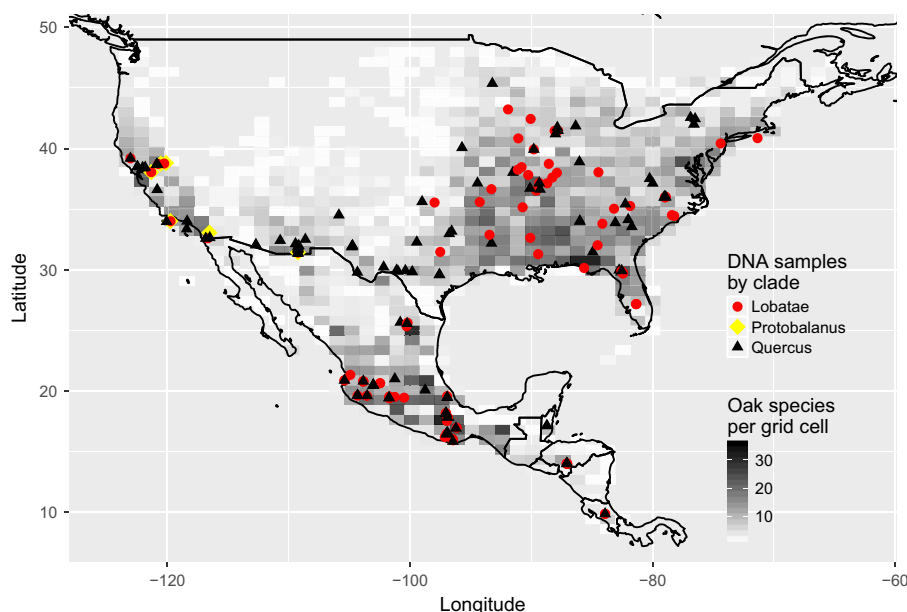
### Sampling

Tissue and herbarium specimens for almost all ingroup samples were collected from wild populations or from botanical garden

specimens that were grown from wild-collected acorns; source populations cover the vast majority of the range of oaks in North America and Mexico (Fig. 1; Supporting Information Table S1). Our sampling covers 79 of the 89 species in the Flora of North America *Quercus* treatment (Nixon, 1997), 68 of an estimated 154 species of Mexico and Central America, and 13 of the estimated 25 white oak species of Eurasia. Nineteen (6.2%) out of 307 collections were collected from material of cultivated or uncertain wild origin, because material of known wild provenance was not available. Seven samples were removed as a result of poor sequence quality or taxonomic problems. The final 300 samples represent 153 ingroup taxa in 146 species and seven outgroups from *Castanopsis*, *Castanea*, *Lithocarpus*, and *Quercus* section *Cerris* (see Methods S1 for additional details).

### Restriction site-associated DNA sequencing and data processing

Restriction site-associated DNA extraction, library preparation, and sequencing were conducted as presented previously by Hipp *et al.* (2014) and Cavender-Bares *et al.* (2015). The RAD sequencing library was prepared at Floragenex, Inc. (Portland, OR, USA) following the methods of Baird *et al.* (2008) with *Pst*I, barcoded by individual, and sequenced on an Illumina Genome Analyzer IIx or HiSeq 2500 at Floragenex. Sequencing reads were 100 bp in length; after removal of the barcode and recognition sequence, analyzed sequences were 85 bp long. Sequence data were assembled into loci for phylogenetic analysis using PyRAD v.1.8 (Eaton, 2014) at 85% sequence similarity, as described in Pham *et al.* (2017) (see Methods S1 for additional details). All sequence data analyzed in this paper are available as FASTQ files from NCBI's Short Read Archive (Table S1), and aligned loci are available from DRYAD (doi: 10.5061/dryad.2qk88) and <https://github.com/andrewhipp/oak-convergence-2017>.



**Fig. 1** Sampling localities and diversity of oaks in North and Central America. Oak species diversity per  $1.2^\circ \times 0.8^\circ$  cell, illustrated in grayscale, was estimated from specimens utilized in analyses of climatic and soils niche presented in this paper. All samples are mapped to the latitude and longitude of the source population, where data were available (Supporting Information Table S1).

## Phylogenetic inference

Maximum likelihood phylogenetic analyses were conducted in RAxML v8.2.4 (Stamatakis, 2014) using the GTRCAT approximation of the general time-reversible model of nucleotide evolution, with branch support assessed using resampling estimated log likelihood (RELL) bootstrapping. Analyses were conducted on datasets clustered with a minimum of 20 individuals per locus; initial trials (not presented here) demonstrated that varying this number from 10 to 40 individuals per locus had little effect on the topology. Searches were constrained with the topology (*Prinoideae*, (*Albae*, *Roburoids*)), as unconstrained searches recover a topology that, owing to ancient hybridization, misleadingly places the Eurasian white oaks sister to the remainder of the white oaks (see analyses and discussion in McVay *et al.*, 2017b).

For computational reasons with this large dataset, diversification times were estimated using penalized likelihood (Sanderson, 2002; Paradis, 2013). Three alternative models of among-lineage substitution rate variation were compared using  $\Phi$ IC (Paradis, 2013). The tuning parameter  $\lambda$  was varied between 0.1 and 1.0 for the correlated and relaxed clock models, while the default setting of 10 rate categories was used for the discrete clock model. On the discrete clock tree, which was favored ( $\Phi$ IC<sub>discrete</sub> = 353.5;  $\Phi$ IC<sub>correlated</sub> = 918.4;  $\Phi$ IC<sub>relaxed</sub> = 925.9), we estimated uncertainty by first optimizing branch lengths in RAxML using a nonclock model (GTR +  $\gamma$ ) on the maximum likelihood tree using 100 delete-half jackknife datasets. We optimized branch lengths under the discrete clock option five times on each input tree, drawing calibrations from the distributions specified in the fossil calibrations below.

Four fossil calibrations were used. Calibration at the crown of the genus *Quercus* follows Cavender-Bares *et al.* (2015), drawing from an exponential distribution with a median of 50.3 Ma and 95% of the distribution between 48 and 60.3 Ma ( $\lambda e^{-\lambda x} + 48$ ,  $\lambda = 0.3$ ). The exponential distribution is used because the fossil ages are estimated at 50 Ma BP (Denk & Grimm, 2009; Denk *et al.*, 2012; Grímsson *et al.*, 2015), but the clade could be older. A normal distribution with a median of 45 Ma and 95% of the distribution between 42 and 48 Ma was used for the root of the American oak clade (McIntyre, 1991; McIver & Basinger, 1999). The crowns of the red oaks and of the white oaks (excluding *Virentes* and *Q. pontica*–*Q. sadleriana*) were set at minimum ages of 33 Ma, using an exponential distribution with a median of 33.3 Ma and 95% of the distribution between 31.1 and 43.3 Ma ( $\lambda e^{-\lambda x} + 31$ ,  $\lambda = 0.3$ ), based on fossils from the Oligocene Catahoula Formation showing separation of the red oaks and white oaks (Daghlian & Crepet, 1983; Elsie & Yancey, 2000). While these fossils look much like modern North American oaks, they cannot be reliably placed among contemporary species. Diversification rate analyses were conducted both with and without these latter two calibrations.

## Specimen and environmental data

Soil and climate variables were estimated from georeferenced occurrence records of herbarium specimens or field records

obtained from 12 public specimen databases, augmented by our own records (Table S2). Data for 136 667 records were cleaned by removal of records outside of published ranges, records of suspected cultivated origin or questionable georeferencing quality, and duplicated or geographically proximate (within 1.3–1.6 km of each other) records (Table S2). The resulting dataset included  $877 \pm 2385$  (SD) unique presence records per species for each of 137 species. Thirteen species had 10 or fewer records, and 98 had at least 50 (Table S2).

For each occurrence point, we sampled 30' resolution soil records (HWSD v.2.1; FAO/IIASA/ISRIC/ISSCAS/JRC 2012) and climate records from current rasters (WorldClim v.1.4) (Hijmans *et al.*, 2005). Moisture index ( $I_m = 100 \times (\text{precipitation} - \text{PET})/(\text{PET})$ ), where PET is potential evapotranspiration) was used to estimate water stress (Edwards, 2006). PET was calculated from WorldClim mean monthly air maximum and minimum temperature values and day length for each location (Thornthwaite, 1948). Climate and soils data averaged by species were used for the analyses discussed later.

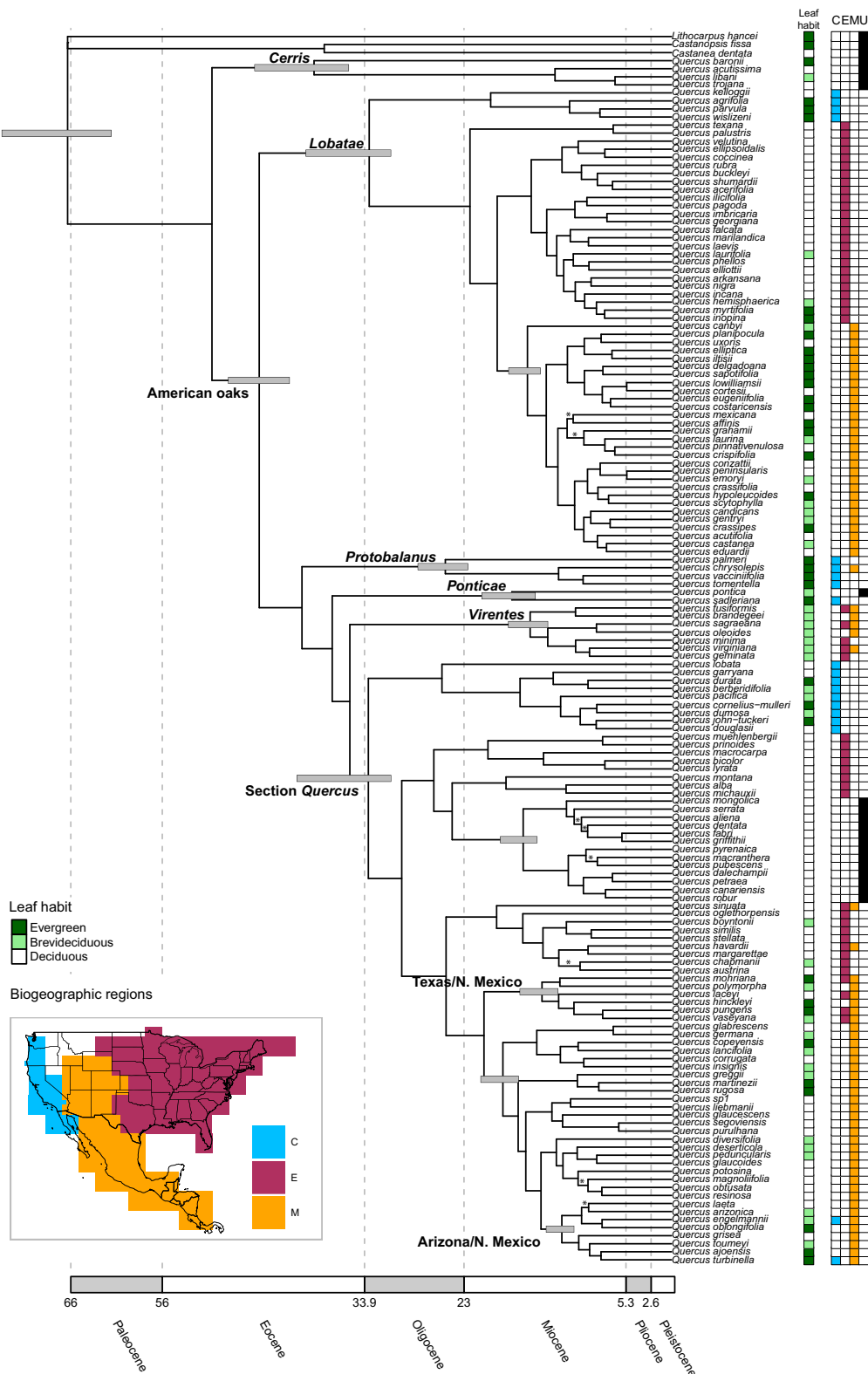
## Biogeography and diversification

Taxa were all scored as being present in one or more of four different areas (Table S3): M, Mexico (Mexico, Central America, Arizona and New Mexico); C, California (the California Floristic Province and the Pacific Northwest); E, eastern North America (North America outside of the areas listed earlier); and U, Eurasia. These bioregions were identified using Infomap Bioregions (Edler *et al.*, 2017) based on specimen records described in the previous section and default clustering parameters (cell size = 1–4°, cell capacity = 10–100, for a total of 750 grid cells; 10 trials conducted; Figs 2, S1). For analyses focused on Mexico and Central America vs all other regions, biogeography was coded as a two-state character, with areas C, T, E, and U combined into one region, and M coded as the second region.

Three estimates of diversification rate were conducted. Magallon & Sanderson's (2001) method was first used to estimate net diversification rates for clades enumerated in Table 1. We then utilized the geographic state speciation and extinction model (GeoSSE) (Goldberg *et al.*, 2011) to jointly reconstruct ancestral distributions, extinction rates, speciation rates, and dispersal between Mexico and North America north of Mexico; for this analysis, we considered Mexico to include the southwestern US, supported by the Bioregions clustering approach (Fig. S1) and corresponding approximately to Megaméxico 3 of Rzedowski (1991). Finally, we compared these analyses conditioned on changes in diversification rate predicted by a particular hypothesis with analyses conducted using reversible-jump Markov chain Monte Carlo (rjMCMC) to identify shifts in diversification rate without specifying transition nodes *a priori* (Rabosky, 2014). Analyses are detailed in Methods S1.

## Niche evolution

We used nonmetric multidimensional scaling (NMDS) to characterize the climatic, edaphic, and combined niche for the species



**Fig. 2** Maximum likelihood phylogeny of the American oak clade, with biogeographic regions, leaf habit, and divergence time estimates. Branches marked with an asterisk (\*) have bootstrap values < 0.95; all others have bootstrap support of 96–100%. Gray bars at the internodes indicate the 95% uncertainty in dating, integrating over branch-length uncertainty and calibration uncertainty, as described in the Methods. Biogeographic regions were inferred using the bipartite network method implemented in Infomap Bioregions. C, California (California Floristic Province and the Pacific Northwest); E, eastern North America; M, Mexico (Mexico, Central America, Arizona and New Mexico); U, Eurasia.

**Table 1** Divergence time estimates (in millions of yr (Ma)) with four fossil constraints, estimated clade diversity (number of species,  $n$ ), and absolute diversification rates assuming extinction rates ( $\epsilon$ ) of 0.9 and 0.0, calculated using Magallon & Sanderson's (2001) estimator

Clade	Crown (Ma)	$n$	$\epsilon = 0.9$	$\epsilon = 0.0$
Oaks ( <i>Quercus</i> )	50.2 (48.1, 60)	435	0.0745	0.1072
American oak clade	45.1 (41.9, 48)	284	0.0738	0.1100
Protobalanus	24.7 (22.7, 27.6)	4	0.0082	0.0281
Ponticae	17.4 (15.3, 20.7)	2	0.0000	0.0000
Virentes	15.4 (14.2, 18.1)	7	0.0268	0.0811
Sect. <i>Quercus</i> sensu stricto	33.1 (31.1, 42)	150	0.0820	0.1303
Sect. <i>Quercus</i> , Eurasian lineage	16.2 (15.1, 18.7)	25	0.0723	0.1559
Sect. <i>Quercus</i> , Mexico lineage	18.4 (17.3, 21)	80	0.1159	0.2002
Sect. <i>Quercus</i> , Mexico + Texas lineages	20.5 (19.2, 23.5)	86	0.1074	0.1836
Sect. <i>Quercus</i> , SW US lineage	12 (11.1, 13.6)	8	0.0397	0.1157
Sect. <i>Lobatae</i>	33 (31.1, 42.4)	120	0.0759	0.1240
Sect. <i>Lobatae</i> , Mexico lineage	15.7 (14.7, 17.7)	74	0.1312	0.2294

we investigated. Thirty-three continuous or ordered categorical soil variables and all 19 BIOCLIM variables were included (Table S2). Ordination was conducted on a Euclidean distance matrix with variables scaled to unit variance. NMDS was conducted from  $K=1$  to  $K=10$ , and stress was plotted to select the best-fit number of dimensions for ordination. For visualization, tree topologies were projected into the two-dimensional ( $K=2$ ) ordination space. Analyses were conducted using the VEGAN and PHYTOOLS packages (Oksanen *et al.*, 2012; Revell, 2012) in R v. 3.3.1 (R Core Team).

To detect transitions in niche, we evaluated the relative support for alternative multi-regime Ornstein–Uhlenbeck (OU) models. Because analyses based on most variable ordination axes are biased toward detecting evolution earlier in a phylogeny (Uyeda *et al.*, 2015), we analyzed both axes of the  $K=2$  MDS ordinations, with analysis on the  $K=5$  ordination axes for comparison. Analysis was performed using the LILOU+IC method (Khabbazian *et al.*, 2016), a model-selection and evaluation approach implemented using the least absolute shrinkage and selection operator ('LASSO') (Tibshirani, 1996, 2011) method and a modified information criterion that accounts for clade sizes entailed by shifts in adaptive regime. We followed the initial detection of shifts using LILOU with an adaptation of the phylogenetic lasso that collapses shifts to identify convergent selective regimes.

Of similar interest is the question of whether there are increases or decreases in the rate of trait evolution on the tree associated with changes in speciation rate. We used an rjMCMC method to estimate rates of evolution in annual  $I_m$  for all branches of the tree (Eastman *et al.*, 2011). This approach models changes in rate evolution by modifying the covariance matrix of a Brownian motion model of evolution (O'Meara *et al.*, 2006), using rjMCMC (Green, 1995) to estimate the posterior distribution of the number of evolutionary rates on a tree (i.e., the

number of parameters in the rate model), the position of rate shifts, and rates of evolution in different subtrees. The average post-burn-in rate of trait evolution for each branch integrates over the model space investigated.

## Evolution of leaf habit

Species were coded as deciduous (leaves drop synchronously at the onset of drought or freezing temperatures, and leaves are absent for a significant portion of each year), brevideciduous (leaves generally present year-round or with a brief period of leaflessness), or evergreen (leaf life span  $> 1$  yr, leaves present year-round) (Table S3). This coding closely follows Schmerler *et al.* (2012). Data were taken from floristic literature (Nixon, 1997; de Beaulieu & Lamant, 2010) and cross-validated where possible among sources.

Stochastic mapping (Huelsenbeck *et al.*, 2003) was utilized to estimate the number of transitions in leaf habit for the phylogeny as a whole as well as for the predominantly Mexican clades vs the remainder of the tree. One thousand MCMC generations were performed under three assumptions about the root state for the American oaks: an evergreen root, a deciduous root, and the root state set at the stationary distribution estimated numerically, as implemented in phytools (Revell, 2012). Leaf habit was coded as a binary trait for this analysis. Phylogenetic regression was used to assess the partial effects of moisture stress ( $I_m$ ), growing season temperature (BIO10), minimum temperature in the coldest month (BIO6), temperature seasonality (BIO4) and moisture seasonality (BIO15) on variance in leaf habit. Leaf habit was coded as a binary character in three ways: treating deciduous and brevideciduous together as a state alternative to evergreen; treating evergreen and brevideciduous together as a state alternative to deciduous; and ignoring brevideciduous taxa. Analyses were performed using phylogenetic logistic regression (Ives & Garland, 2010) and generalized estimating equations (Paradis & Claude, 2002). Cumulative Akaike information criterion (AIC) weights and partial correlation coefficients were to assess the relative importance of each environmental variable in explaining the observed leaf character states.

## Results

### RAD-seq data matrix

Restriction site-associated DNA sequencing runs produced  $8.97 \times 10^5$  to  $3.84 \times 10^6$  sequences per individual, with a mean of  $1.56 \times 10^6 \pm 6.24 \times 10^5$  (SD). Sequencing quality averaged  $36.48 \pm 0.733$  (SD). After removal of seven individuals based on initial analyses, the number of loci returned for single individuals ranged from 500 (1.17% of the total) to  $1.27 \times 10^4$  (29.6%), with a mean ( $\pm$  SD) of  $7.38 \times 10^3$  (17.2%)  $\pm 2.08 \times 10^3$ . Including outgroups, the RAD-seq dataset comprises 300 individuals and  $4.29 \times 10^4$  loci, spanning 80–127 bp, with an average length of  $90.1 \text{ bp} \pm 4.21$ , and total matrix width of  $3.87 \times 10^6$  nucleotides.

## Phylogeny

We recovered six major *Quercus* clades (Figs 2, S1) that correspond closely with traditional taxonomy. One is an outgroup, section *Cerris* of Europe and Asia, which consistently falls outside the American oak clade (Manos *et al.*, 1999; Pearse & Hipp, 2009). The remaining five major clades are *Quercus* section *Lobatae* G.Don, the red oaks; *Quercus* section *Protobalanus* (Trelease) A.Camus, the intermediate or golden-cup oaks; the sister species pair *Quercus sadleriana* R.Br. of California and Oregon, USA, and *Q. pontica* K.Koch of the western Caucasus Mountains and northern Turkey and Armenia (the *Ponticae* clade *sensu* McVay *et al.*, 2017b); *Quercus* series *Virentes* Trel. [= *Q.* section *Virentes* Loudon]; and *Quercus* section *Quercus*, the white oaks, excluding *Virentes* and the *Ponticae* clade.

Within both the red oaks (sect. *Lobatae*) and the white oaks (sect. *Quercus*), the species of the California Floristic Province are sister to a predominantly eastern North American lineage. Within both sections, a Mexican/southwestern US clade is sister to an eastern North American clade and derived from within rather than sister to the section as a whole. Of the 110 species for which two or more samples were available, 84 were monophyletic (Fig. S2). Only four Mexican species had taxonomic disparity index (Pham *et al.*, 2016) values > 9, suggesting taxonomic issues. Whether the c. 24% of species in our sample that are not strictly monophyletic reflect unclear taxonomy (e.g. the examples noted earlier from Mexico) or introgression is not a focus of this paper but has been studied in detail in previous papers (Cavender-Bares *et al.*, 2015; Eaton *et al.*, 2015; Pham *et al.*, 2017; McVay *et al.*, 2017a,b).

## Timing of diversification and biogeography

Infomap Bioregions identified 11 regions (Fig. S1) that we have aggregated into three: the California Floristic Province and the Pacific Northwest; eastern North America; and Mexico, Central America, and the southwestern USA. These regions correspond closely with clades in both the white oaks and the red oaks, with one notable exception: the *Virentes* clade ranges from southeastern North America to Mexico (Muller, 1961; Cavender-Bares *et al.*, 2015). Two white oak species exhibit cross-region distributions: *Q. havardii* of the southeastern US clade that contains *Q. stellata* is endemic to the southwestern high plains and tablelands of Texas and New Mexico; and *Q. engelmannii* of the southern Rocky Mountain lineage sits at the north edge of Baja California. Aside from these species, geographic regions reliably place each species within one geographically defined clade in the red oaks or the white oaks, with minimal overlap among regions.

With the crowns of both the white oak and the red oak clades constrained to a mean age of c. 33 Ma, the crown age of the *Virentes* is inferred to be 15.4 (14.2–18.1) Ma (Table 1), which is somewhat older than the 11 (8.4–14.1) Ma estimate of Cavender-Bares *et al.* (2015). The two-calibration tree estimates a *Virentes* crown age of 11.7 (10.8–13.3) Ma, close to the Cavender-Bares *et al.* (2015) estimate, but it estimates a white oak (section *Quercus*) crown age of 22.2 (20.9–24.7) Ma, far younger

than the Catahoula Formation fossils (Table S4). We consequently consider the four-calibration estimates to be more reliable and present these throughout the paper (Table 1; Fig. 2). The former paper utilized a sparse taxon sampling outside of *Virentes* and a Bayesian relaxed clock method. Our estimates may therefore not be entirely comparable and should be taken as rough estimates of divergence times.

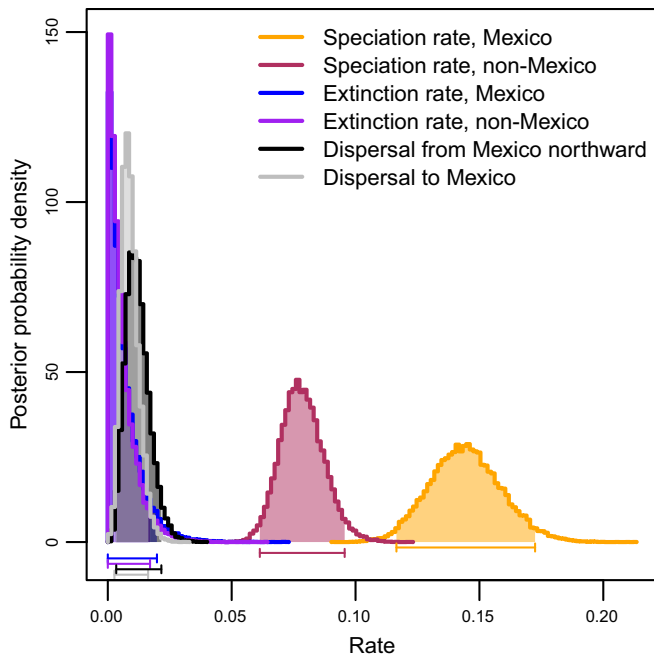
## Estimates of diversification rates

The Mexican white oak and red oak clades exhibit 54% and 85% increases in diversification rate, respectively, over their source clade at  $\epsilon = 0.0$ , as estimated using Magallon & Sanderson's (2001) method (Table 1). This increase is also supported by analysis under the geographic state speciation and extinction model (GeoSSE) model, which strongly rejects a model in which speciation rate is the same in the Mexican lineages as in the remainder of the Americas (AIC = 1138.16; speciation rate in both areas = 0.1026) relative to models in which speciation rate differs between the areas (excluding speciation in lineages split across the regions, AIC = 1121.65; or a model allowing speciation in the shared region, AIC = 1123.65). In the latter models, speciation rate in Mexico and Central America is approximately twice the speciation rate in the remainder of the Americas (0.140 vs 0.075; Fig. 3), and extinction and dispersal rates do not differ significantly between areas. Analysis using rjMCMC similarly supports increases in the rate of diversification at the base of the Mexican oak clades, as well as increases at the base of the eastern North American white oaks and the crown node of the red oaks excluding *Q. palustris* and *Q. texana* (Fig. S3).

## Niche evolution

Nonmetric multidimensional scaling results from  $K = 1$  to  $K = 5$  exhibit steady decreases in stress from 0.183 at  $K = 2$  to 0.043 at  $K = 5$ , with little decrease in stress beyond  $K = 5$  (Fig. S4). All traits except for soil drainage, topsoil sand and clay, and mean temperature of the wettest quarter (BIO8) show strongly significant correlation with ordination axes (at  $P < 0.01$ ). NMDS axis 1 exhibits a strong latitudinal gradient and a gradient in mean temperature of the coldest quarter (BIO11; Fig. 4); NMDS axis 2 exhibits a longitudinal gradient (Fig. S5). Mean temperature of the warmest quarter does not exhibit gradual variation on the ordination surface. Plotting taxa from the  $K = 2$  ordination separates taxa by geography and by geographic clades within sections but does not separate the sections from one another (the red vs the white oaks; Fig. 4). The white and red oaks show nearly identical radiations in geographic and ecological space. Climate alone shows a similar separation between oaks of geographic regions, but soils data alone more weakly distinguish the Mexican/Central American clade from the remaining taxa (Fig. S6).

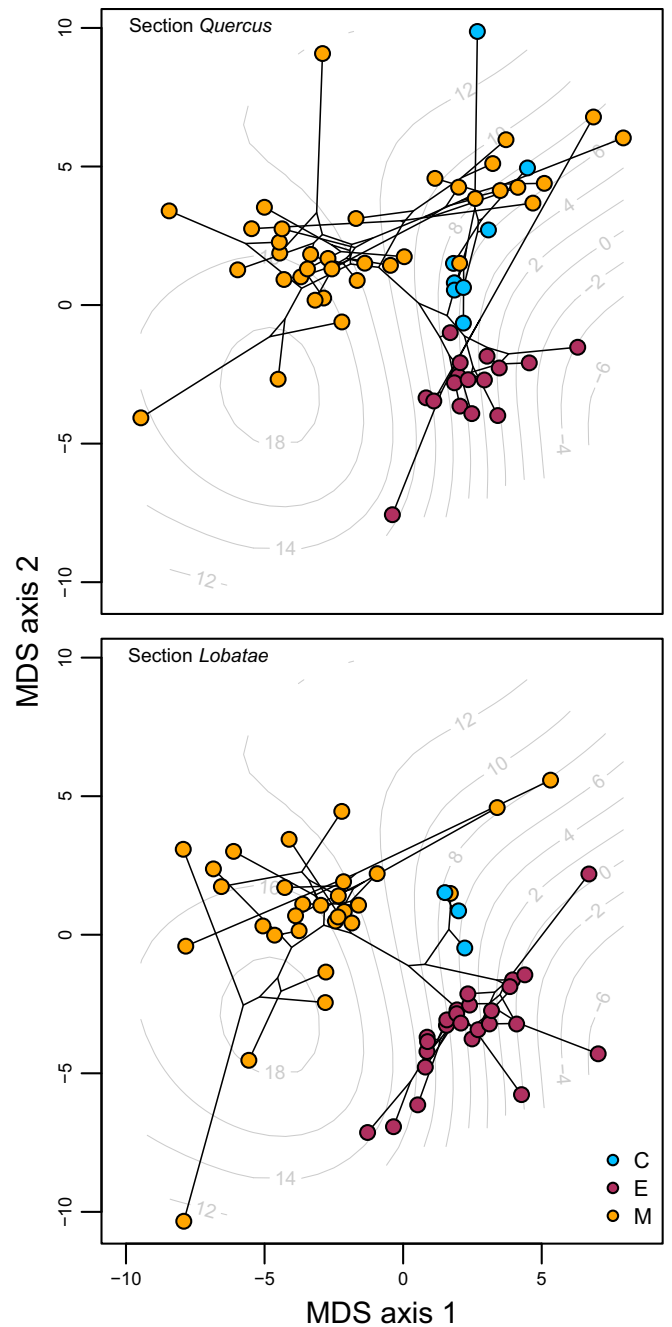
Multivariate OU analyses in LIQU based on the  $K = 2$  ordination identify a total of 18 ecological transitions collapsed into eight selective regimes, of which four regimes entail clades of more than two individuals: one regime comprising two convergent shifts in climatic and soils niche at the base of the Mexican



**Fig. 3** Speciation, extinction, and dispersal rates for lineages of Mexico/Central America and American lineages outside of Mexico and North America, estimated under the geographic state speciation and extinction model (GeoSSE) model, assuming no speciation in species that are present in both Mexico and outside of Mexico ( $sAB = 0$ ); this model (Akaike information criterion (AIC) = 1121.65) is weakly favored over the full (unconstrained) model (AIC = 1123.65), because parameter estimates under the full model are almost identical, in which  $sAB$  is estimated to be zero. Both models are strongly favored over a model in which the speciation rate in Mexico is constrained to be the same as the speciation rate outside of Mexico (AIC = 1138.16). Horizontal bars below the figure represent 95% confidence intervals estimated using Markov chain Monte Carlo. Rates are in events per million yr.

clade in the red oaks and the combined Mexican/Texas clade in the white oaks; one bringing together five of the seven southern live oaks (*Virentes*); and one at the base of the portion of the Mexican white oak clade that is centered on Arizona, southern California, and northern Mexico (Fig. 5). All other transitions entail single species or pairs of tips. Considering that this method as currently implemented does not account for intraspecific measurement error, which may masquerade as meaningful variance outside of the phylogenetically heritable component (Ives *et al.*, 2007), we consider singletons and shifts involving only two taxa to be potentially spurious and do not consider them further in this study.

The rate of adaptation ( $\alpha$ ) of MDS1 is 0.092, which corresponds to an adaptive half-time (the amount of time expected for a trait to evolve halfway from its ancestral value to its optimum, calculated as  $\ln(2)/\alpha$ ) (Hansen, 1997) of 7.53 million yr, roughly half the age of the Mexican oak radiation as estimated on our tree. The rate of adaptation of MDS2 is 0.0042, corresponding to an adaptive half-time of 164 million yr, approximately the age of the angiosperms. While the parameters of OU models, especially  $\alpha$ , are difficult to estimate precisely, alternative models are typically nonetheless readily distinguishable with phylogenies > 50 tips and regimes that differ substantially from one another



**Fig. 4** Ordination of taxa in ecological space, combined soils and climatic data, with phylogeny overlaid. Ordination was conducted with all taxa, but plotted separately by white oaks (section *Quercus*) and red oaks (section *Lobatae*). Each point on the ordination is a tip of the phylogeny, and lines between the points are branches of the phylogeny. Positions of the internodes are estimated under a multivariate Brownian motion model. Contour lines show mean temperature of the coldest quarter in degrees centigrade, ranging from  $-6$  to  $> 18$ . California (California Floristic Province and the Pacific Northwest); E, eastern North America; M, Mexico (Mexico, Central America, Arizona and New Mexico). Eurasian taxa were excluded from this analysis. MDS, multidimensional scaling.

in optima (Cressler *et al.*, 2015). Results based on the  $K = 5$  ordination (Fig. S7) identify similar shifts to the  $K = 2$  analysis: 30 shifts are inferred for this result, collapsing in the convergence

analysis into 14 selective regimes, including a division of the *Virentes*, one of the four major selective regimes discussed earlier, into a tropical group (two species) and a more temperate group (three species), as well as identification of additional shifts in the Mexican red oaks and, somewhat implausibly, convergence between the northern *Q. ellipsoidalis* and *Q. ilicifolia* and the Mexican red oaks. This high degree of differentiation does not make much ecological or biogeographic sense and suggests that the  $K=5$  analysis may be overfitting the data.

Rate of evolution along the moisture gradient ( $I_m$ ; Fig. S8) as estimated by generalized least-squares  $\sigma^2$  (the Brownian motion rate estimator) averages 21.73 (95% quantile = 7.92, 34.82) over the entire tree. In the Mexican and Texas white oak clades,  $\sigma^2 = 20.6$  (5.1, 36.78), while in the Mexican red oaks,  $\sigma^2 = 31.97$  (8.69, 62.27). For all branches outside of these two clades,  $\sigma^2 = 18.71$  (6.83, 32.51), entailing 10.1% and 70.9% increases in the rate of evolution along the moisture gradient in the white oaks and red oaks, respectively.

### Evolution of leaf habit

Under all three root-state assumptions, the ancestors of the white and red Mexican oak clades are reconstructed to be deciduous with relatively high certainty (marginal likelihood of deciduousness = 1.000 for the Mexican white oaks and 0.964 for the Mexican red oaks assuming an evergreen root for the American oak clade, and 1.000 and 0.977, respectively, assuming a deciduous root; Fig. S9). The ancestral state of the white oaks and the red oaks as a whole is uncertain, ranging from a marginal likelihood of deciduousness equal to 1.000 and 0.948, respectively, under a deciduous root assumption, to 0.967 and 0.412, respectively, under an evergreen root assumption. Irrespective of the root state for the American oaks, the Mexican oak clades transition between deciduous and evergreen habit twice as frequently as the remainder of the oaks, increasing from 0.013 (95% quantile = 0.010, 0.018) changes per million yr under an evergreen root assumption or 0.012 (0.008, 0.017) changes under a deciduous root assumption to 0.024 (0.020, 0.032; no difference between reconstructions, assuming different root states) changes per million yr in the Mexico clades. Assuming an evergreen root, the American oaks exhibit 41.50 (17, 26) transitions from deciduousness to the evergreen habit. Assuming a deciduous root, they exhibit 23.17 (20, 28) transitions to the evergreen habit.

Phylogenetic logistic regression models (Ives & Garland, 2010) and generalized estimating equations (Paradis & Claude, 2002) consistently point to two models as the best fit: one in which  $I_m + \text{BIO10}$  (mean temperature in the warmest quarter) are included along with their interaction term (average AIC weight = 0.370), and one in which all predictors are included (average AIC weight = 0.438; Table S5). The Ives and Garland approach seems to favor the more complex model in our study (mean AIC weight = 0.833), while the Paradis and Claude approach less strongly favors the  $I_m + \text{BIO10} + I_m \times \text{BIO10}$  model (mean AIC weight = 0.393, compared with mean AIC weight of 0.234 for the full model). As it is not clear to us that one of these is an inherently preferable regression approach, we

rescaled predictors on the full model to evaluate the relative effect size of each predictor ( $I_m + \text{BIO4} + \text{BIO10} + \text{BIO11} + \text{BIO15}$ ) on leaf phenology. In this model, temperature seasonality (BIO4), mean temperature in the coldest quarter (BIO11), and mean temperature in the warmest quarter (BIO10) all have approximately the same effect size and are all strongly significant, irrespective of which regression method or leaf coding is used (Table S6). The transition from deciduous to evergreen leaves is associated with decrease in temperature seasonality, increase in the mean temperature in the coolest months (i.e., a decrease in intensity of cold), and decrease in the mean temperature in the warmest months (Fig. S10), but not by precipitation seasonality or drought intensity ( $I_m$ ).

### Discussion

Our study provides unprecedented resolution for the phylogeny of the American oaks and the white oaks of Eurasia. Our work demonstrates that white oaks and red oaks both arose and diversified simultaneously in temperate North America, diverging initially between the California Floristic Province and eastern North America, then radiated in Mexico from eastern North American ancestors (Fig. 2). Within Mexico and Central America, rates of ecological and lineage diversification increased in both the red and white oaks (Fig. 3). Mexico is thus not where the American oaks arose; rather, the oaks of Mexico and Central America arose in parallel radiations from eastern North American ancestors between 10 and 20 Ma. Our study reveals, moreover, that speciation rate increased as oaks moved into Mexico, in association with increased rates of evolution along moisture availability gradients and increased rates of leaf habit evolution. Thus, we demonstrate the importance of both sympatric parallel diversification of major clades and ecological diversification within clades as drivers of oak species diversity.

### American oaks arose in the northern temperate zone and diversified southward in parallel

Middle Eocene pollen records from Axel Heiberg Island (79°55'N, 88°58'W) indicate that oaks had arrived at high latitudes in North America by *c.* 45 Ma (McIntyre, 1991). Assuming this to be the minimal age of the American oak clade, the sections recognized in the current study (sections *Lobatae*, *Protobalanus*, *Ponticae*, *Virentes*, and *Quercus*) had arisen by *c.* 33 Ma (Fig. 2). This initial divergence of the crown groups of the American oaks may have been more rapid even than our analysis suggests, depending on interpretation of *Quercus* pollen records from middle Eocene Greenland (Grímsson *et al.*, 2015). In either case, however, as temperatures at high latitudes decreased by *c.* 3–5°C during the Eocene–Oligocene climate transition 34 Ma (Liu *et al.*, 2009), oaks were pushed southward toward their modern-day distributions (Fig. 1) and underwent a parallel vicariance in the red oaks and the white oaks, each section diverging into a western (California Floristic Province) clade sister to an eastern North American clade. The red oaks and white oaks then radiated simultaneously in the California Floristic Province and



eastern North America before diverging from an eastern North American ancestor and diversifying in northern Mexico. While our Mexican and Central American sampling is sparse (*c.* 44% of recognized species; see Methods for sampling), our data suggest that returns north have been limited to two species in the red oaks (*Q. emoryi* and *Q. hypoleucooides*, which range into Arizona, New Mexico and Texas) and one clade of eight species in the white oaks (Fig. 2, 'Arizona/N. Mexico' clade).

This study thus demonstrates that the American oaks have a northern temperate origin and only later colonized Mexico and Central America. While oaks are not altogether unique in their origin in the temperate zone and subsequent diversification in Mexico (Graham, 1999; Moore & Donoghue, 2007; Sloan *et al.*, 2014; Manos & Meireles, 2015), they are perhaps the most spectacular example of this process. Since their arrival in Mexico an estimated 14.5–20.8 Ma, oaks have diversified into *c.* 154 species and radiated into habitats that range from dry chaparral to mesic forests and climates that range from temperate to subtropical, with the strongest diversification in montane regions (Valencia, 2004). The closest any other woody clade we are aware of approaches this disparity between the relatively low diversity of the northern temperate ancestral lineages and higher diversity of the derived Mexican clades (Manos & Meireles, 2015) is *Pinus*, a genus that similarly arose in the north and radiated into primarily temperate regions of Mexico by the early Miocene, at *c.* 33 native species in the US (excluding five species whose range is largely in Mexico) and *c.* 50 in Mexico (Kral, 1993; Perry *et al.*, 1998; Germandt & Pérez-de la Rosa, 2014). The oaks, however, have achieved approximately threefold higher regional diversity since their similarly timed Miocene arrival in Mexico.

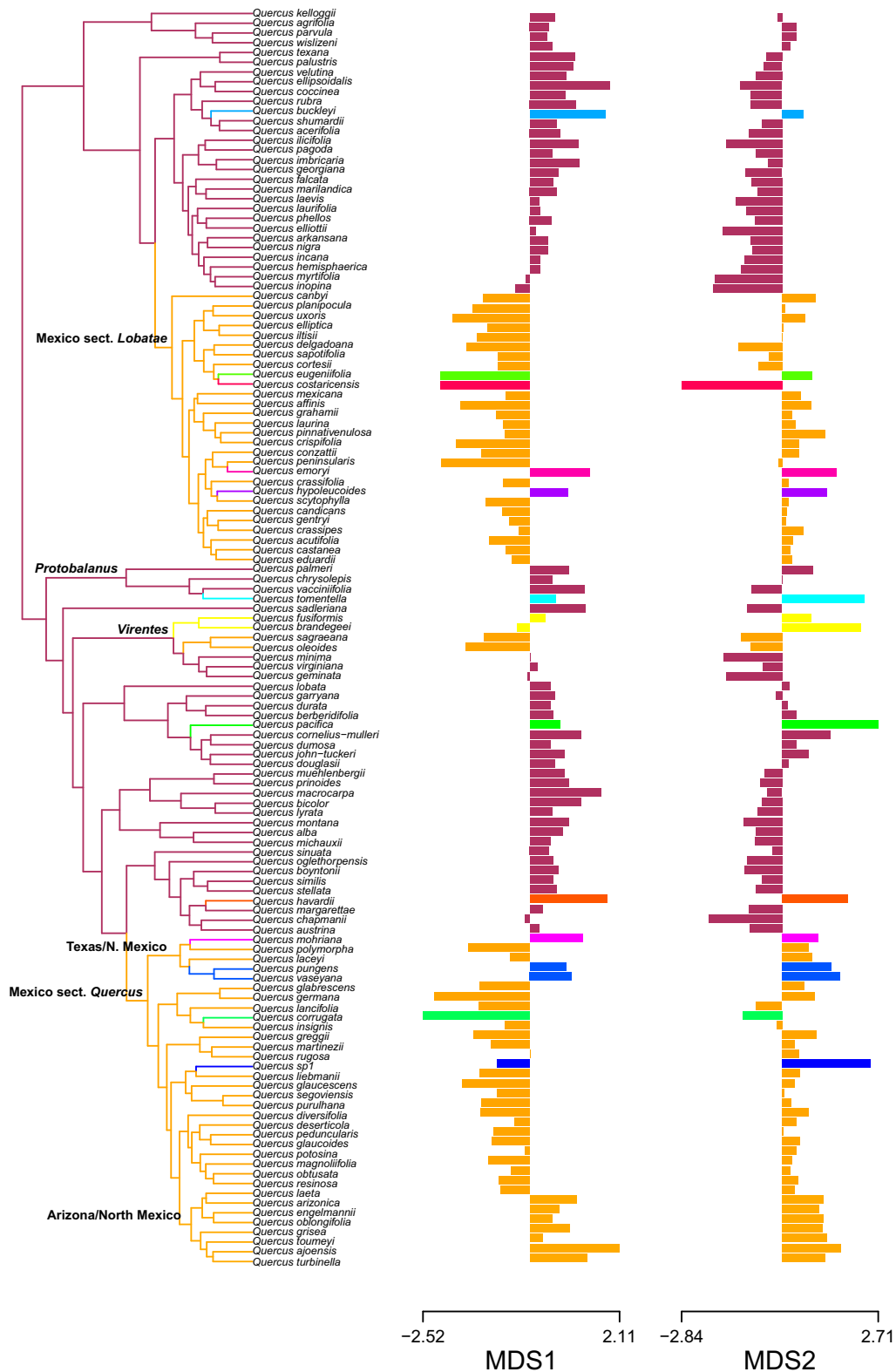
The timing of our estimate of vicariance between the eastern North American and Mexican clades is earlier than the desert incursion of *c.* 10 Ma in Texas and northeastern Mexico, which is associated with eastern Mexico–eastern North American vicariance in many angiosperm clades (Manos & Meireles, 2015). Our crown age estimates for the Mexican clades may therefore be somewhat too old. In either case, however, our reconstruction implies a rapid increase in diversification rate at the base of the Mexican clades. The rates of speciation we observe in the Mexican oaks (Table 1) are comparable to two of the most rapidly diversifying major angiosperm clades (Lamiales and core Asterids, 0.211 and 0.235 species Myr<sup>-1</sup>, assuming no extinction) as well as the core group of the diverse herbaceous angiosperm genus *Carex* (0.224 species Myr<sup>-1</sup>) (Magallon & Sanderson, 2001; Escudero *et al.*, 2012). While clade age poorly predicts species richness across wide phylogenetic scales (Rabosky *et al.*, 2012), all things being equal, we expect older clades to be at least as diverse as closely related younger clades (Wiens & Donoghue, 2004; Escudero & Hipp, 2013). Mexico and Central America are thus not a 'museum' of oak diversity, with diversity shaped by clade age (cf. Fine & Ree, 2006; Antonelli *et al.*, 2015 and citations therein; Schluter, 2015), but a hotbed of comparatively young and rapid oak diversification. High diversity in the Mexican oak clades is driven by increased speciation rate in a topographically heterogeneous environment

(Schemske, 2009; Eiserhardt *et al.*, 2017) marked by elevational stratification of oaks that is unparalleled in the clades to the north (Fig. 1 in Spellenberg *et al.*, 1998).

Our work also demonstrates that oaks have undergone a remarkable parallel sympatric diversification within geographic regions. Simultaneous diversification of the red oaks and white oaks in California, eastern North American, and Mexico resulted in co-occurrence of red and white oaks across much of the continent (with exceptions among the red oaks and white oaks, as described earlier). Considering only the red oaks or the white oaks, however, we generally find representatives of just one subclade in any given region (Figs 2, 5). This finding of parallel sympatric diversification of closely related clades is rare (cf Kozak *et al.*, 2009). The pattern we observe entails phylogenetic niche conservatism within the red oaks and within the white oaks (Donoghue, 2008), but it is important to note that taking into account habitat and clade co-occurrence across communities (e.g. Cavender-Bares *et al.*, 2004) has the potential to shift the emphasis of this story from convergence to divergence (Cavender-Bares *et al.*, 2006; J. Cavender-Bares, unpublished; Swenson *et al.*, 2006, 2007).

### Ecological diversification set the stage for diversity of the American oaks

As discussed earlier, the rate of species diversification increased on the move into Mexico in both the red and the white oak clades (Table 1; Fig. 3). At the same time, the Mexican/Mesoamerican oak clades exhibit approximately a doubling in the rate of transitions among leaf habit states relative to the remainder of the oaks (Fig. S9) and a 10–71% increase in the rate of evolution along the moisture gradient in the white and red oaks, respectively (Fig. S8). Moreover, acorn maturation time appears to evolve more quickly in Mexico and Mesoamerica, where the annual habit is estimated to have been reacquired numerous times in red oaks (Nixon, 2006). This correlation between rates of speciation and ecological diversification suggests that while the high diversity of oaks throughout the Americas is shaped in large part by sympatric convergence in climatic and soils niche between the white and red oaks, the diversity of oaks in Mexico receives a boost from divergence as a result of increased ecological opportunity and/or increased allopatry in the ecologically heterogeneous montane areas of Mexico (Moore & Donoghue, 2007; Kozak & Wiens, 2010; Torres-Miranda *et al.*, 2011, 2013). Our analyses support nonphylogenetic analyses of Mexican oaks that indicate niche diversification may help explain Mexican oak diversity (Rodríguez-Correa *et al.*, 2015). The high phylogenetic niche conservatism suggested by our analysis of transitions among climatic and soils niches contrasts with the relatively high rate of evolution we find along the moisture gradient in Mexican oaks, but without a shift into the tropics in essentially any of the American oak clade outside of the *Virentes* (Cavender-Bares *et al.*, 2015). We infer that niche conservatism along axes that might have allowed transitions into the tropics drove increased allopatry and thus increased opportunity for speciation as oaks climbed the mountains into Mexico, as has been argued



**Fig. 5** Adaptive transitions in ecological space as estimated from the two axes of the  $K = 2$  nonmetric multidimensional scaling (NMDS) ordination. Changes in color on the tree signify transitions in ecological space that exceeded expectations under a Brownian motion model of character evolution. Analysis using the phylogenetic lasso (implemented in the `l1ou` R package) detected 17 transitions in ecological space, and subsequent analysis of these transitions in an adaptive framework identified nine selective regimes, including convergent shifts in the two Mexican clades (white oaks and red oaks). Each of the nine selective regimes is represented by an arbitrary color, except for the two selective regimes restricted primarily to eastern North America and Mexico, for which colors match the map and biogeographic coding in Fig. 2.

for the Cape Biodiversity Hotspot (Schnitzler *et al.*, 2011) and the Dipsacales (Moore & Donoghue, 2007). At the same time, high lability along moisture gradients allowed Mexican oaks to diversify among newly available temperate habitats created by a diverse set of mountain-building events (e.g. the uplift of the Sierra Madre Oriental, Sierra Madre Occidental, and the Trans-Mexican Volcanic Belt) and climate change leading to progressively cooler and moister climates within Mexico. This combination of ecological conservatism and lability may explain the exceptional diversification of Mexican oaks, perhaps along with decreased age at maturity in low-latitude alpine habitats (Verdú, 2002), which we have not tested in this study.

We further demonstrate that transitions from deciduousness and brevideciduousness to evergreenness, like transitions among biogeographic and climatic/soil regions, have occurred numerous times in parallel. Irrespective of what we assume about the root state of the American oak clade, our data consistently support a deciduous root for the Mexican clades and exhibit > 20 transitions to the evergreen habit within the Mexican clades. Leaf habit has most likely evolved in response to warmer temperatures in winter and reduced temperature seasonality; evergreen oak species tend to be associated with climates in which the coldest temperatures are less severe and temperature exhibits less seasonal variation (Fig. S10), suggesting that deciduousness evolved as a strategy to avoid drought and freezing. Moisture regime has been shown in previous studies to exert selection on leaf abscission and drought tolerance (Ramírez-Valiente & Cavender-Bares, 2017) as well as leaf growth rate and nitrogen content (Ramírez-Valiente *et al.*, 2017) in a Neotropical *Virentes* oak, and severity of cold season has similarly been shown to shape convergent transitions in leaf morphology in *Viburnum* (Schmerler *et al.*, 2012).

Our study demonstrates that the convergent transitions we see in oak leaf habit are associated with convergence in biogeography, climate, and soils. It also demonstrates that increases in the rate of transitions among these leaf life-history strategies are associated with an increase in lineage diversification rate in both the red oaks and the white oaks, supporting the hypothesis that increased lineage diversification in the Mexican oaks may have been driven by a higher rate of niche diversification among Mexican oaks than among their ancestors.

## Acknowledgements

Assistance in the laboratory and field was provided by staff at Floragenex; Breane Budaitis; students, postdocs and interns in the Hipp, González-Rodríguez, Manos and Cavender-Bares laboratories; Bethany Brown; Guy Sternberg and staff of Starhill Forest Arboretum; Ian Pearse, Jill Baty, Emily Griswold and staff of UC-Davis Shields Oak Grove; Beth Fallon; Jeanne Romero-Severson; and staff at the Southwestern Research Station. Discussions and feedback on initial drafts of the manuscript were provided by Deren Eaton, Emily Sessa, M. Socorro González Elizondo, Hernando Rodríguez-Correa, the Systematics Lab of The Morton Arboretum, and three anonymous reviewers. Funding for this project was provided by NSF Awards 1146488, 1146102

and 1146380, a Fulbright-Comexus grant to J.C-B. and The Morton Arboretum Center for Tree Science.

## Author contributions

A.L.H., P.S.M., J.C-B and A.G-R. conceived and obtained funding for the study. A.L.H., P.S.M., J.C-B, A.G-R., M.H. and J.D.M. collected samples in the field and generated sequence data. M.K. and J.C-B. generated the ecological data. A.L.H. and J.D.M. analyzed the phylogenetic data, and A.L.H. analyzed the comparative data. S.V.A. verified the species identity of most collections in Mexico and contributed to interpretation. All authors contributed to reviewing results and writing the manuscript.

## References

- Antonelli A, Zizka A, Silvestro D, Scharn R, Cascales-Miñana B, Bacon CD. 2015. An engine for global plant diversity: highest evolutionary turnover and emigration in the American tropics. *Frontiers in Genetics* 6: 130.
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3: e3376.
- Banda K, Delgado-Salinas A, Dexter KG, Linares-Palomino R, Oliveira-Filho A, Prado D, Pullan M, Quintana C, Riina R, Rodríguez GM *et al.* 2016. Plant diversity patterns in neotropical dry forests and their conservation implications. *Science* 353: 1383–1387.
- de Beaulieu A le H, Lamant T. 2010. *Guide Illustré des Chênes*. Geer, Belgium: Edilens.
- Bouchenak-Khelladi Y, Onstein RE, Xing Y, Schwery O, Linder HP. 2015. On the complexity of triggering evolutionary radiations. *New Phytologist* 207: 313–326.
- Cavender-Bares J. 2016. Diversity, distribution, and ecosystem services of the North American oaks. *International Oaks* 27: 37–48.
- Cavender-Bares J, Ackerly D, Baum D, Bazzaz F. 2004. Phylogenetic overdispersion in Floridian oak communities. *American Naturalist* 163: 823–843.
- Cavender-Bares J, Gonzalez-Rodríguez A, Eaton DAR, Hipp AL, Beulke A, Manos PS. 2015. Phylogeny and biogeography of the American live oaks (*Quercus* subsection *Virentes*): a genomic and population genetics approach. *Molecular Ecology* 24: 3668–3687.
- Cavender-Bares J, Keen A, Miles B. 2006. Phylogenetic structure of Floridian plant communities depends on taxonomic and spatial scale. *Ecology* 87: S109–S122.
- Cressler CE, Butler MA, King AA. 2015. Detecting adaptive evolution in phylogenetic comparative analysis using the Ornstein-Uhlenbeck model. *Systematic Biology* 64: 953–968.
- 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.
- 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, Grimm GW. 2010. The oaks of western Eurasia: traditional classifications and evidence from two nuclear markers. *Taxon* 59: 351–366.
- 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.
- Donoghue MJ. 2008. A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences, USA* 105: 11549–11555.
- Eastman JM, Alfaro ME, Joyce P, Hipp AL, Harmon LJ. 2011. A novel comparative method for modeling shifts in the rate of character evolution on trees. *Evolution* 65: 3578–3589.

- Eaton DAR. 2014. PyRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics* 30: 1844–1849.
- Eaton DAR, Hipp AL, González-Rodríguez A, Cavender-Bares J. 2015. Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution* 69: 2587–2601.
- Elder D, Guedes T, Zizka A, Rosvall M, Antonelli A. 2017. Infomap bioregions: interactive mapping of biogeographical regions from species distributions. *Systematic Biology* 66: 197–204.
- Edwards EJ. 2006. Correlated evolution of stem and leaf hydraulic traits in *Pereskia* (Cactaceae). *New Phytologist* 172: 479–789.
- Eiserhardt WL, Couvreur TLP, Baker WJ. 2017. Plant phylogeny as a window on the evolution of hyperdiversity in the tropical rainforest biome. *New Phytologist* 214: 1408–1422.
- 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.
- Escudero M, Hipp A. 2013. Shifts in diversification rates and clade ages explain species richness in higher-level sedge taxa (Cyperaceae). *American Journal of Botany* 100: 2403–2411.
- Escudero M, Hipp AL, Waterway MJ, Valente LM. 2012. Diversification rates and chromosome evolution in the most diverse angiosperm genus of the temperate zone (*Carex*, Cyperaceae). *Molecular Phylogenetics and Evolution* 63: 650–655.
- Fine PV, Ree RH. 2006. Evidence for a time-integrated species-area effect on the latitudinal gradient in tree diversity. *American Naturalist* 168: 796–804.
- Fitz-Gibbon S, Hipp AL, Pham KK, Manos PS, Sork V. 2017. Phylogenomic inferences from reference-mapped and *de novo* assembled short-read sequence data using RADseq sequencing of California white oaks (*Quercus* subgenus *Quercus*). *Genome* 60: 743–755.
- Gernandt DS, Pérez-de la Rosa JA. 2014. Biodiversidad de Pinophyta (coníferas) en México. *Revista Mexicana de Biodiversidad* 85: 126–133.
- Goldberg EE, Lancaster LT, Ree RH. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Systematic Biology* 60: 451–465.
- Graham A. 1999. The Tertiary history of the northern temperate element in the northern Latin American biota. *American Journal of Botany* 86: 32–38.
- Green PJ. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82: 711–732.
- Grimsson F, Zetter R, Grimm GW, Pedersen GK, Pedersen AK, Denk T. 2015. Fagaceae pollen from the early Cenozoic of West Greenland: revisiting Engler's and Chaney's Arcto-Tertiary hypotheses. *Plant Systematics and Evolution* 301: 809–832.
- Hansen TF. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51: 1341–1351.
- 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, 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: e93975.
- Hubert F, Grimm GW, Jousset 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.
- Huelsenbeck JP, Nielsen R, Bollback JP. 2003. Stochastic mapping of morphological characters. *Systematic Biology* 52: 131–158.
- Ives AR, Garland T. 2010. Phylogenetic logistic regression for binary dependent variables. *Systematic Biology* 59: 9–26.
- Ives AR, Midford PE, Theodore Garland J. 2007. Within-species variation and measurement error in phylogenetic comparative methods. *Systematic Biology* 56: 252–270.
- Khabbazian M, Kriebel R, Rohe K, Ané C. 2016. Fast and accurate detection of evolutionary shifts in Ornstein-Uhlenbeck models. *Methods in Ecology and Evolution* 7: 811–824.
- Kozak KH, Mendyk RW, Wiens JJ. 2009. Can parallel diversification occur in sympatry? Repeated patterns of body-size evolution in coexisting clades of North American salamanders. *Evolution* 63: 1769–1784.
- Kozak KH, Wiens JJ. 2010. Accelerated rates of climatic-niche evolution underlie rapid species diversification. *Ecology Letters* 13: 1378–1389.
- Kral R. 1993. *Pinus* Linnaeus. In: Flora of North America Editorial Committee, ed. *Flora of North America, Volume 2: Pteridophytes and gymnosperms*. New York, NY, USA: Oxford University Press, 373–398.
- Larson-Johnson K. 2016. Phylogenetic investigation of the complex evolutionary history of dispersal mode and diversification rates across living and fossil Fagales. *New Phytologist* 209: 418–435.
- Leroy T, Roux C, Villate L, Bodénès C, Romiguier J, Paiva JAP, Dossat C, Aury J-M, Plomion C, Kremer A. 2017. Extensive recent secondary contacts between four European white oak species. *New Phytologist* 214: 865–878.
- Liu Z, Pagani M, Zinniker D, DeConto R, Huber M, Brinkhuis H, Shah SR, Leckie RM, Pearson A. 2009. Global cooling during the Eocene-Oligocene climate transition. *Science* 323: 1187–1190.
- Magallon SA, Sanderson MJ. 2001. Absolute diversification rates in angiosperm clades. *Evolution* 55: 1762–1780.
- 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.
- Manos PS, Meireles JE. 2015. Biogeographic analysis of the woody plants of the Southern Appalachians: implications for the origins of a regional flora. *American Journal of Botany* 102: 780–804.
- Manos PS, Stanford AM. 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.
- McIntyre DJ. 1991. Pollen and spore flora of an Eocene forest, eastern Axel Heiberg Island. *N.W.T. Geological Survey of Canada Bulletin* 403: 83–97.
- McIver EE, Basinger JF. 1999. Early tertiary floral evolution in the Canadian High Arctic. *Annals of the Missouri Botanical Garden* 86: 523–545.
- McVay JD, Hauser D, Hipp AL, Manos P. 2017a. Phylogenomics reveals a complex evolutionary history of lobed-leaf white oaks in Western North America. *Genome* 60: 733–742.
- McVay JD, Hipp AL, Manos PS. 2017b. A genetic legacy of introgression confounds phylogeny and biogeography in oaks. *Proceedings of the Royal Society B* 284: 20170300.
- Moore BR, Donoghue MJ. 2007. Correlates of diversification in the plant clade Dipsacales: geographic movement and evolutionary innovations. *American Naturalist* 170: S28–S55.
- Muller CH. 1961. The live oaks of the Series Virentes. *American Midland Naturalist* 65: 17–39.
- Nixon KC. 1993. The genus *Quercus* in Mexico. In: Ramamoorthi R, Bye R, Lot A, Fa I, eds. *Biological diversity of Mexico: origins and distribution*. Oxford, UK: Oxford University Press, 447–458.
- Nixon KC. 1997. *Quercus*. In: Flora of North America Editorial Committee, eds. *Flora of North America North of Mexico*. New York, NY, USA: Oxford University Press, 445–447.
- Nixon KC. 2006. Global and neotropical distribution and diversity of oak (genus *Quercus*) and oak forests. In: Kappelle M, ed. *Ecology and conservation of neotropical montane oak forests*. Berlin/Heidelberg, Germany: Springer Verlag, 3–13.
- Oh S-H, Manos PS. 2008. Molecular phylogenetics and cupule evolution in Fagaceae as inferred from nuclear CRABS CLAW sequences. *Taxon* 57: 434–451.
- Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara B, Simpson GL, Henry M, Stevens H, Wagner H. 2017. *vegan: Community Ecology Package*. R package v.2.4-3. [WWW document] URL <https://cran.r-project.org/web/packages/vegan/index.html> [accessed 31 August 2017].
- O'Meara BC, Ané C, Sanderson MJ, Wainwright PC. 2006. Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60: 922–933.
- Owusu SA, Sullivan AR, Weber JA, Hipp AL, Gailing O. 2015. Taxonomic relationships and gene flow in four North American *Quercus* species (*Quercus* section *Lobatae*). *Systematic Botany* 40: 510–521.
- Paradis E. 2013. Molecular dating of phylogenies by likelihood methods: a comparison of models and a new information criterion. *Molecular Phylogenetics and Evolution* 67: 436–444.

- Paradis E, Claude J. 2002. Analysis of comparative data using generalized estimating equations. *Journal of Theoretical Biology* 218: 175–185.
- 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, USA* 106: 18097–18102.
- Perry JP Jr, Graham A, Richardson DM. 1998. The history of pines in Mexico and Central America. In: Richardson DM, ed. *Ecology and biogeography of Pinus*. Cambridge, UK: Cambridge University Press, 137–149.
- Pham KK, Hahn M, Lueders K, Brown BH, Bruederle LP, Bruhl JJ, Chung K-S, Dierig NJ, Escudero M, Ford BA *et al.* 2016. Specimens at the center: an informatics workflow and toolkit for specimen-level analysis of public DNA database data. *Systematic Botany* 41: 529–539.
- Pham KK, Hipp AL, Manos PS, Cronn RC. 2017. A time and a place for everything: phylogenetic history and geography as joint predictors of oak plastome phylogeny. *Genome* 60: 720–732.
- R Core Team. 2016. *R: a language and environment for statistical computing, version 3.3.1*. Vienna, Austria: R Foundation for Statistical Computing. [WWW document] URL <http://www.R-project.org/>.
- Rabosky DL. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE* 9: e89543.
- Rabosky DL, Slater GJ, Alfaro ME. 2012. Clade age and species richness are decoupled across the eukaryotic tree of life. *PLoS Biology* 10: e1001381.
- Ramírez-Valiente JA, Cavender-Bares J. 2017. Evolutionary trade-offs between drought resistance mechanisms across a precipitation gradient in a seasonally dry tropical oak (*Quercus oleoides*). *Tree Physiology* 37: 889–901.
- Ramírez-Valiente JA, Center A, Sparks J, Sparks K, Etterson J, Longwell T, Pilz G, Cavender-Bares J. 2017. Population-level differentiation in growth rates and leaf traits in seedlings of the neotropical live oak *Quercus oleoides* grown under natural and manipulated precipitation regimes. *Frontiers in Plant Science* 8: 585.
- Ree RH, Hipp AL. 2015. Inferring phylogenetic history from restriction site associated DNA (RADseq). In: Hoerandl E, Appelhaus M, eds. *Next generation sequencing in plant systematics*. Koenigstein, Germany: Koeltz Scientific Books, 181–204.
- Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3: 217–223.
- 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.
- Rzedowski J. 1991. Diversidad y orígenes de la flora fanerogámica de México. *Acta Botánica Mexicana* 14: 3–21.
- Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* 19: 101–109.
- Schemske DW. 2009. Biotic interactions and speciation in the tropics. In: Butlin RK, Bridle JR, Schluter D, eds. *Speciation and patterns of diversity*. Cambridge, UK: Cambridge University Press, 219–239.
- Schluter D. 2015. Speciation, ecological opportunity, and latitude. *American Naturalist* 187: 1–18.
- Schmerler SB, Clement WL, Beaulieu JM, Chatelet DS, Sack L, Donoghue MJ, Edwards EJ. 2012. Evolution of leaf form correlates with tropical-temperate transitions in *Viburnum* (Adoxaceae). *Proceedings of the Royal Society B: Biological Sciences* 279: 3905–3913.
- Schnitzler J, Barraclough TG, Boatwright JS, Goldblatt P, Manning JC, Powell MP, Rebelo T, Savolainen V. 2011. Causes of plant diversification in the Cape biodiversity hotspot of South Africa. *Systematic Biology* 60: 343–357.
- Sloan S, Jenkins CN, Joppa LN, Gaveau DLA, Laurance WF. 2014. Remaining natural vegetation in the global biodiversity hotspots. *Biological Conservation* 177: 12–24.
- Spellenberg R, Bacon JR, González Elizondo MS. 1998. Los encinos (*Quercus*, Fagaceae) en un transecto sobre la Sierra Madre Occidental. *Boletín de la Sociedad Botánica de México* 5: 357–387.
- Stamatakis A. 2014. RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.
- Swenson NG, Enquist BJ, Pither J, Thompson J, Zimmerman JK. 2006. The problem and promise of scale dependency in community phylogenetics. *Ecology* 87: 2418–2424.
- Swenson NG, Enquist BJ, Thompson J, Zimmerman JK. 2007. The influence of spatial and size scale on phylogenetic relatedness in tropical forest communities. *Ecology* 88: 1770–1780.
- Thornthwaite CW. 1948. An approach toward a rational classification of climate. *Geographical Review* 38: 55–94.
- Tibshirani R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B (Methodological)* 58: 267–288.
- Tibshirani R. 2011. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73: 273–282.
- Torres-Miranda A, Luna-Vega I, Oyama K. 2011. Conservation biogeography of red oaks (*Quercus*, section *Lobatae*) in Mexico and Central America. *American Journal of Botany* 98: 290–305.
- Torres-Miranda A, Luna-Vega I, Oyama K. 2013. New approaches to the biogeography and areas of endemism of red oaks (*Quercus* L., Section *Lobatae*). *Systematic Biology* 62: 555–573.
- Uyeda JC, Caetano DS, Pennell MW. 2015. Comparative analysis of principal components can be misleading. *Systematic Biology* 64: 677–689.
- Valencia AS. 2004. Diversidad del género *Quercus* (Fagaceae) en México. *Boletín de la Sociedad Botánica de México* 75: 33–53.
- Verdú M. 2002. Age at maturity and diversification in woody angiosperms. *Evolution* 56: 1352–1361.
- Wiens JJ, Donoghue MJ. 2004. Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution* 19: 639–644.

## Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

**Fig. S1** Infomap Bioregions map of biogeographic regions.

**Fig. S2** Maximum likelihood phylogeny of all tips sampled, with untransformed branch lengths.

**Fig. S3** Analysis of lineage diversification rates using rjMCMC in BAMM.

**Fig. S4** Plot of nonmetric multidimensional scaling stress against ordination dimensions ( $K$ ).

**Fig. S5** Surface plots, raw ordination data plotted over NMDS plot.

**Fig. S6** NMDS plots for soils and climate separately.

**Fig. S7** Ornstein–Uhlenbeck adaptive transitions on the  $K=5$  NMDS ordination axes.

**Fig. S8** Model-averaged rate of evolution along moisture gradient ( $I_m$ ), reconstructed using reversible-jump Markov chain Monte Carlo (rjMCMC).

**Fig. S9** Reconstruction of leaf habit evolution under stochastic mapping and maximum likelihood.

**Fig. S10** Leaf phenology correlates with selected climatic variables.

**Table S1** Samples utilized in molecular analysis

**Table S2** Specimen records included in this study, including specimens excluded during data clean-up, and BIOCLIM data

**Table S3** Species list, with sectional classification, biogeographic coding, and leaf phenology coding

**Table S4** Divergence time estimates with two fossil constraints

**Table S5** AIC weights for phylogenetic binary regressions evaluated

**Table S6** Coefficients for full phylogenetic binary logistic regression model with deciduous and brevideciduous states binned together

**Methods S1** Supplemental methods.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.



## About *New Phytologist*

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Trust, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <26 days. There are **no page or colour charges** and a PDF version will be provided for each article.
- The journal is available online at Wiley Online Library. Visit **www.newphytologist.com** to search the articles and register for table of contents email alerts.
- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@lancaster.ac.uk)
- For submission instructions, subscription and all the latest information visit **www.newphytologist.com**