Journal of Biogeography (J. Biogeogr.) (2015) 42, 328–338

ORIGINAL ARTICLE

Climatically stable landscapes predict patterns of genetic structure and admixture in the Californian canyon live oak

Joaquín Ortego[1,2]*, Paul F. Gugger[3] and Victoria L. Sork[3,4]

[1]Conservation and Evolutionary Genetics Group, Department of Integrative Ecology, Estación Biológica de Doñana, EBD-CSIC, E-41092 Seville, Spain, [2]Grupo de Investigación de la Biodiversidad Genética y Cultural, Instituto de Investigación en Recursos Cinegéticos – IREC (CSIC, UCLM, JCCM), Ciudad Real, E-13005, Spain, [3]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA, [4]Institute of the Environment and Sustainability, University of California, Los Angeles, CA 90095-1496, USA

*Correspondence: Joaquín Ortego, Conservation and Evolutionary Genetics Group, Department of Integrative Ecology, Estación Biológica de Doñana, EBD-CSIC, Avda. Américo Vespucio s/n, E-41092 Seville, Spain.
E-mail: joaquin.ortego@csic.es

ABSTRACT

Aim We studied which factors shape contemporary patterns of genetic structure, diversity and admixture in the canyon live oak (*Quercus chrysolepis*). Specifically, we tested two alternative hypotheses: (1) that areas with high habitat suitability and stability since the Last Glacial Maximum (LGM) sustain higher effective population sizes, resulting in increased levels of genetic diversity; and (2) that populations from areas with lower habitat stability show higher levels of genetic admixture due to their recurrent colonization by individuals originating from genetically differentiated populations. Furthermore, we analysed the relative importance of past and current habitat suitability and their additive effects on contemporary patterns of genetic structure.

Location California, USA.

Methods We sampled 160 individuals from 33 localities across the distribution range of the canyon live oak in California and then combined information from 13 nuclear microsatellite DNA markers and climate niche modelling to study patterns of genetic variation in this species. We used Bayesian clustering analyses to analyse geographical patterns of genetic structure and admixture, and circuit theory to generate isolation-by-resistance (IBR) distance matrices.

Results We found that the degree of genetic admixture was higher in localities with lower inferred population stability, but that genetic diversity was not associated with habitat suitability or stability. Landscape genetic analyses identified habitat stability as the primary driver of population genetic differentiation.

Main conclusions This study shows that habitat stability can be a major factor shaping genetic variation in wind-pollinated trees and supports the idea that stable regions contribute to genetic connectivity across different climatic periods. To our knowledge, this study is the first to report an association between patterns of genetic admixture and stability of local habitat.

Keywords

California, climatic stability, ecological niche modelling, gene flow, genetic diversity, genetic structure, interglacial refugia, palaeodistribution modelling, *Quercus chrysolepis*.

INTRODUCTION

Understanding the impact of past climate changes on species distribution and geographical patterns of intraspecific genetic variation can provide valuable insight into the demographic and evolutionary trajectories of a species and its responses to future environmental changes (Dudaniec *et al.*, 2012; Devitt *et al.*, 2013; He *et al.*, 2013; Fant *et al.*, 2014; Yannic *et al.*, 2014). Realized gene flow can be affected by many extrinsic factors, including geographical distance, suitability of intervening habitats, and physical barriers to dispersal, as well as local patterns of selection on immigration events (e.g. Shafer & Wolf, 2013; Wang *et al.*, 2013). With the general exception of geographical distance, all the factors mentioned above are dynamic through time. For this reason, the demographic and distributional changes mediated by both past and current

http://wileyonlinelibrary.com/journal/jbi
doi:10.1111/jbi.12419

environmental conditions can leave signatures on contemporary patterns of genetic variation and structure (Dudaniec *et al.*, 2012; Ortego *et al.*, 2012; Poelchau & Hamrick, 2012; Gugger *et al.*, 2013; Fant *et al.*, 2014).

The relative impact of past and current environments on observed patterns of genetic structure and diversity are likely to be closely dependent on effective population sizes and the life-history traits of the organism, such as dispersal capacity, generation time and longevity (Wright, 1943; Crow & Aoki, 1984; Rousset, 1997; Sork *et al.*, 2010; Gugger *et al.*, 2013). Long generation times and large effective population sizes can buffer the impact of contemporary gene flow on genetic structure and result in time lags (Wright, 1943). In comparison to species with limited dispersal capacities, organisms with a high dispersal potential are likely to show faster responses to environmental changes because they are able to track and colonize adequate habitats more rapidly as they become available, and reach an earlier equilibrium between local genetic drift and inter-population gene flow across suitable landscapes (Devitt *et al.*, 2013).

The last Pleistocene glaciation (*c.* 110,000 to 12,000 years ago) and the subsequent Holocene climate warming are probably the last major environmental changes experienced by most organisms in temperate zones (Hewitt, 2000). During glacial periods, species distributions and patterns of gene flow often greatly differed from current dispersal routes (Devitt *et al.*, 2013; He *et al.*, 2013). Regions that have remained climatically suitable and stable throughout the Quaternary have been found to contribute to maintain population connectivity and gene flow over extended periods of time in different organisms (Bell *et al.*, 2010; Devitt *et al.*, 2013; He *et al.*, 2013). Stable areas are also likely to act as refugia and sustain populations that harbour higher levels of genetic diversity (Carnaval & Moritz, 2008; Carnaval *et al.*, 2009; Sork & Werth, 2014; Yannic *et al.*, 2014). However, habitat instability can also contribute to boost local genetic diversity if colonization of empty or low density habitat patches results in increased genetic admixture due to the arrival of immigrants originating from genetically differentiated populations (Fauvelot *et al.*, 2006; Reusch, 2006). This idea is similar to the 'intermediate disturbance hypothesis', which states that species diversity is maximized in areas with moderate levels of ecological disturbance (Connell, 1978). Thus, considering the temporal scale at which the processes shaping contemporary patterns of genetic structure have taken place is essential to better understand species evolutionary and demographic history (He *et al.*, 2013; Yannic *et al.*, 2014). However, not many studies have simultaneously analysed the relative role of past and current climatic conditions on observed patterns of genetic structure (Gugger *et al.*, 2013; He *et al.*, 2013; Lawson, 2013). Currently available tools for integrating palaeodistribution modelling and genetic data allow the study of population connectivity across alternative landscapes and the formulation of hypotheses that can be explicitly tested across taxa with different responses to climate change and landscape features (Poelchau & Hamrick, 2012; Gugger *et al.*, 2013).

In this study, we analysed geographical patterns of genetic variation in the canyon live oak (*Quercus chrysolepis* Liebm.) across most of the species' distribution range in California, USA, an area that has experienced warming and cooling periods during the Pleistocene but no extensive glaciation. The California Floristic Province is characterized by deep topographic gradients and complex but regionally stable climates – factors that have been identified to be responsible for the complicated patterns of genetic structure often observed in this biodiversity hotspot (Raven & Axelrod, 1978; Calsbeek *et al.*, 2003; Davis *et al.*, 2008; Lancaster & Kay, 2013). Thus, this region offers an ideal scenario with which to study the role of local environmental changes and coupled species distributional shifts on patterns of gene flow and diversity (Gugger *et al.*, 2013).

Canyon live oak is one of the most abundant and widespread oaks in California and is believed to sustain very high local effective population sizes, N_e (Burns & Honkala, 1990). This N_e, together with the long generation time of oaks (*c.* 100 years; Gugger *et al.*, 2013) and local colonization via seed dispersal (Grivet *et al.*, 2006), suggests that genetic structure in canyon live oak should respond slowly to climate-mediated changes of gene flow. Conversely, high potential for long-distance dispersal via pollen movement (Buschbom *et al.*, 2011; Ortego *et al.*, 2014) would increase genetic connectivity (Sork *et al.*, 2010) and contribute to quickly erode the genetic signal left by past environmental conditions. Thus, the canyon live oak is an interesting case study to analyse the prevailing time frame at which environmental factors contribute to explain contemporary patterns of genetic variation and the complex landscapes characterizing the California region offer an excellent abiotic template.

Here, we combine information from nuclear microsatellite DNA markers and climatic niche modelling to study the factors associated with contemporary patterns of genetic variation in the canyon live oak. We first analysed geographical patterns of genetic diversity and tested whether areas with high past and present habitat suitability and stability sustain increased levels of local genetic variation (e.g. Carnaval *et al.*, 2009; Gugger *et al.*, 2013; Yannic *et al.*, 2014). Second, we tested the alternative hypothesis that stands from areas with lower habitat stability show higher levels of population genetic admixture due to frequent population turnover and recurrent colonization by individuals from genetically differentiated populations (Fauvelot *et al.*, 2006; Reusch, 2006; Ortego *et al.*, 2012). Finally, we used circuit theory to generate different isolation-by-resistance scenarios (McRae, 2006; McRae & Beier, 2007; McRae *et al.*, 2008) and test the relative importance of geographical distance and past and current habitat suitability and their additive effects (i.e. habitat suitability stability) on contemporary patterns of genetic structure.

MATERIALS AND METHODS

Study species and sampling

Quercus chrysolepis is a diploid, wind-pollinated and monoecious tree species. It is mostly distributed in California, with some relict populations in New Mexico, Arizona, Nevada, southern Oregon (USA) and northern Baja California (Mexico) (Burns & Honkala, 1990; eFloras, 2014). This species is generally found in mountain ridges, canyons, and steep and moist slopes, at elevations ranging between 200 and 2600 m above sea level. The study area covers most of the distribution range of the canyon live oak in California (Fig. 1). A few scattered individuals of this species have been recorded in the past in Arizona and New Mexico (eFloras, 2014; http://ag.arizona.edu/herbarium/), but we were unable to find stands to sample. During 2011–2012, we sampled leaves from 160 adult trees from 33 localities (Fig. 1, and see Appendix S1 in Supporting Information). We define 'locality' as a canyon live oak stand where we sampled adult trees separated by less than 200 m. Spatial coordinates of each locality were registered using a Global Positioning System (GPS; Garmin GPSMap64, Garmin Ltd, Olathe, KS, USA) and leaf samples were stored frozen (−20 °C) until needed for genetic analyses.

Ecological niche modelling

We used ecological niche modelling (ENM) to predict the geographical distribution of climatically suitable areas for canyon live oak and to test whether habitat stability and current and past climatic conditions are responsible for observed patterns of genetic diversity and structure. We modelled the current climate-based distribution of canyon live oak using a maximum entropy algorithm, MAXENT 3.3.3 (Phillips *et al.*, 2006; Phillips & Dudík, 2008). Species occurrence data were obtained from sampling points as well as from herbarium record databases (Consortium of California Herbaria, http://ucjeps.berkeley.edu/consortium/; Consortium of Pacific Northwest Herbaria, http://www.pnwherbaria.org/; University of Arizona Herbarium, http://ag.arizona.edu/herbarium/; and Global Biodiversity Information Facility, http://www.gbif.org/). Prior to modelling, all herbarium records were mapped and examined to identify and exclude records having obvious georeferencing errors and misidentifications. For models, all locations falling within the same grid cell were also removed, resulting in a final data set of 1432 entries. To construct the models, we used 19 bioclimatic variables from the WorldClim dataset interpolated to 30-arcsec (*c.* 1-km) resolution (Hijmans *et al.*, 2005). ENMs generated with the entire set of 19 bioclimatic variables or excluding those highly correlated resulted in qualitatively similar ENMs (data not shown) and all analyses presented here are based on the entire set. Model evaluation statistics were produced from 10 cross-validation replicate model runs. Overall model performance was evaluated using the area under the receiver operating characteristic curve (AUC), which ranges from 0.5 (random prediction) to 1 (maximum prediction).

We obtained the predicted distribution of canyon live oak at the Last Glacial Maximum (LGM; *c.* 21,000 years ago) by projecting contemporary species–climate relationships to the LGM. We used the same 19 climate layers from the Community Climate System Model derived from the PMIP2 database and available at WorldClim (CCSM3, http://www2.cesm.ucar.edu/; Kiehl & Gent, 2004; Otto-Bliesner *et al.*, 2006; Collins *et al.*, 2006). These layers were downloaded at 2.5-arcmin resolution and resampled to 30-arcsec resolution via bilinear interpolation. Current and LGM habitat suitability

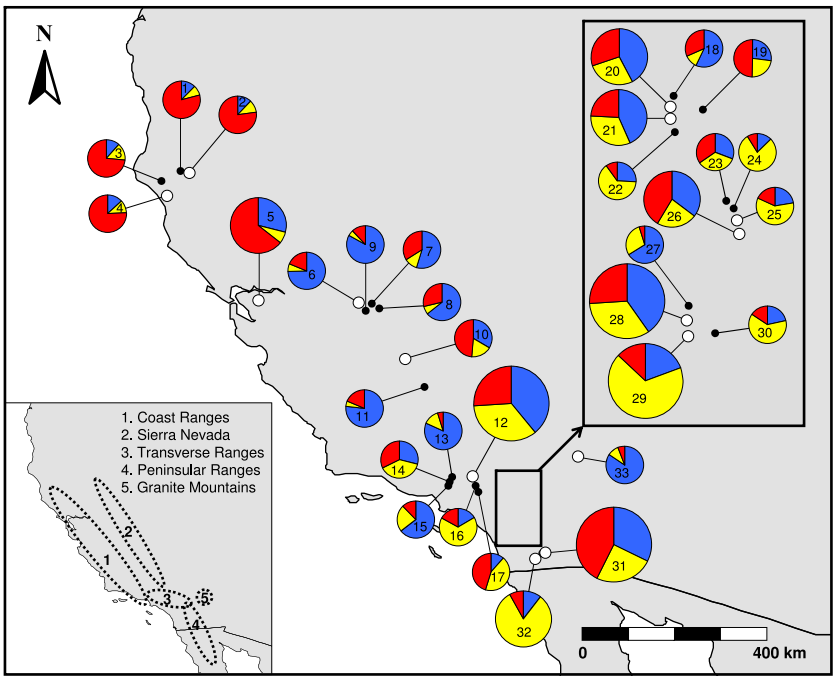

Figure 1 Sampling sites of canyon live oak (*Quercus chrysolepis*) in California and genetic assignment of populations based on the Bayesian method implemented in the program STRUCTURE considering three genetic clusters (n = 160 individuals from 33 localities). The admixture proportions generated by STRUCTURE were represented using pie charts, with each colour indicating a different genotypic cluster. Open dots indicate localities with five or more sampled individuals and numbers correspond to population codes described in Appendix S1. Inset map shows the geographical location of the main mountain ranges in California.

maps based on the logistic output of MAXENT were added to generate maps of habitat suitability stability (*sensu* Devitt *et al.*, 2013; Yannic *et al.*, 2014), with pixel values ranging from 0 (minimum climatic suitability in both periods) to 2 (maximum climatic suitability in both periods). Visualization of model predictions and all GIS calculations and analyses were performed in ARCMAP 10.0 (ESRI, Redlands, CA, USA).

Microsatellite genotyping

We used 13 polymorphic microsatellite markers previously developed for other *Quercus* species to genotype canyon live oaks (see Appendix S2). DNA extraction and microsatellite amplification and genotyping were performed as described in Ortego *et al.* (2014). In each locality, we tested for linkage equilibrium between each pair of loci, departure from Hardy–Weinberg equilibrium (HWE), and the presence of null alleles as described in Ortego *et al.* (2014).

Genetic structure

We investigated population genetic structure by calculating pairwise F_{ST}-values between sampling locations with five or more genotyped individuals and tested significance with Fisher's exact tests after 10,000 permutations as implemented in ARLEQUIN 3.1 (Excoffier *et al.*, 2005). Critical *P*-values for pairwise tests of allelic differentiation were determined using a sequential Bonferroni adjustment. Calculation of F_{ST}-values based on 100 replicates of a random sampling of five individuals (10 alleles per locus) in populations with more than 10 individuals ($n = 7$; Appendix S1) provided comparable and highly correlated F_{ST} estimates across the different replicates (mean Mantel $r \pm SD = 0.742 \pm 0.087$, all *P*-values < 0.001). This indicates that a sample size equal to five individuals provides a reliable estimate of pairwise F_{ST}-values with our set of markers.

We analysed the spatial genetic structure across all sampled individuals using the Bayesian Markov chain Monte Carlo (MCMC) clustering analysis implemented in the program STRUCTURE 2.3.3 (Pritchard *et al.*, 2000; Falush *et al.*, 2003; Hubisz *et al.*, 2009). We ran STRUCTURE assuming correlated allele frequencies and admixture and conducted 10 independent runs for each value of $K = 1$–10 to estimate the number of clusters with 200,000 MCMC cycles, following a burn-in steps of 100,000 iterations. The number of populations best fitting the data set was defined both using log probabilities [Pr $(X|K)$] (Pritchard *et al.*, 2000) and the ΔK method (Evanno *et al.*, 2005). We used CLUMPP to align multiple runs of STRUCTURE for the optimum *K*-value using the Greedy algorithm (Jakobsson & Rosenberg, 2007).

Genetic diversity and admixture

At the population level, we calculated allelic richness (A_R) for sites with at least five genotyped individuals. A_R was standardized for sample size ($n = 5$ individuals; Appendix S1) using HP-RARE (Kalinowski, 2005). At the individual level, we used two metrics to estimate genetic diversity: (1) uncorrected heterozygosity (H_O); and (2) homozygosity by loci (*HL*) (Aparicio *et al.*, 2006). H_O and *HL* were calculated using CERNICALIN, an EXCEL spreadsheet available on request.

We analysed the association between genetic diversity (A_R, H_O and *HL*) and habitat suitability, habitat suitability stability and the geographical location of populations. In particular, we considered five explanatory covariates in the models: (1) current habitat suitability ($HS_{CURRENT}$); (2) LGM habitat suitability (HS_{LGM}); (3) habitat suitability stability (HS_{STA}); (4) latitude; and (5) longitude. $HS_{CURRENT}$, HS_{LGM} and HS_{STA} were estimated at three spatial scales, using buffers of 1 km^2, 10 km^2 and 1000 km^2 around sampling localities and calculating average pixel values for the three variables with ARCMAP 10.0. To analyse A_R we used a general linear model (GLM) with a normal error structure and an identity link function. The precision of A_R estimates may differ among populations because of differences in sample sizes and we took this into account using a weighted least square (WLS) method, where weight equals the sample size for each studied population. Individual-based data (H_O and *HL*) were analysed using generalized linear mixed models (GLMM) also with a normal error structure and an identity link function. Locality was included as a random effect to control for the expected non-independence of data due to genetic and environmental spatial autocorrelation.

We estimated the genetic admixture of the studied populations using a 'genetic admixture index' (G_{Admix}), calculated as the standard deviation (SD) of the probabilities of population membership to each genetic cluster inferred by STRUCTURE analyses (see previous section). We normalized SD values to vary between 0 and 1 and subtracted these values from 1 so that G_{Admix} ranges from 0 to 1, with values equal to 0 indicating no admixture (i.e. genetically pure populations assigned to a single genetic cluster) and values equal to 1 indicating maximum admixture (i.e. genetically admixed populations with an equal probability of membership to each inferred genetic cluster). Thus, this summary statistic provides information on within-population genetic admixture that can be directly compared with different population characteristics. We analysed the association between G_{Admix} and HS_{STA} using GLMs and a weighted least square (WLS) method as explained above for population-based genetic diversity analyses. We also considered latitude and longitude as additional covariates to take into account potential geographical patterns of genetic admixture (see Fig. 1). All models were initially constructed with all explanatory terms fitted and final models were selected following a backward procedure, by progressively eliminating non-significant variables. The significance of the remaining variables was tested again until no additional variable reached significance. All statistical analyses were performed using the R 3.0.2 package LME4 (R Core Team, 2012).

Populations were tested for heterozygosity excess in order to detect recent population bottlenecks using the program BOTTLENECK 1.2.02 (Cornuet & Luikart, 1996; Piry *et al.*, 1999). We ran BOTTLENECK under the two-phase model (TPM) and assuming 70–90% of the stepwise mutation model. The variance of mutations was set to 12 (Piry *et al.*, 1999). We used 10,000 replications and statistical significance was assessed using Wilcoxon signed rank tests (Piry *et al.*, 1999). For these analyses we pooled individuals from populations located in the same mountain range (i.e. 1–5, 6–11, 12–22, 23–32 and 33; Fig. 1, Appendix S1) and repeated the analyses excluding populations genetically differentiated with others within these groups to avoid any potential confounding effect of population stratification (Table 1).

Landscape genetic analyses

We applied circuit theory to model gene flow across spatially heterogeneous landscapes and determine the impact of isolation by distance (IBD) and different isolation by resistance (IBR) scenarios on observed patterns of genetic structure (McRae, 2006; e.g. McRae & Beier, 2007; Poelchau & Hamrick, 2012). We used CIRCUITSCAPE 3.5.8 to calculate resistance distance matrices between all pairs of sites considering an eight-neighbour cell connection scheme (McRae, 2006). We used habitat suitability data obtained from ecological niche models (ENM) to generate three IBR scenarios based on: (1) current habitat suitability ($IBR_{Current}$); (2) LGM habitat suitability (IBR_{LGM}); and (3) habitat stability ($IBR_{Stability}$). Cell size used for all raster layers was 30-arcsec (*c.* 1-km). Habitat suitability and habitat stability surfaces were used as conductance grids. To test the effect of IBD we calculated a matrix of Euclidean geographical distances between sampled populations using GEOGRAPHIC DISTANCE MATRIX GENERATOR 1.2.3 (Ersts, 2011). We used two approaches to model gene flow: (1) a population-based approach, considering pairwise F_{ST}-values for the 15 localities with five or more sampled individuals (see previous section; Appendix S1); (2) an individual-based approach, considering genetic relatedness (*r*) between individuals from the 33 studied localities (Appendix S1). Genetic relatedness values were calculated using Lynch & Ritland's (1999) estimator with MARK (K. Ritland: http://www.genetics.forestry.ubc.ca/ritland/programs.html). This index has been proved to be an adequate marker-based estimator of relatedness in natural populations and outperforms other estimators (Lynch & Ritland, 1999). Given that individuals from the same locality cannot be considered independent samples, we randomly selected one individual per sampling site to calculate pairwise relatedness values without pseudo-replication. We repeated this random selection of individuals and calculations 100 times and averaged for each pair of localities the genetic relatedness values obtained across all these random sub-samples of individuals. Geographical distance and IBR matrices were tested against genetic distance matrices using a multiple matrix regression with randomization (MMRR) approach. We used the MMRR function script (Wang, 2013) implemented in R 3.0.2. The final model was selected following a backward procedure as described above for analyses of genetic diversity.

RESULTS

Niche modelling

The predicted distribution of canyon live oak in the present (Fig. 2a) is consistent with its observed current distribution (Burns & Honkala, 1990; eFloras, 2014). The AUC for the

Table 1 Pairwise F_{ST}-values between the studied localities of canyon live oak (*Quercus chrysolepis*) in California. Localities are described in Appendix S1 and grouped according to their geographical location. Values in bold are statistically significant after sequential Bonferroni correction ($P < 0.05$). F_{ST}-values were only calculated for localities with five or more genotyped individuals.

	Coast Ranges			Sierra Nevada		Transverse Ranges			Peninsular Ranges						Granite Mountains
	2	4	5	6	10	12	20	21	25	26	28	29	31	32	33
2	–														
4	0.001	–													
5	0.011	0.025	–												
6	0.065	**0.125**	0.061	–											
10	0.075	0.102	0.064	0.007	–										
12	**0.093**	**0.109**	**0.079**	0.049	0.050	–									
20	**0.092**	**0.095**	**0.072**	0.064	0.031	0.012	–								
21	0.091	0.105	0.076	0.066	0.058	0.004	0.012	–							
25	0.055	0.097	0.086	**0.129**	0.081	0.060	0.022	0.029	–						
26	**0.101**	0.085	**0.075**	**0.113**	**0.086**	0.017	0.010	0.000	0.026	–					
28	0.070	**0.090**	**0.073**	0.048	0.032	**0.049**	0.036	0.055	0.062	**0.075**	–				
29	0.095	0.095	**0.100**	**0.100**	0.041	0.056	0.034	0.080	0.035	0.084	0.025	–			
31	0.082	0.062	0.058	0.090	0.043	0.042	0.012	0.009	0.017	0.010	0.044	0.037	–		
32	**0.173**	**0.163**	**0.183**	**0.165**	**0.089**	**0.091**	**0.060**	0.064	0.075	**0.085**	**0.086**	**0.075**	0.052	–	
33	**0.136**	**0.161**	**0.138**	0.079	**0.121**	**0.129**	**0.104**	**0.143**	**0.171**	**0.162**	**0.107**	**0.119**	**0.132**	**0.204**	–

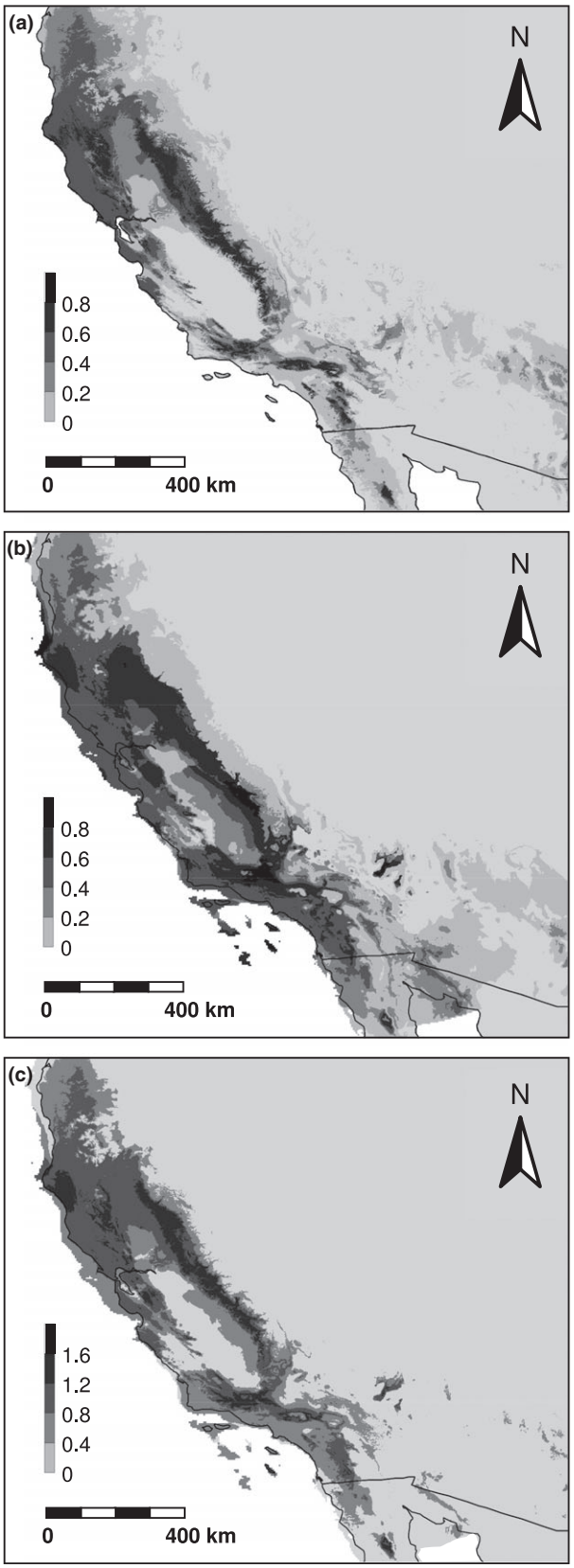

Figure 2 Ecological niche modelling of canyon live oak (*Quercus chrysolepis*) in western North America for (a) the present and (b) the Last Glacial Maximum (LGM; *c.* 21,000 years ago). Niche models are based in 1432 records of the species. The LGM distribution was modelled using the CCSM3 climatic model. Panel (c) shows habitat stability, estimated as the sum of pixel values from current and LGM habitat suitability maps. Greyscales refer to logistic probability (range: 0–1) of species occurrence (panels (a) and (b)) and stability (panel c; range: 0–2), with increasingly darker shades of grey with increasing habitat suitability and stability, respectively. In panels (b) and (c), the lower sea level during the LGM results in some populations that are predicted to have occurred during that period beyond current continental limits.

test data was on average 0.912 (SD = 0.003; n = 10 replicate model runs), indicating a high fit of the modelled and the actually observed current distribution (Fielding & Bell, 1997; Phillips *et al.*, 2006). The estimated potential distribution of canyon live oak during the LGM indicates that this species had a relatively stable distribution range during the last 21,000 years (Fig. 2b,c). However, our models also suggest some important local changes in distribution since the LGM. Populations were more connected during the last glacial period, becoming progressively more isolated and fragmented in the present (Fig. 2a,b). Populations located in the Sierra Nevada and Coast Ranges could have expanded towards the Central Valley and the coast during the LGM and the currently fragmented populations from the southernmost portion of the species' distribution range (southern California and Baja California) appeared to be more connected than in the present (Fig. 2a,b). Populations located at the northern edge of the species' distribution (northern California and southern Oregon) were slightly more fragmented during the LGM than in the present (Fig. 2a,b).

Microsatellite data

All microsatellite markers were highly polymorphic and observed heterozygosity at each locus ranged from 0.50 to 0.93, with 6–53 alleles per locus (Appendix S2). On average ($\pm$ SD), individuals were successfully typed at 12.2 $\pm$ 1.5 microsatellite loci. After applying sequential Bonferroni corrections to compensate for multiple statistical tests, only locus QrZAG20 deviated from HWE in a single population (Laguna Mountain-B; Appendix S1) and MICRO-CHECKER 2.2.3 analyses (van Oosterhout *et al.*, 2004) indicated that null alleles may be present at this locus in this population. We did not find any evidence of genotypic linkage disequilibrium at any pair of loci in any population (exact tests; all P-values > 0.05).

Genetic structure

Pairwise F_{ST}-values ranged from 0.001 to 0.204, and 41 of the 105 pairwise comparisons were significant after sequential Bonferroni correction (Table 1). Comparisons involving populations from Granite and Laguna Mountains showed particularly high differentiation (Table 1), a fact that may be in part explained by the relative isolation of the former and the

peripheral location of the latter (Fig. 1). STRUCTURE analyses and the statistic ΔK indicated an optimal value of $K = 3$ (Appendix S3), but most sampled populations showed a considerable degree of genetic admixture (Fig. 1). The first genetic cluster was the most frequent in the northern and western populations located close to the coast (Fig. 1). The second cluster was more frequently represented in eastern populations along the Sierra Nevada and Granite Mountains (Fig. 1). Finally, populations of the Transverse (San Gabriel and San Bernardino Mountains) and Peninsular Ranges (San Jacinto, Palomar and Laguna Mountains) from southern California showed a particularly high degree of admixture but were generally characterized by a higher probability of membership to the third genetic cluster, which was mostly represented in these populations (Fig. 1).

Genetic diversity and admixture

For those localities with five or more sampled individuals, allelic richness (A_R) standardized for sample size ranged from 2.49 to 3.07 alleles per locus (Appendix S1). A_R or individual-based estimates of genetic diversity (HL, H_O) were not associated with $HS_{CURRENT}$, HS_{LGM}, HS_{STA}, latitude or longitude at any analysed spatial scale (all P-values > 0.15). Population genetic admixture (G_{Admix}) was negatively associated with HS_{STA} at the smallest spatial scale (1 km^2) (estimate ± SE: -0.580 ± 0.167, $t = -3.467$, $P = 0.001$), but was not correlated with latitude ($t = -1.824$, $P = 0.078$) or longitude ($t = 1.403$, $P = 0.171$) (Fig. 3). However, G_{Admix} was not significantly associated with HS_{STA} at the larger spatial scales analysed (10 km^2 and 1000 km^2) (all P-values > 0.4). We found that the degree of population genetic admixture was correlated with genetic diversity ($r = 0.423$, $t = 2.370$, $P = 0.033$), suggesting that admixture bolsters the levels of genetic variation of the studied populations. Quadratic terms and other interactions between independent variables were not significant in any analysis ($P > 0.2$). Wilcoxon signed-rank tests did not reveal a significant excess of heterozygosity in any population (all P-values > 0.9), suggesting that they have not experienced recent and/or strong population bottlenecks.

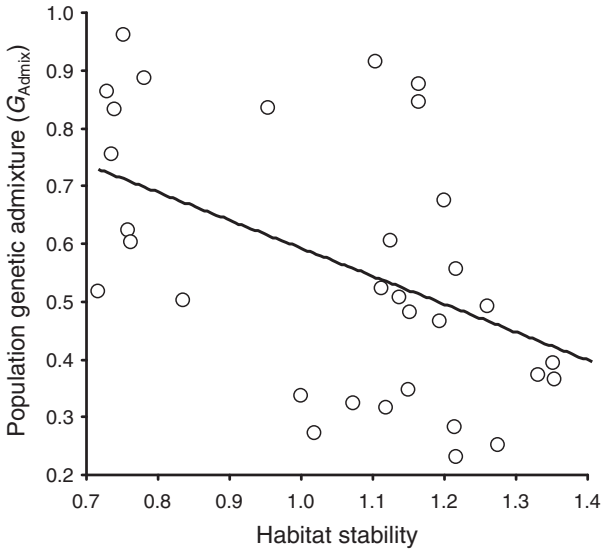

Figure 3 The relationship between genetic admixture (G_{Admix}) and local habitat suitability stability (HS_{STA}) ($r = -0.463$) in the canyon live oak (*Quercus chrysolepis*) in California ($n = 33$ localities). HS_{STA} was estimated in a circular area of 1 km^2 centred on each sampling locality.

Landscape genetic analyses

Population-based analyses showed that genetic distance (F_{ST}) was highly correlated with geographical distance and all the analysed IBR scenarios when these variables were included alone into different models (all P-values < 0.01). However, only habitat stability-based resistance distance was retained in the final model ($r^2 = 0.518$, $\beta = 0.719$, $t = 10.53$, $P = 0.001$; Fig. 4a) and no other variable remained significant after it was included (all P-values > 0.3). Individual-based analyses also showed that average genetic relatedness (r) was highly correlated with all the analysed variables when they were included alone in different models (all P-values < 0.01), but only habitat stability-based resistance distance was retained in the final model ($r^2 = 0.076$, $\beta = -0.286$, $t = -6.60$, $P = 0.001$; Fig. 4b) and no other variable remained significant after it was included (all P-values > 0.1).

DISCUSSION

Niche modelling under current conditions and its projection into the past indicate that the distribution of suitable habitat for canyon live oak has remained relatively stable since the LGM, a pattern similar to that found for other Californian oaks and contrasting with those inferred for species inhabiting geographical regions that have been more severely impacted by Pleistocene glaciations (Fig. 2; e.g. Gugger *et al.*, 2013). This general stability, together with the large population sizes and longevity of this species, can explain the lack of signatures for genetic bottlenecks observed in all the study populations. However, there are some local and regional changes in geographical patterns of habitat suitability and connectivity. The potential distribution of canyon live oak has not experienced major geographical shifts, but overall habitat suitability across the species' range has been reduced since the LGM (Fig. 2a,b). Considering a value for habitat suitability above 0.4 to define a given site as suitable for the species (a threshold based on the 'equal training sensitivity and specificity' value obtained from MAXENT), our species distribution models indicate 44.8% of habitat lost in the transition from the LGM to present, whereas only 38.6% of the territory that was either suitable at the LGM or that is currently suitable has been maintained available for the species in both time periods (i.e. stable over time). Local/regional

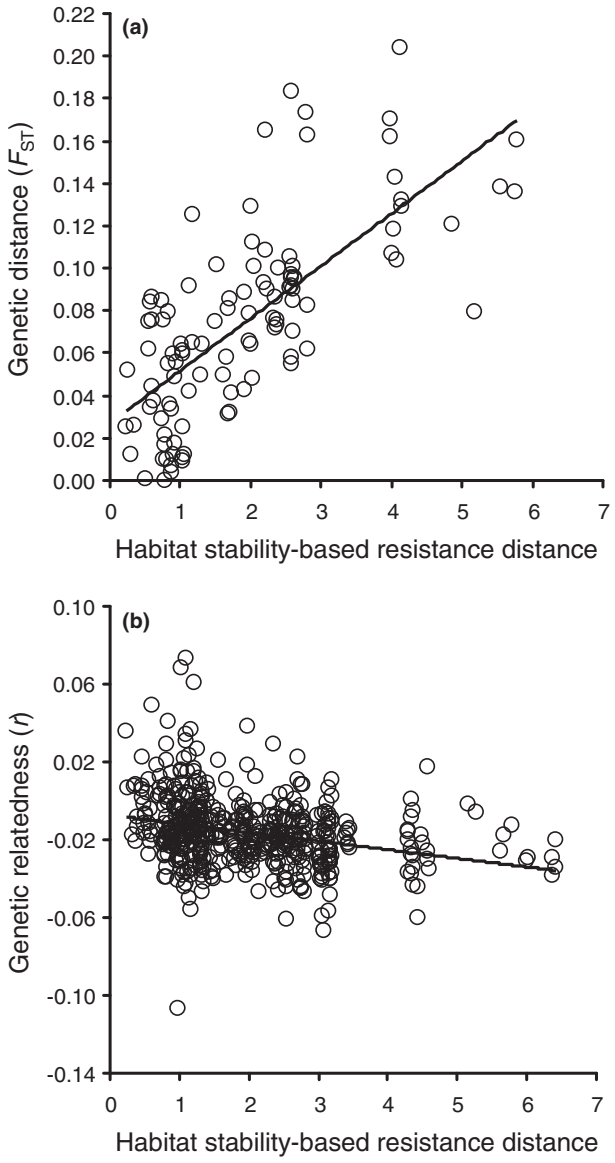

Figure 4 Relationship between (a) pairwise F_{ST} (128 individuals from 15 localities) and (b) genetic relatedness (r) (160 individuals from 33 localities) and resistance distances calculated using CIRCUITSCAPE from habitat suitability stability surfaces in the studied populations of canyon live oaks (*Quercus chrysolepis*) from California.

distribution patterns also seem to have been altered during the last glaciation, including elevational shifts from mountains to valleys, particularly down slope towards the Central Valley, increased fragmentation in the northern edge of the species' distribution and higher connectivity between the Sierra Nevada and Coast Ranges and among the southernmost populations from the Transverse and Peninsular Ranges (Fig. 2a,b). Thus, species distribution models indicate that population connectivity has changed since the LGM, which is likely to have altered patterns of gene flow across the landscape and affected the observed genetic structure in contemporary populations (He *et al.*, 2013).

The spatial pattern of genetic structure and admixture is likely to represent progressive differentiation among major geographical regions (Coast Ranges, Sierra Nevada and Transverse/Peninsular Ranges; Fig. 1) due to increased fragmentation during interglacials (Fig. 2a) and subsequent gene flow and population admixture in contact zones during glacial periods, when the species seems to have mostly experienced range expansions (Fig. 2b). This geographical structure of genetic variation is consistent with previous studies on Californian oaks (Dodd & Kashani, 2003; Grivet *et al.*, 2008; Sork *et al.*, 2010; Gugger *et al.*, 2013) and other organisms distributed in a ring around the Central Valley of California (Moritz *et al.*, 1992; Lapointe & Rissler, 2005; Vandergast *et al.*, 2008). Niche models also indicate considerable patchiness of suitable habitats in the southernmost studied populations (Fig. 2a,b), suggesting that the stronger patterns of genetic structure observed in this region may be mediated by the increased fragmentation experienced by these peripheral populations (Table 1; Hampe & Petit, 2005; e.g. Ortego *et al.*, 2012). This finding also supports previous studies reporting high diversification rates and concentration of isolated lineages in southern California mountain ranges (Davis *et al.*, 2008; Vandergast *et al.*, 2008; Devitt *et al.*, 2013; Sork & Werth, 2014). The high degree of genetic admixture in canyon live oak contrasts with that reported for the co-distributed valley oak (*Quercus lobata*), which shows two geographically well-structured genetic clusters on both sides of the Central Valley (Gugger *et al.*, 2013). These contrasting results indicate that patterns of genetic structure can greatly differ even among co-distributed species with comparable life-history traits (i.e. long-lived, wind pollinated, and with slow rates of seed dispersal) and that would be expected to respond similarly to environmental and geographical barriers to gene flow (Duminil *et al.*, 2007).

We rejected the hypothesis that areas with high habitat suitability and stability since the LGM sustain higher levels of genetic diversity. However, we found support for the alternative hypothesis that populations from less stable areas show higher levels of genetic admixture. The fact that G_{Admix} is positively correlated with population genetic diversity suggests that recurrent population turnover in unstable local habitat patches could increase genetic diversity through the admixture of colonizing individuals originating from genetically differentiated populations. Alternatively, temporally varying selection could also result in higher levels of genetic diversity in climatically unstable patches submitted to changing environmental conditions (e.g. Borash *et al.*, 1998; Ortego *et al.*, 2012). This pattern could be frequent in areas with limited impact from Pleistocene glaciations, where species experience frequent climatic-driven local/regional distribution changes but maintain stable geographical ranges over time (Ortego *et al.*, 2012; Gugger *et al.*, 2013). Although the role of habitat stability on observed patterns of genetic diversity and structure has been addressed in some previous studies (Carnaval *et al.*, 2009; Bell *et al.*, 2010; Ortego *et al.*,

2012; Devitt *et al.*, 2013; Gugger *et al.*, 2013; He *et al.*, 2013; Yannic *et al.*, 2014), admixture is a population genetic trait that has been rarely explored in the context of landscape genetics and, to the best of our knowledge, this is the first study reporting an association between patterns of genetic admixture and habitat stability.

Landscape genetic analyses indicate that dispersal routes defined by stable landscapes seem to be the most relevant factor determining population differentiation (Devitt *et al.*, 2013; He *et al.*, 2013). Our results also suggest that patterns of genetic structure of canyon live oak do not yet reflect ongoing patterns of gene flow mediated by current environmental conditions and indicate a time-lag in the response of this species to changing climatic conditions. This time-lag is likely to reflect a counterbalance between high rates of gene movement mediated by long-distance pollen dispersal, slow colonization of new available habitats via seed dispersal, and the buffering effects of long-generation time and large effective population sizes (Ortego *et al.*, 2012; Gugger *et al.*, 2013).

Overall, our study shows that habitat stability can be also an important factor determining range-wide patterns of genetic structure and variation in a long-lived tree species. Our results, and those provided by previous studies on vertebrates, support the idea that stable regions effectively contribute to maintaining genetic connectivity among populations across different climatic periods (Bell *et al.*, 2010; Devitt *et al.*, 2013; He *et al.*, 2013). This study emphasizes the need to integrate variables that capture dynamic landscape processes, particularly over long time-scales, as they may explain much more variation than static snapshots of current or past environmental conditions (Devitt *et al.*, 2013; He *et al.*, 2013). Future studies integrating resistance surfaces over more time points will also help to improve our understanding of the impact of environmental changes on species distribution, demography and long-term patterns of genetic connectivity (Dudaniec *et al.*, 2013) and contribute to improve predictions about the future responses of species to climate and habitat changes (Carnaval & Moritz, 2008; Carnaval *et al.*, 2009; Yannic *et al.*, 2014).

ACKNOWLEDGEMENTS

For help with field sampling and genetic analysis, we thank Julie Lambert and Conchi Cáliz, respectively. We wish to thank the L. Lacey Knowles Lab for valuable advice about GIS and niche modelling, in particular Marcelo J. Sturaro. The editors and three anonymous referees provided useful discussion and valuable comments on an earlier draft of this manuscript. During this work J.O. was supported by Juan de la Cierva (MICINN), José Castillejo (ME) and Severo Ochoa (EBD) research fellowships. P.F.G. received post-doctoral support from a UCLA research award to V.L.S. This work received financial support from grants CGL2011-25053 (MICINN) and UNCM08-1E-018 (FEDER).

REFERENCES

Aparicio, J.M., Ortego, J. & Cordero, P.J. (2006) What should we weigh to estimate heterozygosity, alleles or loci? *Molecular Ecology*, **15**, 4659–4665.

Bell, R.C., Parra, J.L., Tonione, M., Hoskin, C.J., Mackenzie, J.B., Williams, S.E. & Moritz, C. (2010) Patterns of persistence and isolation indicate resilience to climate change in montane rainforest lizards. *Molecular Ecology*, **19**, 2531–2544.

Borash, D.J., Gibbs, A.G., Joshi, A. & Mueller, L.D. (1998) A genetic polymorphism maintained by natural selection in a temporally varying environment. *The American Naturalist*, **151**, 148–156.

Burns, R.M. & Honkala, B.H. (tech. coords) (1990) *Silvics of North America,* Vol. 2, *Hardwoods*. Agriculture Handbook 654. U.S. Department of Agriculture, Forest Service, Washington, DC.

Buschbom, J., Yanbaev, Y. & Degen, B. (2011) Efficient long-distance gene flow into an isolated relict oak stand. *Journal of Heredity*, **102**, 464–472.

Calsbeek, R., Thompson, J.N. & Richardson, J.E. (2003) Patterns of molecular evolution and diversification in a biodiversity hotspot: the California Floristic Province. *Molecular Ecology*, **12**, 1021–1029.

Carnaval, A.C. & Moritz, C. (2008) Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography*, **35**, 1187–1201.

Carnaval, A.C., Hickerson, M.J., Haddad, C.F.B., Rodrigues, M.T. & Moritz, C. (2009) Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, **323**, 785–789.

Collins, W.D., Bitz, C.M., Blackmon, M.L. *et al.* (2006) The Community Climate System Model version 3 (CCSM3). *Journal of Climate*, **19**, 2122–2143.

Connell, J.H. (1978) Diversity in tropical rain forests and coral reefs. *Science*, **199**, 1302–1310.

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

Crow, J.F. & Aoki, K. (1984) Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. *Proceedings of the National Academy of Sciences USA*, **81**, 6073–6077.

Davis, E.B., Koo, M.S., Conroy, C., Patton, J.L. & Moritz, C. (2008) The California Hotspots Project: identifying regions of rapid diversification of mammals. *Molecular Ecology*, **17**, 120–138.

Devitt, T.J., Devitt, S.E.C., Hollingsworth, B.D., McGuire, J.A. & Moritz, C. (2013) Montane refugia predict population genetic structure in the large-blotched *Ensatina salamander*. *Molecular Ecology*, **22**, 1650–1665.

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

Dudaniec, R.Y., Spear, S.F., Richardson, J.S. & Storfer, A. (2012) Current and historical drivers of landscape genetic structure differ in core and peripheral salamander populations. *PLoS ONE*, **7**, e36769.

Dudaniec, R.Y., Rhodes, J.R., Wilmer, J.W., Lyons, M., Lee, K.E., McAlpine, C.A. & Carrick, F.N. (2013) Using multilevel models to identify drivers of landscape-genetic structure among management areas. *Molecular Ecology*, **22**, 3752–3765.

Duminil, J., Fineschi, S., Hampe, A., Jordano, P., Salvini, D., Vendramin, G.G. & Petit, R.J. (2007) Can population genetic structure be predicted from life-history traits? *The American Naturalist*, **169**, 662–672.

Ersts, P.J. (2011) *Geographic Distance Matrix Generator (version 1.2.3)*. American Museum of Natural History, Center for Biodiversity and Conservation. Available at: http://biodiversityinformatics.amnh.org/open_source/gdmg (downloaded on 18 March 2011).

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

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

Falush, D., Stephens, M. & Pritchard, J.K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Fant, J.B., Havens, K., Keller, J.M., Radosavljevic, A. & Yates, E.D. (2014) The influence of contemporary and historic landscape features on the genetic structure of the sand dune endemic, *Cirsium pitcheri* (Asteraceae). *Heredity*, **112**, 519–530.

Fauvelot, C., Cleary, D.F.R. & Menken, S.B.J. (2006) Short-term impact of disturbance on genetic diversity and structure of Indonesian populations of the butterfly *Drupadia theda* in East Kalimantan. *Molecular Ecology*, **15**, 2069–2081.

Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.

eFloras (2014) *Flora of North America*. Missouri Botanical Garden, St. Louis, Missouri, MO, and Harvard University Herbaria, Cambridge, MA. Available at: http://www.efloras.org/.

Grivet, D., Deguilloux, M.F., Petit, R.J. & Sork, V.L. (2006) Contrasting patterns of historical colonization in white oaks (*Quercus* spp.) in California and Europe. *Molecular Ecology*, **15**, 4085–4093.

Grivet, D., Sork, V.L., Westfall, R.D. & Davis, F.W. (2008) Conserving the evolutionary potential of California valley oak (*Quercus lobata* Née): a multivariate genetic approach to conservation planning. *Molecular Ecology*, **17**, 139–156.

Gugger, P.F., Ikegami, M. & Sork, V.L. (2013) Influence of late Quaternary climate change on present patterns of genetic variation in valley oak, *Quercus lobata* Née. *Molecular Ecology*, **22**, 3598–3612.

Hampe, A. & Petit, R.J. (2005) Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters*, **8**, 461–467.

He, Q.X., Edwards, D.L. & Knowles, L.L. (2013) Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution*, **67**, 3386–3402.

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

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

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

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

Kalinowski, S.T. (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187–189.

Kiehl, J.T. & Gent, P.R. (2004) The Community Climate System Model, version 2. *Journal of Climate*, **17**, 3666–3682.

Lancaster, L.T. & Kay, K.M. (2013) Origin and diversification of the California flora: re-examining classic hypotheses with molecular phylogenies. *Evolution*, **67**, 1041–1054.

Lapointe, F.J. & Rissler, L.J. (2005) Congruence, consensus, and the comparative phylogeography of codistributed species in California. *The American Naturalist*, **166**, 290–299.

Lawson, L.P. (2013) Diversification in a biodiversity hot spot: landscape correlates of phylogeographic patterns in the African spotted reed frog. *Molecular Ecology*, **22**, 1947–1960.

Lynch, M. & Ritland, K. (1999) Estimation of pairwise relatedness with molecular markers. *Genetics*, **152**, 1753–1766.

McRae, B.H. (2006) Isolation by resistance. *Evolution*, **60**, 1551–1561.

McRae, B.H. & Beier, P. (2007) Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences USA*, **104**, 19885–19890.

McRae, B.H., Dickson, B.G., Keitt, T.H. & Shah, V.B. (2008) Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology*, **89**, 2712–2724.

Moritz, C., Schneider, C.J. & Wake, D.B. (1992) Evolutionary relationships within *Ensatina eschscholtzii* complex confirm the ring species interpretation. *Systematic Biology*, **41**, 273–291.

van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M. & Shipley, P. (2004) MICRO-CHECKER: software for identify-

ing and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Ortego, J., Riordan, E.C., Gugger, P.F. & Sork, V.L. (2012) Influence of environmental heterogeneity on genetic diversity and structure in an endemic southern Californian oak. *Molecular Ecology*, **21**, 3210–3223.

Ortego, J., Bonal, R., Muñoz, A. & Aparicio, J.M. (2014) Extensive pollen immigration and no evidence of disrupted mating patterns or reproduction in a highly fragmented holm oak stand. *Journal of Plant Ecology*, **7**, 384–395.

Otto-Bliesner, B.L., Marsha, S.J., Overpeck, J.T., Miller, G.H., Hu, A.X. & Mem, C.L.I.P. (2006) Simulating arctic climate warmth and icefield retreat in the last interglaciation. *Science*, **311**, 1751–1753.

Phillips, S.J. & Dudík, M. (2008) Modelling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Piry, S., Luikart, G. & Cornuet, J.M. (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **90**, 502–503.

Poelchau, M.F. & Hamrick, J.L. (2012) Differential effects of landscape-level environmental features on genetic structure in three codistributed tree species in Central America. *Molecular Ecology*, **21**, 4970–4982.

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

R Core Team (2012) *R: a language and environment for statistical computing*. Version 3.0.2. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.r-project.org.

Raven, P.H. & Axelrod, D.I. (1978) Origin and relationships of the California flora. *University of California Publications in Botany*, **72**, 1–134.

Reusch, T.B.H. (2006) Does disturbance enhance genotypic diversity in clonal organisms? A field test in the marine angiosperm *Zostera marina*. *Molecular Ecology*, **15**, 277–286.

Rousset, F. (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Shafer, A.B.A. & Wolf, J.B.W. (2013) Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters*, **16**, 940–950.

Sork, V.L. & Werth, S. (2014) Phylogeography of *Ramalina menziesii*, a widely distributed lichen-forming fungus in western North America. *Molecular Ecology*, **23**, 2326–2339.

Sork, V.L., Davis, F.W., Westfall, R., Flint, A., Ikegami, M., Wang, H.F. & Grivet, D. (2010) Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change. *Molecular Ecology*, **19**, 3806–3823.

Vandergast, A.G., Bohonak, A.J., Hathaway, S.A., Boys, J. & Fisher, R.N. (2008) Are hotspots of evolutionary potential adequately protected in southern California? *Biological Conservation*, **141**, 1648–1664.

Wang, I.J. (2013) Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution*, **67**, 3403–3411.

Wang, I.J., Glor, R.E. & Losos, J.B. (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters*, **16**, 175–182.

Wright, S. (1943) Isolation by distance. *Genetics*, **28**, 114–138.

Yannic, G., Pellissier, L., Ortego, J. *et al.* (2014) Genetic diversity in caribou linked to past and future climate change. *Nature Climate Change*, **4**, 132–137.

SUPPORTING INFORMATION

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

Appendix S1 Geographical location and genetic variation of the studied populations of canyon live oak (*Quercus chrysolepis*).

Appendix S2 Microsatellite loci used to genotype canyon live oaks (*Quercus chrysolepis*) in California.

Appendix S3 Results of Bayesian clustering analyses in STRUCTURE.

BIOSKETCHES

Joaquín Ortego is currently a postdoctoral fellow at Estación Biológica de Doñana (CSIC) interested in the study of genetic variability in natural populations at different spatio-temporal and phylogenetic scales.

Paul F. Gugger is a postdoctoral scholar interested in the phylogeography, population genetics, genomics and palaeoecology of trees.

Victoria L. Sork studies the ecology, evolution, genomics, and conservation of tree populations, with an emphasis on the ability of California oaks to tolerate climate change.

Editor: Brett Riddle