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

Joaquín Ortego[1,2], Raúl Bonal[2,3], Alberto Muñoz[4] & Josep M. Espelta[3]

1 Conservation and Evolutionary Genetics Group, Department of Integrative Ecology, Estación Biológica de Doñana - EBD (CSIC), Seville, Spain

2 Grupo de Investigación de la Biodiversidad Genética y Cultural, Instituto de Investigación en Recursos Cinegéticos - IREC (CSIC, UCLM, JCCM), Ciudad Real, Spain

3 Centre de Recerca Ecològica i Aplicacions Forestals (CREAF), Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain

4 Departamento de Didáctica de las Ciencias Experimentales. Facultad de Educación. Universidad Complutense de Madrid, Madrid, Spain

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

Running title: Genetic variability in peripheral oak populations

Keywords: ecological niche, genetic diversity, genetic structure, isolation by distance, isolation by ecology, marginal populations, species distribution models

ABSTRACT

Understanding the factors determining genetic diversity and structure in peripheral populations is a long-standing goal of evolutionary biogeography, yet little empirical information is available for tropical species. In this study, we combine information from nuclear microsatellite markers and niche modelling to analyze the factors structuring genetic variation across the southernmost populations of the tropical oak *Quercus segoviensis*. First, we tested the hypothesis that genetic variability decreases with population isolation and increases with local habitat suitability and stability since the Last Glacial Maximum (LGM). Second, we employed a recently developed multiple matrix regression with randomization (MMRR) approach to study the factors associated with genetic divergence among the studied populations and test the relative contribution of environmental and geographic isolation to contemporary patterns of genetic differentiation. We found that genetic diversity was negatively correlated with average genetic differentiation with other populations, indicating that isolation and limited gene flow have contributed to erode genetic variability in some populations. Considering the relatively small size of the study area (<120 km), analyses of genetic structure indicate a remarkable inter-population genetic differentiation. Environmental dissimilarity and differences in current and past climatic niche suitability and their additive effects were not associated with genetic differentiation after controlling for geographic distance, indicating that local climate does not contribute to explain spatial patterns of genetic structure. Overall, our data indicate that geographic isolation, but not current or past climate, is the main factor determining contemporary patterns of genetic diversity and structure within the southernmost peripheral populations of this tropical oak.

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 sub-optimal habitats, which has been often 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; e.g. 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 (e.g. 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 anecdotic remnants around species

range edges, they can be then 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 characterized 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 & Manthey 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 *et al*. 2013; Wang *et al*. 2013; Sexton *et al*. 2014) or result from a combination of both processes. Disentangling the relative role of environment and geography on 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 marginal populations and understanding their evolutionary and demographic dynamics (Eckert *et al*. 2008). Peripheral populations have been mostly studied in temperate regions in which refugia, post-glacial colonization routes and spatial patterns of genetic variability have been already described for many species (Hewitt 2000; Hampe & Petit 2005). However, very little information is still 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 analyze patterns of genetic diversity and structure across the southernmost populations of the tropical oak *Quercus segoviensis* Liebm. 1854

(Fagaceae). This species distributes 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 characterizing this region (Poelchau & Hamrick 2013). This, together with the high potential for gene flow via long-distance pollen dispersal in oaks (e.g. Buschbom *et al*. 2011; Ortego *et al*. 2014), makes *Q. segoviensis* an interesting case study to analyze the role of environment and population isolation on structuring genetic variation. For this purpose, we combine information from nuclear microsatellite markers and climatic niche modelling. We use niche modelling to identify climatically suitable habitats for this species and then project the present-day niche envelope to the climatic conditions present during the Last Glacial Maximum (LGM; ca. 21000 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. First, we assessed which factors contribute to explain the levels of genetic diversity observed in the studied peripheral populations. We expect genetic diversity (i) to be lower in more isolated populations with reduced gene flow with other populations (Ortego *et al*. 2012; Wang *et al*. 2011) and (ii) to increase with habitat suitability and stability since the Last Glacial Maximum (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 *Quercus segoviensis* and test (iii) the relative contribution of environmental and geographic isolation to contemporary patterns of genetic differentiation (Wang 2013; Wang *et al*. 2013; Shafer & Wolf 2013; Sexton *et al*. 2014).

MATERIAL AND METHODS

Study species and sampling

Quercus segoviensis Liebm. 1854 (Fagaceae), is a diploid, wind-pollinated, monoecious and semi-deciduous tropical oak. It is distributed in southern Mexico, Honduras, El Salvador and northern Nicaragua, mostly in the slopes of interior valleys at elevations ranging from of 650 m to 1800 m above sea level (http://www.tropicos.org). In 2010, we sampled 112 adult trees from 11 localities located 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 clearing for agriculture in the region. Spatial coordinates of each individual were registered using a Global Positioning System (GPS) and leaf samples were stored frozen (-20º C) until needed for genetic analyses.

Ecological niche modelling

We used ecological niche modelling (ENM) to predict the geographic distribution of climatically suitable habitats for *Q. segoviensis* within our study area and analyze whether climatic stability and current and past climatic conditions are responsible for observed patterns of genetic diversity and structure. We modelled the current climate-based distribution of *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 proven 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 records having obvious georeferencing errors and misidentifications. For models, all locations falling within the same grid cell were also removed, resulting in a final data set of 54 entries within our study area (Fig. 1). We selected variables from a set of 19 bioclimatic layers from the WorldClim dataset (version 1.4, see http://www.worldclim.org/ for variable descriptions) interpolated to 30-arcsec (c. 1-km) resolution (Hijmans *et al*. 2005). We assessed the correlation between these bioclimatic layers using ENMTOOLS (Warren *et al*. 2010). When two layers were highly correlated ($r > 0.7$) we discarded the layer with the highest number of correlations with other layers. We selected a final set of eight bioclimatic layers to construct the models: isothermality (Bio3), temperature seasonality (Bio4), temperature annual range (Bio7), mean temperature of driest quarter (Bio 9), annual precipitation (Bio12), precipitation seasonality (Bio15), precipitation of warmest quarter (Bio18), and precipitation of coldest quarter (Bio19). Model evaluation statistics were produced from 10 cross-validation replicate model runs. 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 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 *Q. segoviensis* at the Last Glacial Maximum (LGM; c. 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://www.ccsm.ucar.edu/; Kiehl & Gent 2004; Otto-Bliesner *et al*. 2006; Collins *et al*. 2006). Current and LGM habitat suitability maps were summed to generate maps of climatic stability (*sensu* Devitt *et al*. 2013; Yannic *et al*. 2014), with pixel values ranging from 0 (minimum climatic suitability in both periods) to 2 (maximum climatic suitability in both periods). Visualization of model predictions and all GIS calculations and analyses were performed in ARCMAP 10.0 (ESRI, Redlands, CA, USA).

Microsatellite genotyping 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 1X reaction buffer (EcoStart Reaction Buffer, Ecogen), 2 mM $MgCl_2$, 0.2 mM of each dNTP, 0.15 µM of each dye-labelled primer (FAM, PET, VIC or NED) and 0.1 U of *Taq* DNA EcoStart Polymerase (Ecogen). The PCR programme used was 9 min denaturing at 95 ºC followed by 40 cycles of 30 s at 94 ºC, 45 s at the annealing temperature (Table S1) and 45 s at 72 ºC, ending with a 10 min

final elongation stage at 72 ºC. Amplification products were electrophoresed using an ABI 310 Genetic Analyzer (Applied Biosystems) 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 calculated A_R values for each locality standardized 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 analyzed which variables related with niche suitability contributed to explain observed patterns of genetic diversity. We considered four explanatory covariates in the models: 1) Average genetic differentiation (F_{ST}) of each population with all other populations (e.g. Wang *et al*. 2011; Ortego *et al*. 2012); 2) Average genetic differentiation corrected for geographical distance ($F_{ST\text{-}GEO}$), calculated from the standardized residuals of a linear regression of F_{ST} values against inter-population Euclidean geographical distances; 2) current niche suitability ($NS_{CURRENT}$); 3) LGM niche suitability (NS_{LGM}); 4) Niche stability (NS_{STA}). To analyze A_R we used a General Linear Model (GLM) with a normal

error structure and an identity link function. The precision of A_R estimates may differ among populations due to differences in sample sizes and we took this into account using a weighted least square method, where weight equals the sample size for each studied population. Initially, the model was constructed with all explanatory terms fitted and 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 the 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 Development Core Team 2012).

Genetic structure

We investigated population genetic structure among sampling locations calculating pair-wise F_{ST}-values and testing their significance with Fisher's 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 analyzed 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 200000 MCMC cycles, following a burn-in steps of 100000 iterations. The number

of populations best fitting the data set was defined both using log probabilities [Pr($X|K$)] (Pritchard *et al*. 2000) and the ΔK method (Evanno *et al.* 2005), as implemented in STRUCTURE HARVESTER (Earl & vonHoldt 2012).

Landscape genetic analyses

We considered six potential drivers of genetic structure in *Q. segoviensis*: (1) the geographical distance; (2) differences in current niche suitability; (3) differences in LGM niche suitability; (4) differences in niche suitability stability; (5) environmental dissimilarity in the present and (6) 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 the previous section "Ecological niche modelling"). Environmental data during the present and the LGM were obtained from the eight bioclimatic layers used to build the ENMs (see above). We reduced the number of predictor variables performing a principal components analysis (PCA) using STATISTICA 6.0 (Statsoft. Ltd, Sweden). Finally, we calculated the distances between localities plotted on the resulting three first 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 geographical distances between populations using GEOGRAPHIC DISTANCE MATRIX GENERATOR 1.2.3 (Ersts 2011).

We used a Multiple Matrix Regression with Randomization (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 randomized permutation procedure because of the non-independence of the data (Manly 1991; Smouse *et al*. 1986; 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 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 (S.D. = 0.025; n = 10 replicate model runs), indicating a high fit of the modelled and the actually observed current distribution (Fielding & Bell 1997; Phillips *et al*. 2006). The estimated distribution of *Q. segoviensis* during the LGM indicates that the species had a highly stable distribution range during the last 20 000 years (Fig. 2b, c). However, overall habitat suitability within the study area has slightly increased since the LGM, resulting in increased 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 it 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 2-18 alleles per locus (Table 1). After applying sequential Bonferroni corrections to compensate for multiple statistical tests, no locus deviated from HWE 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 (A_R) standardized for sample size ranged from 2.25 to 2.57 alleles per locus (Table 1). Only average $F_{ST\text{-}GEO}$ was retained into the final model for A_R ($t = -4.107$, $P = 0.003$; Fig. 3a) and no other variable remained significant after it was included (all Ps > 0.4). It should be noted that both average F_{ST} and average $F_{ST\text{-}GEO}$ were highly intercorrelated and after the exclusion from the model of the variable $F_{ST\text{-}GEO}$, A_R 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). STRUCTURE analyses and the Evanno *et al*. (2005) method

indicated an optimal value of $K = 3$ (Fig. 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 the 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

MMRR analyses showed that only Euclidean geographical distance ($\beta = 0.582$, $t = 4.79$, $P = 0.025$; Fig. 4) was retained into the final model ($r^2 = 0.302$) and no other variable remained significant after it was 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 (PC1: $t = 7.08$, $P = 0.061$; PC2: $t = 0.59$, $P = 0.670$; PC3: $t = -0.393$, $P = 0.807$) or the LGM (PC1: $t = 6.59$, $P = 0.075$; PC2: $t = 2.63$, $P = 0.120$; PC3: $t = 1.22$, $P = 0.130$) were not significant when they were 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 *Q. segoviensis* has remained highly stable at least during the last 20000 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 a 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 the evolution of local adaptations that may shape patterns of genetic variability and structure in the studied populations of *Q. segoviensis* (e.g. Ortego *et al*. 2012; Wang *et al*. 2013).

Considering the relatively small size of the study area (<120 km), analyses on spatial genetic structure indicate a remarkable genetic differentiation among the southernmost populations of *Q. segoviensis* (Fig. 1; Table 1). Bayesian analyses of genetic structure indicate the presence of three genetic clusters and some pair-wise F_{ST} values were higher than those previously reported for oaks from temperate areas sampled at a similar or much larger spatial scale (e.g. Ramirez-Valiente *et al*. 2009; Alberto *et al*. 2010; Zeng *et al*. 2011; Ortego *et al*. 2012). STRUCTURE analyses also indicate a geographic cline of genetic structure, with the three distinct genetic clusters being distributed in the south, central-east, and north-west 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 presence of inter-population gene flow.

Despite we found 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 IBD pattern of genetic structure suggests equilibrium between gene flow and drift (Hutchinson & 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 summarizing 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 (IBD; Wright 1943) but not adaptation to local climatic environments (i.e. IBE; Shafer *et al*. 2013; Wang *et al*. 2013; Sexton *et al*. 2014) is behind observed patterns of genetic structure. This contrasts with previous studies that have found an important role of environment on structuring genetic variation in oaks after removing the effects of geography (Sork *et al*. 2010; Ortego *et al*. 2012; Gugger *et al*. 2013). Some of these studies have compared populations distributed across a large geographical 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*. 2012). The lack of signal of IBE analyses may be due to different biological reasons, including adaptation to local environments via phenotypic plasticity (Ramírez-Valiente *et al*. 2010), positive selection on immigrant genotypes from distant populations mediated by heterosis (Bensch *et al*. 2006) or consequence of long-distance gene flow counteracts the effects of natural selection and impedes or attenuates local adaptation processes (Buschbom *et al*. 2011). It should be noted that we cannot totally reject the hypothesis of IBE given that other parameters (e.g. soil characteristics,

nutrient availability, etc.) not considered in our study could be potentially shaping the patterns of genetic structure we found (e.g. Macel *et al*. 2007; Freeland *et al*. 2010). Finally, it must be also considered that small sample sizes in some localities (Table 1) may have also reduced our statistical power to detect IBE, which generally explains a lower proportion of variance in genetic divergence than IBD (Wang *et al*. 2013).

Our data indicate that niche stability or current or past climatic suitability were not associated 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 with 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 lost of genetic diversity and/or that inter-population gene flow suggested by observed patterns of admixture is not enough to counterbalance the effects of genetic drift.

Overall, our data points to geographic isolation as the main factor structuring genetic variation within the peripheral populations of *Q. segoviensis*. 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 analyzing 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 been scarcely impacted by

Pleistocene glacial cycles and their geographic patterns of genetic diversity and structure can greatly differ from those reported at the much better studied temperate regions (Hewitt 2000; Eckert *et al*. 2008; Guo 2012).

ACKNOWLEDGEMENTS

We wish to thank to 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-Doc (CSIC) postdoctoral fellowships. This work received financial support from grants D/7592/07 (AECID) and UNCM08-1E-018 (FEDER). The study has also benefited from discussion and information exchange in the IBERO-REDD+ network (P411RT0559-CYTED).

SUPPORTING INFORMATION

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

Table S1. Microsatellite loci used to genotype *Quercus segoviensis*: 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. Values in bold are statistically significant after sequential Bonferroni correction ($P < 0.05$).

Fig. S1. Results of Bayesian clustering analyses in STRUCTURE. Plots show the mean (± S.D.) log probability of the data (ln Pr(X|*K*)) over 10 runs (left axis, black dots and error

bars) for each value of *K*. The magnitude of Δ*K* as a function of *K* indicates the most likely number of genetic clusters (*K* = 3) in the STRUCTURE analyses (right axis, open dots) (Evanno *et al*. 2005).

References

Alberto, F., Niort, J., Derory, J., Lepais, O., Vitalis, R., Galop, D., Kremer, A. (2010) Population differentiation of sessile oak at the altitudinal front of migration in the French Pyrenees. *Molecular Ecology*, **19**, 2626-2639.

Bensch, S., Andren, H., Hansson, B., Pedersen, H.C., Sand, H., Sejberg, D., Wabakken, P., Akesson, M., Liberg, O. (2006) Selection for heterozygosity gives hope to a wild population of inbred wolves. *PLOS One*, **1**, e72.

Brown, J.H., Mehlman, D.W., Stevens, G.C. (1995) Spatial variation in abundance. *Ecology*, **76**, 2028-2043.

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

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

Castilla, A.R., Alonso, C., Herrera, C.M. (2012) Genetic structure of the shrub *Daphne laureola* across the Baetic Ranges, a Mediterranean glacial refugium and biodiversity hotspot. *Plant Biology*, **14**, 515-524.

Collins, W.D., Bitz, C.M., Blackmon, M.L., Bonan, G.B., Bretherton, C.S., Carton, J.A., Chang, P., Doney, S.C., Hack, J.J., Henderson, T.B., Kiehl, J.T., Large, W.G.,

McKenna, D.S., Santer, B.D., Smith, R.D. (2006) The Community Climate System Model version 3 (CCSM3). *Journal of Climate*, **19**, 2122-2143.

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

Dow, B.D., Ashley, M.V., Howe, H.F. (1995) Characterization of highly variable (Ga/Ct)(N) microsatellites in the bur oak, *Quercus macrocarpa*. *Theoretical and Applied Genetics*, **91**, 137-141.

Doyle, J., Doyle, J. (1990) Isolation of plant DNA from fresh tissue. *Focus*, **12**, 13-15.

Durand, J., Bodenes, C., Chancerel, E., Frigerio, J.M., Vendramin, G., Sebastiani, F., Buonamici, A., Gailing, O., Koelewijn, H.P., Villani, F., Mattioni, C., Cherubini, M., Goicoechea, P.G., Herran, A., Ikaran, Z., Cabane, C., Ueno, S., Alberto, F., Dumoulin, P.Y., Guichoux, E., de Daruvar, A., Kremer, A., Plomion, C. (2010) A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. *BMC Genomics*, **11**, 570

Earl, D.A. and vonHoldt, B.M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359-361.

Eckert, C.G., Samis, K.E., Lougheed, S.C. (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology*, **17**, 1170-1188.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E.,

Soberon, J., Williams, S., Wisz, M.S., Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129-151.

Ersts, P.J. (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, J.K. (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, **164**, 1567-1587.

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

Freeland, J.R., Biss, P., Conrad, K.F., 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.

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

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

Guo, Q.F. (2012) Incorporating latitudinal and central-marginal trends in assessing genetic variation across species ranges. *Molecular Ecology*, **21**, 5396-5403.

Guo, S.W. and Thompson, E.A. (1992) A monte-carlo method for combined segregation and linkage analysis. *American Journal of Human Genetics*, **51**, 1111-1126.

Hampe, A. and Bairlein, F. (2000) Modified dispersal-related traits in disjunct populations of bird-dispersed *Frangula alnus* (Rhamnaceae): a result of its Quaternary distribution shifts? *Ecography*, **23**, 603-613.

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

Hampe, A., Pemonge, M.H., Petit, R.J. (2013) Efficient mitigation of founder effects during the establishment of a leading-edge oak population. *Proceedings of the Royal Society B-Biological Sciences*, **280**, 1764.

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

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

Hoffmann, A.A. and Blows, M.W. (1994) Species borders - Ecological and evolutionary perspectives. *Trends in Ecology & Evolution*, **9**, 223-227.

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

Hutchison, D.W. and Templeton, A.R. (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.

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

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

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

Legendre, P., Lapointe, F.J., Casgrain, P. (1994) Modeling brain evolution from behavior - A permutational regression approach. *Evolution*, **48**, 1487-1499.

Lira-Noriega, A., Manthey, J.D. (2014) Relationship of genetic diversity and niche centrality: A survey and analysis. *Evolution*, in press.

Macel, M., Lawson, C.S., Mortimer, S.R., Smilauerova, M., Bischoff, A., Cremieux, L., Dolezal, J., Edwards, A.R., Lanta, V., Bezemer, T.M., van der Putten, W.H., Igual, J.M., Rodriguez-Barrueco, C., Muller-Scharer, H., Steinger, T. (2007) Climate vs. soil factors in local adaptation of two common plant species. *Ecology*, **88**, 424-433.

Mägi, M., Semchenko, M., Kalamees, R., Zobel, K. (2011) Limited phenotypic plasticity in range-edge populations: a comparison of co-occurring populations of two *Agrimonia* species with different geographical distributions. *Plant Biology*, **13**, 177-184.

Manly, B.F.J. (1991) *Randomization and Monte Carlo methods in biology*. New York: Chapman & Hall.

Miller, M.J., Bermingham, E., Klicka, J., Escalante, P., Winker, K. (2010) Neotropical birds show a humped distribution of within-population genetic diversity along a latitudinal transect. *Ecology Letters*, **13**, 576-586.

Moritz, C. (2002) Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology*, **51**, 238-254.

Ortego, J., Aguirre, M.P., Cordero, P.J. (2012a) Landscape genetics of a specialized grasshopper inhabiting highly fragmented habitats: a role for spatial scale. *Diversity and Distributions*, **18**, 481-492.

Ortego, J., Bonal, R., Munoz, A. (2010) 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., Riordan, E.C., Gugger, P.F., Sork, V.L. (2012b) Influence of environmental heterogeneity on genetic diversity and structure in an endemic southern Californian oak. *Molecular Ecology*, **21**, 3210-3223.

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

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

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

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

Poelchau, M.F. and Hamrick, J.L. (2013) Palaeodistribution modelling does not support disjunct Pleistocene refugia in several Central American plant taxa. *Journal of Biogeography*, **40**, 662-675.

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

Ramirez-Valiente, J.A., Lorenzo, Z., Soto, A., Valladares, F., Gil, L., Aranda, I. (2009) Elucidating the role of genetic drift and natural selection in cork oak differentiation regarding drought tolerance. *Molecular Ecology*, **18**, 3803-3815.

Ramirez-Valiente, J.A., Sanchez-Gomez, D., Aranda, I., Valladares, F. (2010) Phenotypic plasticity and local adaptation in leaf ecophysiological traits of 13 contrasting cork oak populations under different water availabilities. *Tree Physiology*, **30**, 618-627.

Rice, W.R. (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223-225.

Sexton, J.P., McIntyre, P.J., Angert, A.L., Rice, K.J. (2009) Evolution and ecology of species range limits. *Annual Review of Ecology Evolution and Systematics*, **40**, 415-436.

Sexton, J.P., Strauss, S.Y., Rice, K.J. (2011) Gene flow increases fitness at the warm edge of a species' range. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 11704-11709.

Sexton, J.P., Hangartner, S.B., Hoffmann A.A. (2014) Genetic isolation by environment or distance: Which patterns of gene flow is most common? *Evolution*, **68**, 1-15.

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

Smouse, P.E., Long, J.C., Sokal, R.R. (1986) Multiple-regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627-632.

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

Steinkellner, H., Fluch, S., Turetschek, E., Lexer, C., Streiff, R., Kremer, A., Burg, K., Glossl, J. (1997) Identification and characterization of (GA/CT)(n)-microsatellite loci from *Quercus petraea*. *Plant Molecular Biology*, **33**, 1093-1096.

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

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

Wang, I.J., Johnson, J.R., Johnson, B.B., Shaffer, H.B. (2011) Effective population size is strongly correlated with breeding pond size in the endangered California tiger salamander, *Ambystoma californiense*. *Conservation Genetics*, **12**, 911-920.

Warren, D.L., Glor, R.E., Turelli, M. (2010) ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography*, **33**, 607-611.

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

Yannic, G., Pellissier, L., Ortego, J., Lecomte, N., Couturier, S., Cuyler, C., Dussault, C., Hundermark, K.J., Irvine, R.J., Jenkins, D.A., Kolpashikov, L., Mager, K., Musiani, M., Parker, K.L., Roed, K.H., Sipko, T., Þórisson, S.G., Weckworth, B.V.,

Guisan, A., Bernatchez, L., Côté, S.D. (2014) Genetic diversity in caribou linked to past and future climate change. *Nature Climate Change*, **4**, 132-137.

Zeng, Y.F., Liao, W.J., Petit, R.J., Zhang, D.Y. (2011) Geographic variation in the structure of oak hybrid zones provides insights into the dynamics of speciation. *Molecular Ecology*, **20**, 4995-5011.

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

Locality	Code	Latitude	Longitude	Altitude (m)	*N*	A_R
Jalapa	JAL	13.93333	-86.16667	1240	4	2.25
San Juan del Río Coco	SJR	13.56667	-86.15000	950	9	2.49
Telpaneca	TEL	13.55000	-86.20000	1280	10	2.56
Palacagüina	PAL	13.51750	-86.37972	990	11	2.44
Miraflor	MIR	13.21944	-86.25000	1370	24	2.46
Yali	YAL	13.21667	-86.13333	1160	10	2.40
San Rafael del Norte	SRN	13.17722	-86.07583	1050	10	2.33
Lago Apanas	APA	13.16887	-85.92184	1000	4	2.37
Jinotega	JIN	13.08871	-85.99028	1120	5	2.32
Cerro Tomabu	TOM	13.01639	-86.29861	1200	10	2.51
Cerro Tisey	TIS	12.95222	-86.34639	1320	15	2.57

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

Figure legends

Fig. 1. Sampling sites of *Quercus segoviensis* and genetic assignment of populations based on the Bayesian method implemented in the program STRUCTURE considering three genetic clusters. The admixture proportions generated by STRUCTURE were represented using pie charts, with each colour indicating a different genotypic cluster. Pie chart size is proportional to the number of individuals sampled at each location. Population codes are described in Table 1.

Fig. 2. (a) Distribution of *Quercus segoviensis* (dashed line) based on herbarium records (open dots) and the location of the study area in the southernmost portion of the species range (open square). Ecological niche modelling of *Quercus segoviensis* for (a) the present and (b) the Last Glacial Maximum (LGM; c. 21 000 years BP). The LGM distribution was modelled using the CCSM3 climatic model. Panel (c) shows habitat stability estimated from current and LGM habitat suitability maps. Bull eyes in panels b-d indicate sampling locations and dashed lines represent the border between Honduras and Nicaragua.

Fig. 3. Relationship between standardized allelic richness (A_R) and (a) average population differentiation corrected for geographical distance ($F_{ST\text{-}GEO}$) and (b) average population differentiation with all other studied populations (F_{ST}).

Fig. 4. Relationship between genetic distance (F_{ST}) and Euclidean geographic distance.

Figure 1

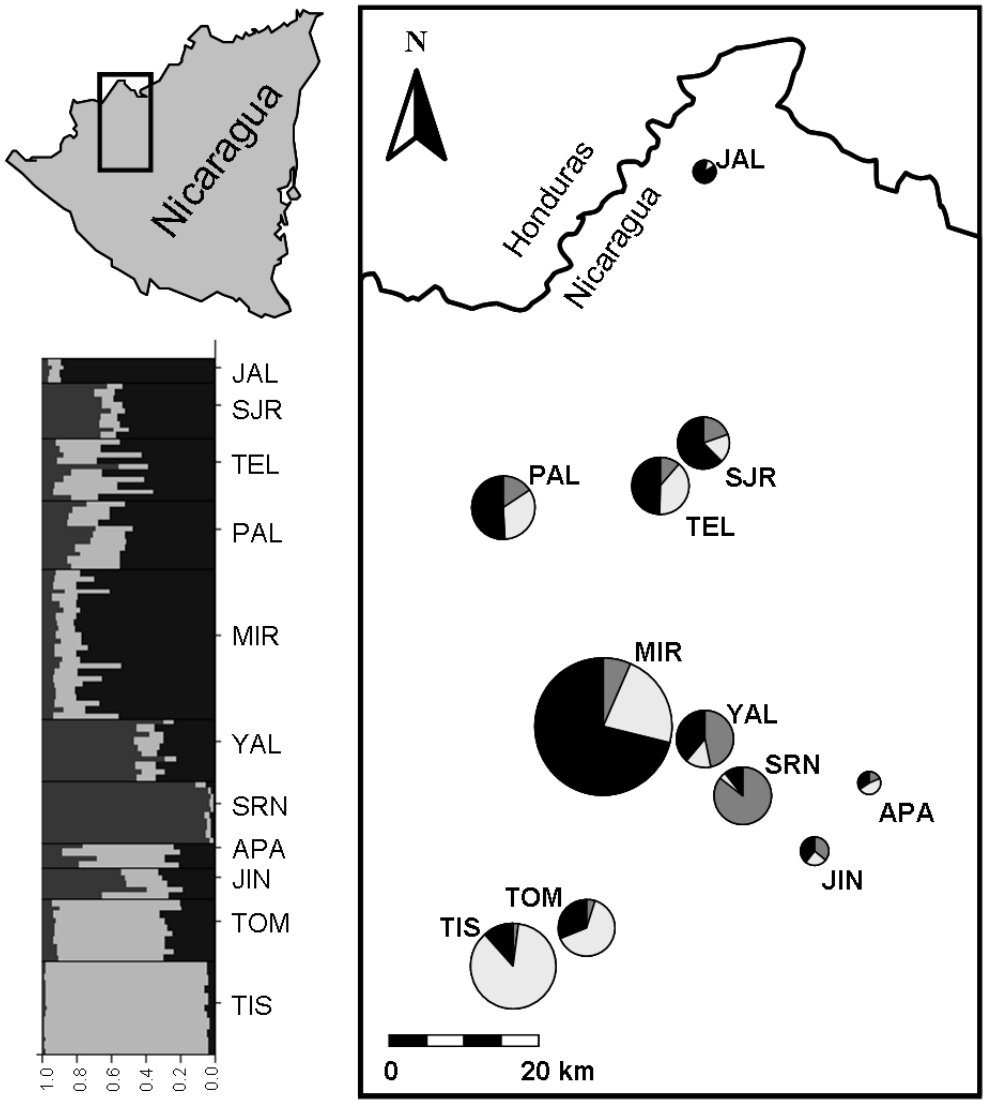

Nicaragua
N
Honduras
Nicaragua
JAL
PAL
SJR
TEL
MIR
YAL
SRN
APA
JIN
TOM
TIS
0
20 km
JAL
SJR
TEL
PAL
MIR
YAL
SRN
APA
JIN
TOM
TIS
1.0
0.8
0.6
0.4
0.2
0.0

Figure 2

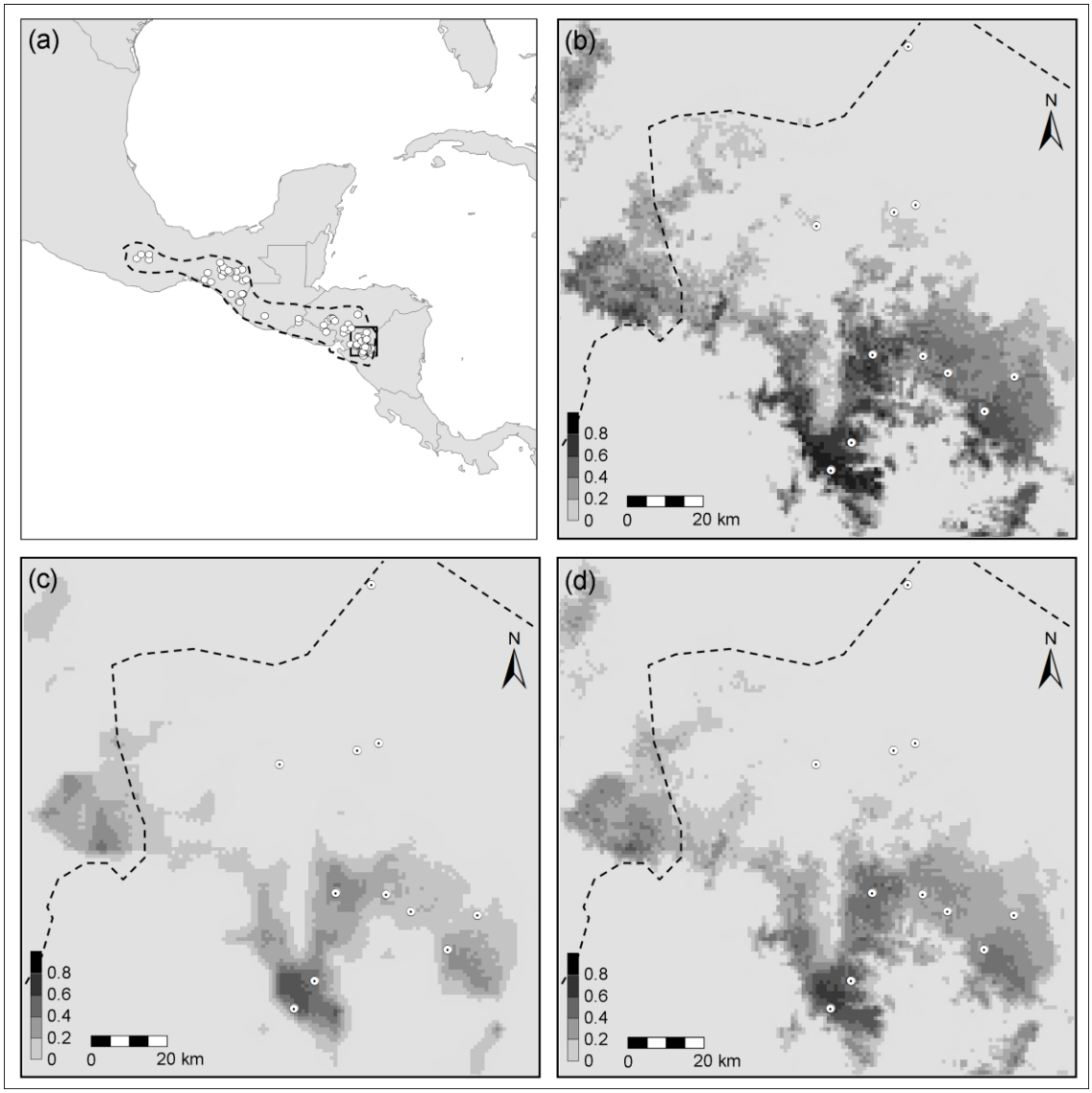


Figure 3

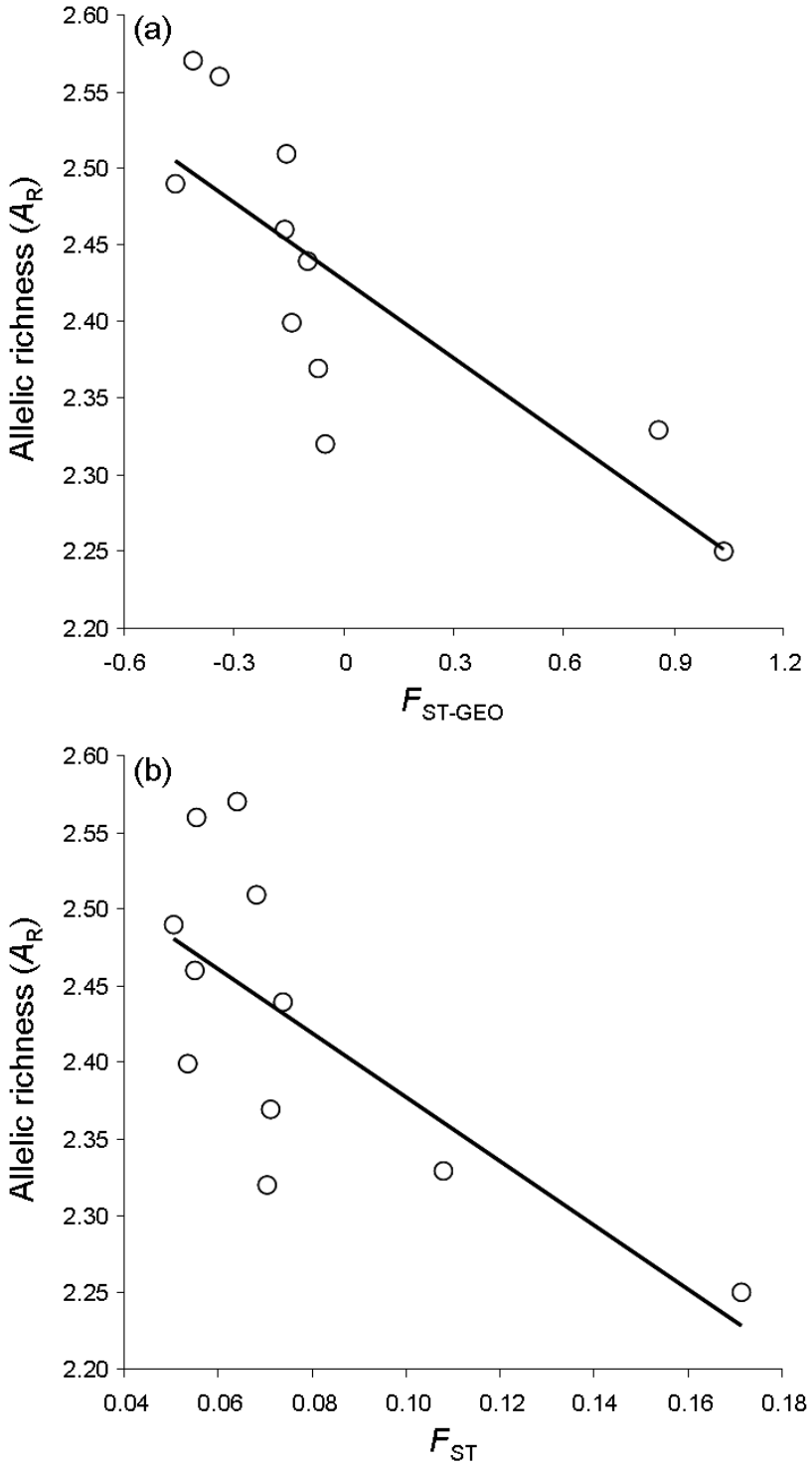
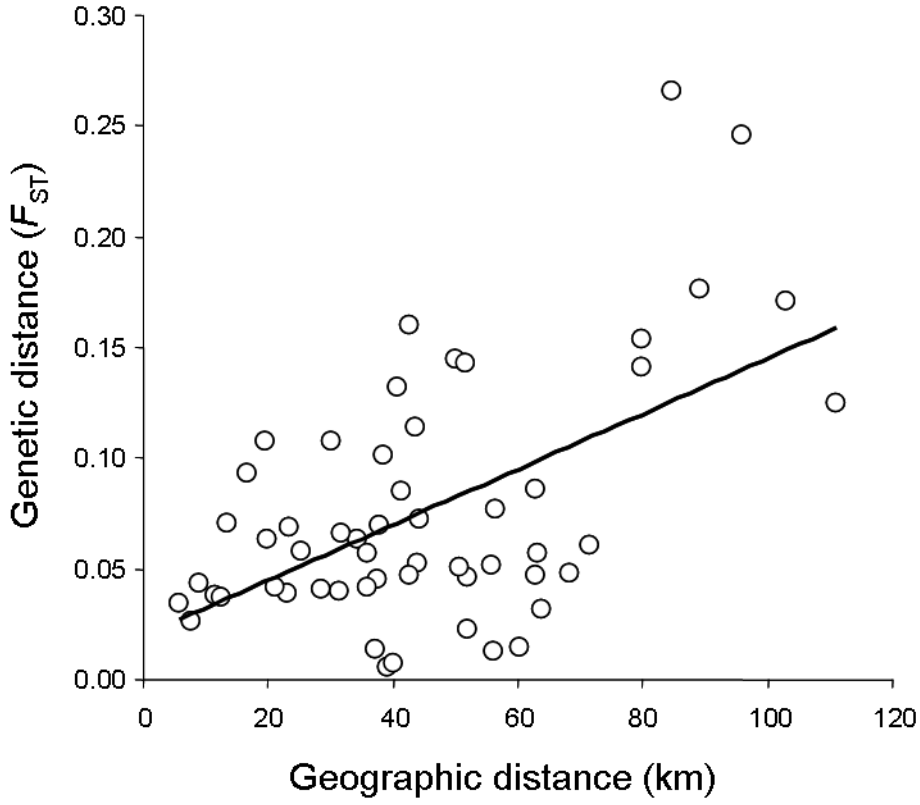

(a)
Allelic richness (A_R)
2.60
2.55
2.50
2.45
2.40
2.35
2.30
2.25
2.20
-0.6
-0.3
0
0.3
0.6
0.9
1.2
F_{ST-GEO}
(b)
Allelic richness (A_R)
2.60
2.55
2.50
2.45
2.40
2.35
2.30
2.25
2.20
0.04
0.06
0.08
0.10
0.12
0.14
0.16
0.18
F_{ST}

Figure 4


Genetic distance (F_{ST})
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
20
40
60
80
100
120
Geographic distance (km)