MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 3210–3223 doi: 10.1111/j.1365-294X.2012.05591.x

Influence of environmental heterogeneity on genetic diversity and structure in an endemic southern Californian oak

JOAQUÍN ORTEGO,*†‡ ERIN C. RIORDAN,* PAUL F. GUGGER* and VICTORIA L. SORK*§

*Department of Ecology and Evolutionary Biology, University of California Los Angeles, Box 957239, Los Angeles, CA 90095-7239, USA, †Grupo de Investigación de la Biodiversidad Genética y Cultural, Instituto de Investigación en Recursos Cinegéticos – IREC (CSIC, UCLM, JCCM), Ronda de Toledo s/n, E-13005 Ciudad Real, Spain, ‡Departamento de Ecología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC), C/José Gutiérrez Abascal 2, Madrid, E-28006, Spain, §Institute of the Environment and Sustainability, University of California Los Angeles, Box 951496, Los Angeles, CA 90095-1496, USA

Abstract

Understanding how specific environmental factors shape gene flow while disentangling their importance relative to the effects of geographical isolation is a major question in evolutionary biology and a specific goal of landscape genetics. Here, we combine information from nuclear microsatellite markers and ecological niche modelling to study the association between climate and spatial genetic structure and variability in Engelmann oak (*Quercus engelmannii*), a wind-pollinated species with high potential for gene flow. We first test whether genetic diversity is associated with climatic niche suitability and stability since the Last Glacial Maximum (LGM). Second, we use causal modelling to analyse the potential influence of climatic factors (current and LGM niche suitability) and altitude in the observed patterns of genetic structure. We found that genetic diversity is negatively associated with local climatic stability since the LGM, which may be due to higher immigration rates in unstable patches during favourable climatic periods and/or temporally varying selection. Analyses of spatial genetic structure revealed the presence of three main genetic clusters, a pattern that is mainly driven by two highly differentiated populations located in the northern edge of the species distribution range. After controlling for geographic distance, causal modelling analyses showed that genetic relatedness decreases with the environmental divergence among sampling sites estimated as altitude and current and LGM niche suitability. Natural selection against nonlocal genotypes and/or asynchrony in reproductive phenology may explain this pattern. Overall, this study suggests that local environmental conditions can shape patterns of genetic structure and variability even in species with high potential for gene flow and relatively small distribution ranges.

Keywords: causal modelling, ecological niche modelling, genetic diversity, genetic structure, landscape genetics, species distribution models

Received 19 May 2011; revision received 13 March 2012; accepted 19 March 2012

Introduction

The study of the mechanisms generating population genetic structure and divergence is a central topic in evolutionary, conservation and landscape genetics research (Manel *et al.* 2003; Storfer *et al.* 2007; Davis *et al.* 2008; Holderegger & Wagner 2008; Sork & Waits 2010). Limited dispersal abilities of most organisms, spatial isolation and geographical barriers to dispersal are considered major drivers of population divergence and speciation across evolutionary scales (Avise 2000; Swenson & Howard 2005; e.g. Pastorini *et al.* 2003; Pyron & Burbrink 2009; Jenkins *et al.* 2010). However, other

Correspondence: Joaquín Ortego, Fax: +34 926 295 451;
E-mail: joaquin.ortego@csic.es

environmental factors are often responsible for the patterns of genetic structure observed at small spatial scales (e.g. Smith *et al.* 1997; Sacks *et al.* 2004) and can be observed even in species with high potential for gene flow (e.g. wind-pollinated plants: Sork *et al.* 2010; Freeland *et al.* 2010; large mammals: Sacks *et al.* 2004; Pease *et al.* 2009; birds: Smith *et al.* 1997). Populations experiencing heterogeneous environmental conditions can show reduced gene flow as a result of natural selection against non-locally adapted genotypes (Endler 1986; Garant *et al.* 2005; Postma & Van Noordwijk 2005) and/or decreased effective dispersal rates as a result of phenological differences among populations (Hendry & Day 2005; Dyer *et al.* 2010; Freeland *et al.* 2010). Environmental factors can thus ultimately shape differences between initial gene movement and the realized gene flow that we observe (e.g. Sork *et al.* 2010). Accordingly, some studies have reported genetic clines across environmental gradients or even discrete genetic breaks between close populations experiencing contrasting ecological conditions (Smith *et al.* 1997; Sacks *et al.* 2004; Freedman *et al.* 2010; Freeland *et al.* 2010; Thomassen *et al.* 2010). For these reasons, understanding the role of specific environmental factors in determining gene flow and disentangling their importance relative to the effects of geographical isolation *per se* is a key evolutionary question that is receiving much attention in recent years (e.g. Pease *et al.* 2009; Freedman *et al.* 2010; Manel *et al.* 2010; Sork *et al.* 2010).

Detailed information on the species' ecological niche can identify environmental gradients potentially associated with processes of local adaptation and genetic divergence (Pease *et al.* 2009; Pyron & Burbrink 2009). Populations located in areas with low habitat suitability are expected to present smaller population sizes and to be more isolated than those populations located in the core of the species niche envelope (García-Ramos & Kirkpatrick 1997; Hampe & Petit 2005; Alleaume-Benharira *et al.* 2006). Different environments experienced by populations at different altitudes can also result in contrasting selective regimes or divergent periods of reproduction even when populations are not widely separated by distance (Schuster *et al.* 1989). Consequently, differences in niche suitability and altitude among populations can be important factors determining gene flow and the potential evolution of local adaptations (Schuster *et al.* 1989; Hampe & Petit 2005). According to the centre–periphery model, populations located in the biogreographic boundaries of species distribution ranges are also expected to become genetically impoverished as a result of small effective population sizes and low population stability (Lammi *et al.* 1999; Shafer *et al.* 2011). This impoverishment can decrease the ability of these 'ecologically peripheral' populations to respond to selection and adapt to novel environmental conditions (Frankham 1996, 2005; Spielman *et al.* 2004; Willi *et al.* 2006). Paradoxically, population isolation and restricted gene flow can preserve genetic variation from the homogenizing effects of gene flow, favouring the maintenance of local adaptations that can potentially increase the chance of population persistence and promote intraspecific diversity (García-Ramos & Kirkpatrick 1997; Pérez-Tris *et al.* 2000; Hampe & Petit 2005). Thus, identifying the factors determining population genetic structure provides essential information for proper management practices aimed to preserve the genetic identity of populations while minimizing the risks associated with reduced genetic variability (Hedrick 2001).

Past environmental conditions also probably have an important influence on the current distribution of genetic variability (Schneider *et al.* 1998; Walker & Avise 1998; Anderson *et al.* 2010). Genetic structure is expected to have a considerable time lag in its response to changes in gene flow (Wright 1943; Nei & Chakravarti 1977; Waples 1998), and accordingly, current patterns of genetic structure have been often explained by environmental conditions in the past (Anderson *et al.* 2010; e.g. Schneider *et al.* 1998; Walker & Avise 1998; Holzhauer *et al.* 2009; Coulon *et al.* 2010). This time lag is probably particularly remarkable for organisms with large effective population sizes and long generation times in which the genetic signal of past patterns of gene flow can persist over extended time periods (Anderson *et al.* 2010; Ortego *et al.* 2010a). Data on past climatic conditions can also provide important information on long-term environmental fluctuation: climatically stable areas are expected to favour diversification processes and sustain genetically more diverse populations in comparison with regions that have experienced greater climatic shifts (Carnaval *et al.* 2009; Rodríguez-Robles *et al.* 2010). Thus, the temporal scale over which the climatic factors determine gene flow and observed patterns of genetic diversity and structure must be considered to get a comprehensive understanding of these processes (Anderson *et al.* 2010).

Here, we study the climatic factors associated with genetic diversity and structure in Engelmann oak (*Quercus engelmannii*), an endemic Californian oak whose entire distribution range consists of southern California (USA) and northern Baja California (Mexico) (Roberts 1995). Like many other species from the California biodiversity hotspot (Myers *et al.* 2000), Engelmann oak is currently classified in the IUCN red list of threatened species because of its small size range and extensive habitat destruction (IUCN 2010). The pronounced ecological gradients and complex climatic and geological history of California have driven an extraordinary

biological diversification and are responsible for one of the most geographically complicated patterns of genetic diversity on Earth (Raven & Axelrod 1978; Calsbeek *et al.* 2003; Davis *et al.* 2008). As a result, the California biota offers an excellent opportunity to study the contribution of climatic heterogeneity to observed patterns of genetic diversity and structure and, at the same time, such research can also provide relevant information to protect the exceptional biodiversity and ongoing evolutionary processes of this region (Raven & Axelrod 1978; Calsbeek *et al.* 2003; Davis *et al.* 2008).

We combine information from nuclear microsatellite markers and ecological niche modelling (ENM) to study the factors associated with the spatial genetic structure and variability in Engelmann oak. To evaluate the relative importance of current and past climatic conditions in observed patterns of genetic variation, we identified the relationship between the current species distribution and climatic predictors and projected the present-day niche envelope to the Last Glacial Maximum (LGM; ca. 21 000 years BP). In particular, we first test the following predictions about patterns of genetic diversity in Engelmann oak: (i) if populations located in ecologically substandard environments have become genetically impoverished as a result of the effects of genetic drift, we expect that the genetic variability should increase with current and LGM habitat suitability and decrease with population genetic isolation; (ii) if areas with changing climates only sustain fluctuating/unstable populations, we predict that within-population genetic variability should be positively associated with climatic stability since the LGM. Second, we study the genetic structure of Engelmann oak and employ causal modelling to identify the most influential variables associated with gene flow within a multiple hypothesis-testing framework (Cushman *et al.* 2006; Cushman & Landguth 2010). According to the centre–periphery model, we predict (iii) increased genetic differentiation in populations along the margins of the species distribution range. We also expect (iv) that gene flow decreases with geographical distance (i.e. isolation-by-distance) if populations are in migration-drift equilibrium; (v) finally, if environmental heterogeneity is an important factor shaping current patterns of genetic structure, we predict that gene flow decreases between populations experiencing contrasting climatic conditions estimated as differences in altitude and current and LGM niche suitabilities.

Methods

Study species and sampling

Engelmann oak, *Quercus engelmanii* Greene (Fagaceae), is a diploid, wind-pollinated and monoecious tree species. Like other oak species, *Quercus engelmannii* should have both the potential for long distant gene flow through pollen movement (e.g. Pluess *et al.* 2009), and the opportunity for local genetic structure because of a predominance of local pollen and seed dispersal (e.g. Sork *et al.* 2002; Grivet *et al.* 2005, 2009). Engelmann oak is an endemic Californian oak whose entire distribution range consists of southern California (USA) and northern Baja California (Mexico), the smallest range for any native Californian oak (Scott 1991; Roberts 1995). The core of the species distribution range is located in San Diego County (USA) where the species is distributed in scattered patches (Fig. 1) (Scott 1991). There are two northern disjunct populations, one located between Hemet and Temecula cities (Riverside County, USA) and another in the southern base of the San Gabriel Mountains around Pasadena city (Los Angeles County, USA) (Roberts 1995) (Fig. 1; Table 1). A third disjunct population is located south of Ensenada city (northern Baja California, Mexico) in the southernmost distribution range of Engelmann oak (Roberts 1995). This species is generally found on mesas mostly between elevations of 700 m and 1250 m above sea level (Scott 1991). It seems to be closely related to the parapatric Arizona oak (*Quercus arizonica*), gray oak (*Quercus grisea*), Sonoran blue oak (*Quercus oblongifolia*) and Sonoran scrub oak (*Quercus turbinella*) found in southwestern USA and northern Mexico, but the exact phylogenetic relationship has not been yet resolved (Pearse & Hipp 2009; eFloras 2011). Engelmann oak is known to hybridize with species of the sympatric scrub oak complex (*Quercus cornelius-mulleri*, *Quercus berberidifolia* and *Quercus durata*) (Roberts 1995; eFloras 2011) and hybrids show intermediate leaf traits and tend to be more shrubby than tree-like (J. Ortego, C. Smith, and V. L. Sork, unpublished data).

We selected 20 localities that cover the entire distribution range of Engelmann oak, with the exception of a disjunct population located in northern Baja California (Mexico) where we were unable to find extant populations to sample (Fig. 1). During 2008–2011, we sampled 165 adult trees from these localities (Table 1; Fig. 1). Our sample goal was to collect at least 10 individuals per site, but in several localities only a few trees were available and/or accessible. A similar sampling scheme with equal or lower sample sizes per site has been previously successfully employed to study genetic diversity and climate–gene flow associations in another Californian oak (Grivet *et al.* 2008; Sork *et al.* 2010). Spatial coordinates were registered using a Global Positioning System (GPS) and leaf samples were stored frozen (−20 °C) until needed for genetic analyses.

To reduce the confounding effects of interspecific gene flow on our analyses, we excluded individuals

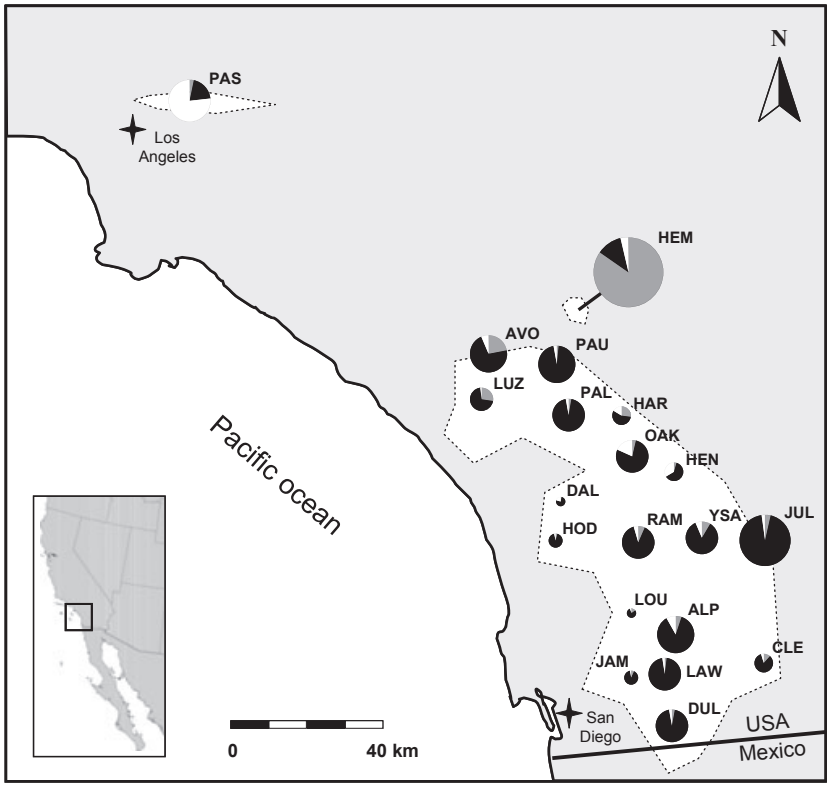

Fig. 1 Sampling sites of Engelmann oak (*Quercus engelmannii*) and genetic assignment of populations based on the Bayesian method implemented in the program STRUCTURE considering three genetic clusters. White areas represent the species distribution range. The admixture proportions generated by STRUCTURE were represented using pie charts, with each colour indicating a different genotypic cluster. Circle size is proportional to the number of individuals sampled at each location. Population codes are described in Table 1.

that showed evidence of introgression from the scrub oak complex (*Q. berberidifolia*, *Q. cornelius-mulleri* and *Q. durata*) according to STRUCTURE 2.3.3 analyses (Pritchard *et al.* 2000; Falush *et al.* 2003) that included samples of this species complex (J. Ortego & V. L. Sork, unpublished data) (see Winkler *et al.* 2011 for a similar approach). We ran STRUCTURE as described in ''Genetic structure'' section below but not considering prior population information (Hubisz *et al.* 2009). We found two main genetic clusters corresponding to Engelmann oak and the scrub oak complex using the method described in Evanno *et al.* (2005). In STRUCTURE, the posterior probability (q) describes the proportion of an individual genotype originating from each cluster. We used a threshold value (Tq) of 0.90 to classify each individual as purebred or hybrid (Vähä & Primmer 2006; Ortego & Bonal 2010; Winkler *et al.* 2011). Thus, we considered that a value of $q \geq 0.90$ indicates a purebred genotype, and a value of $q < 0.90$ indicates an introgressed genotype (e.g. Ortego & Bonal 2010; Winkler *et al.* 2011). According to this criterion, many of the morphologically Engelmann-like trees collected from some of our localities turned out to be introgressed with scrub oaks. These introgressed individuals were discarded for subsequent analyses, which strongly reduced sample size in some localities with high hybridization rates (e.g. LUZ, YSA, LAW; Table 1). Our final sample size consisted of 128 Engelmann oak individuals, with 2 to 15 adult trees per site (see Table 1).

Ecological niche modelling

We used ENM to predict the geographic distribution of climatically suitable habitats for Engelmann oak and analyze whether current and past bioclimatic conditions are responsible for observed patterns of genetic diversity and structure. We first modelled the current climate-based distribution of Engelmann oak using a maximum entropy algorithm, MAXENT 3.3.3 (Phillips *et al.* 2006; Phillips & Dudik 2008). MAXENT calculates probability distributions based on incomplete information and does not require absence data, making it appropriate for modelling species distributions based on presence-only herbarium records (Elith *et al.* 2006; Phillips *et al.* 2006). Species occurrence data were obtained from sampling points as well as from herbarium record databases (Consortium of California Herbaria, http://ucjeps.berkeley.edu/consortium/; Southwest Environmental Information

Table 1 Geographical location and genetic variability for the studied populations of Engelmann oak (*Quercus engelmannii*)

Locality	Code	Latitude	Longitude	Altitude (m)	*N*	A_R
Pasadena	PAS	34.134079	−118.098914	202	9	5.09
Hemet	HEM	33.628262	−117.012864	619	15	4.37
Avocado Mesa	AVO	33.513735	−117.308909	648	8	4.68
Pauba Ranch	PAU	33.508552	−117.088208	381	8	4.87
De Luz	LUZ	33.423553	−117.321380	100	5	4.56
Palomar Mountains	PAL	33.390607	−117.039250	449	7	4.64
Harolds	HAR	33.302025	−116.892959	1023	4	–
Oak Knoll Ranch	OAK	33.298210	−116.922127	731	7	5.20
Lake Henshaw	HEN	33.276442	−116.855038	737	4	–
Daley Ranch	DAL	33.165990	−117.047007	351	2	–
Santa Ysabel	YSA	33.102790	−116.669374	951	7	5.04
Julian	JUL	33.074770	−116.549129	861	11	4.94
Lake Hodges	HOD	33.074700	−117.118072	100	3	–
Ramona	RAM	33.029917	−116.823050	494	7	4.81
Louis A. Stelzer County Park	LOU	32.881655	−116.901225	500	2	–
Alpine	ALP	32.814090	−116.772395	552	8	4.96
Cleveland National Forest	CLE	32.776504	−116.494784	1064	4	–
Lawson Valley Road	LAW	32.744610	−116.805661	585	7	4.88
Jamul	JAM	32.730587	−116.875713	256	3	–
Dulzura	DUL	32.631651	−116.761523	406	7	5.12

N, number of sampled individuals; A_R, standardized allelic richness. A_R was only calculated for populations with five or more genotyped individuals.

Network, http://swbiodiversity.org/seinet/ and Global Biodiversity Information Facility, http://www.gbif.org/) (Fig. 2a). Prior to modelling, all herbarium records were mapped and examined to identify and exclude records having errors in georeferencing, obvious misidentifications and cultivated plants. Only specimens collected from 1950 to present were retained for modelling, resulting in a final data set of 176 entries. To construct the models, we used 19 bioclimatic variables from the WorldClim data set (version 1.4, see http://www.worldclim.org/ for variable descriptions) interpolated to 30-arcsec (ca. 1-km) resolution (Hijmans *et al.* 2005). Final model variables were selected to minimize correlations among them and maximize variable contribution to model predictions. Variable importance was determined from the percent contribution of each variable to the model and the loss of predictive power when each variable was excluded. Model evaluation statistics were produced from 10 replicate model runs, where species records were divided into a 40%/60% test/train random partitions. Overall model performance was evaluated using the area under the receiving operator characteristics curve (AUC), which ranges from 0.5 (random prediction) to 1 (maximum prediction). The logistic output of MAXENT consists of a grid map with each cell having an index of suitability between 0 and 1. Low values indicate that conditions are unsuitable for the species to occur, whereas high values indicate that conditions are suitable. Model predictions were visualized in ARCMAP 9.3 (ESRI, Redlands, CA, USA).

To obtain the predicted distribution of Engelmann oak at the LGM (*c.* 21 000 years BP), we projected contemporary species–climate relationships to the LGM using the Community Climate System Model version 3 (CCSM3, http://www.ccsm.ucar.edu/; Kiehl & Gent 2004; Otto-Bliesner *et al.* 2006; Collins *et al.* 2006) downscaled to 2.5-arcmin resolution. The palaeoclimate layers derived from the CCSM model were developed by R. Hijmans (http://www.worldclim.org). The reconstruction of palaeoclimate by the CCSM model for California shows a pattern broadly consistent with that inferred from palaeopollen records (Pease *et al.* 2009), validating the use of the CCSM for inferring the distribution of Engelmann oak at the LGM. Present and LGM niche suitability scores obtained from models only based on the 123 genetically identified Engelmann oaks considered in this study (i.e. excluding herbarium data) provided analogous results in subsequent analyses of genetic diversity and gene flow (see below).

Microsatellite genotyping and basic genetic statistics

We ground about 50 mg of frozen leaf tissue in tubes with a tungsten ball using a mixer mill, and DNA extraction was performed with the DNeasy Plant Mini Kit (Qiagen). To genotype Engelmann oaks, we used 9

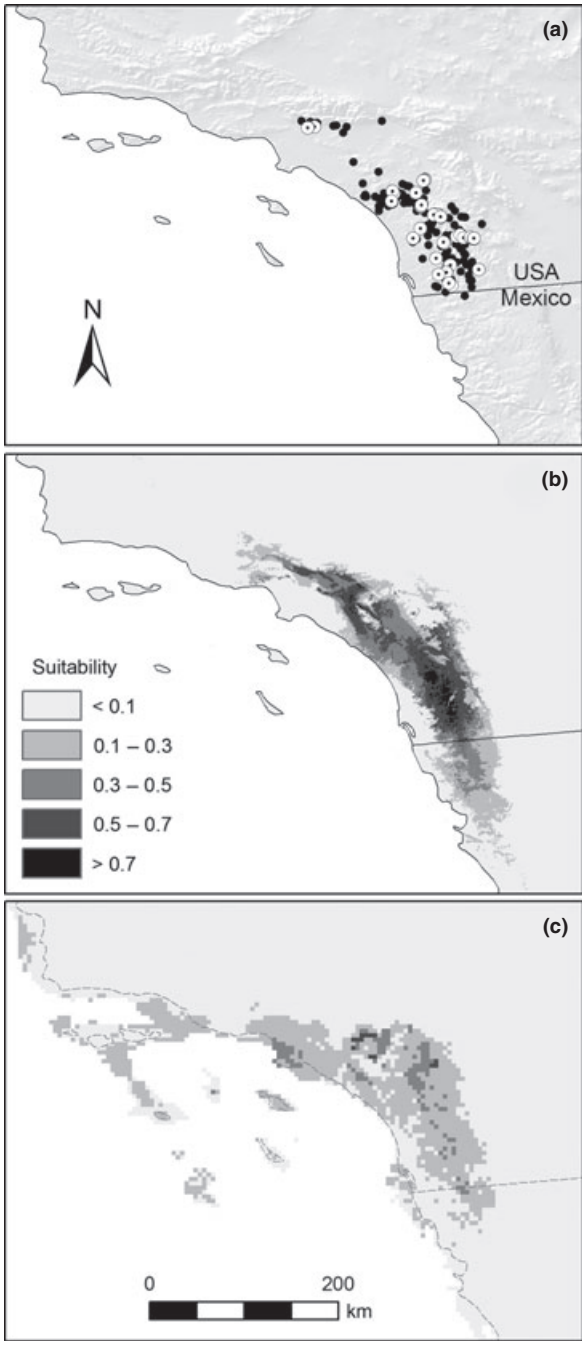

Fig. 2 (a) Point locations (sampling sites: white 'bull eyes'; herbarium records: black dots) and niche models of Engelmann oak (*Quercus engelmannii*) for (b) the present and (c) the Last Glacial Maximum (LGM; *c.* 21 000 years BP). Suitability values indicate logistic probability of presence and range from 0 to 1, with increasingly darker shades of grey with increasing habitat suitability. The LGM distribution was modelled using the CCSM3 climatic model.

polymorphic microsatellite markers previously developed for other *Quercus* species (Table 2) and used in our laboratory (e.g. Sork *et al.* 2002; Grivet *et al.* 2009). Amplifications were carried out in 10-μL reaction mixtures containing 2.5 μL of DNA (~20 ng) and 7.5 μL of PCR mix, which contained the following reagents: 5 μL of Multiplex Mix (Qiagen), 0.4 μL of BSA 10×, 1 μL of dye-labelled primer mix (at 2 μM for QpZAG7, MSQ4, QpZAG9, QpZAG36, QpZAG110, QrZAG20, QM69-2M1 and at 1 μM for QpZAG1/5 and QrZAG11) and 1.1 μL of water. The PCR touchdown profile consisted of an initial denaturing of 15 min at 95 °C, followed by 12 cycles of 30-s denaturing at 94 °C, 90-s annealing from 60 to 55 °C and 60-s extension at 72 °C, followed by 33 cycles of 30-s denaturing at 89 °C, 90-s annealing at 55 °C and 60-s extension at 72 °C, with a final extension step of 60 °C for 30 min. Amplification products were electrophoresed using an ABI PRISM 3700 capillary sequencer (Applied Biosystems) at the University of California, Los Angeles (UCLA) Sequencing and Genotyping Core Facility. Genotypes were scored using GeneMapper 3.7 (Applied Biosystems).

Microsatellite genotypes were tested for departure from Hardy–Weinberg equilibrium within each sampling population at each locus using an exact test (Guo & Thompson 1992) based on 900 000 Markov chain iterations as implemented in the program Arlequin 3.1 (Excoffier *et al.* 2005). We also used Arlequin 3.1 to test for linkage equilibrium between each pair of loci for each sampling population using a likelihood ratio statistic, whose distribution was obtained by a permutation procedure (Excoffier *et al.* 2005). Sequential Bonferroni corrections were applied to account for multiple comparisons (Rice 1989).

Genetic diversity

We calculated allelic richness (A_R) for each locality with at least five genotyped individuals. A_R was standardized for sample size using the program HP-Rare (Kalinowski 2005). We analysed which variables contributed to explain A_R using a General Linear Model (GLM) with spss 19.0. We considered four explanatory covariates in the GLM: (i) Population isolation, estimated as the average genetic differentiation (F_{ST} values) of each population with all other populations in the study area (e.g. Ortego *et al.* 2010b; Wang *et al.* 2011); (ii) current niche suitability ($NS_{CURRENT}$); (iii) LGM niche suitability (NS_{LGM}); (iv) Niche stability (NS_{STA}), estimated as $NS_{STA} = 1 - |NS_{CURRENT} - NS_{LGM}|$. The precision of A_R estimates may differ among populations because of differences in sample sizes. To take this into account, we used a weighted least square method, where weight equals the sample size for each studied population (e.g. Ortego *et al.* 2011). Initially, the GLM was constructed with all explanatory terms fitted, and the final model was selected following a backward procedure, by

Table 2 Microsatellite loci used to genotype Engelmann oaks (*Quercus engelmannii*): number of alleles (A), expected heterozygosity (H_E) and observed heterozygosity (H_O) for each locus

Locus	A	H_E	H_O	Primer origin
MSQ4	16	0.82	0.76	Dow *et al.* 1995
QpZAG1/5	17	0.72	0.58	Steinkellner *et al.* 1997
QpZAG7	28	0.92	0.84	Steinkellner *et al.* 1997
QpZAG9	8	0.28	0.23	Steinkellner *et al.* 1997
QpZAG36	13	0.75	0.61	Steinkellner *et al.* 1997
QpZAG110	23	0.90	0.88	Steinkellner *et al.* 1997
QrZAG11	25	0.81	0.70	Kampfer *et al.* 1998
QrZAG20	18	0.88	0.65	Kampfer *et al.* 1998
QM69-2M1	20	0.65	0.63	Isagi & Suhandono 1997

progressively eliminating nonsignificant variables. The significance of the remaining variables was tested again until no additional variable reached significance. The result is the minimal most adequate model for explaining the variability in the response variable, where only the significant explanatory variables are retained.

Genetic structure

We investigated population genetic structure among sampling locations, calculating pairwise F_{ST}-values and testing their significance by 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 (Rice 1989). The global F_{ST} across all samples was calculated in FSTAT 2.9.3 and its significance was tested using 10 000 randomizations (Goudet 1995, 2001). We assessed the genetic clustering among putatively pure *Q. engelmannii* individuals (see study species section) by using a Bayesian Markov chain Monte Carlo clustering analysis implemented in the program STRUCTURE 2.3.3 (Pritchard *et al.* 2000; Falush *et al.* 2003; Hubisz *et al.* 2009). STRUCTURE assigns individuals to K populations based on their multilocus genotypes. We ran STRUCTURE assuming correlated allele frequencies and admixture (Pritchard *et al.* 2000; Falush *et al.* 2003) and using prior population information (Hubisz *et al.* 2009). We conducted five independent runs for each value of K = 1–10 to estimate the 'true' number of clusters with 10^6 MCMC cycles, following a burn-in period of 10^5 iterations. The number of populations best fitting the data set was defined both using log probabilities [$\Pr(X \mid K)$] and ΔK, as described in Evanno *et al.* (2005).

Causal modelling analyses

Genetic relatedness among all pairs of sampled individuals was 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 often outperforms other estimators (Lynch & Ritland 1999). *A priori*, we considered four potential drivers of genetic structure in Engelmann oak: (i) the geographical distance, (ii) differences in current niche suitability, (iii) differences in LGM niche suitability and (iv) differences in altitude. To generate distance matrices, we calculated the Euclidean distance between altitude and niche suitability scores obtained for each individual. We calculated the matrices of Euclidean geographical distances between individuals using GEOGRAPHIC DISTANCE MATRIX GENERATOR 1.2.3 (Ersts 2011). It should be noted that we are not analysing landscape resistance to gene movement as done in several previous studies on animals (e.g. Cushman *et al.* 2006; Beier *et al.* 2008; Pérez-Espona *et al.* 2008), because the potential effects of the studied environmental factors (current and LGM niche suitability and altitude) on gene flow are not expected to create complex surfaces of pollen or seed movement. Rather, we expect that climatic conditions and altitude limit realized gene flow because of selection against nonlocally adapted genotypes and/or phenological differences (see discussion section), effects that we assume to be linear as done in previous studies (e.g. Ramírez-Valiente *et al.* 2009; see also Funk *et al.* 2009 and Wang & Summers 2009 for call and colour variation in frogs respectively).

We used causal modelling to determine the environmental factors (climate and altitude) influencing genetic relatedness in the study system (Legendre 1993; e.g. Cushman *et al.* 2006; Richards-Zawacki 2009; Bull *et al.* 2011). Causal modelling uses a series of partial Mantel tests to identify the organizational model (i.e. combination of environmental factors) that best explain genetic relatedness. We considered 15 organizational models corresponding to all the patterns of causality among the three variables described above. Each model describes a hypothesis in which certain factors are associated with gene flow and carries a set of statistical predictions. Support for a model can be computed as the number or the proportion of statistical predictions that are fulfilled. A model is fully supported only if the entire set of hypotheses is verified (Legendre & Troussellier 1988; Legendre 1993; e.g. Wang & Summers 2010). We first computed simple and partial Mantel tests between the genetic relatedness matrix and the matrices corresponding to the different factors. Then, we compared the significant and nonsignificant tests to the expectations under the 15 organizational models. Finally, we identified the organizational model with the greatest support, calculating the ratio of the number of nonrejected hypotheses to the total number of tested hypotheses for each model. All

Mantel tests were performed using ZT software with 10 000 permutations (Bonnet & Van de Peer 2002).

Results

Niche modelling

The predicted current distribution (Fig. 2b) is consistent with the observed current distribution (Scott 1991; Roberts 1995). The climatic variables included in the final habitat suitability model were temperature seasonality (BIO4), maximum temperature of the warmest month (BIO5), minimum temperature of the coldest month (BIO6), mean annual precipitation (BIO12), precipitation seasonality (BIO15) and precipitation of the driest quarter (BIO17). The AUC for the test data was on average 0.963 (SD = 0.005; 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). Model AUC was significantly greater than that of a random (AUC = 0.5) prediction (one-tailed Wilcoxon signed rank test; P = 0.002) (Phillips *et al.* 2006). The importance of the variables (measured as the percent drop in training AUC when the variable is excluded from the model) was BIO6 (39.4%), BIO17 (26.8%), BIO15 (11.9%), BIO12 (10.7%), BIO5 (8.9%) and BIO4 (2.4%). The estimated distribution of Engelmann oak during the LGM suggests a range contraction and a general decrease in habitat suitability over the species distribution range during that period (Fig. 2c).

Genetic diversity

After applying sequential Bonferroni corrections to compensate for multiple statistical tests, no locus deviated from HWE in any of the sampled populations. We did not find any evidence of genotypic linkage disequilibrium at any pair of loci in any population either (all $P > 0.05$). Observed heterozygosity at each locus ranged from 0.23 to 0.88, with 8–28 alleles per locus (Table 2). For those localities with five or more sampled individuals, allelic richness (A_R) standardized for sample size ranged from 4.37 to 5.2 alleles per locus (Table 1). A_R was not associated with population isolation (i.e. average pairwise population differentiation, F_{ST}), current niche suitability or LGM niche suitability (Table 3). However, A_R was negatively correlated with niche stability (Table 3; Fig. 3).

Genetic structure

Pairwise F_{ST} values ranged from −0.001 to 0.115, and 18 of the 78 pairwise comparisons were significant after sequential Bonferroni correction (Table S1, Supporting information). Comparisons involving PAS, HEM or PAU populations showed particularly high differentiation (Table S1, Supporting information). Global F_{ST} was 0.038 (95% confidence interval from 15000 bootstrap replicates: 0.027–0.048). STRUCTURE analyses showed a maximum value of $\Pr(X|K)$ for K = 3. However, the Evanno *et al.* (2005) method indicated an optimal value of K = 2, and ΔK dropped sharply with K = 3. At K = 2, the first cluster only included HEM population and the second one comprised the remainder localities. When K = 3 was considered, the first genetic cluster included PAS, the second to HEM and the third grouped the rest of the populations (Fig. 1).

We explored the possibility that the genetic cluster represented by the HEM population is a consequence of the comparatively large sample size available for this locality. We randomly selected a subset of individuals

Table 3 GLMs for allelic richness (A_R) in relation to average pairwise population differentiation (F_{ST}), current niche suitability, Last Glacial Maximum (LGM) niche suitability and niche stability

	Estimate ± SE	Test	P
Explanatory terms			
Intercept	5.455 ± 0.240		
Niche stability	−0.909 ± 0.341	$F_{1,11} = 7.10$	0.022
Rejected terms			
LGM niche suitability		$F_{1,10} = 2.17$	0.171
Current niche suitability		$F_{1,10} = 0.30$	0.593
Average population differentiation		$F_{1,10} = 0.40$	0.540

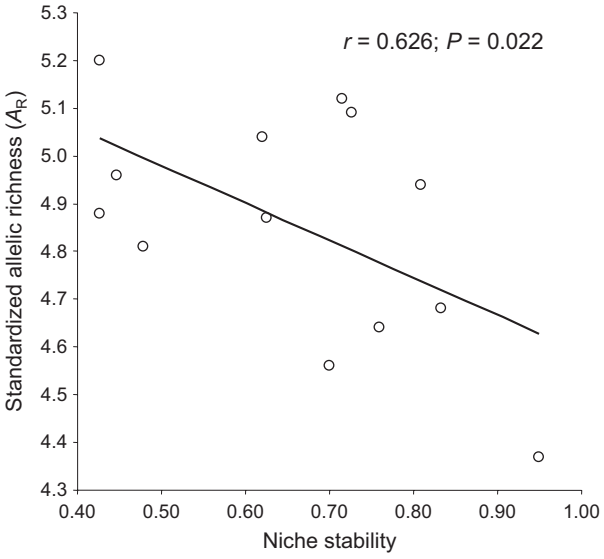

Fig. 3 Relationship between standardized allelic richness (A_R) and niche stability.

from the HEM population to obtain data sets with a reduced sample size ranging between 8 and 18 individuals. We repeated this random selection of individuals 10 times for each sample size category and ran STRUCTURE for these reduced data sets. For any sample size considered, STRUCTURE analyses always identified two genetic clusters. When we considered sample sizes higher than 10 individuals, STRUCTURE always showed that the first genetic cluster only included HEM population and the second one comprised the remainder localities. PAS was often classified as an admixed genetic group (probability of population membership, $q = 0.5$) between the two clusters. For sample sizes lower than 11 individuals, STRUCTURE showed that one of the genetic clusters only included the HEM population, considered PAS as an admixed genetic group or joined HEM-PAS localities. Sample sizes of eight or more individuals have been analysed for other populations located across the species distribution range without that has resulted in a distinct genetic cluster for any single population or geographically close group of populations (Table 1; Fig. 1). Thus, the distinctive genetic cluster for the northernmost populations (HEM and PAS) is not likely to be the result of sample sizes higher than those available for other localities.

Causal modelling analyses

Simple Mantel tests for the four analysed factors were highly significant ($P < 0.001$) and showed correlation coefficient values ranging between −0.04 and −0.08 (Euclidean geographical distance: $r = -0.08$; Current niche suitability: $r = -0.06$; LGM niche suitability: $r = -0.04$; Altitude: $r = -0.04$). Causal modelling analyses based on partial Mantel tests and testing 15 hypothetical organizational models showed that the model including all the variables was fully supported (Table 4).

Discussion

Our results indicate that climatic factors are important to explain the observed patterns of genetic diversity and structure in Engelmann oak. We first analysed the relationship between genetic variability and local climatic stability since the LGM. We found that genetic diversity is negatively associated with local climatic stability, but not with current or past habitat suitability. This pattern is counter to the expectation of higher levels of genetic variability in populations located in stable areas in comparison with those from regions

Table 4 Results from causal modelling analyses

Partial Mantel test	*r*	*P*-values	1 D	2 N	3 L	4 A	5 D+N	6 D+L	7 D+A	8 +L	9 N+A	10 L+A	11 D+N +L	12 D+N +A	13 D+L +A	14 N+L +A	15 D+N +L+A
DG.N	−0.076	**0.00010**	S	NS			S	S	S	NS	NS		S	S	S	NS	S
DG.L	−0.072	**0.00010**	S		NS		S	S	S	NS		NS	S	S	S	NS	S
DG.A	−0.070	**0.00010**	S			NS	S	S	S		NS	NS	S	S	S	NS	S
NG.D	−0.052	**0.00010**	NS	S			S	NS	NS	S	S		S	S	NS	S	S
NG.L	−0.056	**0.00010**		S	NS		S	NS		S	S	NS	S	S	NS	S	S
NG.A	−0.056	**0.00010**		S		NS	S		NS	S	S	NS	S	S	NS	S	S
LG.D	−0.025	**0.01160**	NS		S		NS	S	NS	S		S	S	NS	S	S	S
LG.N	−0.039	**0.00050**		NS	S		NS	S		S	NS	S	S	NS	S	S	S
LG.A	−0.036	**0.00050**			S	NS		S	NS	S	NS	S	S	NS	S	S	S
AG.D	−0.022	**0.02480**	NS			S	NS	NS	S		S	S	NS	S	S	S	S
AG.N	−0.042	**0.00030**		NS		S	NS		S	NS	S	S	NS	S	S	S	S
AG.L	−0.037	**0.00050**			NS	S		NS	S	NS	S	S	NS	S	S	S	S
Model support rate			0.50	0.50	0.50	0.67	0.60	0.60	0.60	0.60	0.60	0.60	0.75	0.75	0.75	0.75	1.00

G, genetic relatedness; D, Euclidean distance; N, current niche suitability; L, last glacial maximum niche suitability; A, altitude. The statistical predictions for each partial Mantel test for each model are indicated as S (expected to be significant) and NS (not expected to be significant). Boxes where a particular test is not applicable for a specific model are represented in black. For each Mantel test, a period separates the main matrices on the left from the covariate matrix on the right (e.g. DG.N tests for the correlation between D and G controlling for N). *P*-values and correlation coefficients (*r*) for Mantel tests are indicated. Bold type means *P*-values are statistically significant after sequential Bonferroni correction. The white boxes identify tests where the *P*-value matched the expected result of the test for each model. The grey boxes identify tests where the *P*-value did not match the expected result of the test for each model. The support rate for each model is the proportion of supported hypotheses in relation to the total number of tested hypotheses.

experiencing greater climatic shifts (Hampe & Petit 2005; Rodríguez-Robles *et al.* 2010). Increased immigration rates in unstable patches during favourable climatic periods may have promoted the admixture of genetically differentiated populations after colonization processes or demographic expansions (Hewitt 2000; Petit *et al.* 2003). Temporally varying selection could have also resulted in higher levels of genetic diversity in climatically unstable patches (e.g. Borash *et al.* 1998), particularly if some of the analysed markers are linked to genes that respond to selection (Ramírez-Valiente *et al.* 2010).

California populations of Engelmann oak divided into three main geographical genetic clusters, a pattern that was mainly driven by the two isolated populations (Pasadena and Hemet) located in the northern edge of the species distribution range (Fig. 1). STRUCTURE analyses have also shown that the distinctive genetic differentiation for these northernmost populations is not likely to be biased by their larger sample sizes in comparison with those available for other localities. These two populations are located in areas with lower niche suitability than the other sampled populations (*t*-tests; Pasadena: $t = 4.49$; $P = 0.001$; Hemet: $t = 9.45$; $P < 0.001$), indicating that they may be considered peripheral from both a geographical and an ecological point of view. These findings are consistent with the centre–periphery model that predicts that populations located along the species boundaries show disproportionately higher levels of genetic differentiation in comparison with those populations located in the core of the species distribution range (Hampe & Petit 2005; Alleaume-Benharira *et al.* 2006; e.g. Shafer *et al.* 2011; Gugger *et al.* 2011). In contrast, allelic richness was significantly higher in Pasadena (*t*-test; $t = -3.45$; $P = 0.006$) but lower in Hemet (*t*-test; $t = 8.48$; $P < 0.001$) when compared with the other populations located in the core of the species distribution range. Thus, information on genetic diversity is not consistent with all predictions of the centre–periphery model of genetic variability (Hampe & Petit 2005). The lack of glaciations in southern California has probably allowed most populations to survive colder periods in much of the current distribution range of Engelmann oak, resulting in levels of genetic variability that are probably mostly dependent on local population dynamics (Gugger & Sugita 2010; Sork *et al.* 2010). This relative stability may have shaped the observed genetic structure consistent with the centre–periphery model but different from the geographical patterns of genetic diversity reported for species whose phylogeographic histories are mostly the result of recent post-glacial range expansions (Hampe & Petit 2005).

Causal modelling analyses supported the hypothesis of decreased gene flow with geographic distance, indicating an isolation-by-distance pattern of genetic structure that in this wind-pollinated tree species probably reflects a migration-drift equilibrium scenario (Slatkin 1993; Hutchison & Templeton 1999). Nonetheless, our findings indicate that environmental factors shape genetic variability as well. We found that genetic relatedness decreases with difference in altitude and current and LGM niche suitability, indicating that gene flow is lower among populations experiencing contrasting environmental conditions. Multiple factors could be behind the observed limitation of gene flow in relation with these variables. First, populations experiencing different climates could have evolved particular adaptations, and reduced gene flow may be reflecting selection against the establishment of nonlocal genotypes even when neutral microsatellite markers are not involved in these evolutionary processes (Aitken *et al.* 2008; Ramírez-Valiente *et al.* 2009, 2010; Eckert *et al.* 2010). A second possibility is that asynchrony in reproductive phenology resulting from either phenotypic plasticity or adaptive divergence has limited gene flow among populations with different climatic conditions, despite long-distance pollen dispersal and lack of selection against nonlocal genotypes (isolation-by-time *sensu* Hendry & Day 2005; e.g. Burczyk & Prat 1997; Dyer *et al.* 2010; Freeland *et al.* 2010). Accordingly, some studies have reported a phenological and genetic divergence across elevational gradients (Schuster *et al.* 1989; Dyer *et al.* 2010) and among populations experiencing contrasting climatic conditions, even when the geographical distances involved are very short (Hirao & Kudo 2008; Freeland *et al.* 2010; Knowles & Alvarado-Serrano 2010). Finally, evidence from other oaks shows that despite the long-distance pollen flow, local pollen and seed movement can create significant genetic structure associated with habitat suitability (Sork *et al.* 2002, 2010; Grivet *et al.* 2005, 2009).

Overall, our results suggest that climate-based habitat suitability, as measured through ecological niche models, shapes patterns of gene flow even within a species with a small distribution range and high potential for gene flow. This finding is consistent with the association of genetic structure for neutral genetic markers and climatic variables in another Californian oak (Sork *et al.* 2010). The fact that climatic heterogeneity shapes geographic genetic structure of tree populations should be incorporated into conservation plans and restoration ecology projects (Sork *et al.* 2009). In addition, the genetic and ecological distinctiveness reported for peripheral populations indicates that such populations should be considered as independent management units for conservation (Leppig & White 2006). This study illustrates the benefits of taking into account spatially explicit information about the environment when interpreting patterns of geographic structure, an

approach emphasized by landscape genetic studies (Sork & Waits 2010). Future studies on phenotypic divergence in combination with common garden and translocation experiments will further elucidate the relative role of natural selection and phenology on the observed patterns of gene flow (Ramírez-Valiente *et al*. 2009; Dyer *et al*. 2010; Freeland *et al*. 2010).

Acknowledgements

During this work, JO was supported by postdoctoral José Castillejo (ME), JAE-Doc (CSIC) and Juan de la Cierva (MICINN) fellowships. For help with field sampling and/or genetic analysis, we thank Keith Gaddis, Edina Losonczi-Gorshtein, Karen Lundy, Pam Thompson and Stephanie Steele. VLS received support from UCLA and National Science Foundation (NSF-DEB-0516529).

References

Aitken SN, Yeaman S, Holliday JA, Wang TL, Curtis-McLane S (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, **1**, 95–111.

Alleaume-Benharira M, Pen IR, Ronce O (2006) Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. *Journal of Evolutionary Biology*, **19**, 203–215.

Anderson CD, Epperson BK, Fortin MJ *et al.* (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology*, **19**, 3565–3575.

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.

Beier P, Majka DR, Spencer WD (2008) Forks in the road: choices in procedures for designing wildland linkages. *Conservation Biology*, **22**, 836–851.

Bonnet E, Van de Peer Y (2002) ZT: a software tool for simple and partial Mantel tests. *Journal of Statistical Software*, **10**, 1–12.

Borash DJ, Gibbs AG, Joshi A, Mueller LD (1998) A genetic polymorphism maintained by natural selection in a temporally varying environment. *American Naturalist*, **151**, 148–156.

Bull RAS, Cushman SA, Mace R *et al.* (2011) Why replication is important in landscape genetics: American black bear in the Rocky Mountains. *Molecular Ecology*, **20**, 1092–1107.

Burczyk JL, Prat D (1997) Male reproductive success in *Pseudotsuga menziesii* (Mirb.) France: the effects of spatial structure and flowering characteristics. *Heredity*, **79**, 638–647.

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

Carnaval AC, Hickerson MJ, Haddad CFB, Rodrigues MT, Moritz C (2009) Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, **323**, 785–789.

Collins WD, Bitz CM, Blackmon ML *et al.* (2006) The community climate system model version 3 (CCSM3). *Journal of Climate*, **19**, 2122–2143.

Coulon A, Fitzpatrick JW, Bowman R, Lovette IJ (2010) Effects of habitat fragmentation on effective dispersal of Florida scrub-jays. *Conservation Biology*, **24**, 1080–1088.

Cushman SA, Landguth EL (2010) Spurious correlations and inference in landscape genetics. *Molecular Ecology*, **19**, 3592–3602.

Cushman SA, McKelvey KS, Hayden J, Schwartz MK (2006) Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *American Naturalist*, **168**, 486–499.

Davis EB, Koo MS, Conroy C, Patton JL, Moritz C (2008) The California Hotspots Project: identifying regions of rapid diversification of mammals. *Molecular Ecology*, **17**, 120–138.

Dow BD, Ashley MV, Howe HF (1995) Characterization of highly variable (Ga/Ct)(N) microsatellites in the bur oak, *Quercus macrocarpa*. *Theoretical and Applied Genetics*, **91**, 137–141.

Dyer RJ, Nason JD, Garrick RC (2010) Landscape modeling of gene flow: improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology*, **19**, 3746–3759.

Eckert AJ, van Heerwaarden J, Wegrzyn JL *et al.* (2010) Patterns of population structure and environmental associations to aridity across the range of Loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics*, **185**, 969–982.

Elith J, Graham CH, Anderson RP *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Endler JA (1986) *Natural Selection in the Wild*. Princeton University Press, Princeton, New Jersey.

Ersts PJ 2011 [Internet] *Geographic Distance Matrix Generator (version 1.2.3)*. American Museum of Natural History, Center for Biodiversity and Conservation. 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 JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

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

eFloras. (2011) *Flora of North America*. Missouri Botanical Garden, St. Louis, Missouri, USA, and Harvard University Herbaria, Cambridge, Massachusetts, USA. http://www.efloras.org

Frankham R (1996) Relationship of genetic variation to population size in wildlife. *Conservation Biology*, **10**, 1500–1508.

Frankham R (2005) Genetics and extinction. *Biological Conservation*, **126**, 131–140.

Freedman AH, Thomassen HA, Buermann W, Smith TB (2010) Genomic signals of diversification along ecological gradients in a tropical lizard. *Molecular Ecology*, **19**, 3773–3788.

Freeland JR, Biss P, Conrad KF, Silvertown J (2010) Selection pressures have caused genome-wide population

differentiation of *Anthoxanthum odoratum* despite the potential for high gene flow. *Journal of Evolutionary Biology*, **23**, 776–782.

Funk WC, Cannatella DC, Ryan MJ (2009) Genetic divergence is more tightly related to call variation than landscape features in the Amazonian frogs *Physalaemus petersi* and *P. freibergi*. *Journal of Evolutionary Biology*, **22**, 1839–1853.

Garant D, Kruuk LEB, Wilkin TA, McCleery RH, Sheldon BC (2005) Evolution driven by differential dispersal within a wild bird population. *Nature*, **433**, 60–65.

García-Ramos G, Kirkpatrick M (1997) Genetic models of adaptation and gene flow in peripheral populations. *Evolution*, **51**, 21–28.

Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate *F*-statistics. *Journal of Heredity*, **86**, 485–486.

Goudet J (2001) *Fstat, a Program to Estimate and Test Gene Diversities and Fixation Indices (Version 2.9.3)*. Institut d'Ecologie, Universite' de Lausanne, Dorigny, Switzerland.

Grivet D, Smouse PE, Sork VL (2005) A novel approach to an old problem: tracking dispersed seeds. *Molecular Ecology*, **14**, 3585–3595.

Grivet D, Sork VL, Westfall RD, Davis FW (2008) Conserving the evolutionary potential of California valley oak (*Quercus lobata* Nee): a multivariate genetic approach to conservation planning. *Molecular Ecology*, **17**, 139–156.

Grivet D, Robledo-Arnuncio JJ, Smouse PE, Sork VL (2009) Relative contribution of contemporary pollen and seed dispersal to the effective parental size of seedling population of California valley oak (*Quercus lobata* Née). *Molecular Ecology*, **18**, 3967–3979.

Gugger PF, Sugita S (2010) Glacial populations and postglacial migration of Douglas-fir based on fossil pollen and macrofossil evidence. *Quaternary Science Reviews*, **29**, 2052–2070.

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

Guo SW, Thompson EA (1992) A Monte-Carlo method for combined segregation and linkage analysis. *American Journal of Human Genetics*, **51**, 1111–1126.

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

Hedrick PW (2001) Conservation genetics: where are we now? *Trends in Ecology & Evolution*, **16**, 629–636.

Hendry AP, Day T (2005) Population structure attributable to reproductive time: isolation by time and adaptation by time. *Molecular Ecology*, **14**, 901–916.

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

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.

Hirao AS, Kudo G (2008) The effect of segregation of flowering time on fine-scale spatial genetic structure in an alpine-snowbed herb *Primula cuneifolia*. *Heredity*, **100**, 424–430.

Holderegger R, Wagner HH (2008) Landscape genetics. *BioScience*, **58**, 199–207.

Holzhauer SIJ, Wolff K, Wolters V (2009) Changes in land use and habitat availability affect the population genetic structure of *Metrioptera roeselii* (Orthoptera: Tettigoniidae). *Journal of Insect Conservation*, **13**, 543–552.

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

Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53**, 1898–1914.

Isagi Y, Suhandono S (1997) PCR primers amplifying microsatellite loci of *Quercus myrsinifolia* Blume and their conservation between oak species. *Molecular Ecology*, **6**, 897–899.

IUCN. (2010) *IUCN Red List of Threatened Species. Version 2010.4. http://www.iucnredlist.org*. Downloaded on 17 April 2011.

Jenkins DG, Carey M, Czerniewska J *et al.* (2010) A meta-analysis of isolation by distance: relic or reference standard for landscape genetics? *Ecography*, **33**, 315–320.

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

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

Kiehl JT, Gent PR (2004) The community climate system model, version 2. *Journal of Climate*, **17**, 3666–3682.

Knowles LL, Alvarado-Serrano DF (2010) Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: insights from coupled ecological, demographic and genetic models in montane grasshoppers. *Molecular Ecology*, **19**, 3727–3745.

Lammi A, Siikamaki P, Mustajarvi K (1999) Genetic diversity, population size, and fitness in central and peripheral populations of a rare plant *Lychnis viscaria*. *Conservation Biology*, **13**, 1069–1078.

Legendre P (1993) Spatial autocorrelation-trouble or new paradigm. *Ecology*, **74**, 1659–1673.

Legendre P, Troussellier M (1988) Aquatic heterotrophic bacteria – modelling in the presence of spatial auto-correlation. *Limnology and Oceanography*, **33**, 1055–1067.

Leppig G, White JW (2006) Conservation of peripheral plant populations in California. *Madroño*, **53**, 264–274.

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

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular Ecology*, **19**, 3824–3835.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Nei M, Chakravarti A (1977) Drift variances of Fst and Gst statistics obtained from a finite number of isolated populations. *Theoretical Population Biology*, **11**, 307–325.

Ortego J, Bonal R (2010) Natural hybridisation between kermes (*Quercus coccifera* L.) and holm oaks (*Q. ilex* L.) revealed by microsatellite markers. *Plant Biology*, **12**, 234–238.

Ortego J, Bonal R, Muñoz A (2010a) Genetic consequences of habitat fragmentation in long-lived tree species: the case of the Mediterranean holm oak (*Quercus ilex*, L.). *Journal of Heredity*, **101**, 717–726.

Ortego J, Aguirre MP, Cordero PJ (2010b) Population genetics of *Mioscirtus wagneri*, a grasshopper showing a highly fragmented distribution. *Molecular Ecology*, **19**, 472–483.

Ortego J, Yannic G, Shafer ABA *et al.* (2011) Temporal dynamics of genetic variability in a mountain goat (*Oreamnos americanus*) population. *Molecular Ecology*, **20**, 1601–1611.

Otto-Bliesner BL, Marsha SJ, Overpeck JT *et al.* (2006) Simulating arctic climate warmth and icefield retreat in the last interglaciation. *Science*, **311**, 1751–1753.

Pastorini J, Thalmann U, Martin RD (2003) A molecular approach to comparative phylogeography of extant Malagasy lemurs. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 5879–5884.

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

Pease KM, Freedman AH, Pollinger JP *et al.* (2009) Landscape genetics of California mule deer (*Odocoileus hemionus*): the roles of ecological and historical factors in generating differentiation. *Molecular Ecology*, **18**, 1848–1862.

Pérez-Espona S, Pérez-Barbería FJ, McLeod JE *et al.* (2008) Landscape features affect gene flow of Scottish Highland red deer (*Cervus elaphus*). *Molecular Ecology*, **17**, 981–996.

Pérez-Tris J, Carbonell R, Telleria JL (2000) Abundance distribution, morphological variation and juvenile condition of robins, *Erithacus rubecula* (L.), in their Mediterranean range boundary. *Journal of Biogeography*, **27**, 879–888.

Petit RJ, Aguinagalde I, de Beaulieu JL *et al.* (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.

Phillips SJ, Dudik M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Pluess AR, Sork VL, Dolan B *et al.* (2009) Short distance pollen movement in a wind-pollinated tree, *Quercus lobata* (Fagaceae). *Forest Ecology and Management*, **258**, 735–744.

Postma E, Van Noordwijk AJ (2005) Gene flow maintains a large genetic difference in clutch size at a small spatial scale. *Nature*, **433**, 65–68.

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

Pyron RA, Burbrink FT (2009) Lineage diversification in a widespread species: roles for niche divergence and conservatism in the common kingsnake, *Lampropeltis getula*. *Molecular Ecology*, **18**, 3443–3457.

Ramírez-Valiente JA, Lorenzo Z, Soto A *et al.* (2009) Elucidating the role of genetic drift and natural selection in cork oak differentiation regarding drought tolerance. *Molecular Ecology*, **18**, 3803–3815.

Ramírez-Valiente JA, Lorenzo Z, Soto A *et al.* (2010) Natural selection on cork oak: allele frequency reveals divergent selection in cork oak populations along a temperature cline. *Evolutionary Ecology*, **24**, 1031–1044.

Raven PH, Axelrod DI (1978) Origin and relationships of the California flora. *University of California Publications in Botany*, **72**, 1–134.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

Richards-Zawacki CL (2009) Effects of slope and riparian habitat connectivity on gene flow in an endangered Panamanian frog, *Atelopus varius*. *Diversity and Distributions*, **15**, 796–806.

Roberts FM (1995) *The Oaks of the Southern California Floristic Province*. F. M. Roberts Publications, Encinitas, California.

Rodríguez-Robles JA, Jezkova T, Leal M (2010) Climatic stability and genetic divergence in the tropical insular lizard *Anolis krugi*, the Puerto Rican 'Lagartijo Jardinero de la Montana'. *Molecular Ecology*, **19**, 1860–1876.

Sacks BN, Brown SK, Ernest HB (2004) Population structure of California coyotes corresponds to habitat-specific breaks and illuminates species history. *Molecular Ecology*, **13**, 1265–1275.

Schneider CJ, Cunningham M, Moritz C (1998) Comparative phylogeography and the history of endemic vertebrates in the Wet Tropics rainforests of Australia. *Molecular Ecology*, **7**, 487–498.

Schuster WS, Alles DL, Mitton JB (1989) Gene flow in limber pine – evidence from pollination phenology and genetic differentiation along an elevational transect. *American Journal of Botany*, **76**, 1395–1403.

Scott TA (1991) The distribution of Engelmann oak (*Quercus engelmannii*) in California. In: *Proceedings of the Symposium on Oak Woodlands and Hardwood Rangeland Management* (ed. Standiford RB), pp. 351–359. US Deptartment of Agriculture, Forest Service Pacific Southwest Forest & Range Experimental Station, Berkeley.

Shafer ABA, Côté SD, Coltman DW (2011) Hot spots of genetic diversity descended from multiple *Pleistocene refugia* in an Alpine ungulate. *Evolution*, **65**, 125–138.

Slatkin M (1993) Isolation by distance in equilibrium and nonequilibrium populations. *Evolution*, **47**, 264–279.

Smith TB, Wayne RK, Girman DJ, Bruford MW (1997) A role for ecotones in generating rainforest biodiversity. *Science*, **276**, 1855–1857.

Sork VL, Waits L (2010) Contributions of landscape genetics – approaches, insights, and future potential. *Molecular Ecology*, **19**, 3489–3495.

Sork VL, Davis FW, Smouse PE *et al.* (2002) Pollen movement in declining populations of California Valley oak, *Quercus l*obata: where have all the fathers gone? *Molecular Ecology*, **11**, 1657–1668.

Sork VL, Davis FW, Grivet D (2009) Incorporating genetic information into conservation planning for California valley oak. Gen. Tech. Rep. PSW-GTR-217 497–509.

Sork VL, Davis FW, Westfall R *et al.* (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.
Spielman D, Brook BW, Frankham R (2004) Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 15261–15264.
Steinkellner H, Fluch S, Turetschek E *et al.* (1997) Identification and characterization of (GA/CT)(n)-microsatellite loci from *Quercus petraea*. *Plant Molecular Biology*, **33**, 1093–1096.
Storfer A, Murphy MA, Evans JS *et al.* (2007) Putting the 'landscape' in landscape genetics. *Heredity*, **98**, 128–142.
Swenson NG, Howard DJ (2005) Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *American Naturalist*, **166**, 581–591.
Thomassen HA, Cheviron ZA, Freedman AH, Harrigan RJ, Wayne RK, Smith TB (2010) Spatial modeling and landscape-level approaches for visualizing intra-specific variation. *Molecular Ecology*, **19**, 3532–3548.
Vähä JP, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, **15**, 63–72.
Walker D, Avise JC (1998) Principles of phylogeography as illustrated by freshwater and terrestrial turtles in the southeastern United States. *Annual Review of Ecology and Systematics*, **29**, 23–58.
Wang IJ, Summers K (2010) Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Molecular Ecology*, **19**, 447–458.
Wang IJ, Johnson JR, Johnson BB, Shaffer HB (2011) Effective population size is strongly correlated with breeding pond size in the endangered California tiger salamander, *Ambystoma californiense*. *Conservation Genetics*, **12**, 911–920.
Waples RS (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, **89**, 438–450.
Willi Y, Van Buskirk J, Hoffmann AA (2006) Limits to the adaptive potential of small populations. *Annual Review of Ecology Evolution and Systematics*, **37**, 433–458.
Winkler KA, Pamminger-Lahnsteiner B, Wanzenbock J, Weiss S (2011) Hybridization and restricted gene flow between native and introduced stocks of Alpine whitefish (*Coregonus* sp.) across multiple environments. *Molecular Ecology*, **20**, 456–472.
Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.

J.O. is currently a postdoctoral fellow at IREC interested in the study of different aspects of genetic variability in natural populations. E.C.R. is a graduate student at UCLA studying potential impacts of climate change on plant diversity and distribution. P.F.G. is a postdoctoral scholar interested in the phylogeography, population genetics, genomics and palaeoecology of trees. V.L.S. studies the ecological and evolutionary processes of tree populations in temperate and tropical habitats with a landscape genetic and environmental genomics approach to study gene flow, adaptation and conservation in California oaks.

Data accessibility

Data deposited in the Dryad repository: doi: 10.5061/dryad.rd645561.

Supporting information

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

Table S1 Pair-wise population FST-values.

Please note: Wiley-Blackwell is 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 corresponding author for the article.