SCIENTIFIC REPORTS

OPEN

Using high-throughput sequencing to investigate the factors structuring genomic variation of a Mediterranean grasshopper of great conservation concern

María José González-Serna [1], **Pedro J. Cordero**[1] **& Joaquín Ortego**[2]

Inferring the demographic history of species is fundamental for understanding their responses to past climate/landscape alterations and improving our predictions about the future impacts of the different components of ongoing global change. Estimating the time-frame at which population fragmentation took place is also critical to determine whether such process was shaped by ancient events (e.g. past climate/geological changes) or if, conversely, it was driven by recent human activities (e.g. habitat loss). We employed genomic data (ddRAD-Seq) to determine the factors shaping contemporary patterns of genetic variation in the Iberian cross-backed grasshopper *Dociostaurus crassiusculus*, an endangered species with limited dispersal capacity and narrow habitat requirements. Our analyses indicate the presence of two ancient lineages and three genetic clusters resulted from historical processes of population fragmentation (~18–126 ka) that predate the Anthropocene. Landscape genetic analyses indicate that the limits of major river basins are the main geographical feature explaining large-scale patterns of genomic differentiation, with no apparent effect of human-driven habitat fragmentation. Overall, our study highlights the importance of detailed phylogeographic, demographic and spatially-explicit landscape analyses to identify evolutionary significant units and determine the relative impact of historical *vs.* anthropogenic factors on processes of genetic fragmentation in taxa of great conservation concern.

Inferring the evolutionary and demographic history of species and populations is fundamental for understanding how they were impacted by past environmental and landscape alterations and anticipating their responses to different components of global change such as climatic variations[1–3], habitat loss[4] or the emergence of infectious diseases[5]. Many organisms show nowadays highly fragmented distributions due to a natural patchy distribution of their particular habitats[6,7] or as consequence of their originally continuous populations became isolated due to habitat fragmentation driven by human activities[8,9] or past climatic/geological events[10,11]. The genetic, ecological and evolutionary consequences of severe population fragmentation are numerous, including alteration of selective pressures, genetic erosion, inbreeding, accumulation of deleterious mutations, reduced evolutionary potential and, ultimately, increased risk of extinction[12,13]. For these reasons, the study of population fragmentation is a major area of research for conservation biologists and geneticists, and particular attention has been paid to taxa forming small populations and presenting narrow distributions, low dispersal capabilities, and specific habitat requirements[14,15].

Estimating the time-frame at which population fragmentation took place is critical to determine whether such process was driven by historical processes that predate the Anthropocene or if, conversely, it is a direct consequence of human activities[16,17]. This has important implications to inform on ground conservation practices[18]. If recent human-induced habitat fragmentation is identified as the main driver of population genetic structure

[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. Correspondence and requests for materials should be addressed to M.J.G.-S. (email: mariajose.gonzalez@uclm.es)

and disruption of gene flow, then conservation practices should focus on restoring population connectivity either creating corridors to dispersal or assisting gene flow to avoid the long-term negative consequences of inbreeding and loss of genetic diversity[19,20]. If, instead, population genetic structure was driven by ancient processes, then the different clades, lineages or genetic clusters might represent evolutionary significant units (ESUs) with particular local adaptations that deserve to be managed independently to maximize the protection of both vicariant and adaptive diversity[18,21,22]. Beyond the temporal scale of population divergence, identifying the proximate factors shaping contemporary patterns of genetic structure is also fundamental to understand how organisms interact with the different components of the landscape[23]. Genetic and spatial information has been successfully integrated to infer dispersal routes across different habitat types[20], identify natural barriers to dispersal (e.g. rivers[24], topography[25], geology[26]) and determine the consequences of human activities on disrupting gene flow of natural populations (e.g. agriculture[27,28], infrastructures[29]). For this reason, testing alternative spatially-explicit scenarios of population connectivity under a landscape genetic framework can help to determine the relative role of human and natural barriers to gene flow on structuring present-day patterns of genetic variation[30]. This takes particular relevance in the case of specialist taxa with patchy distributions, as identifying contemporary barriers to gene flow and cryptic corridors to dispersal is crucial to establish management practices aimed to restore or enhance connectivity among remnant populations[20].

The Iberian Peninsula constitutes an important biodiversity hotspot, with high species richness, rates of endemism and levels of intra-specific genetic diversity[31–33]. Explanations for the high diversity of the Iberian Peninsula include its historically high climatic stability[34], the low impact of Pleistocene glaciations in comparison with northern temperate areas[35], its current proximity and Miocene connection with North Africa and other Mediterranean regions[10,36], and the presence of deep environmental gradients and a complex topography[34,37]. Despite its high biodiversity and conservation value, the Mediterranean region has been exposed to millennia of strong human intervention[27,34] that have reduced the original extent of its primary vegetation by ~95%[33]. This region is also predicted to be impacted by intense effects of climate change and experience distributional shifts and remarkable range contractions in many taxa[38,39]. Both severe habitat loss and climate warming represent serious threats for many taxa with small and highly fragmented distributions that deal with important difficulties for maintaining viable populations and face risk of extinction[40,41]. Thus, understanding the evolutionary history, demographic trends, and interactions with landscape heterogeneity of these taxa is critical for establishing effective conservation policies and informed management practices that ensure their long-term persistence[42,43].

In this study, we use genomic data to infer the processes structuring genetic variation in the Iberian cross-backed grasshopper *Dociostaurus crassiusculus* (Pantel, 1886), a species of great conservation concern that has been recently catalogued as "endangered" in the Red List of European Orthoptera[44]. The taxonomic position of this species was controversial and according to morphological criteria it has been considered for a long time a subspecies of the Asian *Dociostaurus kraussi* (Ingenitskii, 1897)[45–47]. A recent re-evaluation of the taxonomic status of the genus using genetic data has supported the presence of two well recognized species in concordance with their disjunct distributions: *D. crassiusculus* in the Iberian Peninsula and *D. kraussi* in Asia[48]. The full species status of *D. crassiusculus* makes it of higher conservation concern provided that the very few known populations of the species persist in highly isolated pockets of habitat embedded in an expansive matrix of unsuitable areas[3]. The species is currently distributed in central-southeast Iberia, mostly occupying pseudo-steppe habitats with halophytic plant communities linked to gypsum or hypersaline soils[3]. These narrow habitat requirements, together with the reduced flying capacity of the species and the progressive loss of its natural habitat by human activities, has led that all populations of *D. crassiusculus* are nowadays extremely fragmented and at high risk of extinction by stochastic phenomena[3,49].

Here, we employ restriction-site-associated DNA sequencing (ddRAD-seq), coalescent-based simulations and a landscape genetics framework to evaluate alternative demographic scenarios, estimate the timing of population fragmentation, and infer the processes shaping contemporary patterns of genetic structure across all known populations of *D. crassiusculus*. Specifically, we first used genomic data to analyse the spatial genetic structure of extant populations of the species, identify main lineages and establish their phylogenomic relationships, and define hierarchical units for conservation and management[18,21]. Second, we tested alternative coalescent-based demographic and migration models to infer spatial patterns and rates of inter-population gene flow, estimate the timing of population fragmentation at different spatial scales and, ultimately, determine whether the genetic structure of the species is a consequence of ancient events (e.g. past climate or geological changes) or if, conversely, it is compatible with human-driven population fragmentation[16]. Finally, we generated alternative isolation-by-resistance (IBR) scenarios of population connectivity within a spatially-explicit framework to identify the specific landscape features shaping genetic differentiation in the species and unravel the relative importance of natural (topography, lithology, limits of main river basins) *vs*. anthropogenic (habitat loss) processes of genetic fragmentation.

Results

Genomic data and genetic statistics. A total of 91,666,732 reads were obtained for the 35 genotyped individuals of *D. crassiusculus*. The number of reads per individual (mean $\pm$ SD $= 2,619,049 \pm 841,054$ reads) before and after different quality filtering steps is shown in Supplementary Fig. S1. The datasets obtained with STACKS for $p = 2$ and $p = 4$ contained 80,534 and 65,459 unlinked SNPs, respectively. The datasets obtained with PYRAD for $W_{clust} = 90\%$ and $MinCov = 11$ and 23 contained 32,424 and 18,442 unlinked SNPs, respectively; and for $W_{clust} = 95\%$ and $MinCov = 11$ and 23 contained 42,053 and 23,628 unlinked SNPs, respectively. Population genetic statistics (P, π, H_O, H_E and F_{IS}) calculated with STACKS for all positions (polymorphic and non-polymorphic) and considering loci that were represented in at least two ($p = 2$) and four populations ($p = 4$) and the 50% of individuals within populations ($r = 0.5$) are presented in Supplementary Table S1. Pair-wise F_{ST} values ranged from 0.063 to 0.237 and all were significantly different from zero based on 100 permutations ($P < 0.05$; Supplementary Table S2).
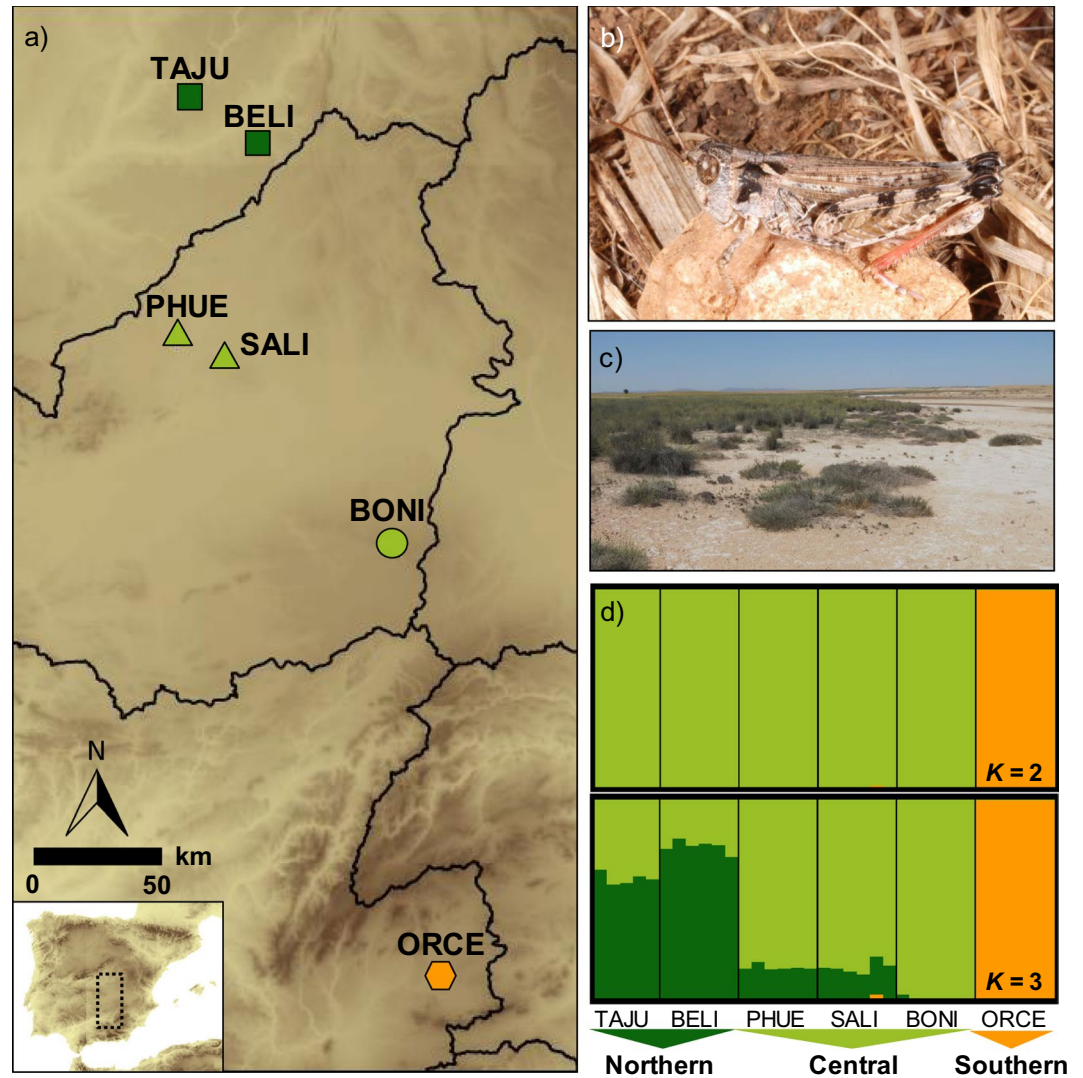
Figure 1. (**a**) Geographical location and genetic structure of the studied populations of the Iberian cross-backed grasshopper *Dociostaurus crassiusculus*. Brown shading in the map represents elevation, with darker areas corresponding to higher altitudes. Black lines show the boundaries of the main river basins separating the three population groups: Northern (dark green), Central (light green) and Southern (orange). (**b**) Male of *D. crassiusculus*. (**c**) Typical habitat of the species, with gypsophilous grounds and wastelands with halophytic vegetation. (**d**) Barplots showing the genetic assignments of the different individuals based on the Bayesian method implemented in the program STRUCTURE for $K = 2$ and $K = 3$. 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. These analyses are based on a random subset of 10,000 unlinked SNPs obtained with STACKS ($p = 4$).

Population genetic structure. STRUCTURE analyses based on a random subset of unlinked 10,000 SNPs from six different datasets obtained with STACKS and PYRAD considering different filtering/clustering parameters (see Supplementary Methods for further details), always identified $K = 2$ as the most likely clustering solution according with the ΔK criterion (Supplementary Figs 2 and 3). The two clusters presented no signature of genetic admixture and split the southernmost population (ORCE) from the remainder of the populations (Fig. 1d). STRUCTURE analyses for $K = 3$ divided Northern (TAJU-BELI) and Central populations (PHUE-SALI-BONI) in two different genetic clusters, but in this case the geographically closer populations (TAJU-BELI and PHUE-SALI) showed a considerable degree of genetic admixture (~25%) (Fig. 1d and Supplementary Fig. S2). The results obtained with STRUCTURE were in agreement with those obtained from Principal Component Analyses (PCA), in which PC1 split the Southern population (ORCE) from the remainder of the populations, and PC2 separated Northern (TAJU-BELI) from Central populations (PHUE-SALI-BONI) (Fig. 2 and Supplementary Fig. S4).

Phylogenomic inference. Phylogenomic relationships among populations inferred by SNAPP were well-resolved and nodes presented high posterior probabilities (Fig. 3a). In agreement with analyses of genetic structure (STRUCTURE and PCAs), SNAPP analyses supported an earlier split of ORCE from the remainder of the
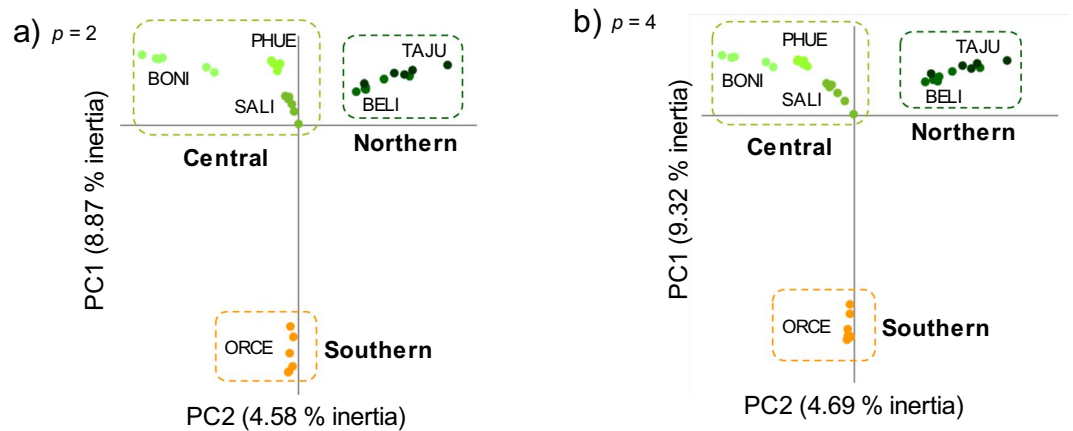
Figure 2. Principal component analyses (PCA) of genetic variation for populations of *D. crassiusculus*. Analyses are based on SNP datasets obtained with Stacks considering different filtering parameters: (**a**) 80,534 unlinked SNPs for $p = 2$; and (**b**) 65,459 unlinked SNPs for $p = 4$. Dotted-line rectangles group main population clusters. Population codes are described in Table 4.
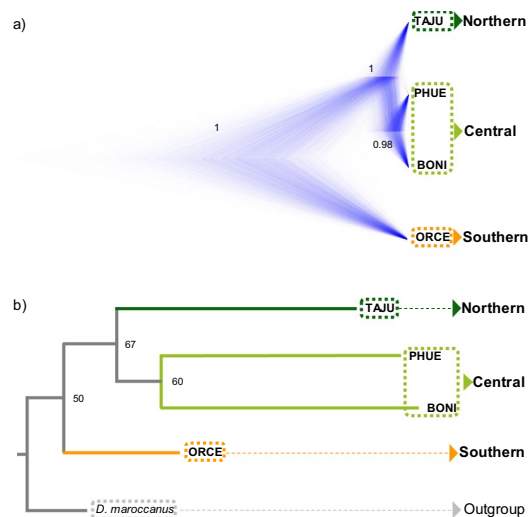

Figure 3. Phylogenetic trees inferred with (**a**) Snapp and (**b**) SVDQuartets considering four populations representative of the main geographical areas (populations separated >80 km) and the three genetic groups identified by Principal Component Analyses (PCA) and Bayesian clustering analyses in Structure (Northern, Central and Southern populations). Bayesian posterior probabilities (for Snapp) and bootstrapping support values (for SVDQuartets) are indicated on the nodes. Population codes are described in Table 4.

populations, which in turn divided into Northern (TAJU) and Central populations (PHUE and BONI) (Fig. 3a). Analyses considering different current and ancestral population sizes ($\alpha = 2$; $\beta = 200$ or $\alpha = 2$; $\beta = 20,000$) and different population combinations for Central and Northern populations (i.e. BELI-SALI, BELI-PHUE, and TAJU-SALI) yielded analogous results (data not shown)[50]. The best tree from SVDQuartets yielded the same topology than Snapp, but the relationships among populations were not well resolved (bootstrap support values <70%) probably as a result of inter-population gene flow or incomplete lineage sorting (Fig. 3b).

Coalescent-based demographic models. Fastsimcoal2 analyses supported models A, B and D (Fig. 4) as the best-fitting and statistically equivalent models (ΔAIC < 2.00; Table 1). These three migration models have in common that all of them consider gene flow between ancestral populations (m_{ANC}) (Fig. 4). Although analogous models without ancestral migration were tested (models C, E and F; Fig. 4), they were poorly supported (Table 1). Demographic parameters estimated under the three best supported models (A, B, and D) and their weighted averages are presented in Table 2. Considering a 1-year generation time for *D. crassiusculus*[3], Fastsimcoal2 analyses showed that the division between the Southern and Northern-Central populations (T_{DIV2}) occurred ~126 ka (95% CIs: 90–197 ka), probably during the Eemian Interglacial period (115–130 ka) (Table 2). The weighted average estimate yielded by Fastsimcoal2 for the more recent split between Northern and Central populations (T_{DIV1}) indicate that this event took place ~17 ka (95% CIs: 11–24 ka), around the last
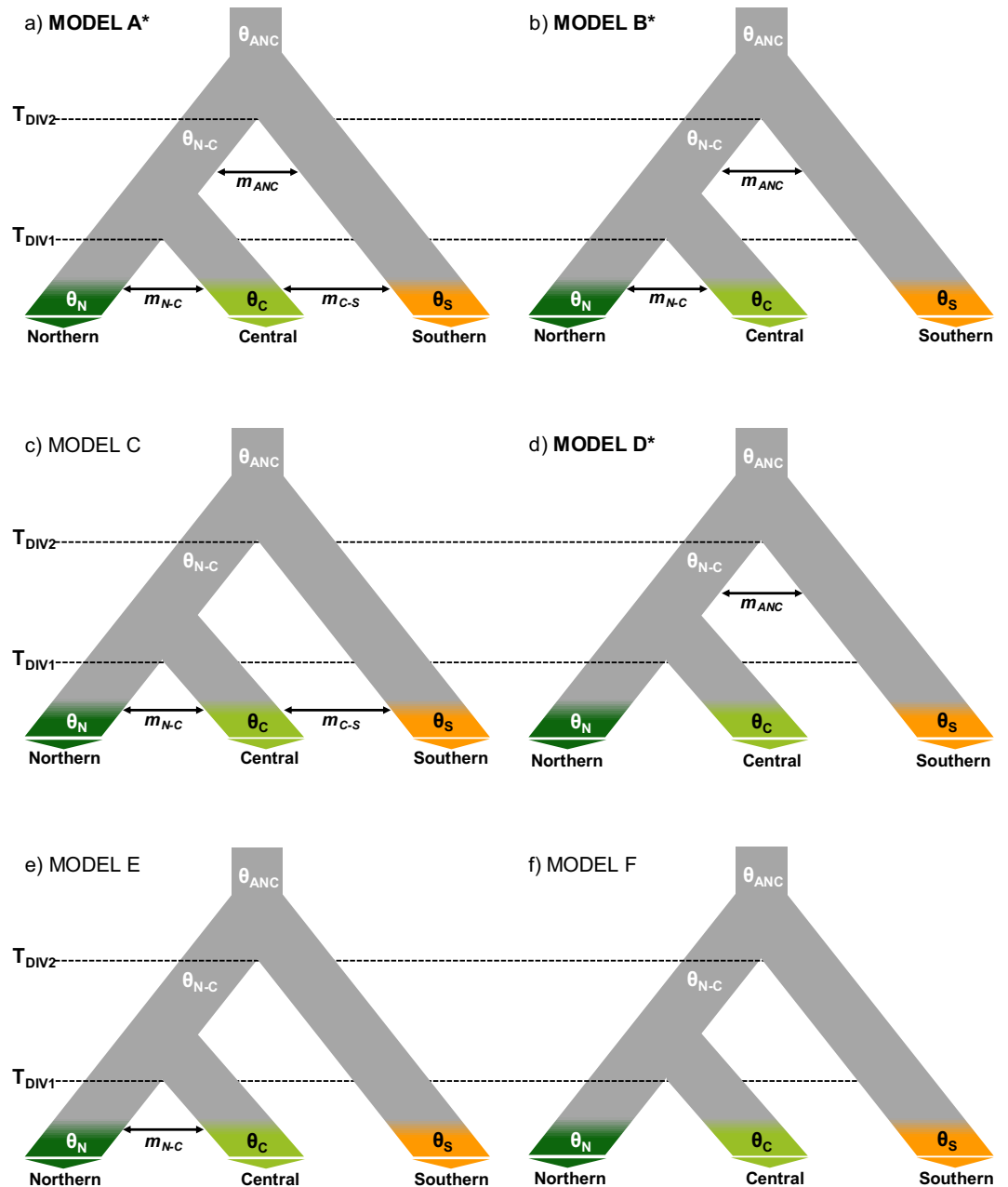
Figure 4. Alternative migration models tested using FastSimCoal2. Parameters include ancestral (θ_{ANC}, θ_{N-C}) and contemporary (θ_N, θ_C, θ_S) effective population sizes, timing of population split (T_{DIV2}, T_{DIV1}), and migration rates (m) between different pairs of populations. An asterisk and bold type indicate the three best supported migration models (see the Results section and Table 1 for more details).

glacial maximum (LGM; 20 ka) (Table 2). Gene-flow estimates were low and the migration rate (m) inferred between Central and Southern populations (m_{C-S}) was nearly an order of magnitude lower than the migration rate between Northern and Central populations (m_{N-C}) and between ancestral populations (m_{ANC}) after the first population split (T_{DIV2}) (Table 2).

Landscape genetic analyses.   Genetic differentiation was significantly and positively correlated with resistance distances obtained under all tested scenarios (Supplementary Table S3). Hypothetical scenarios based on habitat and lithology reached the highest model fit at the lowest resistance value for the non-suitable category, indicating that they do not explain the data better than a flat landscape in which all cells have equal resistance (=1) (Supplementary Table S3; Fig. 5a). In contrast, model fit for the scenario incorporating the resistance offered by the boundaries of main river basins peaked when the resistance value offered by this landscape feature was set to 100 ($r^2 = 0.830$; $P = 0.001$) (Supplementary Table S3; Fig. 5a). A multiple matrix regression with randomization (MMRR) analysis considering simultaneously the best fit resistance value under each scenario showed that the

Model	ln$L$	k	AIC	ΔAIC	ω_i
A	**−19,539.06**	**11**	**39,100.12**	**1.76**	**0.19**
B	**−19,539.18**	**10**	**39,098.36**	**0.00**	**0.46**
C	−19,546.21	10	39,112.43	14.07	0.00
D	**−19,540.48**	**9**	**39,098.97**	**0.61**	**0.34**
E	−19,566.71	9	39,151.42	53.06	0.00
F	−19,567.69	8	39,151.38	53.02	0.00

Table 1. Comparison of alternative migration models (detailed in Fig. 4) tested using Fastsimcoal2. For each model, the table shows the maximum likelihood estimate (ln$L$), the number of parameters (k), the Akaike's information criterion score (AIC), the difference in AIC value of each model from that of the strongest model (ΔAIC), and AIC weight (ω_i). Best-supported equivalent models (ΔAIC < 2) are indicated in bold (Fig. 4).

Parameter	Model A	Model B	Model D	Model average (95% CIs)
θ_{ANC}	99,783	95,231	111,992	101,840 (48,527–121,923)
θ_{N-C}	148,756	150,712	195,197	165,446 (134,574–201,611)
θ_N	163,990	166,040	130,914	153,687 (136,724–200,599)
θ_C	102,756	118,736	78,941	102,010 (91,758–131,587)
T_{DIV1}	31,995	19,268	4,507	16,795 (11,509–23,731)
T_{DIV2}	152,866	156,242	68,431	125,711 (90,218–197,204)
m_{ANC}	2.25×10^{-05}	1.45×10^{-05}	5.30×10^{-06}	1.30×10^{-05} (8.37×10^{-06}–1.50×10^{-04})
m_{N-C}	3.86×10^{-05}	3.30×10^{-05}	—	3.47×10^{-05} (2.69×10^{-05}–3.99×10^{-05})
m_{C-S}	1.83×10^{-06}	—	—	1.83×10^{-06} (3.57×10^{-07}–2.97×10^{-06})

Table 2. Parameters inferred from coalescent simulations with Fastsimcoal2 under the three best supported migration models (see Fig. 4 for details). The effective population size of one population (ORCE: θ_S) was calculated from nucleotide diversity estimates and fixed in the different models to enable the estimation of other parameters (see the Methods section). Table shows point estimates under each model and model averaged estimates with lower and upper 95% confidence intervals in parenthesis. Estimates of time are given in units of generations.

scenario incorporating the resistance offered by the boundaries of main river basins was the only one retained into the final model (Table 3; Fig. 5b). This indicates that isolation in different river basins is the main factor explaining genetic differentiation in the species, with no apparent effect of topographic roughness, lithology or habitat (Table 3).

Discussion

Genomic data revealed that populations of the endangered Iberian grasshopper *D. crassiusculus* show a marked hierarchical genetic structure, with the presence of two highly divergent cryptic lineages (Fig. 3) that comprise three genetic clusters (Figs 1 and 2). One of the lineages is only represented by the highly isolated population (ORCE) located in the southernmost limit of species distribution (Southern cluster), whereas the other includes the remainder of the populations and is sub-structured into two genetic clusters (Northern and Central clusters). Our phylogenomic and coalescent-based analyses supported an early split of the two lineages and estimated that their divergence took place during the Upper Pleistocene (~126 ka), probably around the Eemian interglacial stage. The Northern (TAJU-BELI) and Central (SALI-PHUE-BONI) genetic clusters were estimated to diverge much more recently (~17 ka), probably after the LGM. Note, however, that these estimates of divergence time must be interpreted with caution. In particular, it is remarkable the different estimates of divergence time obtained for Model D *vs*. Models A–B (Table 2). The fact that Model D does not consider gene flow among contemporary populations is expected to have resulted in younger estimates of population split than in Models A–B. Thus, Model D and Models A–B are statistically indistinguishable but find two different solutions that fit equally well our genomic data (Models A–B: presence of contemporary gene flow and older estimates of divergence times; Model D: lack of contemporary gene flow and younger estimates of divergence times) (Tables 1 and 2). Statistical evaluation of alternative migration models showed that the most likely scenarios were always those considering ancient gene flow between ancestral populations and contemporary gene flow between populations from Northern and Central genetic clusters, although with very low absolute values for migration rates per generation ($m_{ANC} = 1.30 \times 10^{-05}$; $m_{N-C} = 3.47 \times 10^{-05}$; Table 2 and Fig. 4). The consistent support for models including gene flow between ancestral populations (m_{ANC}), indicate that vicariance with multiple contacts (probably during glacial-interglacial cycles) is likely to have led to the current genetic structure of the species (i.e. isolation with gene flow). The best supported scenario (Model B) is the one considering gene flow between recently split populations across time, with higher contemporary migration rates among closer populations from Northern and Central genetic clusters (Fig. 4). Migration models involving gene flow with the Southern lineage were either not supported (Model C) or yielded point estimates of migration rates an order of magnitude lower ($m_{C-S} = 1.83 \times 10^{-06}$) than those inferred between Northern and Central populations (Model A) (Table 1; Fig. 4).
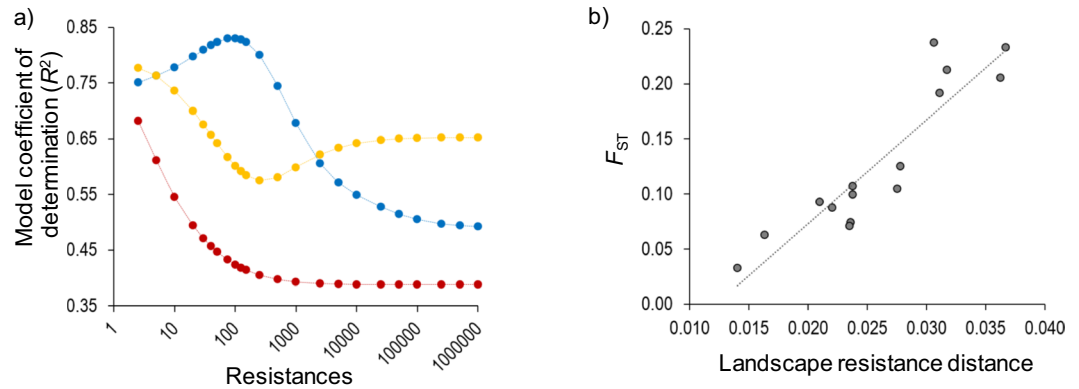
Figure 5. (**a**) Coefficient of determination (R^2) for models analysing genetic differentiation (F_{ST}) in relation with resistance distances defined by limits of main river basins (blue dots/line), habitat (yellow dots/line), and lithology (red dots/line). Each scenario considered a range (2.5–1,000,000) of 23 hypothetical resistance values offered by the barrier (limit of main river basins) or the areas not occupied by the species (non-suitable habitats/lithologies). Resistance values for different scenarios (*x*-axis) are log-transformed for illustrative purposes. (**b**) Relationship between genetic differentiation (F_{ST}) and resistance distances calculated using Circuitscape for the best fitting scenario (resistance offered by the boundaries of main river basins set to 100; see Table 3).

Variable	β	*t*	*p*
Explanatory terms			
Constant	−0.246	−2.117	0.067
Limits of main river basins (100)	1.382	7.973	0.001
Rejected terms			
Lithology (2.5)		0.648	0.827
Habitat (2.5)		2.366	0.312
Topography (slope)		0.304	0.836
Flat landscape		−1.102	0.257

Table 3. Multiple matrix regression with randomization (MMRR) for genetic differentiation (F_{ST}) in relation with resistance distances defined by (i) a flat landscape, (ii) topographic roughness (slope), (iii) limits of main river basins, (iv) habitat, and (v) lithology. The last three scenarios initially considered a range (2.5–1,000,000) of hypothetical resistance values offered by the barrier (limit of main river basins) or the areas not occupied by the species (non-suitable habitats/lithologies), but only resistance values (indicated in parentheses) yielding their respective best fitting models were included in this multivariate analysis (see Supplementary Table S3).

These results are in agreement with Bayesian clustering analyses, which showed considerable genetic admixture (~25%) among nearby populations from Northern and Central genetic clusters but no signature of admixed ancestry for the Southern lineage (Fig. 1d). Thus, despite the small distribution range and the relatively short geographical distances separating the extant populations of *D. crassiusculus*, our results indicate that this species shows a remarkable genetic structure that is comparable to that reported for other Orthoptera taxa with patchy distributions and forming highly fragmented populations[27,28,51,52].

Our landscape genetic analyses indicate that geographical distance, the spatial distribution of suitable habitats, lithology or topography do not explain *per se* the degree of genetic differentiation among populations and revealed that the limits of major river basins are the main factor explaining large-scale patterns of genetic structure in *D. crassiusculus* (Fig. 1a). These results are in agreement with inferences from Structure and PCA analyses, which showed that the populations of the species are clustered according to the limits of main river basins: Northern genetic cluster in Tagus river basin, Central genetic cluster in Guadiana river basin, and Southern genetic cluster in Guadalquivir river basin (Fig. 1a). Apart from numerous freshwater fishes[53,54], the importance of palaeo- and modern drainages in structuring genetic variation has been also reported in another steppe specialist grasshopper (*Mioscirtus wagneri*) presenting highly fragmented populations and inhabiting a similar geographic area[51] and in geckos (genus *Rhynchoedura*) from arid regions of Australia[55]. These results indicate the importance of this landscape feature on the evolutionary histories of terrestrial organisms from steppe and arid landscapes[55]. Rivers themselves do not seem to be an important barrier to dispersal in our study system, as populations located within the same basin but at different sides of main river stems or their tributaries (e.g. TAJU and BELI) show low levels of genetic differentiation in comparison with populations located in different basins[56]. Estimates of divergence time among contemporary populations of *D. crassiusculus* (~17–126 ka; Table 2) and the timing of species split from its sister taxon *D. kraussi* (1.01 Ma[48]) indicate that the origin of the species and its different lineages is probably posterior to the formation of the main river basins from the central-south Iberia, which are thought to have acquired their current configuration during the Oligocene-Pliocene[54,55,57]. Thus, the different genetic

Locality	Province	Code	n males	n females	Latitude	Longitude
Perales de Tajuña	Madrid	TAJU	4	1	40.2064	−3.3181
Belinchón	Cuenca	BELI	3	3	40.0793	−3.0446
Laguna de Peña Hueca	Toledo	PHUE	3	3	39.5158	−3.3486
Laguna de Salicor	Ciudad Real	SALI	3	3	39.4637	−3.1808
El Bonillo	Albacete	BONI	3	3	38.8779	−2.4834
Orce	Granada	ORCE	3	3	37.7515	−2.4259

Table 4. Geographical location and number of samples (*n*) for the studied populations of the Iberian cross-backed grasshopper *Dociostaurus crassiusculus*.

clusters and lineages are not likely to have resulted from population isolation in different palaeodrainages or ancient geological surfaces[26], but probably reflect the role of river drainages and lowlands as corridors of suitable habitat facilitating connectivity among populations located within the same basin[51]. Given that a predominantly flat landscape characterizes the distribution area of *D. crassiusculus* and the main drainages are not separated by an abrupt topography (i.e. mountain systems), our results suggest that populations of the species have probably remained linked to lowlands (e.g. pseudo-steppe saline low grounds) from different river basins[51,58] rather than physically separated by ridges representing the divides between drainages. Accordingly, our analyses indicate that other landscape features such as topographic roughness (slope) or the distribution of the typical habitats and lithological formations occupied by the species are not important factors explaining spatial patterns of genetic structure in *D. crassiusculus* (Table 3). Previous studies have identified topographic roughness as a relevant factor shaping genetic differentiation in two montane grasshoppers inhabiting areas with abrupt landscapes[25,59], a situation contrasting with the predominantly flat areas characterizing the distribution range of *D. crassiusculus*[3]. The widespread presence of sedimentary lithologies (evaporites, limestones, and conglomerates) across the distribution range of the species could have reduced our ability to identify barriers to dispersal linked to unsuitable geological formations or, alternatively, might reflect the capacity of the species to cross them. In any case, we must point out that our landscape genetic analyses should be interpreted with extreme caution, given that the very few extant populations of the species (*n* = 6) strongly limit the power of our analyses and the scope of the obtained inferences.

Coalescent-based analyses support the fact that range-wide patterns of genetic structure in *D. crassiusculus* are a consequence of ancient processes of population fragmentation (~17–126 ka; Table 2) that predate the Anthropocene. Accordingly, landscape genetic analyses suggest that land clearing for agriculture is not likely to explain large-scale patterns of genetic fragmentation (Fig. 5; Table 3). Based on the degree of divergence between the different lineages and genetic clusters, we recommend that the Northern, Central and Southern groups are recognized as Evolutionarily Significant Units (ESU)[60], Designatable Units (DU)[61] or Conservation Significant Units (CSU)[62]. These entities are likely to be substantially reproductively isolated from each other, represent an important component in the evolutionary legacy of the species, and include all discrete genetic and geographic subunits below the species level for status assessment, establishing conservation priorities and setting on-ground management strategies[60,63]. Of particular concern is the highly divergent Southern lineage because, as far as we know, it is currently represented by a single small population (ORCE) within the Guadalquivir river basin and Andalucía region (Fig. 1a,d). The correspondence between the identified units (lineages and genetic clusters) and the circumscription of different government administrations (Madrid, Castilla-La Mancha and Andalucía regions) could facilitate the establishment of regional conservation plans aimed at implementing the most efficient management strategies within each territory. Although nearby populations (TAJU-BELI and PHUE-SALI) showed no apparent signatures of genetic fragmentation (Fig. 1d), we must point that several lines of evidence suggest that this finding is not incompatible with a dramatic impact of human activities on the decline of the species at local and regional scales[3]. For instance, historical museum records indicate that many populations from the Northern cluster (specifically, in Madrid province) have been extirpated in the last decades[3,49]. All remaining populations are extremely small and submitted to severe impacts of human intervention (e.g. land ploughing, urbanization) and stochastic phenomena (e.g. flash flooding) that have been linked to sharp population declines[3,49]. The expected time-lag between population fragmentation/declines and the genetic consequences of such processes (disruption of gene flow, genetic differentiation, loss of genetic diversity, etc.) might explain why recent human impacts have not been yet reflected in spatial patterns of genetic variation[64]. Unfortunately, the small number of extant populations at local/regional scales (1–2 populations/genetic cluster), makes difficult to perform detailed analyses to evaluate the role of current landscape structure (e.g. land clearing for agriculture, urbanization, etc.) on the genetic connectivity of contemporary populations[27,65]. Future genomic analyses of specimens available in museum collections[3] could help to determine temporal changes in genetic diversity and study past patterns of gene flow in relation with historical landscape composition[52].

Overall, our genomic data support that the different lineages and genetic clusters of *D. crassiusculus* can be regarded as independent units that require adequate conservation and management strategies to preserve their idiosyncratic evolutionary histories. Conservation actions for *D. crassiusculus* should be focused on the preservation of areas with sensitive habitat occupied by the main lineages and units delineated by our genomic analyses. These should include the control of negative human interventions and the monitoring of local populations, actions that could also benefit other co-distributed and poorly-known species with similar ecological requirements and fragmented populations linked to gypsum and salt steppes of the Iberian Peninsula[7,10,27,66,67]. Given the extremely low number and size of extant populations of the species, *ex-situ* conservation plans and

reintroduction/translocations programmes in restored habitats could help to reduce the chances of species/lineage extinction[68,69]. These conservation actions should always consider the genomic singularity of the different units identified in this study and be accompanied with long-term habitat management and population monitoring[68,70]. Future studies including detailed ecological information (e.g. diet analyses[71]) and genome scans to detect potential loci under selection implicated in ecological adaptation[72,73] would be of great help to get a better understanding of the processes underlying the evolutionary history of the different lineages and refine the conservation actions for this endangered species.

Methods

Study area and sampling.
During 2008–2015, we sampled six populations of *Dociostaurus crassiusculus* (Pantel, 1886) (Fig. 1; Table 4). All the populations were found in areas with a particular lithological composition (evaporites, limestones, and conglomerates) and with plant communities linked to gypsum or hypersaline soils. We are confident that these populations cover the entire distribution range of the species, as other areas with potentially adequate habitats (i.e. pseudo-steppe saline grounds, wastelands with halophytic vegetation and surroundings of hypersaline/saline lagoons with marl-gypsum outcrops) have been extensively prospected without any records of the species[3,48]. *Dociostaurus crassiusculus* has been recently assigned to the category "endangered" in the IUCN red list of threatened species due to the high fragmentation of its very small size populations[3,44] and, for this reason, we only collected 5–6 adult individuals per population. We aimed at collecting an equal number of males and females in each locality, but samples sizes are often male-biased due to very low female numbers in some populations. Monitoring of some of the studied populations indicates that the abundance of *D. crassiusculus* in years before and after sampling was qualitatively similar, which suggests that the removal of only 5–6 individuals per locality had little impact on the population dynamics of the species. Fresh whole specimens were stored in 2,000 µL ethanol 96% at −20 °C until used for genomic analyses.

DNA extraction and genomic library preparation.
We used NucleoSpin Tissue kits (Macherey-Nagel, Durën, Germany) to extract and purify genomic DNA from the hind femur of each individual. Genomic DNA was individually barcoded and processed into one genomic library using the double-digestion restriction-fragment-based procedure (ddRADSeq) described in Peterson, *et al.*[74]. In brief, DNA was doubly digested with the restriction enzymes *Mse*I and *Eco*RI (New England Biolabs, Ipswich, MA, USA) and Illumina adaptors including unique 7-bp barcodes were ligated to the digested fragments. Ligation products were pooled, size-selected between 475–580 bp with a Pippin Prep (Sage Science, Beverly, MA, USA) machine and amplified by PCR with 12 cycles using the iProof™ High-Fidelity DNA Polymerase (BIO-RAD, Hercules, CA, USA). The library was sequenced in a single-read 150-bp lane on an Illumina HiSeq. 2500 platform at The Centre for Applied Genomics (SickKids, Toronto, ON, Canada).

Genomic data processing and bioinformatics.
We used both Stacks v. 1.35[75–77] and PyRAD v. 3.0.66[78] to assemble our sequences into *de novo* loci and call genotypes. This allowed us to examine the robustness of our analyses based on SNP datasets obtained using two of the most popular programs currently available to assemble RAD-seq data[76,78]. The choice of different filtering thresholds using either Stacks or PyRAD had little impact on the obtained inferences[50]. For this reason, unless otherwise indicated, all downstream analyses were performed using a SNP dataset obtained with Stacks including only those loci that were represented in at least four populations ($p = 4$). See Supplementary Methods for additional details on sequence assembling and data filtering.

Population genetic statistics.
Population genetics statistics, including major allele frequency (P), nucleotide diversity (π), observed (H_O) and expected (H_E) heterozygosity, and the Wright's inbreeding coefficient (F_{IS}), were calculated using the program *populations* from Stacks[75]. For biallelic RADSeq loci, π is an estimate of expected heterozygosity and is therefore a useful measure of the genetic diversity of populations. Furthermore, F_{IS} measures the reduction in observed heterozygosity as compared to expected heterozygosity for an allele in a population, with positive values indicating non-random mating or cryptic population structure[79–82]. Pair-wise F_{ST} values of genetic differentiation were calculated between all pairs of populations in Arlequin v.3.5[83]. We used PGDSpider v. 2.1.0.3[84] to convert Variant Call Format (VCF) files provided by Stacks into the correct format needed for Arlequin.

Population genetic structure.
We analysed population genetic structure and identified groups of individuals with similar ancestral gene pools using the Bayesian clustering method implemented in the program Structure v.2.3.3[85–87]. We ran Structure using a random subset of 10,000 unlinked SNPs from six different datasets obtained with Stacks and PyRAD considering different filtering/clustering parameters (see Supplementary Methods for further details). For each dataset, we ran Structure assuming correlated allele frequencies and admixture and without using prior population information[86]. We conducted 15 independent runs for each value of K, where K ranged from 1 to $n + 1$ for the dataset of n populations, to estimate the "true" number of clusters with a burn-in step of 100,000 iterations followed by 200,000 MCMC cycles. We retained the ten runs having the highest likelihood for each value of K and defined the number of populations best fitting the dataset using log probabilities [Pr(X|K)][87] and the ΔK method[88], as implemented in Structure Harvester[89]. We used Clumpp v. 1.1.2 and the Greedy algorithm to align multiple runs of Structure for the same K value[90] and Distruct v. 1.1[91] to visualize as bar plots the individual's probabilities of membership to each inferred genetic cluster. Complementary to Bayesian clustering analyses and in order to visualize the major axes of population genetic differentiation, we performed individual-based PCA using the R 3.3.3 (R Core Team 2017) package Adegenet[92].

Phylogenomic inference.   We inferred the phylogenetic relationships among the studied populations using the coalescent model implemented in the Snapp v.1.3.0[93] plug-in for Beast v.2.4.5[94]. Due to the large computational demands of this program, Snapp analyses were conducted using a random subset of 2,500 SNPs and including four populations (TAJU, PHUE, BONI, and ORCE) representative of the main geographical areas (i.e. populations separated >80 km; Fig. 1a) and the three genetic groups identified by PCA and Bayesian clustering analyses in Structure (Northern, Central and Southern clusters) (see Results section). We ran these analyses using different theta priors to allow for different current and ancestral population sizes (scenario 1: $\alpha = 2$; $\beta = 200$; and scenario 2: $\alpha = 2$; $\beta = 20,000$). The forward (u) and reverse (v) mutation rates were set to be calculated by Snapp and the remaining parameters were left at default values. We used the phrynomics R script written by Barb Banbury (https://github.com/bbanbury/phrynomics) to remove non-binary and invariant SNPs, code heterozygotes, and format input files for Snapp. We used different starting seed numbers to run two independent runs for each scenario, each with >5 million generations sampled every 1,000 steps. Each run was inspected in Tracer v.1.6[95] in order to check the convergence to stationary of the chains and confirm that Effective Sample Sizes (ESS) for all parameters were always much higher than 200. Afterwards, we combined the two replicate runs for each analysis using LogCombiner v.2.4.5, discarded 10% of trees as burn-in and used TreeAnnotator v.2.4.5 to obtain maximum credibility trees. Phylogenetic trees were displayed with DensiTree v.2.2.5[96]. Complementary to Snapp, we also ran phylogenetic analyses using SVDquartets[97] as implemented in Paup* v.4.0a152[98]. Analyses with SVDquartets included *Dociostaurus maroccanus* (Thunberg, 1815) as outgroup. Phylogenetic trees were constructed by exhaustively evaluating all possible quartets from the dataset and uncertainty in relationships was quantified using 1,000 bootstrapping replicates.

Coalescent-based demographic models.   We used Fastsimcoal2 and the site frequency spectrum (SFS)[99,100] to compare six hypothetical models of gene flow (see Fig. 4), calculate the composite likelihood of the probability of the observed data given a specified model, and estimate divergence times (t), effective population sizes (θ), and migration rates per generation (m)[99,100] under the best supported model/s. For Fastsimcoal2 analyses we considered the three genetic groups inferred by Structure and PCAs (Northern, Central and Southern) and the topology yielded by phylogenomic analyses in Snapp (see Results section)[81,101]. For each of the three population groups considered in the simulations, we selected 11 individuals from the Northern cluster, 12 individuals from the Central cluster, and the 6 individuals from the Southern cluster. A folded joint SFS was calculated considering a single SNP per locus to avoid the effects of linkage disequilibrium[102]. Because we did not consider invariable sites in the SFS (i.e. "removeZeroSFS" option in Fastsimcoal2), we fixed the effective population size for one of the populations (ORCE; θ_S) to enable the estimation of other parameters in Fastsimcoal2[50,81,99,102]. The effective population size fixed in the models was calculated from the level of nucleotide diversity (π) and estimates of mutation rate per site per generation (μ), since Ne $= (\pi/4\mu)$. Nucleotide diversity (π) for the population ORCE was estimated from polymorphic and non-polymorphic loci using Stacks ($\pi = 0.0011$; Supplementary Table S1). We considered an average mutation rate per site per generation[102,103] of 3.50×10^{-9}. To remove all missing data for the calculation of the joint SFS and minimize errors with allele frequency estimates, each population group was downsampled to 8–4 individuals (Northern group: 7 individuals; Central group: 8 individuals; Southern group: 4 individuals) using a custom Python script written by Qixin He and available on Dryad[102]. The final SFS contained information for 10,167 variable SNPs.

Each of the six models was run 100 replicated times using the computing resources provided by CESGA (Galician Supercomputer Center, Spain) and 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[50,102]. We used an information-theoretic model selection approach based on the Akaike's information criterion (AIC) to determine the probability of each model given the observed data[104–106]. After the maximum likelihood was estimated for each model, we calculated the AIC scores[106]. AIC values for each model were rescaled (ΔAIC) calculating the difference between the AIC value of each model and the minimum AIC obtained among all competing models (i.e. the best model has ΔAIC = 0). Confidence intervals of parameter estimates for the best supported models were obtained from 100 parametric bootstrap replicates by simulating SFS from the maximum composite likelihood point estimates and re-estimating parameters each time[81].

Landscape genetic analyses.   We generated alternative spatially-explicit isolation-by-resistance (IBR) scenarios of population connectivity and tested which one is better supported by observed data of genetic differentiation[107]. We applied circuit theory and used Circuitscape 4.0[108,109] to calculate resistance distance matrices between all pairs of populations under five hypothetical scenarios of gene flow: (i) a "flat" landscape in which all cells have equal resistance (resistance = 1), which is analogous to geographical distance but more appropriate for comparison with others competing models also generated with Circuitscape[25]; (ii) topographic roughness (slope); (iii) resistance offered by the boundaries of the main river basins from the study area (Tagus, Guadiana, Guadalquivir, Júcar, and Segura rivers; Fig. 1) (iv) resistance offered by non-natural landscapes and natural habitats not occupied by the species; and (v) resistance offered by areas with lithologies where the species is not present. Topographic roughness (slope) was calculated using a 90-m resolution digital elevation model from NASA Shuttle Radar Topographic Mission (SRTM Digital Elevation Data; http://srtm.csi.cgiar.org/) and the final layer was transformed to 30 arc-sec (c. 1 km) resolution for subsequent analyses. Natural habitats occupied by the species, natural habitats not occupied by the species, and non-natural habitats were defined according to Corine Land Cover maps[110]. We considered as natural habitats occupied by the species the Corine Land Cover categories "Natural grassland" and "Sclerophyllous vegetation", which represent the two habitat classes used by the species according to our own occurrence data[3]. Natural habitats not occupied by the species included all other habitats falling within the category "Forest and semi-natural areas" plus the category "Pastures". Non-natural habitats not occupied by the species grouped all other land cover categories, including agricultural areas and artificial

surfaces[110]. The lithological categories constituting the typical habitats occupied by the species (evaporites, limestones, and conglomerates) were identified according to our own occurrence data[3] and mapped using the spatial dataset OneGeology-Europe (http://info.igme.es/cartografia/oneGeology.asp?mapa = oneGeology). In scenarios iii–v we assigned a range of resistance values (2.5–1,000,000) to the barrier (limit of main river basins) or the areas not occupied by the species (non-suitable habitats/lithologies), which allowed us to identify the resistance value for these landscape features that best fits our data of genetic differentiation (F_{ST})[27,111]. Non-natural habitats (agricultural areas and artificial surfaces) were assumed to offer twice the resistance than natural habitats not occupied by the species (Supplementary Table S3). Background areas (i.e. areas within main river basins and habitats/lithologies occupied by the species) were given a fixed value of 1. All maps and GIS calculations were performed using ArcMap v.10.2.1 (ESRI, Redlands, CA, USA). In Circuitscape, we employed a four-neighbor cell connection scheme in order to make effective the resistance assigned to river basin boundaries, as linear landscape features become permeable through pixel corners under the eight-neighbor cell connection scheme[108]. Finally, we determined how well the different landscape resistance models fit observed data of genetic differentiation (F_{ST}) using multiple matrix regressions with randomization (MMRR) as implemented in R 3.3.3[107]. The final model was selected following a backward procedure, initially fitting all explanatory terms and progressively eliminating non-significant variables until all retained variables were significant. The significance of the variables excluded from the model was tested again until no additional variable reached significance[67].

Data Availability

SNP datasets and all other data generated and analysed in this study are available in Figshare or included in the published article and its Supplementary Information file.

References

1. Espindola, A. *et al.* Predicting present and future intra-specific genetic structure through niche hindcasting across 24 millennia. *Ecol. Lett.* **15**, 649–657 (2012).
2. Jay, F. *et al.* Forecasting changes in population genetic structure of alpine plants in response to global warming. *Mol. Ecol.* **21**, 2354–2368 (2012).
3. Cordero, P. J., Llorente, V., Aguirre, M. P. & Ortego, J. *Dociostaurus crassiusculus* (Pantel, 1886), especie (Orthoptera: Acrididae) rara en la Península ibérica con poblaciones locales en espacios singulares de Castilla-La Mancha (España). *Boletín de la Sociedad Entomológica Aragonesa (S.E.A.)* **46**, 461–465 (2010).
4. Brown, J. L. *et al.* Predicting the genetic consequences of future climate change: The power of coupling spatial demography, the coalescent, and historical landscape changes. *Am. J. Bot.* **103**, 153–163 (2016).
5. Cullingham, C. I., Kyle, C. J., Pond, B. A., Rees, E. E. & White, B. N. Differential permeability of rivers to raccoon gene flow corresponds to rabies incidence in Ontario, Canada. *Mol. Ecol.* **18**, 43–53 (2009).
6. Harrison, S. How natural habitat patchiness affects the distribution of diversity in Californian serpentine chaparral. *Ecology* **78**, 1898–1906 (1997).
7. Ortego, J., Aguirre, M. P. & Cordero, P. J. Population genetics of *Mioscirtus wagneri*, a grasshopper showing a highly fragmented distribution. *Mol. Ecol.* **19**, 472–483 (2010).
8. Lindenmayer, D. B. & Fischer, J. *Habitat Fragmentation and Landscape Change: An Ecological and Conservation Synthesis.* pp. 352. (Island Press, Washington, 2006).
9. Saunders, D. A., Hobbs, R. J. & Margules, C. R. Biological consequences of ecosystem fragmentation - A review. *Conserv. Biol.* **5**, 18–32 (1991).
10. Ribera, I. & Blasco-Zumeta, J. Biogeographical links between steppe insects in the Monegros region (Aragon, NE Spain), the eastern Mediterranean, and central Asia. *J. Biogeogr.* **25**, 969–986 (1998).
11. Wallis, G. P., Waters, J. M., Upton, P. & Craw, D. Transverse alpine speciation driven by glaciation. *Trends Ecol. Evol.* **31**, 916–926 (2016).
12. Willi, Y., Van Buskirk, J. & Hoffmann, A. A. Limits to the adaptive potential of small populations. *Annu. Rev. Ecol. Evol. Syst.* **37**, 433–458 (2006).
13. Zastavniouk, C., Weir, L. K. & Fraser, D. J. The evolutionary consequences of habitat fragmentation: Body morphology and coloration differentiation among brook trout populations of varying size. *Ecol. Evol.* **7**, 6850–6862 (2017).
14. Frankham, R. Conservation Genetics. *Annu. Rev. Genet.* **29**, 305–327 (1995).
15. Frankham, R. & Ralls, K. Conservation biology - Inbreeding leads to extinction. *Nature* **392**, 441–442 (1998).
16. Cunningham, M. & Moritz, C. Genetic effects of forest fragmentation on a rainforest restricted lizard (Scincidae: *Gnypetoscincus queenslandiae*). *Biol. Conserv.* **83**, 19–30 (1998).
17. Kalkvik, H. M., Stout, I. J. & Parkinson, C. L. Unraveling natural versus anthropogenic effects on genetic diversity within the southeastern beach mouse (*Peromyscus polionotus niveiventris*). *Conserv. Genet.* **13**, 1653–1664 (2012).
18. Moritz, C. Strategies to protect biological diversity and the evolutionary processes that sustain it. *Syst. Biol.* **51**, 238–254 (2002).
19. Aitken, S. N. & Whitlock, M. C. Assisted gene flow to facilitate local adaptation to climate change. *Annu. Rev. Ecol. Evol. Syst.* **44**, 367–388 (2013).
20. Wang, I. J., Savage, W. K. & Shaffer, H. B. Landscape genetics and least-cost path analysis reveal unexpected dispersal routes in the California tiger salamander (*Ambystoma californiense*). *Mol. Ecol.* **18**, 1365–1374 (2009).
21. Fraser, D. J. & Bernatchez, L. Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Mol. Ecol.* **10**, 2741–2752 (2001).
22. Yan, F. *et al.* The Chinese giant salamander exemplifies the hidden extinction of cryptic species. *Curr. Biol.* **28** (2018).
23. Manel, S., Schwartz, M. K., Luikart, G. & Taberlet, P. Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol. Evol.* **18**, 189–197 (2003).
24. Quéméré, E., Crouau-Roy, B., Rabarivola, C., Louis, E. E. & Chikhi, L. Landscape genetics of an endangered lemur (*Propithecus tattersalli*) within its entire fragmented range. *Mol. Ecol.* **19**, 1606–1621 (2010).
25. Noguerales, V., Cordero, P. J. & Ortego, J. Hierarchical genetic structure shaped by topography in a narrow-endemic montane grasshopper. *BMC Evol. Biol.* **16**, 96 (2016).
26. Pepper, M., Doughty, P., Arculus, R. & Keogh, J. S. Landforms predict phylogenetic structure on one of the world's most ancient surfaces. *BMC Evol. Biol.* **8**, 152 (2008).
27. Ortego, J., Aguirre, M. P., Noguerales, V. & Cordero, P. J. Consequences of extensive habitat fragmentation in landscape-level patterns of genetic diversity and structure in the Mediterranean esparto grasshopper. *Evol. Appl.* **8**, 621–632 (2015).
28. Tinnert, J., Hellgren, O., Lindberg, J., Koch-Schmidt, P. & Forsman, A. Population genetic structure, differentiation, and diversity in *Tetrix subulata* pygmy grasshoppers: roles of population size and immigration. *Ecol. Evol.* **6**, 7831–7846 (2016).

29. Ruiz-Gonzalez, A. *et al*. Landscape genetics for the empirical assessment of resistance surfaces: The European pine marten (*Martes martes*) as a target-species of a regional ecological network. *PLoS One* **9**(10), e110552 (2014).

30. Segelbacher, G. *et al*. Applications of landscape genetics in conservation biology: concepts and challenges. *Conserv. Genet.* **11**, 375–385 (2010).

31. Brooks, T. M. *et al*. Global biodiversity conservation priorities. *Science (New York, N.Y.)* **313**, 58–61 (2006).

32. Médail, F. & Quézel, P. Biodiversity Hotspots in the Mediterranean Basin: Setting Global Conservation Priorities. *Conserv. Biol.* **13**, 1510–1513 (1999).

33. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).

34. Blondel, J. & Aronson, J. *Biology and wildlife of the Mediterranean region*. Oxford University Press (1999).

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

36. Faille, A., Andújar, C., Fadrique, F. & Ribera, I. Late Miocene origin of an Ibero-Maghrebian clade of ground beetles with multiple colonizations of the subterranean environment. *J. Biogeogr.* **41**, 1979–1990 (2014).

37. Ferrer-Castán, D. & Vetaas, O. R. Pteridophyte richness, climate and topography in the Iberian Peninsula: comparing spatial and nonspatial models of richness patterns. *Glob. Ecol. Biogeogr.* **14**, 155–165 (2005).

38. Barredo, J. I., Caudullo, G. & Dosio, A. Mediterranean habitat loss under future climate conditions: Assessing impacts on the Natura 2000 protected area network. *Appl. Geogr.* **75**, 83–92 (2016).

39. Giorgi, F. & Lionello, P. Climate change projections for the Mediterranean region. *Glob. Planet. Change* **63**, 90–104 (2008).

40. Hanski, I. Habitat loss, the dynamics of biodiversity, and a perspective on conservation. *Ambio* **40**, 248–255 (2011).

41. Remón, N., Galán, P. & Naveira, H. Chronicle of an extinction foretold: genetic properties of an extremely small population of *Iberolacerta monticola*. *Conserv. Genet.* **13**, 131–142 (2012).

42. Rubidge, E. M. *et al*. Climate-induced range contraction drives genetic erosion in an alpine mammal. *Nat. Clim. Change* **2**, 285–288 (2012).

43. Saccheri, I. *et al*. Inbreeding and extinction in a butterfly metapopulation. *Nature* **392**, 491–494 (1998).

44. Hochkirch, A. *et al. European Red List of Grasshoppers*, *Crickets and Bush-crickets*. pp. 94. (Publications Office of the European Union, Luxembourg, 2016).

45. Harz, A. *Die Orthopteren Europas II / The Orthoptera of Europe II*. pp. 941. (Springer, The Netherlands, 1975).

46. Hodjat, S. H. A review of Iranian Dociostaurini (Orthoptera: Gomphocerinae) with keys to their species. *Entomol. Gen.* **35**, 253–268 (2016).

47. Soltani, A. A. Preliminary synonymy and description of new species in the genus *Dociostaurus* Fieber, 1853 (Orthoptera: Acridoidea: Acrididae, Gomphocerinae) with a key to the species in the genus. *Journal of Entomological Society of Iran*, 1–93 (1978).

48. González-Serna, M. J., Ortego, J. & Cordero, P. J. 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. *Syst. Entomol.* **43**, 136–146 (2018).

49. Gangwere, S. K., Viedma, M. G. d. & Llorente, V. *Libro rojo de los ortópteros ibérico*s. pp. 91. (Instituto Nacional para la Conservación de laNaturaleza, 1985).

50. Ortego, J., Gugger, P. F. & Sork, V. L. Genomic data reveal cryptic lineage diversification and introgression in Californian golden cup oaks (section *Protobalanus*). *New Phytol.* **218**, 804–818 (2018).

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

52. Schmid, S. *et al*. Spatial and temporal genetic dynamics of the grasshopper *Oedaleus decorus* revealed by museum genomics. *Ecol. Evol.* **8**, 1480–1495 (2018).

53. Thacker, C. E., Unmack, P. J., Matsui, L. & Rifenbark, N. Comparative phylogeography of five sympatric *Hypseleotris* species (Teleostei: Eleotridae) in south-eastern Australia reveals a complex pattern of drainage basin exchanges with little congruence across species. *J. Biogeogr.* **34**, 1518–1533 (2007).

54. Gómez, A. & Lunt, D. H. *Phylogeography of* Southern *European Refugia* (ed Steven Weiss and Nuno Ferrand). pp. 155–188. (Springer, The Netherlands, 2006).

55. Pepper, M., Doughty, P., Hutchinson, M. N. & Keogh, J. S. Ancient drainages divide cryptic species in Australia's arid zone: Morphological and multi-gene evidence for four new species of Beaked Geckos (*Rhynchoedura*). *Mol. Phylogenet. Evol.* **61**, 810–822 (2011).

56. Tantrawatpan, C. *et al*. Genetic differentiation among populations of *Brachytrupes portentosus* (Lichtenstein 1796) (Orthoptera: Gryllidae) in Thailand and the Lao PDR: the Mekong River as a biogeographic barrier. *Bull. Entomol. Res.* **101**, 687–696 (2011).

57. Doadrio, I. Delimitation of areas in the Iberian Peninsula on the basis of freshwater fishes. *Bonn. Zool. Beitr.* **39**, 113–128 (1988).

58. Peinado, M. *Funcionamiento y variabilidad de los geosistemas de los humedales manchegos*. pp. 480. (PhD Thesis, Universidad Complutense de Madrid, Spain, 1994).

59. Noguerales, V., Cordero, P. J. & Ortego, J. Testing the role of ancient and contemporary landscapes on structuring genetic variation in a specialist grasshopper. *Ecol. Evol.* **7**, 3110–3122 (2017).

60. Waples, R. S., Jones, R. P. J., Beckman, B. R. & Swan, G. A. *Status review for Snake River fall Chinook salmo*n. pp. 73. (U.S. Dep. Commer, NOAA Tech. Memo. NMFS F/NWC-201, 1991).

61. Green, D. M. Designatable units for status assessment of endangered species - Unidades designatables para la evaluación del estatus de especies en peligro. *Conserv. Biol.* **19**, 1813–1820 (2005).

62. Yuan, J. H., Cheng, F. Y. & Zhou, S. L. The phylogeographic structure and conservation genetics of the endangered tree peony, *Paeonia rockii* (Paeoniaceae), inferred from chloroplast gene sequences. *Conserv. Genet.* **12**, 1539–1549 (2011).

63. Moritz, C. Defining 'Evolutionarily Significant Units' for conservation. *Trends Ecol. Evol.* **9**, 373–375 (1994).

64. Cushman, S. A., McKelvey, K. S., Hayden, J. & Schwartz, M. K. Gene flow in complex landscapes: Testing multiple hypotheses with causal modeling. *Amer. Naturalist* **168**, 486–499 (2006).

65. Zellmer, A. J. & Knowles, L. L. Disentangling the effects of historic *vs*. contemporary landscape structure on population genetic divergence. *Mol. Ecol.* **18**, 3593–3602 (2009).

66. Cordero, P. J. & Llorente, V. New data on the 'silver-bell cricket' (Orthoptera, Gryllidae), a forgotten and overlooked cricket subject to a high risk of extinction in western Europe. *Graellsia* **64**, 171–180 (2008).

67. Ortego, J., Gugger, P. F., Sork, V. L. & Riddle, B. Climatically stable landscapes predict patterns of genetic structure and admixture in the Californian canyon live oak. *J. Biogeogr.* **42**, 328–338 (2015).

68. Baur, B., Thommen, G. H. & Coray, A. Dynamics of reintroduced populations of *Oedipoda caerulescens* (Orthoptera, Acrididae) over 21 years. *J. Insect Sci.* **17**(1), 10 (2017).

69. Perl, R. G. B. *et al*. Population genetic analysis of the recently rediscovered Hula painted frog (*Latonia nigriventer*) reveals high genetic diversity and low inbreeding. *Sci. Rep.* **8**, 5588 (2018).

70. Tian, H. Z. *et al*. Genetic diversity in the endangered terrestrial orchid *Cypripedium japonicum* in East Asia: Insights into population history and implications for conservation. *Sci. Rep.* **8**, 6467 (2018).

71. McClenaghan, B., Gibson, J. F., Shokralla, S. & Hajibabaei, M. Discrimination of grasshopper (Orthoptera: Acrididae) diet and niche overlap using next-generation sequencing of gut contents. *Ecol. Evol.* **5**, 3046–3055 (2015).

72. Apple, J. L., Grace, T., Joern, A., Amand, P. S. & Wisely, S. M. Comparative genome scan detects host-related divergent selection in the grasshopper *Hesperotettix viridis*. *Mol. Ecol.* **19**, 4012–4028 (2010).

73. Feng, X. J., Jiang, G. F. & Fan, Z. Identification of outliers in a genomic scan for selection along environmental gradients in the bamboo locust, *Ceracris kiangsu*. *Sci. Rep.* **5**, 13758 (2015).

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

75. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140 (2013).

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

77. Hohenlohe, P. A. *et al.* Population genomics of parallel adaptation in *Threespine Stickleback* using sequenced RAD tags. *PLoS Genet.* **6**, e1000862 (2010).

78. Eaton, D. A. PyRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics* **30**, 1844–1849 (2014).

79. Hartl, D. L. & Clark, A. G. *Principles of population genetics* (4th ed. Sinauer Associates, Inc. Publishers). pp. 545. (Sunderland, Massachusetts, 2007).

80. Holsinger, K. E. & Weir, B. S. Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. *Nat. Rev. Genet.* **10**, 639–650 (2009).

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

82. Nei, M. & Kumar, S. *Molecular evolution and phylogenetics*. pp. 352. (Oxford University Press, New York, 2000).

83. Excoffier, L. & Lischer, H. E. 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 (2010).

84. Lischer, H. E. & Excoffier, L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**, 298–299 (2012).

85. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).

86. Hubisz, M. J., Falush, D., Stephens, M. & Pritchard, J. K. Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* **9**, 1322–1332 (2009).

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

88. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software Structure: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).

89. Earl, D. A. & vonHoldt, B. M. Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361 (2012).

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

91. Rosenberg, N. A. Distruct: a program for the graphical display of population structure. *Mol. Ecol. Notes* **4**, 137–138 (2004).

92. Jombart, T. Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).

93. Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A. & RoyChoudhury, A. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol. Bio. Evol.* **29**, 1917–1932 (2012).

94. Bouckaert, R. *et al.* Beast 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).

95. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. *Tracer v.1.7*, http://tree.bio.ed.ac.uk/software/tracer/ (2018).

96. Bouckaert, R. & Heled, J. DensiTree 2: seeing trees through the forest. *bioRxiv* (2014).

97. Chifman, J. & Kubatko, L. Quartet inference from SNP data under the coalescent. *Bioinformatics* **30**, 3317–3324 (2014).

98. Swofford, D. L. *Paup*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0b10.* (Sinauer Associates, Sunderland, MA, 2002).

99. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. Robust Demographic inference from genomic and SNP data. *PLoS Genet.* **9**, e1003905 (2013).

100. Excoffier, L. & Foll, M. Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* **27**, 1332–1334 (2011).

101. Eaton, D. A., Hipp, A. L., Gonzalez-Rodriguez, A. & Cavender-Bares, J. Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution* **69**, 2587–2601 (2015).

102. Papadopoulou, A. & Knowles, L. L. Species-specific responses to island connectivity cycles: refined models for testing phylogeographic concordance across a Mediterranean Pleistocene Aggregate Island Complex. *Mol. Ecol.* **24**, 4252–4268 (2015).

103. Keightley, P. D. *et al.* Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Res.* **19**, 1195–1201 (2009).

104. Abascal, F. *et al.* Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biol.* **17**, 251 (2016).

105. Burnham, K. P. & Anderson, D. R. *Model selection and inference: a practical information-theoretic approach*. pp. 355. (Springer, New York, 1998).

106. Thome, M. T. & Carstens, B. C. Phylogeographic model selection leads to insight into the evolutionary history of four-eyed frogs. *Proc. Natl. Acad. Sci. USA* **113**, 8010–8017 (2016).

107. Wang, I. J. 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 (2013).

108. McRae, B. H. Isolation by resistance. *Evolution* **60**, 1551–1561 (2006).

109. McRae, B. H. & Beier, P. Circuit theory predicts gene flow in plant and animal populations. *Proc. Natl. Acad. Sci. USA* **104**, 19885–19890 (2007).

110. C*ORINE land cover*. EEA. Commission of the European Communities, Luxembourg, https://land.copernicus.eu/pan-european/corine-land-cover/clc-2012 (2012).

111. Andrew, R. L., Ostevik, K. L., Ebert, D. P. & Rieseberg, L. H. Adaptation with gene flow across the landscape in a dune sunflower. *Mol. Ecol.* **21**, 2078–2091 (2012).

Acknowledgements

Author Contributions

M.G.S., P.J.C. and J.O. conceived the study and collected the samples. M.G.S. and J.O. designed the study and analyses. M.G.S. performed the lab work and analysed the data guided by J.O. M.G.S. led the writing and J.O. and P.J.C. contributed to editing and reviewing the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-31775-x.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.