Insect Systematics and Diversity, (2021) 5(5): 2; 1–14
doi: 10.1093/isd/ixab009
Research

ENTOMOLOGICAL SOCIETY OF AMERICA
SHARING INSECT SCIENCE GLOBALLY

OXFORD

Molecular Phylogenetics, Phylogenomics, and Phylogeography

Broadly Distributed but Genetically Fragmented: Demographic Consequences of Pleistocene Climatic Oscillations in a Common Iberian Grasshopper

Joaquín Ortego,[1,5] Víctor Noguerales,[2] Vanina Tonzo,[1] María José González-Serna,[3] and Pedro J. Cordero[3,4]

[1]Department of Integrative Ecology, Estación Biológica de Doñana (EBD-CSIC), Avda. Américo Vespucio 26, E-41092 Seville, Spain, [2]Instituto de Productos Naturales y Agrobiología (IPNA-CSIC), Avda. Astrofísico Fco. Sánchez 3, E-38206 La Laguna, Tenerife, Canary Islands, Spain, [3]Genetic and Cultural Biodiversity Research Group, Instituto de Investigación en Recursos Cinegéticos (IREC, CSIC-UCLM-JCCM), Ronda de Toledo 12, E-13071 Ciudad Real, Spain, [4]Escuela Técnica Superior de Ingenieros Agrónomos (ETSIA), Universidad de Castilla-La Mancha (UCLM), Ronda de Calatrava 7, E-13071 Ciudad Real, Spain, and [5]Corresponding author, e-mail: joaquin.ortego@csic.es

Subject Editor: Jeffrey Lozier

Received 4 January 2021; Editorial decision 18 March 2021

Abstract

Although the genetic consequences of contemporary landscape composition and range shifts driven Pleistocene climatic oscillations have been studied fairly well in alpine organisms, we know much less about how these factors have shaped the demography of taxa with broader climatic niches and distributions. Here, we use high-throughput sequencing data to study the processes underlying spatial patterns of genomic variation in *Omocestus panteli* (Bolívar, 1887) (Orthoptera: Acrididae), a common Iberian grasshopper distributed across numerous habitat types and a wide elevational range (from sea level to >2,000 m). Although the species is broadly distributed, our analyses support that its contemporary populations show significant genetic fragmentation that dates back to the last glacial period. Accordingly, spatially explicit testing of alternative gene flow scenarios and demographic inference analyses revealed that genetic differentiation between populations and their long-term effective population sizes are best explained by the spatial configuration of environmentally suitable habitats during the last glacial maximum (ca. 21 ka). At that time, the species experienced net demographic expansions but interspersed unsuitable areas might have disrupted gene flow and created opportunity for geographical diversification. Collectively, our analyses indicate that the genetic makeup of contemporary populations is not well explained by current environmental factors or geographical barriers to dispersal but mostly reflects genetic fragmentation during the last glacial period followed by postglacial admixture among previously isolated gene pools. Taken together, these results support that the Pleistocene 'species pump' model might be also useful in explaining demographic dynamics and geographical diversification in taxa characterized by broad climatic niches.

Key words: demographic inference, genetic diversity, genetic structure, landscape genetics, phylogeography

Identifying the ecological and evolutionary phenomena underlying spatial patterns of genetic variation in natural populations is of utmost importance to understanding the processes that fuel geographical diversification and speciation (Knowles et al. 2007, Maier et al. 2019). Contemporary landscape composition (e.g., spatial distribution of suitable habitats and geographical barriers to dispersal; Manel et al. 2003) and selectively driven divergence associated with environmental heterogeneity (e.g., contrasting climates or habitats; Wang and Bradburd 2014) have been identified as important drivers of genetic differentiation among populations in numerous groups of organisms (Wang et al. 2009, Shafer and Wolf 2013, Sexton et al. 2014, González-Serna et al. 2019). However, contemporary patterns of genomic variation often reflect the cues left behind by past demographic dynamics (Lanier et al. 2015, Glover et al. 2018,

Branstetter and Longino 2019, Caterino and Langton-Myers 2019). Remarkably, environmental changes driven by Pleistocene climatic oscillations have drastically altered the distributions and spatial configuration of suitable habitats in many taxa from temperate regions (Hewitt 2000; e.g., Carnaval et al. 2009, He et al. 2013). Range shifts in latitude or elevation have led to recurrent processes of population fragmentation (allopatric divergence) and expansions (secondary contact) with major impacts on patterns of genetic diversity, structure, and admixture (Lanier et al. 2015, Ortego et al. 2015a). For these reasons, the genetic makeup of present-day populations is often the result of the interaction between the contemporary and ancient processes that have impacted the demography of species (He et al. 2013, Inoue et al. 2015, Lanier et al. 2015, Maier et al. 2019).

Interactions between landscape heterogeneity and species-specific life-history traits (body size, dispersal capacity, habitat requirements, etc.) will ultimately determine the degree of geographical diversification (Paz et al. 2015, Massatti and Knowles 2016, Papadopoulou and Knowles 2016) and whether current patterns of genomic variation are dominated by the signals left by either historical or contemporary demographic processes (Zellmer and Knowles 2009, González-Serna et al. 2019). Taxa with limited vagility and narrow habitat requirements are expected to experience marked demographic changes in response to shifts in landscape composition and show deeper genetic structures (Ikeda et al. 2012, Schoville et al. 2012, Ortego et al. 2015b). This is often the case of alpine species, for which the genetic consequences of contemporary landscape composition and range shifts driven Pleistocene climatic oscillations are well known and fairly predictable (e.g., DeChaine and Martin 2006, Knowles and Carstens 2007, Maier et al. 2019). However, there is less consensus about how Pleistocene glacial cycles have shaped the demography of taxa with broad climatic niches, large distributions or high dispersal capacities. These species are expected to experience more extended gene flow, which will likely erase the genetic signatures left by ancient demographic dynamics and reduce the opportunity for local adaptation (Bohonak 1999, Lenormand 2002). The insights that can be gained from their study are often perceived as limited, and this is probably the reason why generalist and widely distributed species have tended to receive less attention than taxa with narrower ecological requirements in the context of Pleistocene phylogeography research (Jacobsen et al. 2014, Inoue et al. 2015, González-Serna et al. 2020).

Here, we use high-throughput sequencing data to study the processes that have shaped spatial patterns of genomic variation in the Pantel's grasshopper *Omocestus panteli* (Bolívar, 1887) (Orthoptera: Acrididae). This species is a small-sized (females: 18–23 mm; males: 13–17 mm) and macropterous (i.e., long-winged) grasshopper endemic to the Iberian Peninsula (Clemente et al. 1991, Cigliano et al. 2019). It is a very common and abundant species, being distributed across a wide elevational range (from sea level to 2,600 m) and numerous habitat types (Clemente et al. 1991, Presa et al. 2016). These include Mediterranean dry grasslands, meadows, agricultural areas, and abandoned fields (Clemente et al. 1991, Presa et al. 2016). The species is univoltine (i.e., a single generation per year), with adults present from April to November and population peaks in September (Clemente et al. 1991). The wide climatic niche and broad distribution of the species raises the question on whether its populations show widespread gene flow or if, on the contrary, present significant genetic structure linked to contemporary landscape composition or historical range fragmentation driven by Pleistocene climatic oscillations. To address this central question, we 1) first employed a suite of complementary methods to quantify spatial patterns of genetic structure and admixture in the Pantel's grasshopper. These analyses supported significant genetic subdivision but also considerable genetic admixture among range-wide populations of the species, rejecting the hypothesis of genetic panmixia. To shed light on the proximate processes underlying spatial patterns of genomic variation in contemporary populations, 2) we estimated the timing of geographical diversification using a coalescent-based simulation approach, 3) complemented this information with spatially-explicit testing of alternative scenarios of current and historical population connectivity within a landscape genetics framework. Specifically, we tested whether genetic differentiation among populations is explained by geographical distances (isolation-by-distance, IBD; Slatkin 1993), topographic complexity (weighted topographic distances, WTD; Wang 2020), environmental dissimilarity (isolation-by-environment, IBE; Sexton et al. 2014, Wang and Bradburd 2014), and resistance distances defined by the distribution of environmentally suitable habitats as inferred from projections of a species-specific environmental niche model (ENM) to present-day and last glacial maximum (LGM, ca. 21 ka) bioclimatic conditions (isolation-by-resistance, IBR; McRae 2006, McRae and Beier 2007). Finally, we 4) used genomic data to reconstruct the demographic history of the study populations and determine whether effective population sizes are predicted by contemporary or LGM estimates of environmental suitability and if their changes through time are compatible with ENM-based inferences of species' distributional shifts (Fordham et al. 2014).

Methods

Population Sampling

Between 2010 and 2014, we sampled 21 populations of the Pantel's grasshopper *O. panteli*, covering the entire distribution of the species (Presa et al. 2016; Fig. 1). We opted for a genotyping strategy based on covering more localities at the expense of reducing the number of analyzed individuals per population (n = 1–5 individuals/population; Table 1). This allowed us to infer in more detail spatial patterns of genetic structure and admixture but resulted in small sample sizes for some populations (Table 1; Fumagalli 2013). For this reason, landscape genetic analyses and demographic reconstructions (see below for details) were based only on those populations with five genotyped individuals (n = 14 populations; Table 1), a sample size that has proven to provide reliable estimates of genetic diversity and differentiation (Nazareno et al. 2017, González-Serna et al. 2020, Li et al. 2020). We registered spatial coordinates using a Global Positioning System (GPS) and preserved whole specimens at –20°C in 1,500 µl ethanol 96% until needed for genomic analyses. Further details on sampling locations are provided in Table 1.

Genomic Library Preparation

We used NucleoSpin Tissue (Macherey-Nagel, Düren, Germany) kits to extract and purify genomic DNA from a hind leg of each individual. We processed DNA into two genomic libraries using the double-digestion restriction-site associated DNA sequencing procedure (ddRADseq) described in Peterson et al. (2012). In brief, we digested DNA with the restriction enzymes MseI and EcoRI (New England Biolabs, Ipswich, MA) and ligated Illumina adaptors, including unique 7-bp barcodes to the digested fragments of each individual. We pooled ligation products and size-selected them between 475 and 580 bp with a Pippin Prep machine (Sage Science, Beverly, MA). We amplified the fragments by PCR with 12 cycles using the iProofTM High-Fidelity DNA Polymerase (BIO-RAD, Veenendaal, The Netherlands) and sequenced each library in a single-read 150-bp lane on an Illumina HiSeq2500 platform at The Centre for Applied Genomics (Toronto, ON, Canada).

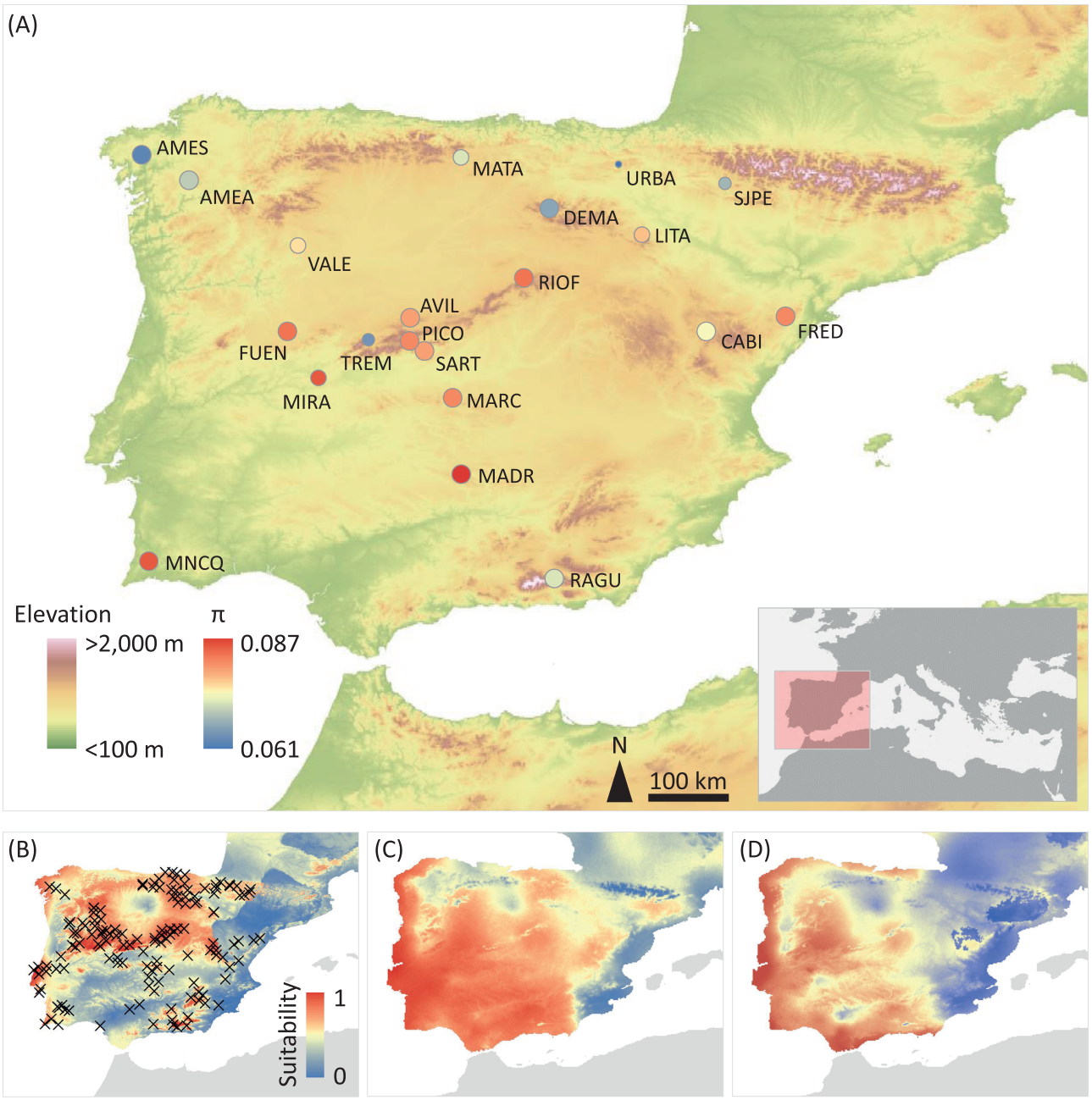

Fig. 1. (A) Map showing the geographical location of sampling populations of Pantel's grasshopper (*Omocestus panteli*), with dot colors indicating their respective levels of genetic diversity (π, in red to blue scale). Dot size is proportional to the number of genotyped individuals (Table 1). (B–D) Projections of the species-specific environmental niche model (ENM) for (B) present and (C–D) last glacial maximum (LGM) bioclimatic conditions under the (C) CCSM4 and (D) MIROC-ESM general atmospheric circulation models. Map in the present shows occurrence points (crosses) used for ENM. Population codes as described in Table 1.

Genomic Data Filtering and Assembling

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 et al. 2013). Reads were de-multiplexed and filtered for overall quality using the program *process_radtags*, retaining reads with a Phred score > 10 (using a sliding window of 15%), no adaptor contamination, and that had an unambiguous barcode and restriction cut site. Raw reads were screened for quality with FASTQC v. 0.11.5 (A. Simon, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and all sequences were trimmed to 129-bp using SEQTK (L. Heng, https://github.com/lh3/seqtk) in order to remove low-quality reads near the 3′ ends. Filtered reads of each individual were assembled *de novo* into putative loci with the *ustacks* program. The minimum stack depth (m) was set to three and we allowed a maximum distance of two nucleotide mismatches (M) to group reads into a 'stack'. We used the 'removal' (r) and 'deleveraging' (d) algorithms to eliminate highly repetitive stacks and resolve over-merged loci, respectively. Single nucleotide polymorphisms (SNPs) were identified at each locus and genotypes were called using a multinomial-based likelihood model that accounts for sequencing errors, with the upper bound of the error rate (ε) set to 0.2 (Catchen et al. 2011,

Table 1. Locality, code, number of genotyped individuals (*n*), latitude, longitude, and elevation for each sampled population.

Locality	Province/district	Country	Code	*n*	Latitude	Longitude	Elevation (m)
Ames	La Coruña	Spain	AMES	5	42.868379	−8.662350	80
Amear	Orense	Spain	AMEA	5	42.544560	−8.016102	720
Vale de Frades	Bragança	Portugal	VALE	4	41.632393	−6.517746	680
Puerto del Tremedal	Ávila	Spain	TREM	3	40.362809	−5.612417	1,620
Sierra de Ávila	Ávila	Spain	AVIL	5	40.655811	−4.983793	1,680
Puerto del Pico	Ávila	Spain	PICO	5	40.346455	−5.014433	1,260
Fuenteguinaldo	Salamanca	Spain	FUEN	5	40.451919	−6.627120	860
Sartajada	Toledo	Spain	SART	5	40.216163	−4.782313	440
Mirabel	Cáceres	Spain	MIRA	4	39.841341	−6.258787	450
Arroyo del Marchés	Toledo	Spain	MARC	5	39.541096	−4.425089	1,010
Sierra Madrona	Ciudad Real	Spain	MADR	5	38.482741	−4.263534	800
Serra de Monchique	Algarve	Portugal	MNCQ	5	37.315478	−8.590395	870
Puerto de La Ragua	Granada	Spain	RAGU	5	37.105101	−3.025233	2,030
Riofrío de Riaza	Segovia	Spain	RIOF	5	41.247832	−3.456831	1,250
Matamorisca	Palencia	Spain	MATA	4	42.842906	−4.322320	980
Sierra de la Demanda	La Rioja	Spain	DEMA	5	42.157401	−3.085410	1,380
Litago	Zaragoza	Spain	LITA	4	41.792922	−1.763370	980
Sierra de Urbasa	Navarra	Spain	URBA	1	42.798817	−2.143702	920
San Juan de la Peña	Huesca	Spain	SJPE	3	42.506952	−0.664140	1,210
Fredes	Tarragona	Spain	FRED	5	40.705840	0.170230	1,090
Puerto de Cabigordo	Teruel	Spain	CABI	5	40.436127	−0.925172	1,560

2013). A catalog of loci was built using the *cstacks* program, with loci recognized as homologous across individuals if the number of nucleotide mismatches between consensus sequences (*n*) was ≤ 2. Each individual data was matched against this catalog using *sstacks* program and output files were exported in different formats for subsequent analyses using the program *populations*. For all downstream analyses, we exported only the first SNP per RAD locus (option *write_single_snp*) and retained loci with a minimum stack depth ≥ 5 (*m* = 5), that were sequenced in at least 50% of the individuals of each locality (parameter *r* = 0.5), represented in ~66% of localities (parameter *p* = 14), and with a minimum minor allele frequency (MAF) ≥ 0.01 to reduce the number of false polymorphic loci due to sequencing errors.

Quantifying Genetic Structure

We employed three complementary approaches to quantify range-wide population genetic structure in the Pantel's grasshopper:

STRUCTURE: We analyzed genetic structure and admixture using the Bayesian Markov chain Monte Carlo clustering method implemented in the program STRUCTURE v. 2.3.3 (Pritchard et al. 2000). We ran STRUCTURE with 200,000 MCMC cycles after a burn-in step of 100,000 iterations, assuming correlated allele frequencies and admixture (Pritchard et al. 2000). We performed 15 independent runs for each value of *K* genetic clusters (*K* = 1 to *K* = 10) to estimate the most likely number of clusters. We retained the 10 runs having the highest likelihood for each value of *K* and determined the number of genetic clusters that best describes our data according to log probabilities of the data (LnPr(X|*K*) (Pritchard et al. 2000) and the Δ*K* method (Evanno et al. 2005), as implemented in STRUCTURE HARVESTER (Earl and 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 and Rosenberg 2007) and DISTRUCT v. 1.1 (Rosenberg 2004) to visualize as bar plots the individual's probabilities of population membership.

Discriminant analysis of principal components (DAPC): We used DAPC to identify clusters of genetically related individuals (Jombart et al. 2010). This method does not lay on the assumptions of STRUCTURE (i.e., HWE and gametic disequilibrium; Pritchard et al. 2000) and could be more efficient to detect complex patterns of genetic differentiation (Jombart et al. 2010). We ran DAPC using the package ADEGENET (Jombart 2008) in R v. 3.6.1 (R Core Team 2020). First, we used the *find.cluster* function using all principal components (PCs) to determine the best-supported number of genetic clusters (from *K* = 1 to *K* = 10) using the Bayesian Information Criterion (BIC). Second, we determined the optimal number of PCs for the DAPC by cross-validation using the *xvalDapc* function with 100 replicates and selected the number of PCs associated with the lowest root mean squared error (RMSE) value. Lastly, we ran DAPC using all available discriminant functions and calculated the individual's probabilities of population membership to each cluster, which were visualized as bar plots using distruct (Rosenberg 2004).

Principal component analysis (PCA): In order to visualize the major axes of genomic variation, we performed a PCA as implemented in the R package ADEGENET (Jombart 2008). Before running the PCA, we replaced missing data with mean allele frequencies using the *scaleGen* function (Jombart 2008).

Divergence Time Estimation

We used the composite-likelihood simulation-based approach implemented in FASTSIMCOAL2 (Excoffier et al. 2013) to estimate divergence times among the three main genetic clusters (northwestern, central-southern, and northeastern) inferred by analyses of genetic structure described in the previous section (see Results section; e.g., Lanier et al. 2015). To avoid potential confounding effects in parameter estimation resulting from post-divergence admixture (revealed by STRUCTURE; see Results section), we selected for FASTSIMCOAL2 analyses the two populations with lowest degree of admixed ancestry for each of the three main genetic clusters (AMES-AMEA, MARC-MADR, and FRED-CABI; i.e., 10 individuals per deme; Table 1; e.g., Tsuda et al. 2015, Zeng et al. 2018). We allowed gene flow between demes and considered that central-southern and northeastern populations share a most recent common ancestor, as supported by the comparatively lower composite likelihood of pilot runs for alternative

topological relationships (Fig. 2). We calculated a folded joint site frequency spectrum (SFS) considering a single SNP per locus to avoid the effects of linkage disequilibrium. To remove all missing data for the calculation of the joint SFS, minimize errors with allele frequency estimates and maximize the number of variable SNPs retained, each population group was downsampled to 70% of individuals (i.e., 7 individuals per deme) using a custom Python script written by Qixin He and available on Dryad (Papadopoulou and Knowles 2015). Final SFS contained 4,105 variable SNPs. Because it is not accurate to estimate the number of monomorphic sites when a single SNP is retained per locus, invariable sites were excluded from likelihood calculations with the '*removeZeroSFS*' option in FASTSIMCOAL2 (Excoffier et al. 2013). For this reason, in order to enable the estimation of all other parameters and improve the performance of simulations by reducing the number of parameters to be estimated from the data (see Excoffier and Foll 2011, Excoffier et al. 2013), we fixed in the analyses the effective population size for one of the demes (northeastern populations, FRED-CABI). Parameters estimates obtained fixing effective population sizes for the two other demes (i.e., AMES-AMEA or MARC-MADR) produced similar results (Supp Table S1 [online only]). The effective population size fixed in the model was calculated directly from empirical data from the level of nucleotide diversity (π) of fixed and variable sites and mutation rate per site per generation estimated for *Drosophila melanogaster* (μ; 2.8×10^{-9}; Keightley et al. 2014; e.g., Tonzo et al. 2020). The model was run 100 replicated times considering 100,000–250,000 simulations for the calculation of the composite likelihood, 10–40 expectation-conditional maximization (ECM) cycles, and a stopping criterion of 0.001 (Excoffier et al. 2013). Point estimates for the different demographic parameters were selected from the replicate with the highest maximum composite likelihood. Finally, we calculated confidence intervals of parameter estimates from 100 parametric bootstrap replicates by simulating SFS from the maximum composite likelihood estimates and re-estimating parameters each time (Excoffier et al. 2013).

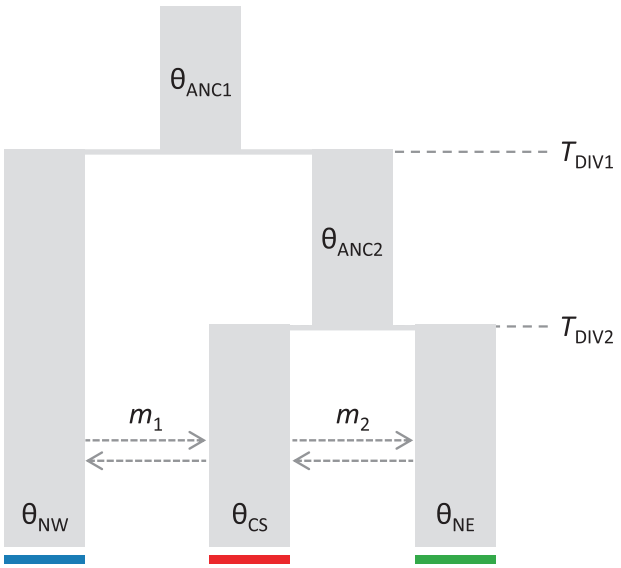

Fig. 2. Demographic model used in FASTSIMCOAL2 analyses to estimate timing of population divergence (T_{DIV1} and T_{DIV2}), mutation-scaled ancestral (θ_{ANC1} and θ_{ANC2}) and contemporary (θ_{NW}, θ_{CS}, and θ_{NE}) effective population sizes, and migration rates per generation (m_1 and m_2).

Environmental Niche Modeling

We built an environmental niche model (ENM) to predict the geographic distribution of environmentally suitable areas for the Pantel's grasshopper both in the present and during the last glacial maximum (LGM, ca. 21 ka). Maps obtained for these two time periods were then used to perform landscape genetic analyses and determine the impact of distributional shifts and environmental suitability on the demography of the species, as detailed in the next sections. To build the ENM, we used the maximum entropy algorithm implemented in MAXENT v. 3.3.3 (Elith et al. 2006, 2011; Phillips et al. 2006, Phillips and Dudík 2008) and the 19 bioclimatic variables from the WORLDCLIM v. 2.1 dataset obtained for the period 1970–2000 (http://www.worldclim.org/) and interpolated to 30-arc-sec resolution (~1 km^2 cell size) (Hijmans et al. 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 and Emori 2004) general atmospheric circulation models. We built the ENM using our own species occurrence data and records available in the literature, the Global Biodiversity Information Facility (GBIF.org, 23 September 2020, GBIF Occurrence Download https://doi.org/10.15468/dl.sxacwk), and iNaturalist (https://www.inaturalist.org/). Prior to modeling, we mapped and examined all records to identify and exclude those corresponding to obvious georeferencing errors. To reduce the problems associated with sampling bias and the strong spatial aggregation of available species records, we applied a systematic sampling correction by randomly selecting a single occurrence record among those falling within the same ~1,500 km^2 grid cell (Fourcade et al. 2014). After this filtering step, we retained a total of 149 occurrence records. We used the package ENMEVAL (Muscarella et al. 2014) in R to conduct parameter tuning and determine the optimal feature class (FC) and regularization multiplier (RM) settings for MAXENT. We tested a total of 128 models of varying complexity by combining a range of regularization multipliers (RM) (from 0 to 15 in increments of 1) with eight different feature classes (FC) combinations (L, LQ, LQP, H, T, LQH, LQHP, LQHPT, where L = linear, Q = quadratic, H = hinge, P = product, and T = threshold) (Muscarella et al. 2014). We compared MAXENT models with different settings using the Akaike Information Criterion corrected for small sample size (AICc) (Warren and Seifert 2011) and followed the approach detailed in González-Serna et al. (2019) for variable selection.

Landscape Genetic Analyses

We applied a landscape genetic approach to analyze a comprehensive suite of factors that could hypothetically explain genetic differentiation (F_{ST}) among populations of the Pantel's grasshopper. Pairwise F_{ST} were calculated using the program *populations* in STACKS (Catchen et al. 2013) and significance determined with Fisher's exact tests after 10,000 permutations as implemented in ARLEQUIN v. 3.5 (Excoffier and Lischer 2010). These analyses were restricted to the 14 populations with five genotyped individuals (Table 1). Explanatory variables include:

(i) Geographical distance: We calculated the geographical distance between sampled populations using GEOGRAPHIC DISTANCE MATRIX GENERATOR v. 1.2.3 (http://biodiversityinformatics.amnh.org/open_source/gdmg).

(ii) Environmental dissimilarity: We used ARCGIS v. 10.3 (ESRI, Redlands, CA, USA) to extract for each sampling locality the values of the 19 present-day bioclimatic variables available in

WORLDCLIM at 30 arc-sec resolution (ca. 1 km^2; Hijmans et al. 2005) and estimate environmental dissimilarity between populations (e.g., Tonzo et al. 2019, González-Serna et al. 2020). To summarize and reduce strong redundancy among the 19 bioclimatic variables, we ran a principal component analysis (PCA) in SPSS v. 26 (IBM, NY, USA). Then, we used the '*dist*' function from R to calculate environmental dissimilarity between each pair of populations based on Euclidean distances for the obtained scores of the first three principal components (PCs; eigenvalues >1), which cumulatively accounted for 88.78% of the variance (PC1: 43.09%; PC2: 32.54%; PC3: 13.15%).

(iii) Weighted topographic distance: We used the elevation layer available in WORLDCLIM to calculate weighted topographic distances between each pair of populations, as implemented in the R package TOPODISTANCE (Wang 2020). We calculated weighted topographic paths using the *topoWeightedDist* function, with a linear function to weight angle of aspect changes and an exponential function to weight the slope between cells, as recommended by Wang (2020).

(iv) IBR: We built spatially-explicit IBR scenarios of population connectivity based on the configuration of environmentally suitable areas as inferred from projections of the ENM to present-day and LGM (both CCSM4 and MIROC-ESM) bioclimatic conditions. Resistance distances between all pairs of populations were calculated under each scenario (current, $LGM_{CCSM,}$ and $LGM_{MIROC-ESM}$) using an eight-neighbor cell connection scheme in CIRCUITSCAPE v. 4.0.5 (McRae 2006, McRae and Beier 2007). We also calculated resistance distances based on a 'flat landscape' (i.e., all cells have equal resistance value), which is analogous to geographical distance but more appropriate for comparison with others competing models also generated with CIRCUITSCAPE (Noguerales et al. 2016, González-Serna et al. 2018).

We used multiple matrix regressions with randomization (MMRR; Wang 2013) to test all distance matrices against a matrix of population genetic differentiation (F_{ST}). The model was initially constructed with all explanatory terms fitted (i.e., a full model) and final model was selected using a backward-stepwise procedure by progressively removing nonsignificant variables (starting with the least significant ones) until all retained terms within the model were significant. Then, we tested the significance of the rejected terms against this model to ensure that no additional variable reached significance. The result is the minimal most adequate model for explaining the variability in the response variable, where only the significant explanatory terms are retained (e.g., Ortego et al. 2015a).

Genetic Diversity and Demographic History

We analyzed whether genetic diversity (nucleotide diversity, π) and long-term effective population sizes (N_e; Sensu Peart et al. 2020) are explained by geography and contemporary or LGM environmental suitability. We calculated nucleotide diversity (π) for each studied population using the program *populations* from STACKS. (Catchen et al. 2013). Long-term N_e, an estimate of the demographic trajectory of populations through evolutionary time, was calculated using STAIRWAY PLOT (Liu and Fu 2015). STAIRWAY PLOT reconstructs changes in N_e through time implementing a SFS-based multi-epoch demographic model that does not require whole-genome sequence data or reference genome information (Liu and Fu 2015). These analyses were restricted to the 14 populations with five genotyped individuals (Table 1), as the calculation of the SFS requires a downsampling procedure to remove missing data. Each population was downsampled to 80% of individuals (i.e., four individuals per population) as detailed for FASTSIMCOAL2 analyses. We ran STAIRWAY PLOT considering a one-year generation time for the species (Clemente et al. 1991), assuming a mutation rate per site per generation of 2.8×10^{-9} (Keightley et al. 2014), and performing 200 bootstrap replicates to estimate 95% confidence intervals (Liu and Fu 2015). Long-term N_e was calculated for each population as the harmonic mean of population sizes inferred by STAIRWAY PLOT at each time interval (for a similar approach, see Peart et al. 2020).

We used generalized linear models (GLM) in SPSS to analyze genetic diversity and long-term N_e in relation to geography (latitude, longitude, and elevation) and environmental suitability as inferred from projections of the ENM to present-day and LGM (both CCSM4 and MIROC-ESM) bioclimatic conditions. Model selection was performed as detailed in the previous section for landscape genetic analyses.

Results

Genomic Data

The average number of reads retained per individual after the different quality filtering steps was 1,810,556 (range= 132,388-2,637,560 reads; Supp Fig. S1 [online only]). On average, this represented 75% (range = 35–86%) of the total number of reads recovered for each individual (Supp Fig. S1 [online only]). The final dataset contained 14,454 SNPs with 33.53% of missing data.

Quantifying Genetic Structure

STRUCTURE analyses identified that the most likely number of genetic clusters was $K = 3$ according to the ΔK criterion, but LnPr(X|K) reached a plateau at $K = 6$ (Supp Fig. S2a [online only]). Analyses for $K = 2$ separated northeastern populations from the rest of the populations (Fig. 3). For $K = 3$, the three genetic clusters separated northwestern, northeastern, and central-southern populations (Fig. 3). Genetic clusters inferred at higher K-values split hierarchically but, in most cases, populations presented a considerable degree of genetic admixture (Fig. 3). Discriminant analysis of principal components (DAPC) presented similarities, but also remarkable differences, with the results yielded by STRUCTURE. First, the most likely number of genetic clusters according to the Bayesian Information Criterion (BIC) was $K = 1$, suggesting lack of genetic structure (Supp Fig. S2b [online only]). However, analyses for $K = 2$–4 separated populations in a geographic fashion and all individuals and populations were assigned with a 100% probability to a single genetic cluster (i.e., lack of genetic admixture). Analyses for $K = 2$ separated northwestern populations from the rest of the populations. For $K = 3$, the three genetic clusters separated northwestern, northeastern, and central-southern populations (Fig. 3). Analyses for $K = 4$ split northeastern populations into two well-defined genetic clusters, one including the nearby populations FRED and CABI and another SJPE, URBA and LITA (Fig. 3). Higher K-values inconsistently assigned populations to different genetic clusters across replicate runs (not shown). Principal component analysis (PCA) confirmed the main results yielded by STRUCTURE and DAPC, separating northwestern, northeastern, and central-southern populations (Fig. 4). The relative placement of populations in the multivariate space also showed a good correspondence with their respective degree of shared ancestry inferred by STRUCTURE (Figs. 3 and 4).

Divergence Time Estimation

Analyses based on the SFS in FASTSIMCOAL2 supported that the three genetic clusters diverged during the last glacial period (Table 2).

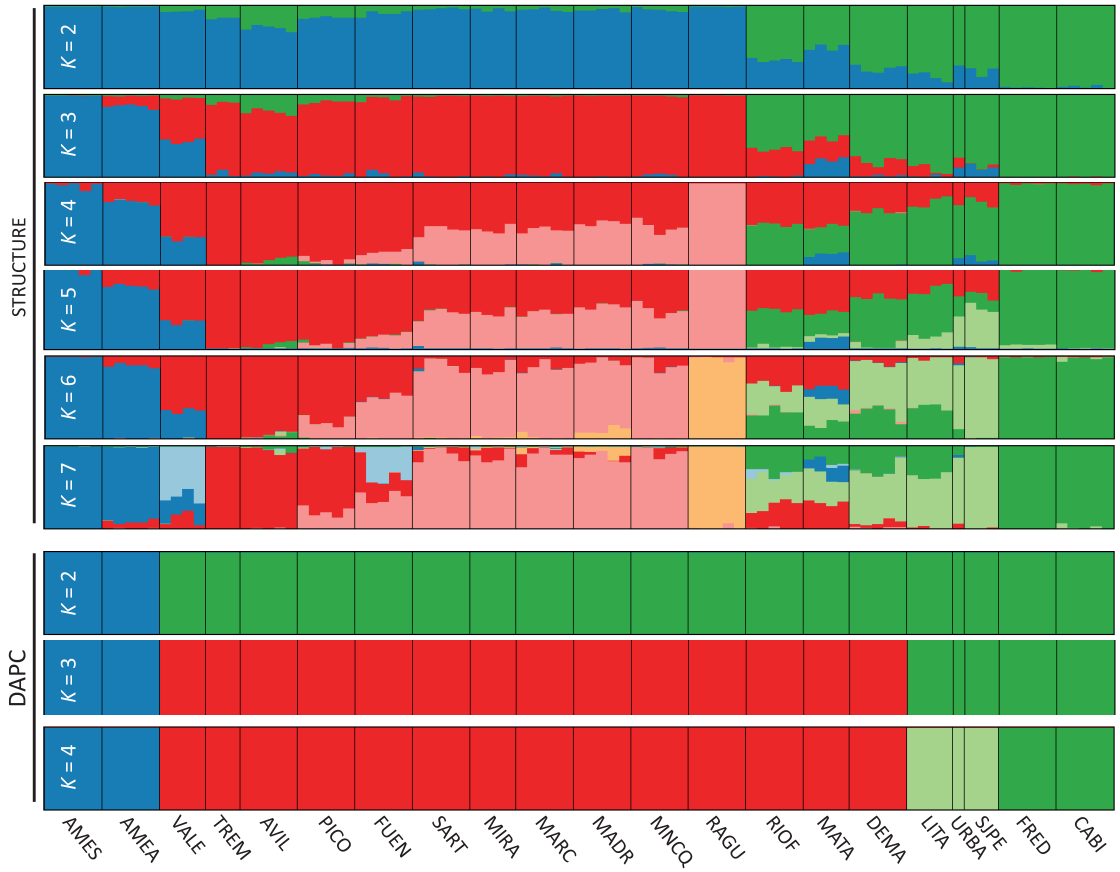

Fig. 3. Results of genetic assignments for populations of the Pantel's grasshopper (*Omocestus panteli*) based on the Bayesian method implemented in the program STRUCTURE and a discriminant analysis of principal components (DAPC). Each individual is represented by a vertical bar, which is partitioned into *K*-colored segments showing the individual's probability of belonging to the cluster with that color. Thin vertical black lines separate individuals from different populations. Analyses are based on a dataset of 14,454 SNPs. Population codes as described in Table 1.

Northwestern populations split from the rest of the populations ca. 112 ka (95% CI: 111–115 ka), whereas the most recent divergence between northeastern and central-southern populations took place ca. 39 ka (95% CI: 38–40 ka) (Table 2). Gene flow among demes was low and estimates of ancestral and contemporary population sizes indicate that the species experienced a demographic expansion after the first genetic split (T_{DIV1}) and a contraction after the most recent split (T_{DIV2}; Table 2).

Environmental Niche Modeling

After parameter tuning, the most supported ENM (i.e., lowest AICc) was the one built considering a regularization multiplier of 2 and a linear-quadratic (LQ) feature class. After removing highly correlated variables ($r \geq 0.9$; BIO7, BIO10, BIO11, BIO13, BIO16, BIO17, and BIO18) and those with zero percent contribution (BIO5 and BIO19), 10 bioclimatic layers were retained to build the final ENM (sorted by percent contribution, BIO1: 35.6%; BIO8: 26.2%; BIO9: 9.5%; BIO14: 8.8%; BIO6: 7.3%; BIO2: 6.3%; BIO4: 2.3%; BIO12: 2.3%; BIO15: 1.0%; BIO2: 0.7%). The map of environmental suitability for current bioclimatic conditions is congruent with the contemporary distribution of *O. panteli*, with higher suitability in the more humid northern half of Iberia and very low suitability around the Mediterranean coast and the Guadalquivir and Ebro depressions where the species becomes rare (Fig. 1B). Projection of the ENM to LGM bioclimatic conditions under the CCSM4 and the MIROC-ESM general circulation models yielded different predictions (Fig. 1C and D). The CCSM4 model predicted higher environmental suitability in absolute terms as well as the presence of the species in some areas (e.g., northeastern Iberia) where the MIROC-ESM model predicted very low probabilities of presence (Fig. 1C and D). In comparison with predictions under current conditions, both CCSSM4 and MIROC-ESM revealed an overall higher availability of environmentally suitable areas in central, southern, and western Iberia and a lower suitability in some areas of northern Iberia (Fig. 1C and D). As shown for current conditions, environmental suitability was predicted to be very low along a large portion of the Mediterranean coast during the LGM (Fig. 1C and D).

Landscape Genetic Analyses

Genetic differentiation (F_{ST}) between populations ranged between 0.093 and 0.176 (Supp Table S2 [online only]). Univariate matrix regressions with randomization showed that all distance matrices were significantly correlated with genetic differentiation ($P < 0.05$ in all cases; Supp Table S3 [online only]). However, the scenario of population connectivity based on the spatial configuration of environmentally suitable areas during the LGM under the CCSM4 model was the best fit to our data ($R^2 = 0.554$; Supp Table S3 [online only]) and the only predictor retained into the final model (Table 3; Fig. 5).

Genetic Diversity and Demographic History

Genetic diversity (π) and long-term N_e were significantly intercorrelated (Pearson's correlation: $r = 0.833$, $P < 0.001$). Genetic

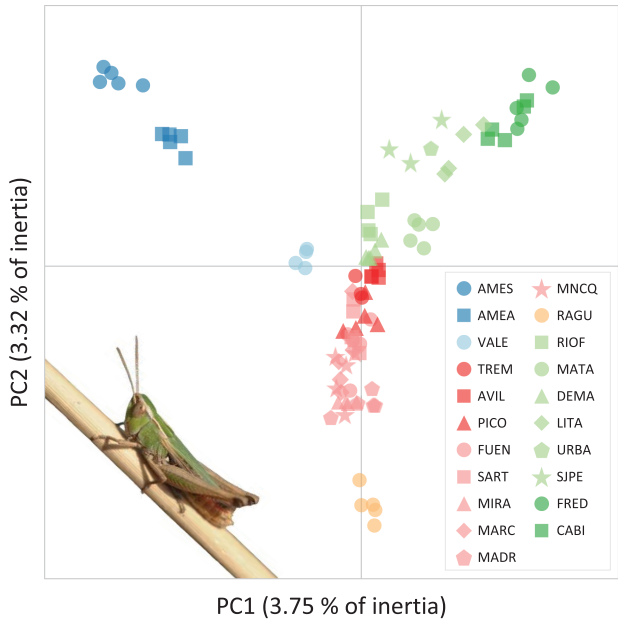

Fig. 4. Principal component analysis (PCAs) of genetic variation for populations of the Pantel's grasshopper (*Omocestus panteli*). Analyses are based on a dataset of 14,454 SNPs. Colors correspond to the genetic cluster at which populations were predominantly assigned according to STRUCTURE analyses for $K = 6$ (Fig. 3). Population codes as described in Table 1. Inset image shows a male of *Omocestus panteli* (photo by Pedro J. Cordero).

diversity was negatively associated with latitude ($t = -3.11$, $P = 0.006$) and positively associated with environmental suitability during the LGM estimated under both the CCSM4 ($t = 2.34$, $P = 0.031$) and the MIROC-ESM ($t = 2.19$, $P = 0.041$) general atmospheric circulation models, which were the only significant variables in univariate analyses ($P > 0.2$ for all other variables). However, only latitude was retained into the final GLM after backward selection (Table 4; Fig. 6A). STAIRWAY PLOT analyses revealed that most populations have experienced parallel demographic responses, undergoing expansions during the last glacial period followed by severe demographic declines starting at the onset of the Holocene (Fig. 7). Only one population (DEMA) deviated from this general pattern, which might be due to the lower number of retained SNPs for this population (Fig. 7; Lapierre et al. 2017). Postglacial declines have reduced effective population sizes (N_e) by >70% (Fig. 7). Long-term N_e was significantly and positively associated with environmental suitability during the LGM under the CCSM4 model, which was the only significant variable in univariate analyses ($P > 0.1$ for all other variables) and the only one retained into the final GLM (Table 4; Fig. 6B).

Discussion

Although the Pantel's grasshopper is broadly distributed over a wide range of altitudes and numerous types of habitats, our analyses revealed significant genetic subdivision dating back to the last glacial age. Accordingly, landscape genetic analyses and demographic reconstructions showed that genetic differentiation between populations and their long-term effective population sizes are best explained by the spatial configuration of environmentally suitable habitats in the LGM. During this time, the species likely had a wider distribution

Table 2. Parameters inferred from coalescent simulations with FASTSIMCOAL2

Parameter	Point estimate	Lower bound	Upper bound
T_{DIV1}	112,135	110,502	114,954
T_{DIV2}	39,268	38,034	40,471
θ_{ANC1}	123,281	110,866	116,321
θ_{ANC2}	1,372,085	856,299	952,438
θ_{NW}	176,500	170,395	178,279
θ_{CS}	220,941	218,161	226,361
m_1	9.30×10^{-6}	9.40×10^{-6}	9.75×10^{-6}
m_2	7.29×10^{-6}	6.71×10^{-6}	7.07×10^{-6}

Table shows point estimates and lower and upper 95% confidence intervals for each parameter, which include timing of population divergence (T_{DIV}), mutation-scaled ancestral and contemporary effective population sizes (θ), and migration rates per generation (m). Estimates of time are given in units of generations. See Fig. 2 for details. Note that effective population size for northwestern populations ($\theta_{NE} = 196{,}429$) was calculated from levels of nucleotide diversity (π) and fixed in FASTSIMCOAL2 analyses to enable the estimation of other parameters (see the Methods section for further details).

but also interspersed unsuitable areas that might have disrupted gene flow for extended periods of time, creating opportunity for geographic diversification. Collectively, these results suggest that the genetic makeup of contemporary populations is not explained by current environmental factors nor geographical barriers to dispersal but primarily reflects genetic fragmentation during the last glacial period followed by postglacial admixture among previously isolated gene pools (see also Inoue et al. 2014, Maier et al. 2019).

Glacial Divergence—Postglacial Admixture

Our comprehensive set of analyses (STRUCTURE, DAPC, and PCA) revealed that genomic variation in the Pantel's grasshopper is organized into three main clusters roughly distributed in northwestern, central-southern, and northeastern Iberia (Figs. 3 and 4). The inference of divergence times based on the coalescence theory indicates that the three genetic clusters most likely split during the last glacial period (Würm glaciation, ca. 12–115 ka), supporting that geographical diversification took place during the coldest stages of the Pleistocene (Knowles 2000, Inoue et al. 2014). It is, however, counterintuitive that genetic fragmentation originated during the last glacial cycle, when the species experienced net population expansions as inferred by palaeodistribution modeling (Fig. 1) and genomic-based demographic reconstructions (Fig. 7). Although the projections of the ENM to LGM bioclimatic conditions indicate an overall higher suitability and a wider distribution of the species in central and southern Iberia than at present time, the areas corresponding to the current location of the other two main genetic clusters might have remained relatively isolated (NW lineage) or embedded within a matrix of environmentally unsuitable habitats (NE lineage) during the last glacial period (Fig. 1C). The long duration of glacial periods (ca. 100,000 yr) and the short generation time of the species (1 yr; Clemente et al. 1991) could have favored the accumulation of genetic differentiation even if local populations experienced net demographic expansions. It should be noted that although genetic differentiation was significant in most pairwise comparisons, the absolute estimates were low (mean $F_{ST} = 0.123$; range = 0.093–0.176; Supp Table S2 [online only]). When compared with other Iberian grasshoppers, these estimates fall within the range of those recovered for pest species with no genetic structure (*Dociostaurus maroccanus* (Thunberg, 1815) (Orthoptera: Acrididae): mean = 0.067; range = 0.051–0.102; González-Serna et al. 2020) and are much lower than the values obtained for

Table 3. Multiple matrix regression with randomization (MMRR) for genetic differentiation (F_{ST}) in relation to: 1) geographical distance; 2) weighted topographical distance; 3) environmental distance; 4) resistance distances defined by a flat landscape (all cells have equal resistance value = 1); and 5) resistance distances defined by environmental suitability as inferred from projections of the species-specific environmental niche model (ENM) to present-day and last glacial maximum bioclimatic conditions under the CCSM4 and MIROC-ESM general atmospheric circulation models

Variable	β	*t*	*P*
Explanatory terms			
Constant	0.112	2.91	0.999
ENM – Last glacial maximum (CCSM4)	0.593	10.52	0.002
Rejected terms			
Geographical distance		1.59	0.479
Flat landscape		0.42	0.881
Environmental distance		2.67	0.123
Weighted topographic distance		1.55	0.466
ENM – Current		0.52	0.863
ENM – Last glacial maximum (MIROC-ESM)		0.52	0.833

β, regression coefficient; *t*, *t*-statistic; *P*, significance level.

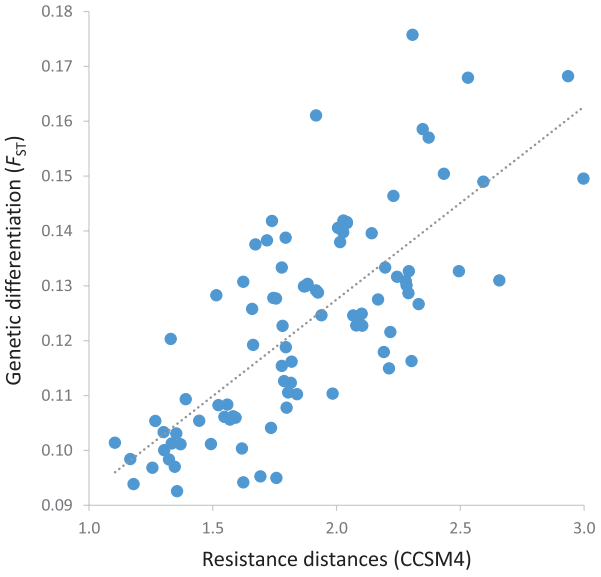

Fig. 5. Relationship between genetic differentiation (F_{ST}) and resistance distances based on environmental suitability during the LGM inferred by projecting the species-specific environmental niche model (ENM) to LGM bioclimatic conditions under the CCSM4 general atmospheric circulation model.

alpine (*Omocestus antigai* (Bolívar, 1897) (Orthoptera: Acrididae); mean = 0.338; range = 0.183–0.457; Tonzo et al. 2019) and highly specialized taxa linked to severely fragmented habitats (*Mioscirtus wagneri* (Eversmann, 1859) (Orthoptera: Acrididae); mean = 0.381; range = 0.075–0.553; Noguerales et al. 2021). Moderately low levels of genetic differentiation between non-admixed populations (F_{ST} < 0.176) suggest that glacial fragmentation was sufficient for geographic diversification even though the large effective population

Table 4. Generalized linear models (GLM) for genetic diversity (nucleotide diversity, π) and long-term effective population size (N_e) in relation to latitude, longitude, elevation, and local environmental suitability as inferred from projections of the species-specific environmental niche model (ENM) to present-day and last glacial maximum bioclimatic conditions under the CCSM4 and MIROC-ESM general atmospheric circulation models

Variable	β	*t*	*P*
(a) Genetic diversity (π)			
Explanatory terms			
Constant	0.173	5.67	<0.001
Latitude	−0.002	−3.11	0.006
Rejected terms			
Longitude		−0.19	0.852
Elevation		0.99	0.336
ENM – Current		−0.88	0.391
ENM – Last glacial maximum (CCSM4)		0.38	0.707
ENM – Last glacial maximum (MIROC-ESM)		0.94	0.358
(b) Long-term effective population size (N_e)			
Explanatory terms			
Constant	22,623	0.80	0.437
ENM – Last glacial maximum (CCSM4)	90,841	2.21	0.047
Rejected terms			
Latitude		−0.31	0.759
Longitude		−0.11	0.913
Elevation		0.35	0.734
ENM – Current		0.13	0.901
ENM – Last glacial maximum (MIROC-ESM)		−0.19	0.852

β, regression coefficient; *t*, *t*-statistic; *P*, significance level.

sizes sustained by the species probably buffered against genetic drift (Charlesworth 2009).

Although STRUCTURE analyses supported considerable genetic admixture in the contact zones, DAPC assigned all individuals and populations to a single genetic group (i.e., probability of membership = 1; Fig. 3). The good correspondence between admixture proportions inferred by the STRUCTURE analyses and the distribution of populations in the multivariate principal component space suggests that the abrupt genetic discontinuities identified by DAPC are not biologically meaningful but rather the outcome of the maximization of differences between groups implemented in this approach that, otherwise, supported a single genetic cluster as the best fit to the data (i.e., lowest BIC for K = 1; Supp Fig. S2 [online only]; see Miller et al. 2020). Admixture among populations that diverged during the last glacial period is thus the most plausible scenario to explain observed patterns of genomic variation (e.g., Inoue et al. 2014, Li et al. 2016, Maier et al. 2019). Despite postglacial genetic admixture is extensive, this process has not been enough to fully erode the genetic structure arisen during the last glacial period. This contrast with pest grasshoppers, which show no or subtle genetic structure across large geographical distances (e.g., Chapuis et al. 2011, Yadav et al. 2019, González-Serna et al. 2020) despite genomic and fossil evidence indicate that some of them have experienced strong demographic declines, likely accompanied by range fragmentation, during the coldest stages of the Pleistocene (Meco et al. 2010, 2011; González-Serna et al. 2020). Although the Pantel's grasshopper is fully winged, its small body size (females <23 mm, males <17 mm; Clemente et al. 1991) has probably reduced the capability of the species to disperse

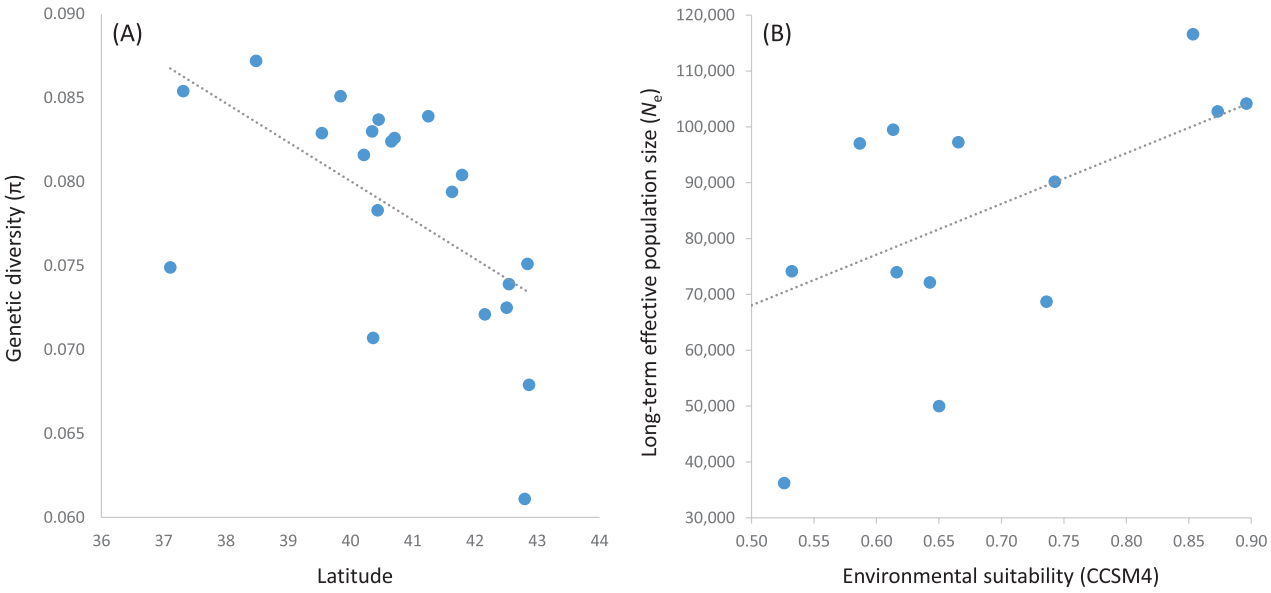

Fig. 6. Relationships (A) between genetic diversity (π) and latitude and (B) between long-term effective population size (N_e) and environmental suitability during the last glacial maximum (LGM) inferred by projecting the species-specific environmental niche model (ENM) to LGM bioclimatic conditions under the CCSM4 general atmospheric circulation model.

across large geographical distances and, thus, limited genetic homogenization (Ortego et al. 2015b).

Demographic History

Demographic reconstructions supported that populations throughout the Pantel grasshopper's range have experienced parallel changes in effective population size over time, suggesting that they have responded similarly to Pleistocene climatic oscillations (Fig. 7). Effective population sizes (N_e) peaked around the LGM and declined at the onset of the Holocene, which is congruent with the generalized reduction of suitable habitats from LGM to present inferred by ENM (Fig. 1). In line with results from landscape genetic analyses, long-term effective population sizes were positively correlated with environmental suitability during the LGM, highlighting the predominant effect of glacial dynamics on the demographic trajectory of the species. Remarkably, only projections under one of the two general atmospheric circulation models (CCSM4) explained contemporary patterns of genetic variation in the species (i.e., long-term N_e and genetic differentiation; Figs. 5 and 6), which is not surprising given the different predictions that they produced (Fig. 1). Differences in predicted distributions between CCSM4 and MIROC-ESM models are expected given their specific assumptions (i.e., CCSM4 assumes a stronger temperature decline than MIROC-ESM; Otto-Bliesner et al. 2007, Alba-Sánchez et al. 2010) and have been reported in numerous previous studies (e.g., Fernández-Mazuecos and Vargas 2013, Zinetti et al. 2013; Ramírez-Barahona and Eguiarte 2014, Wachter et al. 2016). This emphasizes the potential of genomic data to validate distributional shifts inferred from ENMs when independent sources of information are not available (e.g., pollen records or fossils; Nogués-Bravo 2009, Fordham et al. 2014, Metcalf et al. 2014, Bemmels et al. 2019).

Landscape Genetic Analyses

In line with estimations of divergence time among main genetic clusters and correlates of long-term population sizes, our landscape genetic analyses support a predominant role of glacial connectivity/isolation on shaping contemporary patterns of genomic variation in the Pantel's grasshopper. The resistance distances defined by the spatial configuration of environmentally suitable habitats during the LGM provided the best fit to genetic data, with no significant additional effects of topography, current environmental suitability or environmental dissimilarity. The role of current *versus* historical landscape composition on contemporary patterns of genetic structure is exemplified in populations MNCQ (southwestern Iberia) and MARC (central Iberia). These populations belong to the same genetic cluster (Fig. 3) and present one of the lowest values of pairwise genetic differentiation (nonsignificant pairwise F_{ST} = 0.095; Supp Table S2 [online only]) despite being located 440 km apart, a distance comparable to that separating MARC and other populations belonging to different genetic clusters (e.g., AMES = 513 km; CABI = 315 km; Fig. 1). MARC and MNCQ are currently separated by extensive areas with low environmental suitability (Guadiana and Guadalquivir depressions), but a continuous habitat was predicted to connect them during the LGM (Fig. 1), supporting the role of past landscape composition in explaining their contemporary genetic affinities. The much longer duration of glacial periods (70,000–100,000 yr) in comparison with interglacials (10,000–30,000 yr), together with the limited dispersal capacity and large effective population sizes of this abundant species (Clemente et al. 1991, Presa et al. 2016), might have prevented that postglacial gene flow has completely eroded the genetic structure generated during the last glacial period (Table 2) and resulted in the genetic signals of past population subdivision still dominate the genetic background of contemporary populations (e.g., Lanier et al. 2015, Maier et al. 2019, Maigret et al. 2020).

Although topographic complexity has been identified as an important driver of genetic differentiation in several organisms, including grasshoppers, weighted topographic distances did not explain gene flow among populations in the Pantel's grasshopper (Table 3). The wide elevational range occupied by the species could contribute to continuous gene flow among populations located at different altitudes and facilitate genetic connectivity through

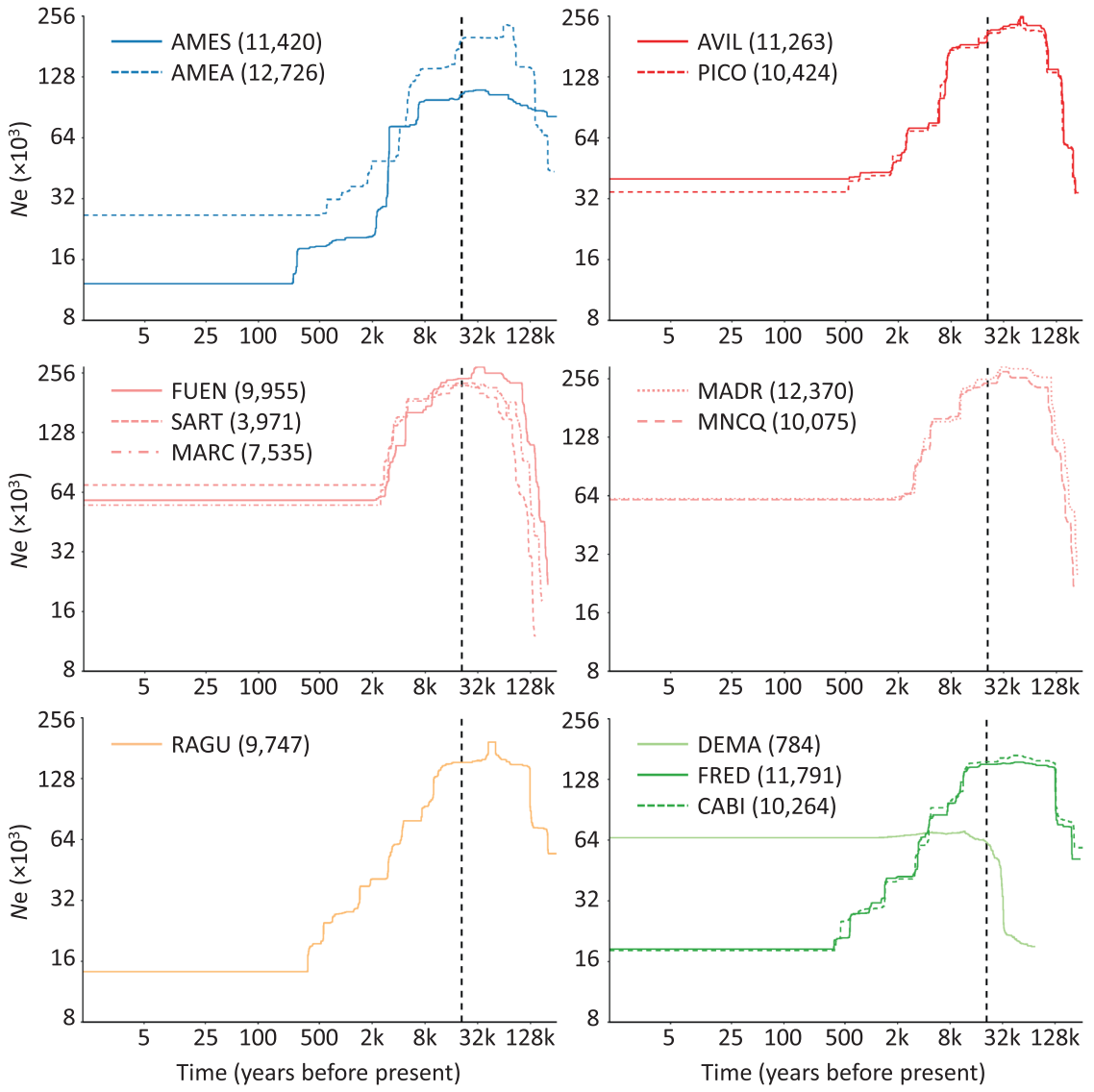

Fig. 7. Demographic history of the studied populations of Pantel's grasshopper (*Omocestus panteli*) inferred using STAIRWAY PLOT. Panels show the median of effective population size (N_e) over time, estimated assuming a mutation rate of 2.8×10^{-9} and a 1-yr generation time (both axes in a logarithmic scale). Vertical dashed lines indicate the Last Glacial Maximum (LGM; ca. 21 ka). Number of polymorphic SNPs used to calculate the site frequency spectrum (SFS) indicated in parentheses. Colors correspond to the main genetic cluster at which populations were predominantly assigned according to STRUCTURE analyses for $K = 6$ (Fig. 3). Population codes as described in Table 1.

topographically complex landscapes. This contrasts with inferences for other grasshopper species exclusively inhabiting lowland or alpine habitats and for which mountain ranges and valley bottoms, respectively, act as impassable barriers to dispersal (Noguerales et al. 2016, González-Serna et al. 2019, Tonzo et al. 2019). Finally, we did not find any evidence for isolation-by-environment, which could be simply due to climate is not an important selective agent in our focal species or, alternatively, that gene flow limits the maintenance of local adaptations in response to spatially-varying selection (Pujolar et al. 2014, Laporte et al. 2016).

Conclusions

Our analyses support that patterns of genomic variation in the Pantel's grasshopper have been predominantly shaped by cycles of population isolation and connectivity driven by Pleistocene glacial cycles. The estimated divergence of the three main genetic clusters around the last glacial period and contemporary admixture in contact zones suggests that the Pleistocene 'species pump' model might be also useful to explain demographic dynamics and geographical diversification in at least some species with broad climatic niches and distributions (Haffer 1969). Our study illustrates the importance of integrating demographic reconstructions, palaeodistribution modeling, and testing of alternative scenarios of genetic connectivity to understand the proximate processes shaping contemporary patterns of genomic variation in species distributed across wide elevational and environmental gradients and whose responses to Pleistocene climatic oscillations are much less predictable than in organisms with narrower environmental requirements.

Supplementary Data

Supplementary data are available at *Insect Systematics and Diversity* online.

Acknowledgments

We are grateful to Conchi Cáliz for her help during field and laboratory work, Sergio Pereira (The Centre for Applied Genomics) for Illumina sequencing, and two anonymous referees for their constructive and valuable comments on an earlier version of the manuscript. We also thank to Centro de Supercomputación de Galicia (CESGA) and Doñana's Singular Scientific-Technical Infrastructure (ICTS-RBD) for access to computer resources. During this work, V.N. was supported by a 'Juan de la Cierva-Formación' postdoctoral fellowship (grant FJC2018-035611-I) funded by the Spanish Ministry of Science and Innovation and the European Regional Development Fund, V.T. was supported by a FPI predoctoral fellowship from the Spanish Ministry of Economy and Competitiveness (grant BES-2015–73159), and M.G.S was supported by a predoctoral fellowship from Junta de Comunidades de Castilla-La Mancha and the European Social Fund. This work was funded by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (grants CGL2011-25053, CGL2016-80742-R, and CGL2017-83433-P).

Author Contributions

J.O. conceived the study. J.O., V.N., and P.J.C. collected the samples. M.G.S. prepared the genomic libraries guided by J.O. J.O., V.N., and V.T. analyzed the data. J.O. wrote the manuscript, with inputs from all the authors.

Data Availability

Raw Illumina reads have been deposited at the NCBI Sequence Read Archive (SRA) under BioProject PRJNA702632. Input files for all analyses are available for download on Figshare (https://doi.org/10.6084/m9.figshare.14071280).

References Cited

Alba-Sánchez, F., J. A. López-Sáez, B. Benito de Pando, and L. López-Merino. 2010. Past and present potential distribution of the Iberian *Abies* species: a phytogeographic approach using pollen data and species distribution models. Divers. Distrib. 16: 214–228.

Bemmels, J. B., L. L. Knowles, and C. W. Dick. 2019. Genomic evidence of survival near ice sheet margins for some, but not all, North American trees. Proc. Natl. Acad. Sci. U. S. A. 116: 8431–8436.

Bohonak, A. J. 1999. Dispersal, gene flow, and population structure. Q. Rev. Biol. 74: 21–45.

Braconnot, P., B. Otto-Bliesner, S. Harrison, S. Joussaume, J. Y. Peterchmitt, A. Abe-Ouchi, M. Crucifix, E. Driesschaert, T. Fichefet, C. D. Hewitt, et al. 2007. Results of PMIP2 coupled simulations of the mid-holocene and last glacial Maximum - Part 1: experiments and large-scale features. Clim. Past 3: 261–277.

Branstetter, M. G., and J. T. Longino. 2019. Ultra-conserved element phylogenomics of New World *Ponera* (Hymenoptera: Formicidae) illuminates the origin and phylogeographic history of the endemic exotic ant *Ponera exotica*. Insect Syst. Divers. 3: 1.

Carnaval, A. C., M. J. Hickerson, C. F. Haddad, M. T. Rodrigues, and C. Moritz. 2009. Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. Science. 323: 785–789.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. STACKS: building and genotyping Loci *de novo* from short-read sequences. G3 (Bethesda). 1: 171–182.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. STACKS: an analysis tool set for population genomics. Mol. Ecol. 22: 3124–3140.

Caterino, M. S., and S. S. Langton-Myers. 2019. Intraspecific diversity and phylogeography in southern Appalachian *Dasycerus carolinensis* (Coleoptera: Staphylinidae: Dasycerinae). Insect Syst. Divers. 3: 8.

Chapuis, M. P., J. A. Popple, K. Berthier, S. J. Simpson, E. Deveson, P. Spurgin, M. J. Steinbauer, and G. A. Sword. 2011. Challenges to assessing connectivity between massive populations of the Australian plague locust. Proc. Biol. Sci. 278: 3152–3160.

Charlesworth, B. 2009. Effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. 10: 195–205.

Cigliano, M. M., H. Braun, D. C. Eades, and D. Otte. 2019. Orthoptera species file. Version 5.0/5.0. Available from http://orthoptera.speciesfile.org/

Clemente, M. E., M. D. García, and J. J. Presa. 1991. Los Gomphocerinae de la península Ibérica: 2. *Omocestus* Bolivar, 1878. (Insecta, Orthoptera, Caelifera). Graellsia 46: 191–246.

DeChaine, E. G., and A. P. Martin. 2006. Using coalescent simulations to test the impact of quaternary climate cycles on divergence in an alpine plant-insect association. Evolution. 60: 1004–1013.

Earl, D. A., and B. M. vonHoldt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4: 359–361.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129–151.

Elith, J., S. J. Phillips, T. Hastie, M. Dudik, Y. E. Chee, and C. J. Yates. 2011. A statistical explanation of MAXENT for ecologists. Divers. Distrib. 17: 43–57.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14: 2611–2620.

Excoffier, L., and M. Foll. 2011. FASTSIMCOAL: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics. 27: 1332–1334.

Excoffier, L., and H. E. Lischer. 2010. ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10: 564–567.

Excoffier, L., I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, and M. Foll. 2013. Robust demographic inference from genomic and SNP data. PLoS Genet. 9: e1003905.

Fernández-Mazuecos, M., and P. Vargas. 2013. Congruence between distribution modelling and phylogeographical analyses reveals Quaternary survival of a toadflax species (*Linaria elegans*) in oceanic climate areas of a mountain ring range. New Phytol. 198: 1274–1289.

Fordham, D. A., B. W. Brook, C. Moritz, and D. Nogués-Bravo. 2014. Better forecasts of range dynamics using genetic data. Trends Ecol. Evol. 29: 436–443.

Fourcade, Y., J. O. Engler, D. Rödder, and J. Secondi. 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: e97122.

Fumagalli, M. 2013. Assessing the effect of sequencing depth and sample size in population genetics inferences. PLoS One 8: e79667.

Glover, M. M., S. P. Egan, G. R. Hood, J. Rull, M. Aluja, and J. L. Feder. 2018. Phylogeography of walnut-infesting *Rhagoletis suavis* (Diptera: Tephritidae) flies. Insect Syst. Divers. 2: 1.

González-Serna, M. J., P. J. Cordero, and J. Ortego. 2018. Using high-throughput sequencing to investigate the factors structuring genomic variation of a Mediterranean grasshopper of great conservation concern. Sci. Rep. 8: 13436.

González-Serna, M. J., P. J. Cordero, and J. Ortego. 2019. 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. Mol. Ecol. 28: 2155–2172.

González-Serna, M. J., P. J. Cordero, and J. Ortego. 2020. Insights into the neutral and adaptive processes shaping the spatial distribution of genomic variation in the economically important Moroccan locust (*Dociostaurus maroccanus*). Ecol. Evol. 10: 3991–4008.

Haffer, J. 1969. Speciation in amazonian forest birds. Science. 165: 131–137.

Hasumi, H., and S. Emori. 2004. K-1 Coupled GCM (MIROC) description. Center for Climate System Research (CCSR), University of Tokyo; National Institute for Environmental Studies (NIES); Frontier Research Center for Global Change (FRCGC), Tokyo, Japan.

He, Q., D. L. Edwards, and L. L. Knowles. 2013. Integrative testing of how environments from the past to the present shape genetic structure across landscapes. Evolution. 67: 3386–3402.

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

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25: 1965–1978.

Ikeda, H., M. Nishikawa, and T. Sota. 2012. Loss of flight promotes beetle diversification. Nat. Commun. 3: 648.

Inoue, K., E. M. Monroe, C. L. Elderkin, and D. J. Berg. 2014. Phylogeographic and population genetic analyses reveal Pleistocene isolation followed by high gene flow in a wide ranging, but endangered, freshwater mussel. Heredity (Edinb). 112: 282–290.

Inoue, K., B. K. Lang, and D. J. Berg. 2015. Past climate change drives current genetic structure of an endangered freshwater mussel species. Mol. Ecol. 24: 1910–1926.

Jacobsen, M. W., J. M. Pujolar, M. T. Gilbert, J. V. Moreno-Mayar, L. Bernatchez, T. D. Als, J. Lobon-Cervia, and M. M. Hansen. 2014. Speciation and demographic history of Atlantic eels (*Anguilla anguilla* and *A. rostrata*) revealed by mitogenome sequencing. Heredity (Edinb). 113: 432–442.

Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 23: 1801–1806.

Jombart, T. 2008. ADEGENET: a R package for the multivariate analysis of genetic markers. Bioinformatics. 24: 1403–1405.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 11: 94.

Keightley, P. D., R. W. Ness, D. L. Halligan, and P. R. Haddrill. 2014. Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. Genetics. 196: 313–320.

Knowles, L. L. 2000. Tests of pleistocene speciation in montane grasshoppers (*genus Melanoplus*) from the sky islands of western North America. Evolution. 54: 1337–1348.

Knowles, L. L., and B. C. Carstens. 2007. Estimating a geographically explicit model of population divergence. Evolution. 61: 477–493.

Knowles, L. L., B. C. Carstens, and M. L. Keat. 2007. Coupling genetic and ecological-niche models to examine how past population distributions contribute to divergence. Curr. Biol. 17: 940–946.

Lanier, H. C., R. Massatti, Q. He, L. E. Olson, and L. L. Knowles. 2015. Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). Mol. Ecol. 24: 3688–3705.

Lapierre, M., A. Lambert, and G. Achaz. 2017. Accuracy of demographic inferences from the site frequency spectrum: the case of the yoruba population. Genetics. 206: 439–449.

Laporte, M., S. A. Pavey, C. Rougeux, F. Pierron, M. Lauzent, H. Budzinski, P. Labadie, E. Geneste, P. Couture, M. Baudrimont, et al. 2016. RAD sequencing reveals within-generation polygenic selection in response to anthropogenic organic and metal contamination in North Atlantic Eels. Mol. Ecol. 25: 219–237.

Lenormand, T. 2002. Gene flow and the limits to natural selection. Trends Ecol. Evol. 17: 183–189.

Li, J., Y. Yang, Q. Chen, L. Fang, Z. He, W. Guo, S. Qiao, Z. Wang, M. Guo, C. Zhong, et al. 2016. Pronounced genetic differentiation and recent secondary contact in the mangrove tree *Lumnitzera racemosa* revealed by population genomic analyses. Sci. Rep. 6: 29486.

Li, H. R., W. M. Qu, J. J. Obrycki, L. Meng, X. G. Zhou, D. Chu, and B. P. Li. 2020. Optimizing sample size for population genomic study in a global invasive lady beetle, *Harmonia axyridis*. Insects 11: 290.

Liu, X. M., and Y. X. Fu. 2015. Exploring population size changes using SNP frequency spectra. Nat. Genet. 47: 555–559.

Maier, P. A., A. G. Vandergast, S. M. Ostoja, A. Aguilar, and A. J. Bohonak. 2019. Pleistocene glacial cycles drove lineage diversification and fusion in the Yosemite toad (*Anaxyrus canorus*). Evolution. 73: 2476–2496.

Maigret, T. A., J. J. Cox, and D. W. Weisrock. 2020. A spatial genomic approach identifies time lags and historical barriers to gene flow in a rapidly fragmenting Appalachian landscape. Mol. Ecol. 29: 673–685.

Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends Ecol. Evol. 18: 189–197.

Massatti, R., and L. L. Knowles. 2016. Contrasting support for alternative models of genomic variation based on microhabitat preference: species-specific effects of climate change in alpine sedges. Mol. Ecol. 25: 3974–3986.

McRae, B. H. 2006. Isolation by resistance. Evolution. 60: 1551–1561.

McRae, B. H., and P. Beier. 2007. Circuit theory predicts gene flow in plant and animal populations. Proc. Natl. Acad. Sci. U. S. A. 104: 19885–19890.

Meco, J., N. Petit-Maire, J. Ballester, J. F. Betancort, and A. J. G. Ramos. 2010. The Acridian plagues, a new Holocene and Pleistocene palaeoclimatic indicator. Glob. Planet. Change 72: 318–320.

Meco, J., D. R. Muhs, M. Fontugne, A. J. G. Ramos, A. Lomoschitz, and D. Patterson. 2011. Late Pliocene and Quaternary Eurasian locust infestations in the Canary Archipelago. Lethaia 44: 440–454.

Metcalf, J. L., S. Prost, D. Nogués-Bravo, E. G. DeChaine, C. Anderson, P. Batra, M. B. Araújo, A. Cooper, and R. P. Guralnick. 2014. Integrating multiple lines of evidence into historical biogeography hypothesis testing: a *Bison bison* case study. Proc. Biol. Sci. 281: 20132782.

Miller, J. M., C. I. Cullingham, and R. M. Peery. 2020. The influence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. Heredity (Edinb). 125: 269–280.

Muscarella, R., P. J. Galante, M. Soley-Guardia, R. A. Boria, J. M. Kass, M. Uriarte, and R. P. Anderson. 2014. ENMEVAL: an R package for conducting spatially independent evaluations and estimating optimal model complexity for MAXENT ecological niche models. Methods Ecol. Evol. 5: 1198–1205.

Nazareno, A. G., J. B. Bemmels, C. W. Dick, and L. G. Lohmann. 2017. Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. Mol. Ecol. Resour. 17: 1136–1147.

Noguerales, V., P. J. Cordero, and J. Ortego. 2016. Hierarchical genetic structure shaped by topography in a narrow-endemic montane grasshopper. BMC Evol. Biol. 16: 96.

Noguerales, V., P. J. Cordero, L. L. Knowles, and J. Ortego. 2021. Genomic insights into the origin of trans-Mediterranean disjunct distributions. J. Biogeogr. 48: 440–452.

Nogués-Bravo, D. 2009. Predicting the past distribution of species climatic niches. Global Ecol. Biogeogr. 18: 521–531.

Ortego, J., V. García-Navas, V. Noguerales, and P. J. Cordero. 2015a. Discordant patterns of genetic and phenotypic differentiation in five grasshopper species codistributed across a microreserve network. Mol. Ecol. 24: 5796–5812.

Ortego, J., P. F. Gugger, and V. L. Sork. 2015b. Climatically stable landscapes predict patterns of genetic structure and admixture in the Californian canyon live oak. J. Biogeogr. 42: 328–338.

Otto-Bliesner, B. L., C. D. Hewitt, T. M. Marchitto, E. Brady, A. Abe-Ouchi, M. Crucifix, S. Murakami, and S. L. Weber. 2007. Last Glacial Maximum ocean thermohaline circulation: PMIP2 model intercomparisons and data constraints. Geophys. Res. Lett. 34: L12706.

Papadopoulou, A., and L. L. Knowles. 2015. Genomic tests of the species-pump hypothesis: recent island connectivity cycles drive population divergence but not speciation in Caribbean crickets across the Virgin Islands. Evolution. 69: 1501–1517.

Papadopoulou, A., and L. L. Knowles. 2016. Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. Proc. Natl. Acad. Sci. U. S. A. 113: 8018–8024.

Paz, A., R. Ibáñez, K. R. Lips, and A. J. Crawford. 2015. Testing the role of ecology and life history in structuring genetic variation across a landscape: a trait-based phylogeographic approach. Mol. Ecol. 24: 3723–3737.

Peart, C. R., S. Tusso, S. D. Pophaly, F. Botero-Castro, C. C. Wu, D. Aurioles-Gamboa, A. B. Baird, J. W. Bickham, J. Forcada, F. Galimberti, et al. 2020. Determinants of genetic variation across eco-evolutionary scales in pinnipeds. Nat. Ecol. Evol. 4: 1095–1104.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS One 7: e37135.

Phillips, S. J., and M. Dudik. 2008. Modeling of species distributions with MAXENT: new extensions and a comprehensive evaluation. Ecography 31: 161–175.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190: 231–259.

Presa, J. J., M. García, M. Clemente, P. Barranco Vega, J. Correas, S. Ferreira, A. Hochkirch, P. Lemos, B. Odé, and F. Prunier. 2016. *Omocestus panteli*. The IUCN Red List of Threatened Species 2016: e.T16084629A75089035. Available from https://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T16084629A75089035.en. Downloaded on 26 November 2020.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics. 155: 945–959.

Pujolar, J. M., M. W. Jacobsen, T. D. Als, J. Frydenberg, K. Munch, B. Jónsson, J. B. Jian, L. Cheng, G. E. Maes, L. Bernatchez, et al. 2014. Genome-wide single-generation signatures of local selection in the panmictic European eel. Mol. Ecol. 23: 2514–2528.

R Core Team. 2020. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/

Ramírez-Barahona, S., and L. E. Eguiarte. 2014. Changes in the distribution of cloud forests during the last glacial predict the patterns of genetic diversity and demographic history of the tree fern *Alsophila firma* (Cyatheaceae). J. Biogeogr. 41: 2396–2407.

Rosenberg, N. A. 2004. DISTRUCT: a program for the graphical display of population structure. Mol. Ecol. Notes 4: 137–138.

Schoville, S. D., G. K. Roderick, and D. H. Kavanaugh. 2012. Testing the 'Pleistocene species pump' in alpine habitats: lineage diversification of flightless ground beetles (Coleoptera: Carabidae: *Nebria*) in relation to altitudinal zonation. Biol. J. Linn. Soc. 107: 95–111.

Sexton, J. P., S. B. Hangartner, and A. A. Hoffmann. 2014. Genetic isolation by environment or distance: which pattern of gene flow is most common? Evolution. 68: 1–15.

Shafer, A. B. A., and J. B. W. Wolf. 2013. Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. Ecol. Lett. 16: 940–950.

Slatkin, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution. 47: 264–279.

Tonzo, V., A. Papadopoulou, and J. Ortego. 2019. Genomic data reveal deep genetic structure but no support for current taxonomic designation in a grasshopper species complex. Mol. Ecol. 28: 3869–3886.

Tonzo, V., A. Papadopoulou, and J. Ortego. 2020. Genomic footprints of an old affair: single nucleotide polymorphism data reveal historical hybridization and the subsequent evolution of reproductive barriers in two recently diverged grasshoppers with partly overlapping distributions. Mol. Ecol. 29: 2254–2268.

Tsuda, Y., K. Nakao, Y. Ide, and Y. Tsumura. 2015. The population demography of *Betula maximowicziana*, a cool-temperate tree species in Japan, in relation to the last glacial period: its admixture-like genetic structure is the result of simple population splitting not admixing. Mol. Ecol. 24: 1403–1418.

Wachter, G. A., A. Papadopoulou, C. Muster, W. Arthofer, L. L. Knowles, F. M. Steiner, and B. C. Schlick-Steiner. 2016. Glacial refugia, recolonization patterns and diversification forces in Alpine-endemic *Megabunus* harvestmen. Mol. Ecol. 25: 2904–2919.

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

Wang, I. J. 2020. Topographic path analysis for modelling dispersal and functional connectivity: calculating topographic distances using the topodistance r package. Methods Ecol. Evol. 11: 265–272.

Wang, I. J., and G. S. Bradburd. 2014. Isolation by environment. Mol. Ecol. 23: 5649–5662.

Wang, I. J., W. K. Savage, and H. B. Shaffer. 2009. Landscape genetics and least-cost path analysis reveal unexpected dispersal routes in the California tiger salamander (*Ambystoma californiense*). Mol. Ecol. 18: 1365–1374.

Warren, D. L., and S. N. Seifert. 2011. Ecological niche modeling in MAXENT: the importance of model complexity and the performance of model selection criteria. Ecol. Appl. 21: 335–342.

Yadav, S., A. J. Stow, and R. Y. Dudaniec. 2019. Detection of environmental and morphological adaptation despite high landscape genetic connectivity in a pest grasshopper (*Phaulacridium vittatum*). Mol. Ecol. 28: 3395–3412.

Zellmer, A. J., and L. L. Knowles. 2009. Disentangling the effects of historic vs. contemporary landscape structure on population genetic divergence. Mol. Ecol. 18: 3593–3602.

Zeng, L., C. Ming, Y. Li, L. Y. Su, Y. H. Su, N. O. Otecko, A. Dalecky, S. Donnellan, K. Aplin, X. H. Liu, et al. 2018. Out of southern East Asia of the brown rat revealed by large-scale genome sequencing. Mol. Biol. Evol. 35: 149–158.

Zinetti, F., L. Dapporto, A. Vovlas, G. Chelazzi, S. Bonelli, E. Balletto, and C. Ciofi. 2013. When the rule becomes the exception. no evidence of gene flow between two *Zerynthia* cryptic butterflies suggests the emergence of a new model group. PLoS One 8: e65746.