Spatiotemporally explicit demographic modelling supports a joint effect of historical barriers to dispersal and contemporary landscape composition on structuring genomic variation in a red-listed grasshopper

María José González-Serna*¹, Pedro J. Cordero¹, and Joaquín Ortego²

¹ Grupo de Investigación de la Biodiversidad Genética y Cultural, Instituto de Investigación en Recursos Cinegéticos – IREC – (CSIC, UCLM, JCCM), Ronda de Toledo 12, E-13071 Ciudad Real, Spain.

² Department of Integrative Ecology, Estación Biológica de Doñana – EBD – (CSIC), Avda. Américo Vespucio 26, E-41092 Seville, Spain.

*** Author for correspondence:**

María José González-Serna
Grupo de Investigación de la Biodiversidad Genética y Cultural
Instituto de Investigación en Recursos Cinegéticos – IREC – (CSIC, UCLM, JCCM)
Ronda de Toledo, 12, E-13071 Ciudad Real, Spain.
E-mail: mariajose.gonzalez@uclm.es
Tel: +34 926 29 54 50
Fax: +34 926 59 54 51

Short running title: **Landscape heterogeneity and genomic variation**

ABSTRACT

Inferring the processes underlying spatial patterns of genomic variation is fundamental to understand how organisms interact with landscape heterogeneity and to identify the factors determining species distributional shifts. Here, we employ genomic data (ddRADSeq) to test biologically-informed models representing historical and contemporary demographic scenarios of population connectivity for the Iberian cross-backed grasshopper *Dociostaurus hispanicus*, a species with a narrow distribution that currently forms highly fragmented populations. All models incorporated biological aspects of the focal taxon that could hypothetically impact its geographical patterns of genomic variation, including (a) spatial configuration of impassable barriers to dispersal defined by topographic landscapes not occupied by the species, (b) distributional shifts resulted from the interaction between the species bioclimatic envelope and Pleistocene glacial cycles, and (c) contemporary distribution of suitable habitats after extensive land clearing for agriculture. Spatiotemporally-explicit simulations under different scenarios considering these aspects and statistical evaluation of competing models within an Approximate Bayesian Computation (ABC) framework supported spatial configuration of topographic barriers to dispersal and human-driven habitat fragmentation as the main factors explaining the geographical distribution of genomic variation in the species, with no apparent impact of hypothetical distributional shifts linked to Pleistocene climatic oscillations. Collectively, this study supports that both historical (i.e., topographic barriers) and contemporary (i.e., anthropogenic habitat fragmentation) aspects of landscape composition have shaped major axes of genomic variation in the studied species and emphasizes the potential of model-based approaches to gain insights into the temporal scale at which different processes impact the demography of natural populations.

KEYWORDS: Approximate Bayesian Computation, ddRadSeq, *Dociostaurus hispanicus*, environmental niche modelling, landscape genetics, SPLATCHE2

INTRODUCTION

The neutral genetic makeup of populations is modulated by intrinsic demographic parameters, which in turn are determined by the interaction between species-specific biological traits (e.g. dispersal capacity, physiological tolerances, environmental niche, etc.) and the different spatial and temporal components that define landscape heterogeneity (e.g. current and past climate, geological changes, spatial configuration of barriers to dispersal, etc.) (Ortego, Aguirre, Nogueras & Cordero, 2015; Paz, Ibáñez, Lips & Crawford, 2015; Sukumaran & Knowles, 2018). Range size expansions/contractions and population fragmentation/coalescence linked to past or contemporary phenomena are demographic processes experienced by most organisms that can ultimately have important evolutionary consequences, from species extinction to rapid speciation (Knowles, 2000; Lorenzen et al., 2011; Rand, 1948; Talbot, Vonhof, Broders, Fenton & Keyghobadi, 2016; van der Mescht, Matthee, & Matthee, 2015; Weir, Haddrath, Robertson, Colbourne & Baker, 2016). For this reason, understanding how organisms have been impacted by past environmental changes and interact with landscape structure is fundamental to elucidate the evolutionary and ecological processes underlying their current distribution, abundance and spatial patterns of genetic variation. The development of spatiotemporally-explicit analytical approaches based in coalescent theory have made it possible to simulate complex demographic processes (e.g. bottlenecks and expansions) under a suite of contemporary and historical scenarios that can be statistically evaluated (Currat, Ray & Excoffier, 2004; e.g. Estoup et al., 2010; He, Edwards & Knowles, 2013; Neuenschwander et al., 2008; Ray, Currat, Foll & Excoffier, 2010). Recently, much emphasis in the literature has been also put on testing biologically-informed models to gain insights into the taxon-specific causal mechanisms structuring a species' gene pool (Papadopoulou & Knowles, 2016; Paz et al., 2015; Sukumaran & Knowles, 2018). These conceptual and analytical advances have proven useful for generating refined hypotheses and

testing the impacts of microhabitat preferences (Massatti & Knowles, 2016), physiological trade-offs (Bemmels, Title, Ortego & Knowles, 2016), and climate-induced range shifts (He et al., 2013; Reid et al., 2018) on the spatial distribution of genetic variation in numerous organisms (Hoban, 2014). Ultimately, this can allow better inference of biological processes that in turn are useful to predict and mitigate threats associated with various components of global change, such as warming climate and habitat fragmentation (Brown et al., 2016; Espindola et al., 2012; Estrada, Delgado, Arroyo, Traba & Morales, 2016; Morrison, Estrada & Early, 2018; Prates et al., 2016).

Climate shifts during the late Quaternary have altered the distribution and spatial patterns of genetic variation in many taxa (Hewitt, 2000), are recognized as an important engine of speciation (Weir et al., 2016), and have shaped large-scale gradients of species diversity (Sandel et al., 2011). Although the consequences of Pleistocene glaciations are well documented in numerous organisms from northern latitudes or linked to montane or alpine environments (Carstens & Knowles, 2007; Hewitt, 2000), their impacts on species distributed at lower latitudinal and elevational ranges, such as those inhabiting flat landscapes from the Mediterranean region, are more difficult to predict and much less well understood (Gómez & Lunt, 2007; Hewitt, 2000). Phylogeographic studies in southern Mediterranean peninsulas have identified multiple species-specific cryptic refugia (Gómez & Lunt, 2007; Hewitt, 2011; Stewart, Lister, Barnes & Dalén, 2010), which suggests that organisms from this region have probably responded to glacial and interglacial periods through local distributional shifts (Abellan & Svenning, 2014; Gómez & Lunt, 2007). Such local-scale distributional shifts can be difficult to predict from environmental niche models and currently available palaeoclimatic layers (Waltari et al., 2007). Such layers effectively capture drastic temperature and precipitation oscillations characterizing steep elevational gradients and northern latitudes that became covered by ice, but may fail to reflect changes in microclimatic conditions that are relevant for species distributed in other areas such as flat lowland landscapes from the

climatically more stable Mediterranean region (Waltari et al., 2007; Worth, Williamson, Sakaguchi, Nevill & Jordan, 2014). This region has been also subjected to millennia of intensive logging and land clearing for agriculture (Blondel & Aronson, 1999), which has reduced the original extent of its primary vegetation by around 95% (Myers, Mittermeier, Mittermeier, da Fonseca & Kent, 2000). Although anthropogenic activities have altered every ecosystem from the Mediterranean region, the consequences of habitat fragmentation are particularly remarkable for organisms occupying flat landscapes at low elevations, as these areas are highly suitable for crop production (e.g. Ortego et al., 2015). It is consequently worth investigating whether the current spatial distribution of genetic variation in these species has been predominantly shaped by population fragmentation resulted from human-driven habitat destruction or if the genetic signals left by historical processes such as Pleistocene climate-change distributional shifts and the disruption of gene flow driven by geographical barriers to dispersal still prevail (Cunningham & Moritz, 1998; Vandergast, Bohonak, Weissman & Fisher, 2007; Zellmer & Knowles, 2009).

Despite the fact that understanding how organisms interact with landscape heterogeneity at different time scales is critical for establishing effective conservation and informed management (Rubidge et al., 2012; Saccheri et al., 1998), disentangling the historical and contemporary processes shaping their spatial distribution of genetic variation is challenging for several reasons (Zellmer & Knowles, 2009). First, the time-lag between the point in which habitat fragmentation takes place and when signatures of disrupted gene flow become detectable is often long (Keyghobadi, Roland, Matter & Strobeck, 2005; Landguth et al., 2010). Second, the amount of time during which historical and contemporary processes impact the genetic makeup of a species generally differ in several orders of magnitude and their study has traditionally required the application of different genetic markers (e.g. mitochondrial DNA vs. microsatellite markers; Wang, 2010). Finally, historical and contemporary landscape structure can be spatially autocorrelated, which can make difficult to

tease apart the effects of past and recent changes in landscape composition (Zellmer & Knowles, 2009). In the last decade, the advent of high-throughput sequencing technology, computational advances and new modelling approaches have made possible to bridge landscape genetics and phylogeography (Rissler, 2016), address questions at different evolutionary levels without being constrained by the nature of the genetic markers employed (Villaverde et al., 2018), and improve our capacity to disentangle processes taking place at contrasting spatiotemporal scales within unified, hypothesis-testing frameworks (He et al., 2013).

Here, we employ genomic data obtained via restriction site-associated DNA sequencing (ddRAD-seq; Peterson, Weber, Kay, Fisher & Hoekstra, 2012) to test alternative demographic models representing historical and contemporary scenarios of population connectivity in the cross-backed grasshopper *Dociostaurus hispanicus* Bolívar, 1898. This narrow-endemic species is exclusively distributed in dehesas, open Mediterranean forests, scrublands and pastures scattered in the flat landscapes characterizing the central Iberian Peninsula (García, Larrosa, Clemente & Presa, 2005; Presa et al., 2016). Human-driven habitat destruction drove severe population declines and population fragmentation in this species and, as a result, it has been recently included in the IUCN Red List of Threatened Species with the category “Near Threatened” (Hochkirch et al., 2016; Presa et al., 2016). Specifically, we first used genomic data to infer the spatial patterns of genetic diversity and structure of the studied species and reconstruct historical demographic profiles of its different populations (Figure 1). Second, we generated a suite of biologically-informed models representing hypothetical processes that might have shaped the spatial distribution of genomic variation in our focal taxon and tested them within an Approximate Bayesian Computation framework (He et al., 2013; e.g. Neuenschwander et al., 2008) (Figure 1). All tested models incorporated biological aspects of the focal taxon that could have hypothetically impacted its geographical patterns of genetic variation, including the spatial configuration of barriers to dispersal defined

by rugged landscapes not occupied by the species, distributional shifts resulting from the interaction between the species-specific bioclimatic envelope and Pleistocene glacial cycles, and the contemporary distribution of remnant habitat patches after large-scale processes of land clearing for agriculture. This information was used to inform competing demographic models and infer, through the evaluation of their relative statistical support, the role of historical and contemporary components of landscape heterogeneity on structuring genomic variation in the focal species (He et al., 2013). Our study exemplifies how a unified, model-based framework integrating genomic data and different sources of spatial information can help to tease apart the role of contemporary and historical processes on shaping geographical patterns of genetic variation. This can improve our ability to reach biologically-meaningful insights about the proximate ecological and evolutionary processes governing the distribution of species and, ultimately, understand how organisms respond to changing and heterogeneous landscapes at contrasting temporal scales (Massatti & Knowles, 2016).

METHODS

POPULATION SAMPLING

During 2012-2016, we prospected potentially adequate habitats (i.e., dry pastures in steppe lands, dehesas, wastelands with scattered slate formations, and lands with sparse vegetation) for *Dociostaurus hispanicus* and collected samples from ten populations (Table 1; Figure 2) spanning the entire distribution range of the species (Presa et al., 2016). *Dociostaurus hispanicus* is classified as “Near Threatened” according to the IUCN Red List of Threatened Species due to the high fragmentation of its small-size populations (Hochkirch et al., 2016; Presa et al., 2016). For this reason, we analysed only 6-7 adult individuals per population (Table 1). We placed fresh whole adult specimens in vials with 2,000 µL ethanol 96% and stored them at -20° C until needed for genomic analyses.

GENOMIC LIBRARY PREPARATION AND PROCESSING OF GENOMIC DATA

We used NucleoSpin Tissue kits (Macherey-Nagel, Dürren, Germany) to extract and purify genomic DNA from the hind femur of each individual. Genomic DNA was processed into two genomic libraries using the double-digestion restriction-fragment-based procedure (ddRAD-seq) described in Peterson et al. (2012) (see Supporting Information Methods S1). We used the different programs distributed as part of the STACKS v.1.35 pipeline (*process_radtags*, *ustacks*, *cstacks*, *sstacks*, and *populations*) to assemble our sequences into *de novo* loci and call genotypes (Catchen, Hohenlohe, Bassham, Amores & Cresko, 2013; Catchen, Amores, Hohenlohe, Cresko & Postlethwait, 2011; Hohenlohe et al., 2010). See Supporting Information Methods S2 for details on sequence assembling and data filtering.

POPULATION GENETIC STRUCTURE

We calculated genetic differentiation among populations and used Bayesian clustering and principal component analyses to determine the geographical patterns of genetic structure in the studied species. We then used this information to evaluate the degree of genetic fragmentation of the populations at different spatial scales, and to interpret inferences obtained from spatiotemporally explicit demographic model testing (Figure 1). We calculated genetic differentiation (F_{ST}) between each pair of populations in ARLEQUIN v.3.5 (Excoffier & Lischer, 2010). We inferred genetic structure using the Bayesian Markov chain Monte Carlo clustering method implemented in the program STRUCTURE v.2.3.3 (Falush, Stephens & Pritchard, 2003; Hubisz, Falush, Stephens & Pritchard, 2009; Pritchard, Stephens & Donnelly, 2000). We conducted STRUCTURE analyses hierarchically, initially analysing data from all populations jointly and, subsequently, running independent analyses for subsets of populations assigned to the same genetic cluster in the previous hierarchical level analysis (e.g. Noguerales, Cordero & Ortego, 2016). We performed 15 independent runs for each value of K ,

where K ranged from 1 to $n+1$ for each dataset of n populations, to estimate the “true” number of clusters. We ran STRUCTURE using 200,000 MCMC iterations after a burn-in step of 100,000 iterations, assuming correlated allele frequencies and admixture, and without using prior population information (Hubisz et al., 2009). We retained the ten runs having the highest likelihood for each value of K and identified the number of genetic clusters best fitting the data set using log probabilities $[\Pr(X|K)]$ (Pritchard et al., 2000) and the ΔK method (Evanno, Regnaut & Goudet, 2005), as implemented in STRUCTURE HARVESTER (Earl & vonHoldt, 2012). We used CLUMPP v.1.1.2 and the Greedy algorithm to align multiple runs of STRUCTURE for the same K value (Jakobsson & Rosenberg, 2007) and DISTRUCT v.1.1 (Rosenberg, 2004) to visualize as bar plots the individual’s posterior probabilities of population membership to K genetic clusters. Complementary to Bayesian clustering analyses and in order to visualize the major axes of population genetic differentiation, we performed an individual-based principal components analysis (PCA) using the R v.3.3.3 (R Core Team 2017) package ADEGENET (Jombart, 2008).

GENETIC DIVERSITY AND PAST DEMOGRAPHIC HISTORY

We used genomic data to reconstruct the time and scale of historical changes in effective population size (N_e) of each studied population and assess whether they have experienced parallel or contrasting demographic histories that might explain spatial heterogeneity in contemporary levels of genetic diversity (Figure 1). First, we calculated different estimates of genetic diversity for each studied population (Table 1) using the program *populations* from STACKS (Catchen et al., 2013) and employed: (i) one-way ANOVAS to test for significant differences in genetic variability between the main clusters inferred by STRUCTURE and PCA analyses (see RESULTS section); and (ii) linear regressions to analyse geographical clines (i.e., latitude and longitude) of genetic diversity. We performed these analyses in SPSS v.24 software (IBM, NY, USA). Second, we reconstructed the past demographic history of the studied

populations using the program STAIRWAY PLOT, a method based on the site frequency spectrum (SFS) that does not require whole-genome sequence data or reference genome information (Liu & Fu, 2015). Given the marked genetic structure and significant genetic differentiation of all studied populations (see RESULTS section), we ran STAIRWAY PLOT analyses for each sampled population individually (Liu & Fu, 2015; e.g. Mattila, Tyrmi, Savolainen & Pyhäjärvi, 2017; Miles et al., 2017). To compute the SFS for each population, we ran the program *populations* from STACKS (Catchen et al., 2013) in order to export only the first SNP per RAD locus and retain loci with a minimum stack depth ≥ 5 ($m = 5$) and that were represented in at least 50% of the individuals of the population ($r = 0.5$). To remove all missing data for the calculation of the SFS and minimize errors with allele frequency estimates, each population was down-sampled to five individuals using a custom Python script written by Qixin He and available on Dryad (Papadopoulou & Knowles, 2015). We ran STAIRWAY PLOT for each population fitting a flexible multi-epoch demographic model, considering a 1-year generation time, assuming the mutation rate per site per generation of 2.8×10^{-9} estimated for *Drosophila melanogaster* (Keightley, Ness, Halligan & Haddrill, 2014), and performing 200 bootstrap replicates to estimate 95% confidence intervals.

ENVIRONMENTAL NICHE MODELLING

We built an environmental niche model (ENM) to predict the geographic distribution of climatically suitable habitats for *D. hispanicus* both in the present and during the last glacial maximum (LGM, 21 ka). We used this information to create maps of environmental suitability in these two time periods and test dynamic demographic models considering hypothetical distributional shifts experienced by the species in response to Pleistocene climatic oscillations (e.g. He et al., 2013; Massatti & Knowles, 2016) (see details in section *TESTING ALTERNATIVE DEMOGRAPHIC MODELS*). To build the ENM, we used the maximum entropy algorithm

implemented in MAXENT v.3.3.3 (Elith et al., 2006; Elith et al., 2011; Phillips, Anderson & Schapire, 2006; Phillips & Dudík, 2008) and the 19 bioclimatic variables from the WORLDCLIM dataset (<http://www.worldclim.org/>) interpolated to 30-arcsec resolution ($\sim 1 \text{ km}^2$ cell size) (Hijmans, Cameron, Parra, Jones & Jarvis, 2005). To generate climate suitability maps during the LGM, we projected the ENM onto LGM bioclimatic conditions derived from the CCSM4 (Community Climate System Model; Braconnot et al., 2007) and the MIROC-ESM (Model of Interdisciplinary Research on Climate; Hasumi & Emori, 2004) general atmospheric circulation models. Palaeoclimatic reconstructions under both models yielded qualitatively similar predictions about the distribution of the species during the LGM, and we performed all subsequent analyses using the MIROC-ESM model (e.g. Wachter et al., 2016). We built the ENM using our own species occurrence data and records available in the literature, the Global Biodiversity Information Facility (<http://www.gbif.org/>), and entomological collections from the Spanish National Museum of Natural Sciences (MNCN, Madrid) (see Table S1). Prior to modelling, we mapped and examined all records to identify and exclude those having obvious geo-referencing errors. To reduce the problems associated with sampling bias and the spatial aggregation of records, we applied a systematic sampling correction by randomly selecting a single occurrence record among those falling within the same $\sim 16\text{-km}^2$ grid cell (Fourcade, Engler, Rödder & Secondi, 2014). After this data filtering step, we retained a total of 57 occurrence records. ENMs built using all available records ($n = 103$; Table S1) or performing systematic sampling corrections at other spatial scales ($\sim 1\text{-km}^2$, $n = 96$ records; $\sim 4\text{-km}^2$, $n = 89$ records; $\sim 8\text{-km}^2$, $n = 80$ records) yielded qualitatively identical results. We used the R package ENMEVAL (Muscarella et al., 2014) and the Akaike's Information Criterion corrected for small sample size (AICc) (Burnham & Anderson, 2002; Warren & Seifert, 2011) to conduct parameter tuning and determine the optimal feature class (FC) and regularization multiplier (RM) settings for MAXENT. Complementary, we evaluated the performance of the retained model using the "block" method for data partitioning into training and testing datasets (Muscarella et al.,

2014). Specifically, we calculated the area under the receiver-operating characteristic plot on the testing data (AUC_{TEST}) and the minimum training presence omission rate (OR_{MTP}). An AUC_{TEST} value > 0.9 suggests a high discriminatory ability of the model (Peterson et al., 2011) whereas an OR_{MTP} close to zero is indicative of a low degree of model overfitting (Radosavljevic & Anderson, 2014). Further details on model and variable selection are presented in Supporting Information Methods S3.

TESTING ALTERNATIVE DEMOGRAPHIC MODELS

We used the integrative distributional, demographic and coalescent (iDDC) approach (He et al., 2013) and an Approximate Bayesian Computation (ABC) framework (Beaumont, Zhang & Balding, 2002; Csilléry, Blum, Gaggiotti & François, 2010) to test alternative models representing different hypotheses about how landscape heterogeneity has impacted the demography and spatial distribution of genomic variation in our focal taxon (Figure 1). This approach is described in detail in He et al. (2013) and consists of three main steps: (i) constructing alternative demographic models representing different hypotheses about how carrying capacities (k) vary across space and through time (i.e., spatiotemporally explicit models); (ii) running demographic and genetic simulations under each model using the software SPLATCHE2 (see Ray et al., 2010); (iii) evaluating the fit of observed genomic data (i.e., empirical data obtained after genotyping sampled populations) to the genetic expectations under each model, identifying the most probable model/s, and estimating demographic parameters (e.g. He et al., 2013; Massatti & Knowles, 2016) (Figure 1). Below we indicate the most relevant aspects for each of these three steps and present an extended explanation in Supporting Information Methods S4.

Constructing demographic models. We generated three basic demographic models: (i) a static model in which carrying capacities (k) are homogeneous across space and time. This scenario is

analogous to a flat landscape or an isolation-by-distance model (e.g. He et al., 2013); (ii) a
 dynamic model representing distributional shifts resulted from the interaction between the
 species bioclimatic envelope and Pleistocene glacial cycles (e.g. He et al., 2013; Massatti &
 Knowles, 2016). In this scenario, carrying capacities change over time according to climatic
 suitability maps obtained from projections of the ENM to the present and the LGM bioclimatic
 conditions (see section *ENVIRONMENTAL NICHE MODELLING*). This dynamic model considered
 landscapes from three consecutive time periods (LGM, intermediate, current) reflecting
 temporal shifts in the spatial distribution of environmentally suitable areas for the species in
 response to climate changes since the LGM (e.g. He et al., 2013; Massatti & Knowles, 2016). As
 done in previous studies, carrying capacities were scaled proportionally to logistic climatic
 suitability scores obtained from the ENM for each time period (e.g. He et al., 2013; Knowles &
 Massatti, 2017; Massatti & Knowles, 2016). Thus, we assumed that the carrying capacity for
 each grid cell was proportional to the estimated probability of presence of the species in that
 grid cell; (iii) a dynamic model starting with a flat landscape (i.e., non-fragmented) that shifted
 to a fragmented landscape in which carrying capacities for habitats unsuitable for the species
 were n times lower than those assigned to suitable habitats (Figure 3; Table 2). The
 distribution of suitable habitats was estimated from Corine Land Cover maps (CORINE, 2012).
 We considered as suitable habitats for the species the Corine Land Cover classes “broad-leaved
 forest”, “transitional woodland-shrub”, “land principally occupied by agriculture, with
 significant areas of natural vegetation”, “sclerophyllous vegetation”, “agro-forestry areas”,
 “pastures” and “natural grasslands”, which represent the habitats used by the species
 according to occurrence data (Table S1). We considered as unsuitable habitats all other land
 cover classes that are never occupied by the species. Habitat specialist grasshoppers are tightly
 linked to their specific microhabitats and generally show low rates of inter-patch dispersal (e.g.
 Ortego, Aguirre & Cordero, 2010; Ortego et al., 2015; Reinhardt, Köhler, Maas, Detzel &
 Spence, 2005). In the case of *D. hispanicus* this is evidenced by the strong genetic

fragmentation of its populations at both local and regional spatial scales (see Results section).

Thus, the land use classes occupied by *D. hispanicus* are likely to be an accurate proxy of suitable habitats that sustain comparatively much larger carrying capacities than those habitat categories where the species has never been recorded. Dynamic models considering the impacts of human-driven habitat fragmentation started with a hypothetically non-fragmented landscape that 250 years ago shifted to a fragmented landscape with a spatial configuration of suitable/unsuitable habitats that has not changed until present. Thus, this model assumes that habitat fragmentation took place 250 years ago, the spatial configuration of suitable habitats after fragmentation took place has remained unaltered until present times, and the carrying capacities of unsuitable habitats are n times smaller than those of habitats suitable for the species. We fine-tuned this model considering different relative values of carrying capacities for unsuitable and suitable habitats (1:10, 1:100 and 1:1000) (e.g. Ortego et al., 2015; Paquette, Talbot, Garant, Mainguy, & Pelletier, 2014). The three orders of magnitude considered for the relative carrying capacities of unsuitable and suitable habitats span a range that has been found to be informative in another specialist grasshopper from the Iberian Peninsula inhabiting similar flat lowland landscapes (Ortego et al., 2015). Finally, we generated two variants of the three basic models described above by incorporating information from geographical features that might hypothetically hinder (topographic roughness) or impede (impassable barriers to dispersal) gene flow among populations. The first variant incorporated the hypothetical negative impact of topographic roughness on migration rates with neighbouring demes (specified as a friction parameter in SPLATCHE2; Ray et al., 2010). The second variant incorporated the presence of impassable barriers to dispersal ($k = 0$). We considered grid cells with a slope > 20 % as impassable barriers to dispersal, as these areas are not occupied by the focal taxon according to occurrence data (Table S1). We calculated topographic roughness (slope) using a 90-m resolution digital elevation model from NASA Shuttle Radar Topographic Mission (SRTM Digital Elevation Data; <http://srtm.csi.cgiar.org/>).

Overall, we generated a total of nine demographic models resulted from the combinations of three basic static/dynamic models with the two hypothetical effects of topography (topographic roughness and the presence/absence of barriers to dispersal) (Figure 1 and 3; Table 2).

Demographic and genetic simulations. We used SPLATCHE2 to perform forward-in-time demographic simulations followed by backward-in-time genetic (coalescent) simulations under each model (see Ray et al., 2010), which are expected to produce contrasting patterns of genetic variation due to differences among scenarios in the way that carrying capacities vary across the landscape and through time (see Massatti & Knowles, 2016) (Figure 3). For each model, we ran 200,000 simulations using the same uniform priors for the three demographic parameters employed: migration rate per deme per generation (m ; range of $\log(m)$: -2.3, -1.2), carrying capacity of the deme with highest suitability ($K_{\max}$; range of $\log(K_{\max})$: 5.2, 7.2), and ancestral population size (N_{ANC} ; range of $\log(N_{\text{ANC}})$: 2.0, 4.0). Following each time-forward demographic simulation, a spatially-explicit time-backward coalescent model informed by the deme-specific demographic parameters (K , m and N_{ANC}) was used to generate genetic data (Currat et al., 2004; Ray et al., 2010). We used ARLSUMSTAT v.3.5.2 to calculate a total of 67 summary statistics for simulated datasets under each model, including mean heterozygosity across loci for each population and across populations (H), number of segregating sites for each population and across populations (S), and pairwise population F_{ST} values (Excoffier & Lischer, 2010). The same 67 summary statistics were extracted from observed empirical data (Figure 1).

Model choice and parameter estimation. We used an Approximate Bayesian computation (ABC) framework to perform model selection and parameter estimation (for an overview of ABC, see Beaumont et al., 2002), as implemented in the programs TRANSFORMER and

ABCESTIMATOR and R scripts (*findPLS*) distributed as part of ABCTOOLBOX (Wegmann, Leuenberger, Neuenschwander & Excoffier, 2010; e.g. He et al., 2013; Massatti & Knowles, 2016). In order to account for correlations between summary statistics and reduce the “curse of dimensionality” associated with using a large number of statistics (Boulesteix & Strimmer, 2007), we used the R package PLS v.2.6-0 (Mevik & Wehrens, 2007) and the *findPLS* script to extract partial least squares (PLS) components with Box-Cox transformation from the summary statistics of the first 10,000 simulations for each model (Boulesteix & Strimmer, 2007; Wegmann, Dupanloup & Excoffier, 2008; Wegmann et al., 2010; e.g. He et al., 2013). We used the linear combinations of summary statistics obtained from the first 10,000 simulations for each model to transform all datasets (observed empirical and simulated datasets) with the program TRANSFORMER (for details about this procedure, see Wegmann et al., 2010). For each model, we retained the 1,000 simulations (0.5%) closest to observed empirical data and used them to approximate marginal densities and posterior distributions of the parameters with a postsampling regression adjustment using the ABC-GLM (general linear model) procedure detailed in Leuenberger & Wegmann (2010) and implemented in ABCTOOLBOX (see also Csilléry et al., 2010). We used Bayes factors (BF) for model selection, defined as the ratio between marginal densities of the model with the highest marginal density and the alternative model (Jeffreys, 1961). The higher the ratio is, the more supported the first model is. A $BF > 20$ indicates strong relative support for the first model, while those >150 indicate very strong support (Jeffreys, 1961; Kass & Raftery, 1995; Leuenberger & Wegmann, 2010). To evaluate the ability of each model to generate the observed empirical data, we calculated the Wegmann’s p -value from the 1,000 retained simulations (Wegmann et al., 2010). The p -value is calculated as the fraction of the retained simulations with a smaller or equal likelihood than the observed empirical data, with low values indicating that a model is highly unlikely (Wegmann et al., 2010). We also assessed the potential for a parameter to be correctly estimated by computing the proportion of parameter variance that was explained (i.e., the

coefficient of determination, R^2) by the retained PLSs (Neuenschwander et al., 2008). Finally, for the most supported model/s, we determined the accuracy of parameter estimation using a total of 1,000 pseudo-observation datasets (PODs) generated from prior distributions of the parameters. If the estimation of the parameters is unbiased, posterior quantiles of the parameters obtained from PODs should be uniformly distributed (Cook, Gelman & Rubin, 2006; Wegmann et al., 2010). As done for observed empirical data, we calculated the posterior quantiles of true parameters for each pseudo run based on the posterior distribution of the regression-adjusted 1,000 simulations closest to each pseudo-observation (e.g. Massatti & Knowles, 2016).

RESULTS

GENOMIC DATASET

After quality filtering, a total of 136,580,290 reads were retained across all genotyped individuals (mean $\pm$ SD = 2,239,021 $\pm$ 657,973) (Figure S1). The final exported dataset obtained with STACKS after removing loci that did not meet the different filtering requirements ($p = 10$, $r = 0.8$, and $min_maf = 0.01$) contained 18,249 SNPs, with a proportion of missing data of 1.68%.

POPULATION GENETIC STRUCTURE

Pair-wise F_{ST} values ranged from 0.082 to 0.307 and all estimates were significantly different from zero based on 100 permutations ($P < 0.05$; Table S2). STRUCTURE analyses revealed a good congruence between geography and the genetic clusters inferred at different hierarchical levels (Figure 4). Analyses considering all populations showed the presence of two main genetic clusters corresponding to populations located north (TRAB, NAVA, PNEG, AVIL; hereafter, Northern populations) and south (ALDE, VENT, TRUJ, BELA, ALCU, SANT; hereafter,

Southern populations) of the Central Mountain System (Figure 4 and Figures S2-S3). STRUCTURE analyses performed at lower hierarchical levels for Southern populations also showed a good correspondence between geography and the identified genetic clusters, with two well defined genetic clusters separating southernmost populations (BELA, ALCU, SANT) from the rest of Southern populations (ALDE, VENT, TRUJ) (Figure 4). However, the congruence between geography and genetic substructure among Northern populations was less obvious and some populations (AVIL, PNEG) were often assigned to different genetic clusters than those grouping nearby and spatially interspersed populations (Figure 4). The results obtained with STRUCTURE were in good agreement with those obtained from Principal Component Analyses (PCA) in which PC1 split Northern and Southern populations, and PC2 separated southernmost populations from the rest of Southern populations (Figure S4).

GENETIC DIVERSITY AND PAST DEMOGRAPHIC HISTORY

Population genetic statistics (π , H_o , H_e , P and F_{is}) calculated with STACKS for all positions (polymorphic and non-polymorphic) and only considering variant positions are presented in Table 1. One-way ANOVAs showed that Northern populations have lower levels of genetic diversity than Southern populations (π : $F_{1,8} = 35.34$, $P < 0.001$; H_o : $F_{1,8} = 13.43$, $P = 0.006$). Accordingly, genetic diversity was negatively correlated with latitude (π : $r = 0.853$, $t = -4.62$, $P = 0.002$; H_o : $r = 0.818$, $t = -4.02$, $P = 0.004$) but showed no association with longitude (π : $r = 0.052$, $t = 0.15$, $P = 0.887$; H_o : $r = 0.043$, $t = 0.12$, $P = 0.906$). STAIRWAY PLOT analyses revealed that contemporary effective population sizes for Northern populations are lower than those inferred for Southern populations (Figure 5), which is in agreement with the observed negative latitudinal cline of genetic diversity (Figure 2). Two Northern populations (AVIL and PNEG) experienced a deep bottleneck starting 40ka BP. These severe bottlenecks reduced N_e by ~75 %, from ~100,000 to ~25,000 diploid individuals, and lasted until the LGM (21 ka), when the two populations abruptly expanded (Figure 5a). Two populations (VENT and BELA) showed

signals of small demographic bottlenecks (N_e reduction by ~10-15%) occurring at different time periods between 50-30 ka BP (Figure 5b). With the exception of AVIL and PNEG, all populations have experienced demographic expansions of different magnitude after long stable periods with negligible fluctuations in population size. Demographic expansions in most of these populations (NAVA, VENT, TRU, BELA, ALCU) started 40-20 ka BP from an ancestral population of ~80,000-100,000 diploid individuals. Two populations (TRAB and ALDE) experienced earlier (100-120 ka BP) abrupt expansions followed by a long period of demographic stability spanning until the present. These expansions extended for over 10-20 ka and strong variability in their magnitude is reflected in the remarkable differences in N_e and estimates of genetic diversity among contemporary populations (Figures 2 and 5; Table 1).

ENVIRONMENTAL NICHE MODELLING

Parameter tuning performed using ENMEVAL showed that a LQH feature class (FC) combination and a regularization multiplier (RM) of 2 minimized AICc. After removing highly correlated variables and those with a zero percent contribution, nine bioclimatic layers were retained to construct the final ENM. These variables, sorted by their percent contribution (in parentheses), included: temperature seasonality (BIO4; 26.2%); mean temperature of wettest quarter (BIO8; 23.2%); precipitation of wettest quarter (BIO16; 18.0%); precipitation seasonality (BIO15; 9.7%); precipitation of driest month (BIO14; 6.8%); mean temperature of coldest quarter (BIO11; 6.6%); isothermality (BIO3; 4.8%); mean diurnal range (BIO2; 2.7%); and mean temperature of driest quarter (BIO9; 1.9%). The high AUC_{TEST} ($AUC_{TEST} = 0.905$) and low OR_{MTP} ($OR_{MTP} = 0.017$) estimates for the most supported model indicate that it has high discriminatory power and a low degree of overfitting, respectively. Accordingly, inspection of predicted distributions for the present confirmed that the ENM yielded a distribution pattern coherent with the observed current distribution of the species (see panel c from Model D in Figure 3). The projection of the present-day climate niche envelope to LGM climatic conditions

suggests some important temporal changes in the distribution and patterns of population connectivity of the species (Figure 3a). Climatic suitability maps predicted a more continuous distribution and overall higher suitability during the LGM than in the present and suggest that the species has experienced a northward range contraction and considerable fragmentation across its entire distribution after the last glacial period (Figure 3).

TESTING ALTERNATIVE DEMOGRAPHIC MODELS

Based on the relative information content contained in the PLS components, the first five PLSs were selected and extracted from the summary statistics for ABC analyses (Figure S5).

According to marginal densities calculated from the 0.5% simulations closest to observed empirical data, Model H had the highest support (Table 2). The high Wegmann p -value (p -value = 0.957) of this model suggests a strong correspondence between observed empirical data and simulated genetic data (Table 2). This model incorporates barriers to dispersal (areas with slopes > 20 % not occupied by the focal taxon are assigned a $k = 0$) and starts with a flat landscape with homogeneous carrying capacities (i.e., non-fragmented) that 250 years ago shifted to a fragmented landscape in which carrying capacities for habitats unsuitable for the species are ten times smaller than those assigned to suitable habitats. Fine-tuning of models G-I considering different ratios of carrying capacities for unsuitable and suitable habitats (1:10, 1:100 and 1:1000) did not have a strong impact on relative model fitting to empirical data and Model H was consistently the most supported model in all cases (Table 2). The relative values of carrying capacities for unsuitable and suitable habitats that better fitted empirical data differed among models (1:100 for model G, 1:10 for model H and 1:1000 for model I; Table 2), indicating that the best solution under each specific model was reached through different combinations of demographic parameter values and ratios of carrying capacities between the two habitat classes (Table 2). Analogous models considering alternative times for the onset of habitat fragmentation had consistently lower support (125 years: BF = 14.53; 500 years: BF =

2.62; 1000 years: BF = 3.33; 2000 years: BF = 7.49). The second most supported model (Model B; BF = 11.26; Table 2) was the one considering a flat landscape incorporating barriers to dispersal (Figure 3). This model was plausibly similar to Model H (BF < 20; Table 2). All other models had differences in BF > 20 with the most supported model (Table 2), indicating strong relative support for Model H (Jeffreys, 1961; Kass & Raftery, 1995).

Posterior distributions of parameter estimates were considerably distinct from the prior, indicating the simulated data contained information relevant to estimating the parameters (Figure S6). The comparison of posterior distributions before and after the ABC-GLM also showed the improvement that this procedure had on parameter estimates (blue shading vs. solid black line in Figure S6). The posterior distribution of maximum carrying capacity ($K_{\max}$) was flatter than those obtained for ancestral population size (N_{ANC}) and migration rates (m) (Figure S6). Accordingly, the coefficients of determination (R^2) from a multiple regression between each demographic parameter and the five retained PLS indicated that the employed summary statistics had a high potential to correctly estimate all the parameters except $K_{\max}$ (Table 2). Histograms of the posterior quantile of m is uniformly distributed, indicating that this parameter is estimated without bias (Figure S7b). In contrast, the Kolmogorov-Smirnov test indicated that the posterior quantiles for $K_{\max}$ and N_{ANC} deviated from an uniform distribution, suggesting a potential bias in the estimation of these parameters (Figure S7a,c).

DISCUSSION

In a first (observational) step, we employed genomic data to infer the spatial patterns of genetic diversity and structure of the studied species and reconstruct historical demographic profiles of the different populations. In a second (hypothesis-testing) step, we used

information on different aspects of the ecology of the focal taxon (environmental niche envelope, habitat-associations, topographic roughness, and barriers to dispersal) to generate biologically-informed models representing alternative hypotheses linking patterns of genetic variation inferred in the first step with the ecological and evolutionary processes shaping them. Statistical evaluation of the different competing scenarios revealed that the marked spatial genetic structure of the focal taxon has been primarily determined by topographic barriers to dispersal and the contemporary configuration of suitable habitats for the species, with no apparent effect of distributional shifts linked to Pleistocene glacial cycles.

Spatial distribution of genetic variation

Despite the narrow distribution of the cross-backed grasshopper *D. hispanicus* and the short geographical distances separating some of the studied populations, Bayesian clustering analyses revealed a marked hierarchical genetic structure in the species. Specifically, these analyses showed the presence of two main genetic clusters separating populations located north and south of the Central Mountain System that, in turn, are substructured at finer geographical scales. Such pronounced genetic structure is comparable to that reported for other Mediterranean grasshoppers showing low dispersal capacities and/or a highly fragmented distribution of suitable habitats (Noguerales et al., 2016; Ortego et al., 2010). Beyond the assignment to different genetic clusters, several lines of evidence indicate that the two main population groups have experienced contrasting demographic histories. Whereas expansions following long stable demographic periods with negligible fluctuations in population size characterize southern populations, some northern populations (PNEG, AVIL) have passed through severe demographic bottlenecks during the last glaciation. Increased genetic drift in northern populations is reflected in their higher levels of genetic differentiation and lower levels of genetic diversity in comparison with southern populations, which leads to a latitudinal decline of genetic diversity. Genetic drift and strong bottlenecks in northern

populations might have led to some incongruences between geography and the inferred genetic substructure and explain why some nearby populations do not cluster together in STRUCTURE analyses (see also Ortego, Gugger & Sork, 2018). Collectively, these results indicate considerable geographic heterogeneity in the distribution of genetic diversity across populations, suggesting that different temporal and spatial aspects of landscape composition have plausibly impacted the demography of the species and shaped its present-day patterns of genetic variation.

Testing alternative demographic models

Model comparison using an Approximate Bayesian Computation framework indicated that the spatiotemporally explicit demographic scenario most supported by genomic data was the one considering a landscape incorporating both topographic barriers to dispersal and the contemporary distribution of suitable habitats (Model H). The scenario based on a static flat landscape with impassable barriers to dispersal (Model B) was also relatively well supported ($BF < 20$; Kass & Raftery, 1995) and provided a reasonably good fit to the data. Remarkably, among models that did not incorporate the barrier, only the one based on the contemporary distribution of suitable habitats (Model G) was able to generate data in agreement with observed empirical data whereas the other models had low to very low relative supports (Wegmann p -values < 0.1).

The predominant role of topographic barriers to dispersal on structuring genomic variation in the studied species is in agreement with its strong preferences for flat landscapes and is in line with previous studies documenting the importance of topographic features on shaping genetic divergence across multiple terrestrial organisms (e.g. Benham & Witt, 2016; Castillo, Epps, Davis & Cushman, 2014). The effects of topographic barriers on structuring genetic variation can be explained by multiple non-mutually exclusive mechanisms, including the greater energetic expenditure associated with moving across topographically

heterogeneous landscapes (Castillo et al., 2014), dispersal preferences for a particular environment (flat landscapes), behavioural reluctance to cross novel habitats (i.e., steep slopes) (Wang & Bradburd, 2014) or physiological constraints (i.e., thermal tolerances and sensitivity) and/or competition disadvantage under the specific microclimatic conditions prevailing at high elevational ranges (Slatyer, Nash & Hoffmann, 2016; Strangas, Navas, Rodrigues & Carnaval, 2018). Remarkably, fitting topographic roughness as a friction parameter (Ray et al., 2010) yielded very poor model performance under any scenario ($BF > 190$; Wegmann p -value < 0.1), which indicates that the failure of the species to cross landscapes with slopes above a certain threshold ($> 20\%$), rather than topographic roughness *per se*, is the proximate factor structuring genetic variation in the focal taxon (see also González-Serna, Ortego & Cordero, 2018). The strong preferences of the species for flat areas might explain the observed “threshold effect”, in contrast to the “cumulative effects” of topographic roughness often documented in organisms inhabiting abrupt landscapes (Benham & Witt, 2016; Castillo et al., 2014; Wang, 2012), including some montane grasshoppers (e.g. Nogueras et al., 2016). Limited gene flow through steep slopes is in strong agreement with inferences from STRUCTURE and PCA analyses, which showed that the different genetic clusters group populations separated by the main mountain systems in the study area (Central Mountain System, Montes de Toledo, and Sierra Morena) and that the hierarchy of such genetic splits is congruent with the extension and elevation of the different ranges. The fact that the uplift of these mountain ranges largely predates (Vera, 2004) the estimated divergence of *D. hispanicus* from its sister taxon *D. brevicollis* Eversmann 1848 (~1.9 Ma; González-Serna et al., 2018) points to range expansion followed by genetic drift, rather than fragmentation of an originally continuous population, as the most likely explanation for the observed link between contemporary patterns of genetic structure and the spatial configuration of barriers to dispersal (Short & Petren, 2011). The most supported model suggested that the contemporary distribution of suitable habitats derived from land cover

maps has impacted the spatial distribution of genetic variation in the studied species. The effect of contemporary landscape composition makes biological sense considering that land clearing for extensive agriculture (vineyards, olive groves and cereal crops) has dramatically fragmented the species habitats (Presa et al., 2016). However, ABC model selection suggests that the impacts of human-driven habitat fragmentation on observed patterns of genetic variation are very small in comparison with the effects of topographic barriers to dispersal. In fact, the most supported model (Model H) was statistically followed by the one based on a flat landscape (Model I), indicating the difficulties to discern the effects of a continuous vs. a fragmented landscape. This suggests that our range-wide sampling scheme was mostly able to capture the effects of historical processes (Oyler-McCance, Fedy & Landguth, 2013; Schwartz & McKelvey, 2008) or, more likely, that the time elapsed since human-driven habitat fragmentation took place has not been enough to strongly impact the genetic makeup of the species (Landguth et al., 2010). Accordingly, our demographic reconstructions do not support population declines during the Anthropocene and, on the contrary, indicate that all populations have actually experienced considerable expansions starting at different time periods followed by a phase of demographic stability until present days. Overall, these results are in line with previous studies on Orthoptera showing that population fragmentation linked to anthropogenic habitat destruction can leave detectable genetic signatures but, generally, with very small effect sizes (Keller et al., 2013; Lange, Durka, Holzhauer, Wolters & Diekötter, 2010; Ortego et al., 2015) unlikely to blur phylogeographic structure resulted from historical processes (Cunningham & Moritz, 1998; Keller et al., 2013).

Although the ENM seemed to predict well the current distribution of the focal taxon, several lines of evidence indicate that estimates of climatic suitability derived from it across different time periods had a low ability to recover the demographic history of the species. Firstly, projection of the present-day bioclimatic envelop to the LGM predicted high climatic suitability across all populations during this period, which is incompatible with the

comparatively lower levels of genetic diversity of northern populations and the severe bottlenecks experienced by some of them during the last glaciation according to demographic reconstructions (e.g. Carnaval, Hickerson, Haddad, Rodrigues & Moritz, 2009; Cristofari et al., 2018; Noguerales et al., 2018; Yannic et al., 2018). Secondly, spatially-explicit simulations and ABC model choice gave a much lower support to the dynamic model based on ENM than to the static model based on a flat landscape representing a baseline isolation-by-distance scenario (e.g. He et al., 2013). These results emphasize the importance of using independent sources of information (i.e., genomic data and, when available, fossil records and ancient DNA) to validate distributional shifts inferred from ENMs, an aspect that might be particularly relevant for species inhabiting Mediterranean lowlands showing less predictable demographic responses to past climatic oscillations than organisms distributed at high latitudes or in alpine/montane environments (Davis, McGuire & Orcutt, 2014; Fordham, Brook, Moritz & Nogués-Bravo, 2014; Metcalf et al., 2014).

Conclusions and final remarks

Our study highlights the power of integrating genomic data and model-based approaches to infer complex demographic and evolutionary processes taking place at contrasting temporal scales. Such processes would be difficult to unravel through traditional phylogeographic or landscape genetic approaches (Vandergast et al., 2007; Zellmer & Knowles, 2009). By integrating these approaches, we have been able to use a unified hypothesis-testing framework to analyse the relative role of historical (i.e., topography and Pleistocene climatic oscillations) and contemporary (i.e., human-driven habitat fragmentation) processes in shaping the distribution of genetic variation in a red-listed species. Our results suggest a low performance of ENM to predict the past demographic history of the studied species and indicate a predominant role of historical landscape composition (i.e., spatial configuration of topographic barriers to dispersal) and contemporary anthropogenic habitat fragmentation on

determining its major axes of genomic variation. At this point, we must note that similar genetic patterns may result from different demographic processes and, thus, our most supported model, even if it is able to generate data in agreement with empirical data, is likely to have ignored many relevant aspects shaping the evolutionary history of focal species (He et al., 2013; Massatti & Knowles, 2016). For instance, the marked spatial heterogeneity in genetic diversity (i.e., negative latitudinal cline) and the heterogeneous demographic profiles (bottlenecks vs. expansions) of the different populations are probably the outcome of complex ecological/evolutionary processes that are unlikely to be covered by the most supported scenarios. Acknowledging the limitations inherent to any model-based approach, our study highlights the potential of testing biologically-informed models to better understand the predominant processes shaping spatial patterns of genetic variation in organisms inhabiting regions severely impacted by changes in landscape composition acting at contrasting temporal scales (He et al., 2013; Papadopoulou & Knowles, 2016).

ACKNOWLEDGEMENTS

We wish to thank to Carlos Muñoz-Alcón for providing us information about the location of some populations, Anna Papadopoulou and L. Lacey Knowles Lab for their great support with data analyses, David Aragonés (LAST-EBD) for his help with GIS analyses, Charles B. van Rees for kindly correcting the English language of the manuscript, and three anonymous referees for helpful and constructive comments on an earlier version of this article. We also thank to Centro de Supercomputación de Galicia (CESGA) and Doñana's Singular Scientific-Technical Infrastructure (ICTS-RBD) for access to computational resources. The respective administrative authorities from each study area (Castilla y León, Madrid, Castilla – La Mancha, Extremadura, and Andalucía) provided us the corresponding permits for sampling. MJG was supported by a pre-doctoral scholarship from Junta de Comunidades de Castilla–La Mancha and European Social Fund. This work received financial support from research grants CGL2011-25053,

CGL2014-54671-P, CGL2016-80742-R and CGL2017-83433-P (co-funded by the Dirección General de Investigación y Gestión del Plan Nacional I+D+i and European Social Fund); PEII-2014-023-P (co-funded by Junta de Comunidades de Castilla–La Mancha and European Social Fund).

AUTHOR CONTRIBUTIONS

MGS and JO conceived and designed the study and analyses. All authors collected the samples. MGS performed the lab work and analysed the data guided by JO. MGS and JO led the writing with inputs from PJC.

DATA ACCESSIBILITY

SNP datasets, input files for STRUCTURE, PCA, and STAIRWAY PLOT analyses, output raster maps from MAXENT, and raster layers used for demographic and genetic simulations in SPLATCHE2 are available for download from the Dryad Digital Repository (<https://datadryad.org/review?doi=doi:10.5061/dryad.ck6q52n>).

REFERENCES

- Abellán, P., & Svenning, J.-C. (2014). Refugia within refugia – patterns in endemism and genetic divergence are linked to Late Quaternary climate stability in the Iberian Peninsula. *Biological Journal of the Linnean Society*, 113(1), 13-28. doi:10.1111/bij.12309
- Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162(4), 2025-2035.
- Bemmels, J. B., Title, P. O., Ortego, J., & Knowles, L. L. (2016). Tests of species-specific models reveal the importance of drought in postglacial range shifts of a Mediterranean-climate tree: insights from integrative distributional, demographic and coalescent modelling and ABC model selection. *Molecular Ecology*, 25(19), 4889-4906. doi:10.1111/mec.13804
- Benham, P. M., & Witt, C. C. (2016). The dual role of Andean topography in primary divergence: functional and neutral variation among populations of the hummingbird, *Metallura tyrianthina*. *BMC Evolutionary Biology*, 16, 22. doi:10.1186/s12862-016-0595-2
- Blondel, J., & Aronson, J. (1999). *Biology and Wildlife of the Mediterranean Region*. New York, NY: Oxford University Press.

- Boulesteix, A. L., & Strimmer, K. (2007). Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics*, 8(1), 32-44. doi:10.1093/bib/bbl016
- Braconnot, P., Otto-Bliesner, B., Harrison, S., Joussaume, S., Peterchmitt, J. Y., Abe-Ouchi, A., . . . Zhao, Y. (2007). Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum - Part 1: experiments and large-scale features. *Climate of the Past*, 3(2), 261-277. doi:10.5194/cp-3-261-2007
- Brown, J. L., Weber, J. J., Alvarado-Serrano, D. F., Hickerson, M. J., Franks, S. J., & Carnaval, A. C. (2016). Predicting the genetic consequences of future climate change: The power of coupling spatial demography, the coalescent, and historical landscape changes. *American Journal of Botany*, 103(1), 153-163. doi:10.3732/ajb.1500117
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference a practical information-theoretic approach*. New York, NY: Springer-Verlag.
- Carnaval, A. C., Hickerson, M. J., Haddad, C. F., Rodrigues, M. T., & Moritz, C. (2009). Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, 323(5915), 785-789. doi:10.1126/science.1166955
- Carstens, B. C., & Knowles, L. L. (2007). Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. *Systematic Biology*, 56(3), 400-411. doi:10.1080/10635150701405560
- Castillo, J. A., Epps, C. W., Davis, A. R., & Cushman, S. A. (2014). Landscape effects on gene flow for a climate-sensitive montane species, the American pika. *Molecular Ecology*, 23(4), 843-856. doi:10.1111/mec.12650
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). STACKS: an analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124-3140. doi:10.1111/mec.12354
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). STACKS: Building and genotyping loci *de novo* from short-read sequences. *G3: Genes/Genomes/Genetics*, 1(3), 171-182. doi:10.1534/g3.111.000240
- Cook, S. R., Gelman, A., & Rubin, D. B. (2006). Validation of software for Bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics*, 15(3), 675-692. doi:10.1198/106186006X136976
- CORINE (2012). CORINE Land Cover, EEA, Commission of the European Communities. Retrieved from <https://land.copernicus.eu/pan-european/corine-land-cover/clc-2012>
- Cristofari, R., Liu, X., Bonadonna, F., Cherel, Y., Pistorius, P., Le Maho, Y., . . . Trucchi, E. (2018). Climate-driven range shifts of the king penguin in a fragmented ecosystem. *Nature Climate Change*, 8(3), 245-251. doi:10.1038/s41558-018-0084-2
- Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., & François, O. (2010). Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, 25(7), 410-418. doi:10.1016/j.tree.2010.04.001
- Cunningham, M., & Moritz, C. (1998). Genetic effects of forest fragmentation on a rainforest restricted lizard (Scincidae: *Gnypetoscincus queenslandiae*). *Biological Conservation*, 83(1), 19-30. doi:10.1016/s0006-3207(97)00046-3
- Curat, M., Ray, N., & Excoffier, L. (2004). SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity. *Molecular Ecology Notes*, 4(1), 139-142. doi:10.1046/j.1471-8286.2003.00582.x
- Davis, E. B., McGuire, J. L., & Orcutt, J. D. (2014). Ecological niche models of mammalian glacial refugia show consistent bias. *Ecography*, 37(11), 1133-1138. doi:10.1111/ecog.01294
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359-361. doi:10.1007/s12686-011-9548-7

- Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., . . . Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129-151. doi:10.1111/j.2006.0906-7590.04596.x
- Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MAXENT for ecologists. *Diversity and Distributions*, 17(1), 43-57. doi:10.1111/j.1472-4642.2010.00725.x
- Espindola, A., Pellissier, L., Maiorano, L., Hordijk, W., Guisan, A., & Alvarez, N. (2012). Predicting present and future intra-specific genetic structure through niche hindcasting across 24 millennia. *Ecology Letters*, 15(7), 649-657. doi:10.1111/j.1461-0248.2012.01779.x
- Estoup, A., Baird, S. J., Ray, N., Currat, M., Cornuet, J. M., Santos, F., . . . Excoffier, L. (2010). Combining genetic, historical and geographical data to reconstruct the dynamics of bioinvasions: application to the cane toad *Bufo marinus*. *Molecular Ecology Resour*, 10(5), 886-901. doi:10.1111/j.1755-0998.2010.02882.x
- Estrada, A., Delgado, M. P., Arroyo, B., Traba, J., & Morales, M. B. (2016). Forecasting large-scale habitat suitability of European bustards under climate change: The role of environmental and geographic variables. *PLOS ONE*, 11(3), e0149810. doi:10.1371/journal.pone.0149810
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14(8), 2611-2620. doi:10.1111/j.1365-294X.2005.02553.x
- Excoffier, L., & Lischer, H. E. (2010). ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resour*, 10(3), 564-567. doi:10.1111/j.1755-0998.2010.02847.x
- Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164(4), 1567-1587.
- Fordham, D. A., Brook, B. W., Moritz, C., & Nogués-Bravo, D. (2014). Better forecasts of range dynamics using genetic data. *Trends in Ecology & Evolution*, 29(8), 436-443. doi:https://doi.org/10.1016/j.tree.2014.05.007
- Fourcade, Y., Engler, J. O., Rödder, D., & Secondi, J. (2014). Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PLOS ONE*, 9(5), e97122. doi:10.1371/journal.pone.0097122
- García, M. D., Larrosa, E., Clemente, M. E., & Presa, J. J. (2005). Contribution to the knowledge of genus *Dociostaurus* Fieber, 1853 in the Iberian Peninsula, with special reference to its sound production (Orthoptera: Acridoidea). *Anales de Biología*, 27, 155-189.
- Gómez, A., & Lunt, D. H. (2007). Refugia within refugia: Patterns of phylogeographic concordance in the Iberian Peninsula. In S. Weiss, & N. Ferrand (Eds.), *Phylogeography of Southern European Refugia* (pp. 155-188). Dordrecht, The Netherlands: Springer. doi.org/10.1007/1-4020-4904-8
- González-Serna, M. J., Ortego, J., & Cordero, P. J. (2018). A review of cross-backed grasshoppers of the genus *Dociostaurus* Fieber (Orthoptera: Acrididae) from the western Mediterranean: insights from phylogenetic analyses and DNA-based species delimitation. *Systematic Entomology*, 43(1), 136-146. doi:10.1111/syen.12258
- Hasumi, H., & Emori, S. (2004). *K-1 Coupled GCM (MIROC) Description*. Tokyo: Center for Climate System Research (CCSR), University of Tokyo; National Institute for Environmental Studies (NIES); Frontier Research Center for Global Change (FRCGC).
- He, Q., Edwards, D. L., & Knowles, L. L. (2013). Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution*, 67(12), 3386-3402. doi:10.1111/evo.12159

- 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(15), 1965-1978. doi:10.1002/joc.1276
- Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907-913. doi:10.1038/35016000
- Hewitt, G. (2011). Mediterranean Peninsulas: The Evolution of Hotspots. In F. E. Zachos & J. C. Habel (Eds.), *Biodiversity Hotspots: Distribution and Protection of Conservation Priority Areas* (pp. 123-148). New York, NY: Springer-Verlag.
- Hoban, S. (2014). An overview of the utility of population simulation software in molecular ecology. *Molecular Ecology*, 23(10), 2383-2401. doi:10.1111/mec.12741
- Hochkirch, A., Nieto, A., García-Criado, M., Cálix, M., Braud, Y., Buzzetti, F. M., . . . Tumbrinck, J. (2016). *European Red List of Grasshoppers, Crickets and Bush-crickets*. Luxembourg: Publications Office of the European Union.
- Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLOS Genetics*, 6(2), e1000862. doi:10.1371/journal.pgen.1000862
- 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(5), 1322-1332. doi:10.1111/j.1755-0998.2009.02591.x
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14), 1801-1806. doi:10.1093/bioinformatics/btm233
- Jeffreys, H. (1961). *Theory of Probability*. Oxford, UK: Oxford University Press.
- Jombart, T. (2008). *Adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403-1405. doi:10.1093/bioinformatics/btn129
- Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430), 773-795. doi:10.1080/01621459.1995.10476572
- Keightley, P. D., Ness, R. W., Halligan, D. L., & Haddrill, P. R. (2014). Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. *Genetics*, 196(1), 313-320. doi:10.1534/genetics.113.158758
- Keller, D., van Strien, M. J., Herrmann, M., Bolliger, J., Edwards, P. J., Ghazoul, J., & Holderegger, R. (2013). Is functional connectivity in common grasshopper species affected by fragmentation in an agricultural landscape? *Agriculture, Ecosystems & Environment*, 175, 39-46. doi:https://doi.org/10.1016/j.agee.2013.05.006
- Keyghobadi, N., Roland, J., Matter, S. F., & Strobeck, C. (2005). Among- and within-patch components of genetic diversity respond at different rates to habitat fragmentation: an empirical demonstration. *Proceedings of the Royal Society B-Biological Sciences*, 272(1562), 553-560. doi:10.1098/rspb.2004.2976
- Knowles, L. L. (2000). Tests of pleistocene speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of western North America. *Evolution*, 54(4), 1337-1348.
- Knowles, L. L., & Massatti, R. (2017). Distributional shifts – not geographic isolation – as a probable driver of montane species divergence. *Ecography*, 40(12), 1475-1485. doi:10.1111/ecog.02893
- Landguth, E. L., Cushman, S. A., Schwartz, M. K., McKelvey, K. S., Murphy, M., & Luikart, G. (2010). Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, 19(19), 4179-4191. doi:10.1111/j.1365-294X.2010.04808.x
- Lange, R., Durka, W., Holzhauer, S. I., Wolters, V., & Diekötter, T. (2010). Differential threshold effects of habitat fragmentation on gene flow in two widespread species of bush crickets. *Molecular Ecology*, 19(22), 4936-4948. doi:10.1111/j.1365-294X.2010.04877.x
- Leuenberger, C., & Wegmann, D. (2010). Bayesian computation and model selection without likelihoods. *Genetics*, 184(1), 243-252. doi:10.1534/genetics.109.109058

- Liu, X., & Fu, Y.-X. (2015). Exploring population size changes using SNP frequency spectra. *Nature Genetics*, 47(5), 555-559. doi:10.1038/ng.3254
- Lorenzen, E. D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K. A., . . . Willerslev, E. (2011). Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, 479, 359-365. doi:10.1038/nature10574
- Massatti, R., & Knowles, L. L. (2016). Contrasting support for alternative models of genomic variation based on microhabitat preference: species-specific effects of climate change in alpine sedges. *Molecular Ecology*, 25(16), 3974-3986. doi:10.1111/mec.13735
- Mattila, T. M., Tyrmi, J., Savolainen, O., & Pyhäjärvi, T. (2017). Genome-wide analysis of colonization history and concomitant selection in *Arabidopsis lyrata*. *Molecular Biology and Evolution*, 34(10), 2665-2677. doi:10.1093/molbev/msx193
- Metcalf, J. L., Prost, S., Nogués-Bravo, D., DeChaine, E. G., Anderson, C., Batra, P., . . . Guralnick, R. P. (2014). Integrating multiple lines of evidence into historical biogeography hypothesis testing: a *Bison bison* case study. *Proceedings of the Royal Society B-Biological Sciences*, 281(1777), 20132782. doi:10.1098/rspb.2013.2782
- Mevik, B.-H., & Wehrens, R. (2007). The *pls* Package: Principal component and partial least squares regression in R. *Journal of Statistical Software*, 18(2), 1-23.
- Miles, A., Harding, N. J., Bottà, G., Clarkson, C. S., Antão, T., Kozak, K., . . . Kwiatkowski, D. P. (2017). Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature*, 552, 96-102. doi:10.1038/nature24995
- Morrison, L., Estrada, A., & Early, R. (2018). Species traits suggest European mammals facing the greatest climate change are also least able to colonize new locations. *Diversity and Distributions*, 24(9), 1321-1332. doi:10.1111/ddi.12769
- Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMEVAL: An R package for conducting spatially independent evaluations and estimating optimal model complexity for MAXENT ecological niche models. *Methods in Ecology and Evolution*, 5(11), 1198-1205. doi:10.1111/2041-210X.12261
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403, 853-858. doi:10.1038/35002501
- Neuenschwander, S., Lurgiader, C. R., Ray, N., Currat, M., Vonlanthen, P., & Excoffier, L. (2008). Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. *Molecular Ecology*, 17(3), 757-772. doi:10.1111/j.1365-294X.2007.03621.x
- Noguerales, V., Cordero, P. J., & Ortego, J. (2016). Hierarchical genetic structure shaped by topography in a narrow-endemic montane grasshopper. *BMC Evolutionary Biology*, 16, 96. doi:10.1186/s12862-016-0663-7
- Noguerales, V., Cordero, P. J., & Ortego, J. (2018). Inferring the demographic history of an oligophagous grasshopper: Effects of climatic niche stability and host-plant distribution. *Molecular Phylogenetics and Evolution*, 118, 343-356. doi:10.1016/j.ympev.2017.10.012
- Ortego, J., Aguirre, M. P., & Cordero, P. J. (2010). Population genetics of *Mioscirtus wagneri*, a grasshopper showing a highly fragmented distribution. *Molecular Ecology*, 19(3), 472-483. doi:10.1111/j.1365-294X.2009.04512.x
- Ortego, J., Aguirre, M. P., Noguerales, V., & Cordero, P. J. (2015). Consequences of extensive habitat fragmentation in landscape-level patterns of genetic diversity and structure in the Mediterranean esparto grasshopper. *Evolutionary Applications*, 8(6), 621-632. doi:10.1111/eva.12273
- Ortego, J., Gugger, P. F., & Sork, V. L. (2018). Genomic data reveal cryptic lineage diversification and introgression in Californian golden cup oaks (section *Protobalanus*). *New Phytologist*, 2, 804-818. doi:10.1111/nph.14951

- Oyler-McCance, S. J., Fedy, B. C., & Landguth, E. L. (2013). Sample design effects in landscape genetics. *Conservation Genetics*, 14(2), 275-285. doi:10.1007/s10592-012-0415-1
- Papadopoulou, A., & Knowles, L. L. (2015). Species-specific responses to island connectivity cycles: refined models for testing phylogeographic concordance across a Mediterranean Pleistocene Aggregate Island Complex. *Molecular Ecology*, 24(16), 4252-4268. doi:10.1111/mec.13305
- Papadopoulou, A., & Knowles, L. L. (2016). Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. *Proceedings of the National Academy of Sciences of the United States of America*, 113(29), 8018-8024. doi:10.1073/pnas.1601069113
- Paquette, S. R., Talbot, B., Garant, D., Mainguy, J., & Pelletier, F. (2014). Modelling the dispersal of the two main hosts of the raccoon rabies variant in heterogeneous environments with landscape genetics. *Evolutionary Applications*, 7(7), 734-749. doi:10.1111/eva.12161
- Paz, A., Ibáñez, R., Lips, K. R., & Crawford, A. J. (2015). Testing the role of ecology and life history in structuring genetic variation across a landscape: a trait-based phylogeographic approach. *Molecular Ecology*, 24(14), 3723-3737. doi:10.1111/mec.13275
- Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., & Araújo, M. B. (2011). *Ecological Niches and Geographic Distributions*. Princeton, NJ: Princeton University Press.
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLOS ONE*, 7(5), e37135. doi:10.1371/journal.pone.0037135
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3), 231-259. doi:https://doi.org/10.1016/j.ecolmodel.2005.03.026
- Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with MAXENT: new extensions and a comprehensive evaluation. *Ecography*, 31(2), 161-175. doi:10.1111/j.0906-7590.2008.5203.x
- Prates, I., Xue, A. T., Brown, J. L., Alvarado-Serrano, D. F., Rodrigues, M. T., Hickerson, M. J., & Carnaval, A. C. (2016). Inferring responses to climate dynamics from historical demography in neotropical forest lizards. *Proceedings of the National Academy of Sciences of the United States of America*, 113(29), 7978-7985. doi:10.1073/pnas.1601063113
- Presa, J. J., García, M., Clemente, M., Barranco Vega, P., Correias, J., Ferreira, S., . . . Prunier, F. (2016). *Dociopterus hispanicus*: the IUCN Red List of Threatened Species (Publication no. http://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T16084433A75088044.en). Retrieved from https://www.iucnredlist.org/species/16084433/75088044
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945-959.
- R Core Team. (2018). R: A Language and Environment for Statistical Computing. Retrieved from https://www.R-project.org
- Radosavljevic, A., & Anderson, R. P. (2014). Making better MAXENT models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography*, 41(4), 629-643. doi:10.1111/jbi.12227
- Rand, A. L. (1948). Glaciation, an isolating factor in speciation. *Evolution*, 2(4), 314-321.
- Ray, N., Currat, M., Foll, M., & Excoffier, L. (2010). SPLATCHE2: a spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics*, 26(23), 2993-2994. doi:10.1093/bioinformatics/btq579
- Reid, B. N., Kass, J. M., Wollney, S., Jensen, E. L., Russello, M. A., Viola, E. M., . . . Naro-Maciel, E. (2019). Disentangling the genetic effects of refugial isolation and range expansion in

- a trans-continentially distributed species. *Heredity (Edinb)*. doi:10.1038/s41437-018-0135-5
- Reinhardt, K., Köhler, G., Maas, S., Detzel, P., & Spence, J. R. (2005). Low dispersal ability and habitat specificity promote extinctions in rare but not in widespread species: the Orthoptera of Germany. *Ecography*, 28(5), 593-602.
- Rissler, L. (2016) Union of phylogeography and landscape genetics. *Proceedings of the National Academy of Sciences of the United States of America*, 113(29), 8079-8086.
- Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, 4(1), 137-138. doi:10.1046/j.1471-8286.2003.00566.x
- Rubidge, E. M., Patton, J. L., Lim, M., Burton, A. C., Brashares, J. S., & Moritz, C. (2012). Climate-induced range contraction drives genetic erosion in an alpine mammal. *Nature Climate Change*, 2, 285-288. doi:10.1038/nclimate1415
- Saccheri, I., Kuussaari, M., Kankare, M., Vikman, P., Fortelius, W., & Hanski, I. (1998). Inbreeding and extinction in a butterfly metapopulation. *Nature*, 392, 491-494. doi:10.1038/33136
- Sandel, B., Arge, L., Dalsgaard, B., Davies, R. G., Gaston, K. J., Sutherland, W. J., & Svenning, J.-C. (2011). The influence of late Quaternary climate-change velocity on species endemism. *Science*, 334(6056), 660-664. doi:10.1126/science.1210173
- Schwartz, M. K., & McKelvey, K. S. (2008). Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, 10(2), 441-452. doi:10.1007/s10592-008-9622-1
- Short, K. H., & Petren, K. (2011). Fine-scale genetic structure arises during range expansion of an invasive gecko. *PLOS ONE*, 6(10), e26258. doi:10.1371/journal.pone.0026258
- Slatyer, R. A., Nash, M. A., & Hoffmann, A. A. (2016). Scale-dependent thermal tolerance variation in Australian mountain grasshoppers. *Ecography*, 39(6), 572-582. doi:10.1111/ecog.01616
- Stewart, J. R., Lister, A. M., Barnes, I., & Dalén, L. (2010). Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B-Biological Sciences*, 277(1682), 661-671. doi:10.1098/rspb.2009.1272
- Strangas, M. L., Navas, C. A., Rodrigues, M. T., & Carnaval, A. C. (2018). Thermophysiology, microclimates, and species distributions of lizards in the mountains of the Brazilian Atlantic Forest. *Ecography*, 0(0). doi:10.1111/ecog.03330
- Sukumaran, J., & Knowles, L. L. (2018). Trait-dependent biogeography: (re)integrating biology into probabilistic historical biogeographical models. *Trends in Ecology & Evolution*, 33(6), 390-398. doi:10.1016/j.tree.2018.03.010
- Talbot, B., Vonhof, M. J., Broders, H. G., Fenton, B., & Keyghobadi, N. (2016). Range-wide genetic structure and demographic history in the bat ectoparasite *Cimex adjunctus*. *BMC Evolutionary Biology*, 16, 268. doi:10.1186/s12862-016-0839-1
- van der Mescht, L., Matthee, S., & Matthee, C. A. (2015). Comparative phylogeography between two generalist flea species reveal a complex interaction between parasite life history and host vicariance: parasite-host association matters. *BMC Evolutionary Biology*, 15, 105. doi:10.1186/s12862-015-0389-y
- Vandergast, A. G., Bohonak, A. J., Weissman, D. B., & Fisher, R. N. (2007). Understanding the genetic effects of recent habitat fragmentation in the context of evolutionary history: phylogeography and landscape genetics of a southern California endemic Jerusalem cricket (Orthoptera: Stenopelmataidae: *Stenopelmatus*). *Molecular Ecology*, 16(5), 977-992. doi:10.1111/j.1365-294X.2006.03216.x
- Vera, J. A. (2004). *Geología de España* (J. A. Vera Ed.). Madrid, Spain: SGE-IGME, Madrid.
- Villaverde, T., Pokorný, L., Olsson, S., Rincón-Barrado, M., Johnson, M. G., Gardner, E. M., . . . Sanmartín, I. (2018). Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytologist*, 220(2), 636-650. doi:10.1111/nph.15312

- Wachter, G. A., Papadopoulou, A., Muster, C., Arthofer, W., Knowles, L. L., Steiner F. M., & Schlick-Steiner, B. C. (2016). Glacial refugia, recolonization patterns and diversification forces in Alpine-endemic *Megabunus* harvestmen. *Molecular Ecology*, 25(12), 2904-2919. doi:10.1111/mec.13634
- Waltari, E., Hijmans, R. J., Peterson, A. T., Nyári, Á. S., Perkins, S. L., & Guralnick, R. P. (2007). Locating Pleistocene refugia: Comparing phylogeographic and ecological niche model predictions. *PLOS ONE*, 2(7), e563. doi:10.1371/journal.pone.0000563
- Wang, I. J. (2010). Recognizing the temporal distinctions between landscape genetics and phylogeography. *Molecular Ecology*, 19(13), 2605-2608. doi:10.1111/j.1365-294X.2010.04715.x
- Wang, I. J. (2012). Environmental and topographic variables shape genetic structure and effective population sizes in the endangered Yosemite toad. *Diversity and Distributions*, 18(10), 1033-1041. doi:10.1111/j.1472-4642.2012.00897.x
- Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, 23(23), 5649-5662. doi:10.1111/mec.12938
- Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in MAXENT: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, 21(2), 335-342.
- Wegmann, D., Dupanloup, I., & Excoffier, L. (2008). Width of gene expression profile drives alternative splicing. *PLOS ONE*, 3(10), e3587. doi:10.1371/journal.pone.0003587
- Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010). ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, 11, 116. doi:10.1186/1471-2105-11-116
- Weir, J. T., Haddrath, O., Robertson, H. A., Colbourne, R. M., & Baker, A. J. (2016). Explosive ice age diversification of kiwi. *Proceedings of the National Academy of Sciences of the United States of America*, 113(38), E5580-E5587. doi:10.1073/pnas.1603795113
- Worth, J. R. P., Williamson, G. J., Sakaguchi, S., Nevill, P. G., & Jordan, G. J. (2014). Environmental niche modelling fails to predict Last Glacial Maximum refugia: niche shifts, microrefugia or incorrect palaeoclimate estimates? *Global Ecology and Biogeography*, 23(11), 1186-1197. doi:10.1111/geb.12239
- Yannic, G., Ortego, J., Pellissier, L., Lecomte, N., Bernatchez, L., & Côté, S. D. (2018). Linking genetic and ecological differentiation in an ungulate with a circumpolar distribution. *Ecography*, 41(6), 922-937. doi:10.1111/ecog.02995
- Zellmer, A. J., & Knowles, L. L. (2009). Disentangling the effects of historic vs. contemporary landscape structure on population genetic divergence. *Molecular Ecology*, 18(17), 3593-3602. doi:10.1111/j.1365-294X.2009.04305.x

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

1084 **Table 1** Geographical location and genetic statistics (π , H_O , H_E , P and F_{IS}) for the studied populations of *Dociostaurus hispanicus*. Genetic statistics were
1085 calculated in STACKS for all positions (polymorphic and non-polymorphic) and only variant (polymorphic) positions considering loci that were represented in
1086 all populations ($p = 10$) and the 80% of individuals within populations ($r = 0.8$). Number of analysed individuals (n); average values across loci are presented
1087 for nucleotide diversity (π), observed (H_O) and expected (H_E) heterozygosity, major allele frequency (P), and the Wright's inbreeding coefficient (F_{IS}).
1088

Locality (Province)	Code	n	Latitude	Longitude	All positions (variant and fixed)					Variant positions				
					π	H_O	H_E	P	F_{IS}	π	H_O	H_E	P	F_{IS}
Trabanca (Salamanca)	TRAB	7	41.242753	-6.402215	0.0010	0.0009	0.0010	0.9993	0.0003	0.1330	0.1188	0.1233	0.9094	0.0373
Navas de San Antonio (Segovia)	NAVA	6	40.744560	-4.299340	0.0010	0.0009	0.0010	0.9993	0.0003	0.1354	0.1216	0.1237	0.9088	0.0327
Puerto de Peña Negra (Ávila)	PNEG	6	40.432180	-5.316690	0.0010	0.0010	0.0009	0.9993	0.0000	0.1253	0.1237	0.1146	0.9150	0.0029
Ávila (Ávila)	AVIL	6	40.624106	-4.689135	0.0009	0.0008	0.0008	0.9994	0.0003	0.1132	0.0985	0.1037	0.9224	0.0326
Aldea del Fresno (Madrid)	ALDE	6	40.391270	-4.209260	0.0011	0.0010	0.0010	0.9992	0.0003	0.1455	0.1292	0.1327	0.9028	0.0388
Casa Ventorro (Cáceres)	VENT	6	39.685520	-5.763620	0.0012	0.0011	0.0011	0.9992	0.0004	0.1561	0.1359	0.1428	0.8954	0.0496
Trujillo (Cáceres)	TRUJ	6	39.474870	-5.892462	0.0012	0.0011	0.0011	0.9992	0.0002	0.1556	0.1426	0.1424	0.8960	0.0318
Belalcázar (Córdoba)	BELA	6	38.594329	-5.157927	0.0013	0.0011	0.0011	0.9992	0.0004	0.1617	0.1392	0.1478	0.8919	0.0550
Valle de Alcudia (Ciudad Real)	ALCU	6	38.566058	-4.310914	0.0012	0.0011	0.0011	0.9992	0.0003	0.1554	0.1394	0.1420	0.8961	0.0369
Santa Elena (Jaén)	SANT	6	38.332111	-3.529301	0.0012	0.0011	0.0011	0.9992	0.0004	0.1597	0.1397	0.1462	0.8934	0.0495

Table 2 Statistics from the Approximate Bayesian Computation (ABC) procedure used for evaluating the relative support of each demographic model. A higher marginal density corresponds to a higher model support and non-significant p -values indicate that the model is able to reproduce data in agreement with observed empirical data (Wegmann et al., 2010). Bayes factors (BF) represent the degree of relative support for the most supported model (in bold) over the other models. BF >20 indicate strong support, while those >150 indicate very strong support (Kass and Raftery, 1995). Models G-I were fine-tuned considering different relative values of carrying capacities for unsuitable and suitable habitats (1:10, 1:100 and 1:1000) (Ratio). The coefficient of determination (R^2) between each parameter and the five partial least squares (PLS) components retained is indicative of the power of estimating the parameters.

Model	Description	Ratio	Marginal density	p -value	Bayes Factor	R^2		
						$K_{\max}$	m	N_{ANC}
A	Flat landscape	-	9.06×10^{-07}	0.078	448.08	0.275	0.826	0.924
B	Flat landscape + barriers	-	3.60×10^{-05}	0.357	11.26	0.417	0.837	0.916
C	Flat landscape + roughness	-	7.11×10^{-08}	0.005	5708.09	0.385	0.812	0.927
D	ENM	-	4.02×10^{-08}	0.006	10089.70	0.267	0.826	0.929
E	ENM + barriers	-	7.03×10^{-06}	0.335	57.78	0.386	0.836	0.927
F	ENM + roughness	-	1.15×10^{-07}	0.002	3534.66	0.331	0.818	0.923
G	Habitat	1:10	1.85×10^{-05}	0.774	21.95	0.301	0.822	0.918
G	Habitat	1:100	2.89×10^{-05}	0.908	14.07	0.313	0.824	0.918
G	Habitat	1:1000	1.93×10^{-05}	0.804	20.99	0.288	0.826	0.920
H	Habitat + barriers	1:10	4.06×10^{-04}	0.957	1.00	0.397	0.845	0.922
H	Habitat + barriers	1:100	1.88×10^{-04}	0.988	2.16	0.372	0.825	0.924
H	Habitat + barriers	1:1000	7.72×10^{-05}	0.804	5.26	0.366	0.820	0.921
I	Habitat + roughness	1:10	1.38×10^{-07}	0.010	2943.72	0.371	0.834	0.923
I	Habitat + roughness	1:100	1.50×10^{-06}	0.032	270.10	0.374	0.844	0.929
I	Habitat + roughness	1:1000	2.12×10^{-06}	0.060	191.87	0.348	0.817	0.923

Figure 1 Workflow summarizing the integrative approach employed in this study to infer the processes shaping the spatial distribution of genomic variation in the cross-backed grasshopper *Dociostaurus hispanicus*. iDDC: Integrative distributional, demographic and coalescent modelling; ABC: Approximate Bayesian computation; RM: Regularization multiplier in MAXENT; FC: Feature class in MAXENT; LGM: Last Glacial Maximum.

Figure 2 Map displaying the geographical location of sampling populations of *Dociostaurus hispanicus* in the Iberian Peninsula. Warmer colours and larger circle sizes denote higher levels of population genetic diversity (π , nucleotide diversity). The dashed polygon indicates the area considered to run SPLATCHE2 analyses (see Figure 3). Population codes are described in Table 1.

Figure 3 Layers used to run alternative demographic models in SPLATCHE2. Model A is a static model defined by a flat landscape in which carrying capacities of demes (k) are homogeneous across space and time. Model D is a dynamic model in which carrying capacities change over three time periods and are scaled proportionally to logistic environmental suitability scores (red: high; blue: low) obtained from projections of the species-specific environmental niche model (ENM) to the present (panel c) and the last glacial maximum (panel a). Model G is a dynamic model starting with a flat landscape (panels a and b) that shifted to a contemporary fragmented landscape in which carrying capacities for habitats unsuitable for the species (in blue) are smaller than those assigned to suitable habitats (in red) (panel c). Each of these three basic models was also run considering barriers to dispersal (areas with slopes > 20 % not occupied by the focal taxon; in black) (Models B, E, and H, respectively) and incorporating a friction layer defined by topographic roughness (Models C, F, and I, respectively; not shown). Green dots in initial layers of each model indicate the locations of ancestral populations used to initiate the simulations. Panel (c) for Model D shows the species records (black crosses) used to build the ENM. The geographical location of the layers used to run the models is indicated in Figure 2.

Figure 4 Results of STRUCTURE analyses showing posterior probability plots of individual assignments to the different genetic clusters inferred at each hierarchical level. Subsets of populations included in each analysis were defined according to their assignment to the genetic clusters inferred at the previous hierarchical level. Each individual is represented by a vertical bar, which is partitioned into k coloured segments showing the individual's probability of belonging to the cluster with that colour. Thin vertical black lines separate individuals from different populations. Population codes are described in Table 1.

1141 **Figure 5** Demographic history of each studied population of *Docioptaurus hispanicus* inferred
1142 using STAIRWAY PLOT. Lines show the median estimate of N_e over time for (a) Northern
1143 populations (in blue) and (b) Southern populations (in red). N_e is the effective population size,
1144 assuming a mutation rate of 2.8×10^{-9} and 1-year generation time. Population codes shown in
1145 the legend of each panel are described in Table 1.