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
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; Thomsassen 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 biogeographic 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 suitability.

Methods

Study species and sampling

Engelmann oak, Quercus engelmannii 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. Fluess 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...
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 (T\(_q\)) of 0.90 to classify each individual as purebred or hybrid (Vahä & 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
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 characteristic 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

<table>
<thead>
<tr>
<th>Locality</th>
<th>Code</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Altitude (m)</th>
<th>N</th>
<th>$A_R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pasadena</td>
<td>PAS</td>
<td>34.134079</td>
<td>−118.098914</td>
<td>202</td>
<td>9</td>
<td>5.09</td>
</tr>
<tr>
<td>Hemet</td>
<td>HEM</td>
<td>33.628262</td>
<td>−117.012864</td>
<td>619</td>
<td>15</td>
<td>4.37</td>
</tr>
<tr>
<td>Avocado Mesa</td>
<td>AVO</td>
<td>33.513735</td>
<td>−117.308909</td>
<td>648</td>
<td>8</td>
<td>4.68</td>
</tr>
<tr>
<td>Pauba Ranch</td>
<td>PAU</td>
<td>33.508552</td>
<td>−117.088208</td>
<td>381</td>
<td>8</td>
<td>4.87</td>
</tr>
<tr>
<td>De Luz</td>
<td>LUY</td>
<td>33.423533</td>
<td>−117.321380</td>
<td>100</td>
<td>5</td>
<td>4.56</td>
</tr>
<tr>
<td>Palomar Mountains</td>
<td>PAL</td>
<td>33.390607</td>
<td>−117.039250</td>
<td>449</td>
<td>7</td>
<td>4.64</td>
</tr>
<tr>
<td>Harolds</td>
<td>HAR</td>
<td>33.302025</td>
<td>−116.892959</td>
<td>1023</td>
<td>4</td>
<td>–</td>
</tr>
<tr>
<td>Oak Knoll Ranch</td>
<td>OAK</td>
<td>33.298210</td>
<td>−116.922127</td>
<td>731</td>
<td>7</td>
<td>5.20</td>
</tr>
<tr>
<td>Lake Henshaw</td>
<td>HEN</td>
<td>33.276442</td>
<td>−116.855038</td>
<td>737</td>
<td>4</td>
<td>–</td>
</tr>
<tr>
<td>Daley Ranch</td>
<td>DAL</td>
<td>33.169990</td>
<td>−116.047007</td>
<td>351</td>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td>Santa Ysabel</td>
<td>YSA</td>
<td>33.102790</td>
<td>−116.669374</td>
<td>951</td>
<td>7</td>
<td>5.04</td>
</tr>
<tr>
<td>Julian</td>
<td>JUL</td>
<td>33.074770</td>
<td>−116.549129</td>
<td>861</td>
<td>11</td>
<td>4.94</td>
</tr>
<tr>
<td>Lake Hodges</td>
<td>HOD</td>
<td>33.074700</td>
<td>−117.118072</td>
<td>100</td>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>Ramona</td>
<td>RAM</td>
<td>33.029917</td>
<td>−116.823050</td>
<td>494</td>
<td>7</td>
<td>4.81</td>
</tr>
<tr>
<td>Louis A. Stelzer County Park</td>
<td>LOU</td>
<td>32.881655</td>
<td>−116.901225</td>
<td>500</td>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td>Alpine</td>
<td>ALP</td>
<td>32.814090</td>
<td>−116.772395</td>
<td>552</td>
<td>8</td>
<td>4.96</td>
</tr>
<tr>
<td>Cleveland National Forest</td>
<td>CLE</td>
<td>32.776504</td>
<td>−116.494784</td>
<td>1064</td>
<td>4</td>
<td>–</td>
</tr>
<tr>
<td>Daley Valley Road</td>
<td>LAW</td>
<td>32.744610</td>
<td>−116.805661</td>
<td>585</td>
<td>7</td>
<td>4.88</td>
</tr>
<tr>
<td>Jamul</td>
<td>JAM</td>
<td>32.730587</td>
<td>−116.875713</td>
<td>256</td>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>Dulzura</td>
<td>DUL</td>
<td>32.631651</td>
<td>−116.761523</td>
<td>406</td>
<td>7</td>
<td>5.12</td>
</tr>
</tbody>
</table>

$N$, number of sampled individuals; $A_R$, standardized allelic richness. $A_R$ was only calculated for populations with five or more genotyped individuals.
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*<sub>R</sub>) for each locality with at least five genotyped individuals. *A*<sub>R</sub> was standardized for sample size using the program HP-RARE (Kalinowski 2005). We analysed which variables contributed to explain *A*<sub>R</sub> 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*<sub>ST</sub> 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 (**NSLGM**); (iv) Niche stability (**NS STA**), estimated as **NS STA** = 1 − |**NS CURRENT** − **NSLGM**|. The precision of *A*<sub>R</sub> 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

![Fig. 2](image-url)
Table 2 Microsatellite loci used to genotype Engelmann oaks (Quercus engelmannii): number of alleles (A), expected heterozygosity (HE) and observed heterozygosity (HO) for each locus

<table>
<thead>
<tr>
<th>Locus</th>
<th>A</th>
<th>HE</th>
<th>HO</th>
<th>Primer origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSQ4</td>
<td>16</td>
<td>0.82</td>
<td>0.76</td>
<td>Dow et al. 1995</td>
</tr>
<tr>
<td>QpZAG1/5</td>
<td>17</td>
<td>0.72</td>
<td>0.58</td>
<td>Steinkellner et al. 1997</td>
</tr>
<tr>
<td>QpZAG7</td>
<td>28</td>
<td>0.92</td>
<td>0.84</td>
<td>Steinkellner et al. 1997</td>
</tr>
<tr>
<td>QpZAG9</td>
<td>8</td>
<td>0.28</td>
<td>0.23</td>
<td>Steinkellner et al. 1997</td>
</tr>
<tr>
<td>QpZAG36</td>
<td>13</td>
<td>0.75</td>
<td>0.61</td>
<td>Steinkellner et al. 1997</td>
</tr>
<tr>
<td>QpZAG110</td>
<td>23</td>
<td>0.90</td>
<td>0.88</td>
<td>Steinkellner et al. 1997</td>
</tr>
<tr>
<td>QpZAG11</td>
<td>25</td>
<td>0.81</td>
<td>0.70</td>
<td>Kampfer et al. 1998</td>
</tr>
<tr>
<td>QpZAG20</td>
<td>18</td>
<td>0.88</td>
<td>0.65</td>
<td>Kampfer et al. 1998</td>
</tr>
<tr>
<td>QM69-2M1</td>
<td>20</td>
<td>0.65</td>
<td>0.63</td>
<td>Isagi &amp; Suhandono 1997</td>
</tr>
</tbody>
</table>

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($X|K$)] and $\Delta K$, 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 tstat software with 10000 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 (AR) standardized for sample size ranged from 4.37 to 5.2 alleles per locus (Table 1). AR was not associated with population isolation (i.e. average pairwise population differentiation, FST), current niche suitability or LGM niche suitability (Table 3). However, AR was negatively correlated with niche stability (Table 3; Fig. 3).

Genetic structure

Pairwise FST 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 FST 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 AK 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
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>
<thead>
<tr>
<th>Partial Mantel test</th>
<th>$r$</th>
<th>$P$-values</th>
<th>D</th>
<th>N</th>
<th>L</th>
<th>A</th>
<th>D+N</th>
<th>D+L</th>
<th>D+A</th>
<th>+L</th>
<th>N+A</th>
<th>L+A</th>
<th>D+N</th>
<th>D+L</th>
<th>N+L</th>
<th>D+N+L+A</th>
</tr>
</thead>
<tbody>
<tr>
<td>DG.N</td>
<td>$-0.076$</td>
<td>$0.00010$</td>
<td>S</td>
<td>NS</td>
<td>N</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>DG.L</td>
<td>$-0.072$</td>
<td>$0.00010$</td>
<td>S</td>
<td>NS</td>
<td>N</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>DG.A</td>
<td>$-0.070$</td>
<td>$0.00010$</td>
<td>S</td>
<td>NS</td>
<td>N</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>NG.D</td>
<td>$-0.052$</td>
<td>$0.00010$</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>NG.L</td>
<td>$-0.056$</td>
<td>$0.00010$</td>
<td>S</td>
<td>NS</td>
<td>N</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>NG.A</td>
<td>$-0.056$</td>
<td>$0.00010$</td>
<td>S</td>
<td>NS</td>
<td>N</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>LG.D</td>
<td>$-0.025$</td>
<td>$0.01160$</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>LG.N</td>
<td>$-0.039$</td>
<td>$0.00050$</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>LG.A</td>
<td>$-0.036$</td>
<td>$0.00050$</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>AG.D</td>
<td>$-0.022$</td>
<td>$0.02480$</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>AG.N</td>
<td>$-0.042$</td>
<td>$0.00050$</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>AG.L</td>
<td>$-0.037$</td>
<td>$0.00050$</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>NS</td>
<td>S</td>
<td></td>
</tr>
</tbody>
</table>

Model support rate

0.50 0.50 0.50 0.67 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; Rodriguez-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 (Ramirez-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


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


Guoudet J (2001) Fstat, a Program to Estimate and Test Gene Diversities and Fixation Indices (Version 2.9.3). Institut d’Ecologie, Universite’ de Lausanne, Dorigy, Switzerland.


Ortego J, Bonal R (2010) Natural hybridization between kermes 


Sork VL, Davis FW, Westfall R et al. (2010) Gene movement and genetic association with regional climate gradients in

© 2012 Blackwell Publishing Ltd
California valley oak (Quercus lobata Née) in the face of climate change. *Molecular Ecology, 19*, 3806–3823.


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


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.