Living on the edge: the role of geography and environment in structuring genetic variation in the southernmost populations of a tropical oak

J. Ortego1,2, R. Bonal2,3, A. Muñoz4 & J. M. Espelta3
1 Conservation and Evolutionary Genetics Group, Department of Integrative Ecology, Estación Biológica de Doñana, Seville, Spain
2 Grupo de Investigación de la Biodiversidad Genética y Cultural, Instituto de Investigación en Recursos Cinegéticos, Ciudad Real, Spain
3 Centre de Recerca Ecològica i Aplicacions Forestals, Universitat Autònoma de Barcelona, Barcelona, Spain
4 Departamento de Didáctica de las Ciencias Experimentales, Facultad de Educación, Universidad Complutense de Madrid, Madrid, Spain

Keywords
Ecological niche; genetic diversity; genetic structure; isolation by distance; isolation by ecology; marginal populations; species distribution models.

Correspondence
E-mail: joaquin.ortego@csic.es

Editor
T. Peeters

Received: 24 June 2014; Accepted: 26 September 2014
doi:10.1111/plb.12272

INTRODUCTION
Peripheral populations represent the geographic limit of species natural ranges and have been the focus of a considerable amount of research due to their ecological and evolutionary singularity (Hoffmann & Blows 1994; Sexton et al. 2009). These populations are generally located in fragmented and suboptimal habitats, which has often been associated with their low population densities and fitness in comparison with central populations (Hoffmann & Blows 1994; Brown et al. 1995; Sexton et al. 2011; Castilla et al. 2012). The relative importance of inter-population gene flow and local adaptation processes are the most relevant factors determining the evolutionary, ecological and demographic trajectories of peripheral populations (Hampe & Petit 2005; Sexton et al. 2011). Some marginal populations persist thanks to recurrent immigration and gene flow from core populations (i.e. source–sink metapopulation dynamics), which can increase their effective population sizes and ensure their long-term viability (Hoffmann & Blows 1994; Buschbom et al. 2011; Hampe et al. 2013). In contrast, other peripheral populations have evolved local adaptations to the idiosyncratic environmental conditions prevailing at the species’ range edges (Hoffmann & Blows 1994; Hampe & Bairlein 2000; Mägi et al. 2011; Sexton et al. 2011). These two scenarios lead to very different outcomes from a conservation point of view (Eckert et al. 2008; Guo 2012). Marginal populations that have evolved unique adaptations often sustain an important portion of the species evolutionary potential and may be better adapted than central populations to face some future environmental changes, making them of great conservation concern (Eckert et al. 2008). If marginal populations show no evidence of local adaptation and only constitute anecdotal remnants around species range edges, they can then be considered of limited conservation interest due to their lack of evolutionary significance (Moritz 2002; Eckert et al. 2008; Guo 2012).

From a genetic point of view, marginal populations are generally characterised by a high degree of genetic differentiation...
among them and impoverished within-population genetic diversity in comparison with those located at the core of the species range (Eckert et al. 2008; Guo 2012; Lira-Noriega & Manthev 2014; Yannic et al. 2014). These patterns of genetic diversity and structure can arise from disrupted gene flow and severe genetic drift due to small population sizes and geographic isolation (isolation by distance, IBD; Wright 1943), be consequence of reduced realized gene flow due to selection against non-locally adapted genotypes (isolation by ecology,IBE; sensu Shafer & Wolf 2013; Wang et al. 2013; Sexton et al. 2014) or result from a combination of the two processes. Disentangling the relative role of environment and geography in shaping contemporary patterns of genetic variation can help to discern local adaptation processes from a simple scenario of spatial isolation (Wang 2013; Wang et al. 2013), which can have important implications for the conservation of marginal populations and understanding their evolutionary and demographic dynamics (Eckert et al. 2008). Peripheral populations have mostly been studied in temperate regions in which refugia, post-glacial colonisation routes and spatial patterns of genetic variability have already been described for many species (Hewitt 2000; Hampe & Petit 2005). However, very little information is available for tropical species that show highly stable distribution ranges and have less predictable spatial patterns of genetic variability than those found for species from temperate regions (Eckert et al. 2008; Guo 2012; see Miller et al. 2010 for an exception).

In this study, we analyse patterns of genetic diversity and structure across the southernmost populations of the tropical oak *Quercus segoviensis* Liebm. 1854 (Fagaceae). This species is distributed from southern Mexico to Nicaragua and its peripheral southernmost populations, on which we focus this study, are not likely to have experienced major range changes due to the climatic stability characterising this region (Poelchau & Hamrick 2013). This, together with the high potential for gene flow via long-distance pollen dispersal in oaks (Bushbom et al. 2011; Ortego et al. 2014), makes *Q. segoviensis* an interesting case study to analyse the role of environment and population isolation on structuring genetic variation. For this purpose, we combine information from nuclear microsatellite markers and climate niche modelling. We use niche modelling to identify climatically suitable habitats for this species and then project the present-day niche envelope to the climate conditions present during the Last Glacial Maximum (LGM; ca. 21,000 years BP). This information was used to generate habitat suitability maps in both time periods and study the role of present and past climate on observed patterns of genetic diversity and structure. We expect genetic diversity to (i) be lower in more isolated populations with reduced gene flow with other populations (Wang et al. 2011; Ortego et al. 2012a) and (ii) increase with habitat suitability and stability since the LGM (Carnaval et al. 2009; Yannic et al. 2014). Second, we employed a novel multiple matrix regression approach (Wang 2013) to study the factors structuring genetic variation among the studied populations of *Q. segoviensis* and test (iii) the relative contribution of environmental and geographic isolation on contemporary patterns of genetic differentiation (Shafer & Wolf 2013; Wang 2013; Wang et al. 2013; Sexton et al. 2014).

### MATERIAL AND METHODS

#### Study species and sampling

*Quercus segoviensis* Liebm. 1854 (Fagaceae) is a diploid, wind-pollinated, monoeccious and semi-deciduous tropical oak. It is found in southern Mexico, Honduras, El Salvador and northern Nicaragua, mostly on the slopes of interior valleys at elevations ranging from 650 to 1800 m asl. (http://www.tropicos.org). In 2010, we sampled 112 adult trees from 11 localities in Nicaragua, the southernmost portion of the species range (Table 1, Fig. 1). We aimed to sample 20 adult individuals per population, but very few remnant trees were available in most localities, probably due to low population densities at the limits of the species range in combination with extensive land clearance for agriculture in the region. Spatial coordinates of each individual were recorded using a Global Positioning System (GPS), and leaf samples were stored frozen (−20°C) until needed for genetic analyses.

#### Ecological niche modelling

We used ecological niche modelling to predict the geographic distribution of climatically suitable habitats for *Q. segoviensis* within our study area and analyse whether climatic stability and current and past climate conditions are responsible for observed patterns of genetic diversity and structure. We modelled the current climate-based distribution of *Q. segoviensis* 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 records (Elith et al. 2006; Phillips et al. 2006). The Maxent approach has proved to be very effective for bioclimatic modelling, and performs better with presence-only data than most other available methods (Elith et al. 2006). Species occurrence data were obtained from sampling points as well as from herbarium records available at the Global Biodiversity Information Facility (http://www.gbif.org/). Prior to modelling, all herbarium records were mapped and examined to identify and exclude

#### Table 1. Geographic location and genetic variability for the studied populations of *Quercus segoviensis*.

<table>
<thead>
<tr>
<th>locality</th>
<th>code</th>
<th>latitude</th>
<th>longitude</th>
<th>altitude (m)</th>
<th>n</th>
<th>Ar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jalapa</td>
<td>JAL</td>
<td>13.93333</td>
<td>−86.16667</td>
<td>1240</td>
<td>4</td>
<td>2.25</td>
</tr>
<tr>
<td>San Juan del</td>
<td>SJR</td>
<td>13.56667</td>
<td>−86.15000</td>
<td>950</td>
<td>9</td>
<td>2.49</td>
</tr>
<tr>
<td>Rio Coco</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telpaneca</td>
<td>TEL</td>
<td>13.55000</td>
<td>−86.20000</td>
<td>1280</td>
<td>10</td>
<td>2.56</td>
</tr>
<tr>
<td>Palacaguina</td>
<td>PAL</td>
<td>13.51750</td>
<td>−86.37972</td>
<td>990</td>
<td>11</td>
<td>2.44</td>
</tr>
<tr>
<td>Miraflor</td>
<td>MIR</td>
<td>13.21944</td>
<td>−86.25000</td>
<td>1370</td>
<td>24</td>
<td>2.46</td>
</tr>
<tr>
<td>Yali</td>
<td>YAL</td>
<td>13.21667</td>
<td>−86.13333</td>
<td>1160</td>
<td>10</td>
<td>2.40</td>
</tr>
<tr>
<td>San Rafael del</td>
<td>SRN</td>
<td>13.17722</td>
<td>−86.07583</td>
<td>1050</td>
<td>10</td>
<td>2.33</td>
</tr>
<tr>
<td>Norte</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lago Apanas</td>
<td>APA</td>
<td>13.16887</td>
<td>−85.92184</td>
<td>1000</td>
<td>4</td>
<td>2.37</td>
</tr>
<tr>
<td>Jinoteca</td>
<td>JIN</td>
<td>13.08871</td>
<td>−85.99028</td>
<td>1120</td>
<td>5</td>
<td>2.32</td>
</tr>
<tr>
<td>Cerro Tomabu</td>
<td>TOM</td>
<td>13.01639</td>
<td>−86.28961</td>
<td>1200</td>
<td>10</td>
<td>2.51</td>
</tr>
<tr>
<td>Cerro Tisay</td>
<td>TIS</td>
<td>12.95222</td>
<td>−86.34639</td>
<td>1320</td>
<td>15</td>
<td>2.57</td>
</tr>
</tbody>
</table>

n = number of sampled individuals; Ar = standardised allelic richness.
We used these estimates of habitat suitability for subsequent model validation. Low values indicate conditions are unsuitable for the species to occur, whereas high values indicate that conditions are suitable. We used these estimates of habitat suitability for subsequent analyses of genetic diversity and structure (see below).

We obtained the predicted distribution of *Quercus segoviensis* at the LGM (ca. 21,000 years BP) projecting contemporary species–climate relationships to the LGM. We used the same eight bioclimatic layers from the Community Climate System Model derived from PMIP2 database and available at WorldClim (CCSM3, http://www2.cesm.ucar.edu/; Kiehl & Gent 2004; Collins et al. 2006; Otto-Bliesner et al. 2006). Current and LGM habitat suitability maps were summed to generate maps of climate stability (*sensu* Devitt et al. 2013; Yannic et al. 2014), with pixel values ranging from 0 (minimum climate suitability in both periods) to 2 (maximum climate suitability in both periods). Visualisation of model predictions and all GIS calculations and analyses were performed in ArcMap 10.0 (ESRI, Redlands, CA, USA).

**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 CTAB protocol (Doyle & Doyle 1990). We used 11 polymorphic microsatellite markers previously developed for other *Quercus* species (Table S1). Approximately 5 ng of template DNA was amplified in 10-µl reaction volumes containing 1× reaction buffer (EcoStart Reaction Buffer; Ecogen, Barcelona, Spain), 2 mM MgCl₂, 0.2 m M of each dNTP, 0.15 µM of each dye-labelled primer (FAM, PET, VIC or NED) and 0.1 U *Taq* DNA EcoStart Polymerase (Ecogen). The PCR programme used was 9 min denaturing at 95 °C followed by 40 cycles of 30 s at 94 °C, 45 s at the annealing temperature (Table S1) and 45 s at 72 °C, ending with a 10 min final elongation stage at 72 °C. Amplification products were electrophoresed using an ABI 310 Genetic Analyzer (Applied Biosystems, Carlsbad, CA, USA) and 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). We applied sequential Bonferroni corrections to account for multiple comparisons (Rice 1989).

**Genetic diversity**

To make estimates of allelic richness (A_r) comparable across populations with different sample sizes, we standardised A_r values for each locality to our smallest sample size (four individuals; Table 1). For this purpose, we used the rarefaction procedure implemented in the program HP-RARE (Kalinowski 2005). We analysed which variables related to niche suitability contributed to explain observed patterns of genetic diversity. We considered four explanatory covariates in the models: (i) average genetic differentiation (*F_st*) of each population with all other populations (Wang et al. 2011; Ortego et al. 2012b); (ii) average genetic differentiation corrected for geographic distance (*F_st-gEO*), calculated from the standardised residuals of a linear regression of *F_st* values against inter-population Euclidean geographic distances; (iii) current niche suitability (*NS_CURRENT*); (iv) LGM niche suitability (*NS_LGM*), and (v) niche stability (*NS_STA*). To analyse A_r, we used a General...
Linear Model (GLM) with a normal error structure and an identity link function. The precision of AR estimates may differ among populations due to differences in sample sizes, and we took this into account using a weighted least squares method, where weight equals the sample size for each studied population. Initially, the model was constructed with all explanatory terms fitted and a final model was selected following a backward procedure, by progressively eliminating non-significant variables. The significance of the remaining variables was tested again until no additional variable reached significance. The result is the minimal most adequate model for explaining variability in the response variable, where only the significant explanatory variables are retained. All statistical analyses were performed using the R 3.0.0 package lme4 (R Foundation for Statistical Computing, Vienna, Austria).

Genetic structure
We investigated population genetic structure among sampling locations, calculating pair-wise $F_{ST}$ values and testing their significance with Fisher exact tests after 10,000 permutations, as implemented in ARLEQUIN 3.1 (Excoffier et al. 2005). Critical $P$-values for pair-wise tests of allelic differentiation were determined using a sequential Bonferroni adjustment (Rice 1989). We also analysed the spatial genetic structure using the 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 and using prior population information (Hubisz et al. 2009). We conducted ten independent runs for each value of $K = 1$–10 to estimate the ‘true’ number of clusters with 200,000 MCMC cycles, following burn-in steps of 100,000 iterations. The number of populations best fitting the data set was defined using both log probabilities [Pr(X|K)] (Pritchard et al. 2000; Falush et al. 2000) and the ΔK method (Evanno et al. 2005), as implemented in STRUCTURE HARVESTER (Earl & von Holdt 2012).

Landscape genetic analyses
We considered six potential drivers of genetic structure in Q. segoviensis: (i) geographic distance; (ii) differences in current niche suitability; (iii) differences in LGM niche suitability; (iv) differences in niche suitability stability; (v) environmental dissimilarity in the present; and (vi) the LGM. These six variables were tested against matrices of pair-wise $F_{ST}$ values (see previous section). To generate distance matrices, we calculated the Euclidean distance between niche suitability and stability scores extracted for each population from niche suitability maps obtained from ecological niche models (see also previous section, Ecological niche modelling). Environmental data during the present and the LGM were obtained from the eight bioclimatic layers used to build the ecological niche model (see above). We reduced the number of predictor variables performing a principal components analysis (PCA) using STATISTICA 6.0 (Statsoft Inc., Tulsa, OK, USA). Finally, we calculated the distances between localities plotted on the resulting first three axes, which explained a high proportion of the variance for both the present (84.23%) and the LGM (83.05%; see Wang et al. 2013 for a similar approach). We calculated the matrices of Euclidean geographic distances between populations using GEOGRAPHIC DISTANCE MATRIX GENERATOR 1.2.3 (Ersts 2011).

We used a Multiple Matrix Regression with Randomisation (MMRR) approach to evaluate the factors influencing genetic structure in our study system (Wang 2013). This method allows quantifying how the dependent variable (genetic distance) responds to multiple independent variables that can be simultaneously included into the model. MMRR uses standard multiple regression techniques but performs tests of significance using a randomised permutation procedure because of the non-independence of the data (Smouse et al. 1986; Manly 1991; Legendre et al. 1994). All models were initially constructed with all explanatory terms fitted and final models were selected following a backward procedure as described for analyses of genetic diversity. We used the ‘MMRR function’ as implemented in R version 3.0.0 (Wang 2013).

RESULTS
Niche modelling
The predicted distribution of Q. segoviensis in the present (Fig. 2a) is consistent with its observed current distribution (http://www.tropicos.org). The AUC for the test data was on average 0.901 ± 0.025 (±SD; n = 10 replicate model runs), indicating a high fit of the modelled and actually observed current distribution (Fielding & Bell 1997; Phillips et al. 2006). The estimated distribution of Q. segoviensis during the LGM indicates that the species has had a highly stable distribution range during the last 20,000 years (Fig. 2b and c). However, overall habitat suitability within the study area has slightly increased since the LGM, resulting in higher population connectivity in the present (Fig. 2). Focusing on the studied populations, we found that habitat suitability was highly correlated across both time periods (Pearson’s correlation: $r = 0.968$, $P < 0.001$) but has significantly increased since the LGM (paired t-test: $t = -5.571$, $P < 0.001$). We also found strong differences in predicted habitat suitability of the studied populations: southern populations were located in highly suitable areas whereas northern populations occupied areas with very low habitat suitability scores (Fig. 2).

Microsatellite data and genetic diversity
All microsatellite markers were polymorphic, and observed heterozygosity at each locus ranged from 0.24 to 0.80, with two to 18 alleles per locus (Table 1). After applying sequential Bonferroni corrections to compensate for multiple statistical tests, no locus deviated from Hardy–Weinberg equilibrium in any of the studied populations (all $P > 0.05$). We only found evidence of genotypic linkage disequilibrium between loci PIE020–PIE258 and QpZAG9–QpZAG110 in populations TIS and MIR, respectively. Allelic richness (AR) standardised for sample size ranged from 2.25 to 2.57 alleles per locus (Table 1). Only average $F_{ST}$-GEO was retained into the final model for AR ($t = -4.107$, $P = 0.003$; Fig. 3a), and no other variable remained significant after it was included (all $P > 0.4$). It should be noted that both average $F_{ST}$ and average $F_{ST}$-GEO were highly inter-correlated, and after the exclusion from the model of the variable $F_{ST}$-GEO, AR was negatively associated with average $F_{ST}$ ($t = -2.607$, $P = 0.028$; Fig. 3b). Quadratic
terms and other interactions between independent variables were not significant in any analysis \((P > 0.2)\).

**Genetic structure**

Pair-wise \(F_{ST}\) values ranged from 0.006 to 0.266, and 16 of the 55 pair-wise comparisons were significant after sequential Bonferroni correction (Table S2). Comparisons involving JAL and SRN populations showed particularly high differentiation (Table S2). \textit{STRUCTURE} analyses and the Evanno et al. (2005) method indicated an optimal \(K\) value 3 (Figure S1), but most sampled populations showed a considerable degree of genetic admixture (Fig. 1). The first genetic cluster was mostly represented in the southern populations (TIS and TOM), the second genetic cluster was most frequent in the eastern populations (SRN and YAL), and the probability of population membership to the third cluster was higher in central–northern populations (MIR, SJR, TEL, PAL and JAL; Fig. 1).

**Landscape genetic analyses**

The MMRR analyses showed that only Euclidean geographic distance \((\beta = 0.582, t = 4.79, P = 0.025; \text{Fig. 4})\) was retained in the final model \((r^2 = 0.302)\), and no other variable remained significant when included (all \(Ps > 0.2\)). Current habitat suitability \((t = 2.14, P = 0.17)\), LGM habitat suitability \((t = 1.16, P = 0.390)\), habitat stability \((t = 1.75, P = 0.207)\) or environmental dissimilarity in the present \((\text{PC1}: t = 7.08, P = 0.061; \text{PC2}: t = 0.59, P = 0.670; \text{PC3}: t = -0.393, P = 0.807)\) or the LGM \((\text{PC1}: t = 6.59, P = 0.075; \text{PC2}: t = 2.63, P = 0.120; \text{PC3}: t = 1.22, P = 0.130)\) were not significant when included alone into different models, indicating that the lack of correlation between genetic distance and these predictors was not due to interactions among independent variables.

**DISCUSSION**

Climate niche modelling indicates that the distribution of the southernmost populations of \textit{Q. segoviensis} has remained highly stable for at least the last 20,000 years (Fig. 2). Despite this regional stability, niche modelling also revealed that habitat suitability in the study area has slightly increased since the LGM, and showed remarkable geographic heterogeneity, with the four northernmost studied populations (JAL, SJR, TEL and PAL) having particularly low suitability scores in comparison with the southern localities (Fig. 2). The climatic spatial heterogeneity and long-term stability of this tropical region offers the ideal template for evolution of local adaptations that may shape patterns of genetic variability and structure in the studied populations of \textit{Q. segoviensis} (Ortego et al. 2012a, b; Wang et al. 2013).

Considering the relatively small size of the study area (<120 km), analyses on spatial genetic structure indicate remarkable genetic differentiation among the southernmost populations of \textit{Q. segoviensis} (Fig. 1, Table 1). Bayesian analyses of genetic structure indicate the presence of three genetic clusters and some pair-wise \(F_{CT}\) values were higher than those previously reported for oaks from temperate areas sampled at a similar or much larger spatial scale (Ramirez-Valiente et al. 2009; Alberto et al. 2010; Zeng et al. 2011; Ortego et al. 2012a, b). \textit{STRUCTURE} analyses also indicate a geographic cline of genetic structure, with the three distinct genetic clusters being
distributed in the south, central–east, and northwest portions of the study area. Bayesian analyses also revealed a considerable degree of genetic admixture, and several populations showed a high probability of population membership to different clusters (Fig. 1), suggesting that observed genetic differentiation is maintained in the presence of inter-population gene flow.

Despite the significant spatial genetic structure (Fig. 1, Table S2) and important environmental heterogeneity across the study area (Fig. 2), MMRR analyses revealed that geographic distance is the only factor explaining observed patterns of genetic differentiation. The observed isolation by distance pattern of genetic structure suggests equilibrium between gene flow and drift (Hutchison & Templeton 1999), which in the case of oaks, is likely to be driven by long-distance pollen movement (Buschbom et al. 2011; Ortego et al. 2014) and local seed dispersal (Grivet et al. 2006). Niche suitability and environmental dissimilarity summarising the current and past climatic conditions experienced by the studied populations had no significant effect on genetic differentiation after controlling for the effects of geographic distance, indicating that geographic isolation (isolation by distance; Wright 1943) but not adaptation to local climatic environments (i.e. isolation by ecology; Shafer & Wolf 2013; Wang et al. 2013; Sexton et al. 2014) is behind observed patterns of genetic structure. This contrasts with previous studies that found an important role of environment in structuring genetic variation in oaks after removing the effects of geography (Sork et al. 2010; Ortego et al. 2012a; Gugger et al. 2013). Some of these studies have compared populations distributed across a large geographic area, which is likely to increase the range of environmental conditions experienced by the different populations, attenuate the homogenising effects of gene flow and favour genetic divergence by local adaptation (Sork et al. 2010; Gugger et al. 2013). However, other studies have found environmental correlates of genetic structure across geographically close populations, suggesting that local adaptation and subsequent selection against immigrant genotypes could occur at relatively small spatial scales, even in wind-pollinated species with extraordinary dispersal potential (Alberto et al. 2010; Ortego et al. 2012a). The lack of signal in isolation by ecology analyses may occur for several biological reasons, including adaptation to local environments via phenotypic plasticity (Ramirez-Valiente et al. 2010), positive selection on immigrant genotypes from distant populations mediated by heterosis (Bensch et al. 2006) or as a consequence of long-distance gene flow counteracting the effects of natural selection and impeding or attenuating local adaptation processes (Buschbom et al. 2011). It should be noted that we cannot totally reject the hypothesis of isolation by ecology, given that other parameters (e.g. soil characteristics, nutrient availability, etc.) not considered in our study could potentially shape the patterns of genetic structure found here (Macel et al. 2007; Freeland et al. 2010). Finally, it must also be considered that small sample sizes in some localities (Table 1) may have also reduced our statistical power to detect isolation by ecology, which generally explains a lower proportion of
variance in genetic divergence than isolation by distance (Wang et al. 2013).

Our data indicate that niche stability or current or past climatic suitability were not correlated with intra-population genetic diversity, suggesting that these variables are not directly associated with local effective population sizes. The fact that the study area is climatically highly stable may explain the lack of association between genetic diversity and habitat stability, a pattern that has been previously reported in species from regions with more fluctuating climates (Carnaval et al. 2009; Yannic et al. 2014). However, genetic diversity was negatively correlated with average genetic differentiation in all other populations, indicating that isolation and limited gene flow have contributed to erode genetic variability in some populations (Ortego et al. 2010; Wang et al. 2011). This indicates that effective population sizes of the studied populations are not above a threshold that prevents the loss of genetic diversity and/or that inter-population gene flow suggested by observed patterns of admixture is not sufficient to counterbalance the effects of genetic drift.

Overall, our data point to geographic isolation as the main factor structuring genetic variation within the peripheral populations of Q. segovierensis. We have found strong genetic subdivision within our relatively small study area, supporting the hypothesis of fragmentation of peripheral populations in this tropical oak species. Further studies analysing the complete distribution range of this and other tropical species could help to further understand the demographic and evolutionary dynamics of peripheral populations. In these biomes, species have scarcely been impacted by Pleistocene glacial cycles, and their genetic patterns of genetic diversity and structure can be highly different from those reported in the much better studied temperate regions (Hewitt 2000; Eckert et al. 2008; Guo 2012).

ACKNOWLEDGEMENTS

We wish to thank Conchi Cáliz for sample genotyping and Marcelo J. Sturaro for valuable advice about GIS and niche modelling. Two anonymous referees provided useful discussion and valuable comments on an earlier draft of this manuscript. JO and RB were supported by Severo Ochoa, Juan de la Cierva (MICINN) and JAE-Doctor (CSIC) post-doctoral fellowships. This work received financial support from grants D/7592/07 (AECID) and UNCM08-1E-018 (FEDER). The study also benefited from discussion and information exchange in the IBER-REDD+ network (P411RT0559-CYTED).

SUPPORTING INFORMATION

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

Figure S1. Results of Bayesian clustering analyses in STRUCTURE.

Table S1. Microsatellite loci used to genotype Quercus segovierensis: number of alleles ($A$), expected heterozygosity ($H_e$), observed heterozygosity ($H_o$) and annealing temperature ($T_a$, in °C) for each locus.

Table S2. Pair-wise population $F_{ST}$ values.

REFERENCES


