Journal of Biogeography (*J. Biogeogr.*) (2011) **38**, 2006–2020

ORIGINAL ARTICLE

Tracing the origin of disjunct distributions: a case of biogeographical convergence in *Pyrgus* butterflies

Juan L. Hernández-Roldán[1,2], Cesc Múrria[3], Helena Romo[2], Gerard Talavera[1], Evgeny Zakharov[4], Paul D. N. Hebert[4] and Roger Vila[1,5]*

[1]*Institut de Biologia Evolutiva (CSIC-UPF), Passeig Marítim de la Barceloneta, 37-49, E-08003 Barcelona, Spain,* [2]*Departamento de Biología (Zoología), Facultad de Ciencias de la Universidad Autónoma de Madrid, C/Darwin, 2, E-28049 Madrid, Spain,* [3]*Department of Entomology, The Natural History Museum, London SW7 5BD, UK,* [4]*Biodiversity Institute of Ontario, University of Guelph, Guelph, ON, Canada N1G 2W1,* [5]*Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, E-08010 Barcelona, Spain*

*Correspondence: Roger Vila, Institut de Biologia Evolutiva (CSIC-UPF), Passeig Marítim de la Barceloneta, 37-49, E-08003 Barcelona, Spain.
E-mail: roger.vila@ibe.upf-csic.es

ABSTRACT

Aim To study the biogeographical factors responsible for the current disjunct distributions of two closely related species of butterflies (*Pyrgus cinarae* and *Pyrgus sidae*, Lepidoptera: Hesperioidea). Both species have small populations in the Iberian Peninsula that are isolated by more than 1000 km from their nearest conspecifics. Because these species possess similar ecological preferences and geographical distributions, they are excellent candidates for congruent biogeographical histories.

Location The Palaearctic region, with a special focus on the Mediterranean peninsulas as glacial refugia.

Methods We integrated phylogeography and population genetic analyses with ecological niche modelling. The mitochondrial gene cytochrome *c* oxidase subunit 1 (COI) and the non-coding nuclear marker internal transcribed spacer 2 (ITS2) were analysed for 62 specimens of *P. cinarae* and for 80 of *P. sidae* to infer phylogeography and to date the origin of disjunct distributions. Current and ancestral [Last Glacial Maximum using MIROC (Model for Interdisciplinary Research on Climate) and CCSM (Community Climate System Model) circulation models] distribution models were calculated with Maxent. Using present climatic conditions, we delimited the ecological space for each species.

Results The genetic structure and potential ancestral distribution of the two species were markedly different. While the Iberian population of *P. cinarae* had an old origin (*c.* 1 Ma), that of *P. sidae* was closely related to French and Italian lineages (which jointly diverged from eastern populations *c.* 0.27 Ma). Ecological niche modelling showed that minor differences in the ecological preferences of the two species seem to account for their drastically different distributional response to the last glacial to post-glacial environmental conditions. Although the potential distribution of *P. cinarae* was largely unaffected by climate change, suitable habitat for *P. sidae* strongly shifted in both elevation and latitude. This result might explain the early origin of the disjunct distribution of *P. cinarae*, in contrast to the more recent disjunction of *P. sidae*.

Main conclusions We show that convergent biogeographical patterns can be analysed with a combination of genetic and ecological niche modelling data. The results demonstrate that species with similar distributional patterns and ecology may still have different biogeographical histories, highlighting the importance of including the temporal dimension when studying biogeographical patterns.

Keywords
Biogeography, COI, disjunct distribution, ecology, ITS2, Lepidoptera, niche modelling, Palaearctic region, palaeoclimate, phylogeography.

http://wileyonlinelibrary.com/journal/jbi
doi:10.1111/j.1365-2699.2011.02547.x

Journal of Biogeography

INTRODUCTION

Disjunct distributions (i.e. populations separated by a wide area where the species does not occur) represent extreme biogeographical patterns characterized by unusual evolutionary histories (Schmitt & Hewitt, 2003; Schmitt *et al.*, 2006; Garcia Collevatti *et al.*, 2009). Such distributions may reflect range fragmentation of a formerly widely distributed species due to changes in environmental conditions that affected suitable habitat distribution (Cox & Moore, 2005), as in the case of isolation of populations in refugia from glacial periods (e.g. Stehlik *et al.*, 2000). Alternatively, they can arise by extraordinary long-distance dispersal events, which result in the colonization of new suitable habitats (Davis & Shaw, 2001; Cox & Moore, 2005; Garcia Collevatti *et al.*, 2009). Determining how historical and present-day factors interact to initiate and maintain disjunct distributions in different evolutionary scenarios is a challenge for biogeographers and evolutionary biologists.

The distribution of organisms in the Palaearctic region has been strongly influenced by historical and current factors, especially the cyclic glacial and interglacial periods during the Pleistocene. These events affected large geographical areas and produced cycles of demographic contraction and expansion in species with low dispersal capacity, or recurrent shifts in the distributions of species with good dispersal abilities (Hewitt, 2000, 2004; Schmitt, 2007). Present-day factors have an important role in the maintenance of disjunct distributions by reducing connectivity among isolated populations. As a result, the current geographical distribution of genetic structure within species may reflect the impacts of Pleistocene climatic oscillations on refugial location, the level of gene flow between refugia during interglacial periods and varying connectivity linked to dispersal capacity (Avise, 2000; Hewitt, 2004). As a consequence, it is anticipated that most current disjunct distributions in the Palaearctic region arose from populations isolated in glacial refugia where varied species found suitable conditions for their survival during glacial maxima and that there has been low or no gene flow among these habitats.

In Europe, the main glacial refugia were located in the Mediterranean area: the Iberian, Italian and Balkan peninsulas (Hewitt, 1996). Phylogeographical studies based on species variability have shown a different molecular biogeographical pattern for this area compared with Continental and Arctic European regions (Schmitt, 2007). This pattern is characterized by one or more genetic lineages that began to diverge in Mediterranean refugia. In many species of animals and plants (Taberlet *et al.*, 1998; Hewitt, 1999, 2000) gene flow between populations in these former refugia is at present absent or very limited, and they apparently evolved independently (e.g. Schmitt & Seitz, 2002; Schmitt & Krauss, 2004; Habel *et al.*, 2005). Studying gene flow during interglacial periods, which depends on dispersal traits and habitat connectivity, is thus critical for understanding the biogeographical history of species and to reveal the origin of disjunct distributions.

Molecular genetic techniques are now used widely in phylogeographical studies of both animals and plants (e.g. Hewitt, 2004; Garcia Collevatti *et al.*, 2009). Recently, new methodologies based on environmental variables have been developed to estimate the potential geographical distributions of species (Guisan & Zimmerman, 2000; Elith *et al.*, 2006; Phillips *et al.*, 2006). Among them, the program MAXENT has been shown to perform better than other methods [for example GARP (Stockwell & Peters, 1999) and BIOCLIM (Nix, 1986)] in predicting species distributions from presence-only data (Phillips *et al.*, 2004, 2006; Elith *et al.*, 2006; Wisz *et al.*, 2008). MAXENT is based on maximum entropy modelling of species geographical distributions, and computes a probability distribution of habitat suitability over the geographical area of the units considered. The integration of phylogeographical and distribution modelling seems to be a promising way to unravel the biogeographical history behind disjunct distributions (Weaver *et al.*, 2006; Jakob *et al.*, 2007; Alsos *et al.*, 2009).

In this study we trace the genesis of disjunct distributions in two members of the genus *Pyrgus* (Lepidoptera: Hesperioidea: Hesperiidae: Pyrginae). *Pyrgus cinarae* (Rambur, 1839) and *Pyrgus sidae* (Esper, 1784) have very similar distributions: in Europe they are limited to the north Mediterranean, extending deeply into central Asia, and both display a disjunct distribution with small isolated populations in the Iberian Peninsula (Kudrna, 2002; García-Barros *et al.*, 2004). The extremely localized Iberian population is restricted to central Spain, *c.* 1800 km (*P. cinarae*) and *c.* 1000 km (*P. sidae*) from the nearest conspecific populations (see Fig. 1). Despite the apparent presence of suitable habitats and host plants (*Potentilla recta* and *Potentilla hirta*) between the Iberian and non-Iberian populations, no studies have suggested their connectivity. *Pyrgus cinarae* and *P. sidae* display similar ecological preferences (they frequently co-occur in the same habitats and share host plants), and both have low dispersal ability (Hernández-Roldán *et al.*, 2009; Wagner, 2009), especially across water surfaces (they are not present on islands or in North Africa), a trait shared by all members of the genus *Pyrgus*.

Species with similar current distributions and ecological traits are good candidates for having congruent biogeographical histories, i.e. similar responses to the same environmental changes or geographical events (Bocxlaer *et al.*, 2006; Noonan & Chippindale, 2006). Recent calls to integrate the temporal and spatial dimensions in biogeographical studies, especially when studying patterns across taxa, have been made (Hunn & Upchurch, 2001; Donoghue & Moore, 2003). In this regard, it is worth noting that two species could display spatially congruent but temporally incongruent biogeographical histories (Loader *et al.*, 2007). In this paper, by combining molecular data (both phylogeography and population genetics) with ecological niche modelling, we study two closely related species that are apparently equivalent in most regards to test whether they share parallel evolutionary histories.

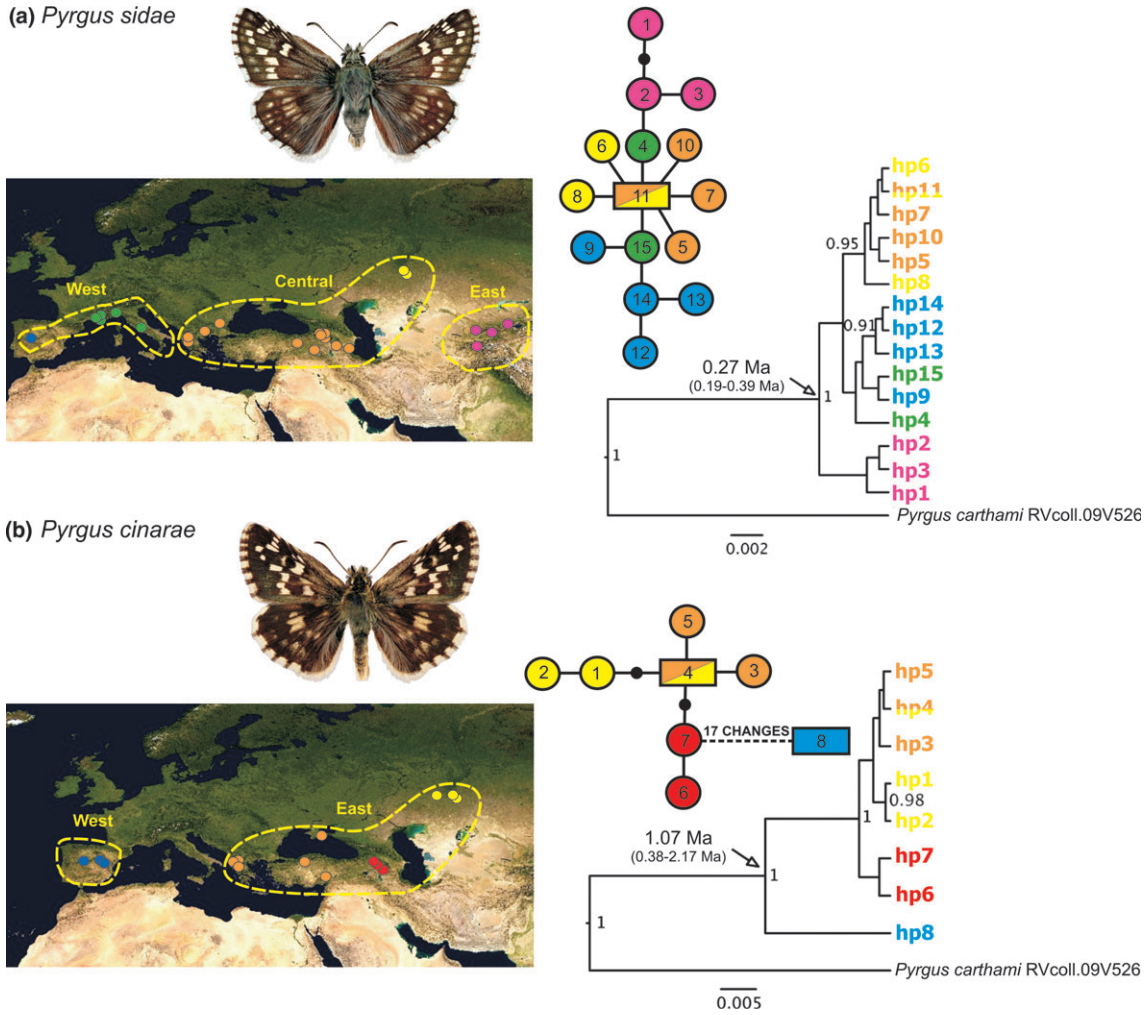

Figure 1 Maps showing localities of studied specimens, COI haplotype networks and phylogenetic trees for (a) *Pyrgus sidae* and (b) *P. cinarae*. TCS v.121 with a 95% connection limit was used to reconstruct the haplotype networks. In the maps, the three main haplotype clades for *P. sidae* and the two for *P. cinarae* are indicated by discontinuous lines. Colours indicate five populations for *P. sidae* (magenta, Kyrgyzstan; yellow, south Urals; orange, Caucasus–Black Sea–Balkans; green, France–Italy; blue, Iberian Peninsula) and four populations for *P. cinarae* (yellow, south Urals; red, Caucasus; orange, Black Sea–Balkans; blue, Iberian Peninsula). BEAST v.1.5.8 under a coalescent model was used for Bayesian tree inference. Scale bars in the trees show divergence in substitutions/site and only posterior probabilities > 0.9 are shown in the nodes. The divergence time of the oldest split within each species calculated by MDIV models is indicated, with the confidence interval in the form of 95% highest posterior densities in parentheses.

MATERIALS AND METHODS

Sampling and gathering of the molecular data set

Samples

Sixty-two specimens of *P. cinarae* and 80 specimens of *P. sidae* were collected from 18 and 32 localities, respectively, covering the known ranges of both species (Fig. 1). The 48 sampling sites were partitioned into five populations for *P. sidae* (Kyrgyzstan, south Urals, Caucasus–Black Sea–Balkans, France–Italy, Iberian Peninsula) and into four populations for *P. cinarae* (south Urals, Caucasus, Black Sea–Balkans, Iberian Peninsula) that were separated by more than 1000 km without any record for the species (Fig. 1). Samples were preserved in 100% ethanol for molecular analysis. Identification codes and collection localities for the samples used are listed in Appendix S1 in Supporting Information. Voucher specimens were deposited in the collection of the Institut de Biologia Evolutiva (CSIC-UPF), Barcelona, Spain.

Cytochrome c oxidase subunit I (COI) amplification

DNA was extracted using a glass fibre protocol (Ivanova *et al.*, 2006) from a single leg of each specimen. A 658-bp fragment of the mitochondrial gene cytochrome *c* oxidase subunit I (COI) was targeted for polymerase chain reaction (PCR) amplification using the primers LepF (5′-ATTCAACCAATCATAAAGATATTGG-3′) and LepR (5′-TAAACTTCTGGATGTCCAAAAAATCA-3′) (Hajibabaei *et al.*, 2005; deWaard *et al.*, 2008).

Samples that did not produce a PCR product with the primers LepF and LepR were reamplified with the primers LepF and Enh_LepR (5′-CTCCWCCAGCAGGATCAAAA-3′), which amplify a 609-bp fragment of COI. All specimens were successfully amplified for this marker. One specimen of *Pyrgus carthami* (Hübner, 1813) was amplified and used as an outgroup. One COI sequence of *Pyrgus communis* (Grote, 1872) obtained from GenBank (accession number AF170857) was also added to the dataset as an outgroup. Sequences were obtained with an ABI 3730 × 1 sequencer (Applied Biosystems, Carlsbad, CA, USA) following the manufacturer's recommendations.

Internal transcribed spacer 2 (ITS2) amplification

A total of 23 *P. sidae* and 12 *P. cinarae* samples representing all the COI haplotype clades, plus a specimen of *Pyrgus armoricanus* (Oberthür, 1910) used as an outgroup, were sequenced for the non-coding nuclear marker internal transcribed spacer 2 (ITS2). Total genomic DNA was extracted using Chelex 100 resin, 100–200 mesh, sodium form (Bio-Rad, Richmond, CA, USA), under the following protocol: one leg was removed and introduced into a tube with 100 μL of Chelex 10% solution and 5 μL of Proteinase K (20 mg mL^{-1}). The samples were incubated overnight at 55 °C and then incubated at 100 °C for 15 min. The samples were subsequently centrifuged for 10 s at 1500 *g* and the supernatant was used for PCR amplification. A 684-bp fragment at the 5′-end of ITS2 was amplified using the primers ITS3 (5′-GCATCGATGAAGAACGCAGC-3′) and ITS4 (5′-TCCTCCGCTTATTGATATGC-3′) (White *et al.*, 1990). Double-stranded DNA was amplified in 25 μL volume reactions: 16.7 μL ultra pure (high-performance liquid chromatography quality) water, 2.5 μL 10× buffer, 1 μL 100 mM $MgCl_2$, 0.25 L 100 mM dNTP, 1.2 μL of each primer (10 mM), 0.15 μL Taq DNA Polymerase (Bioron GmbH, Ludwigshafen, Germany) and 2 μL of extracted DNA. The typical thermal cycling profile was: 95 °C for 45 s, 47 °C for 60 s and 72 °C for 60 s, for 40 cycles. Finally, ITS2 PCR products were purified and sequenced by Macrogen Inc. (Seoul, Korea).

Sequences, specimen photographs and associated data for the COI sequences are available in the 'Butterflies of Spain' project Barcode of Life Data Systems (http://www.barcodinglife.org, downloaded 2 November 2010). COI and ITS2 sequences are also available in GenBank (see Appendix S1 for accession numbers).

Data analyses

Phylogenetic inference

COI and ITS2 sequences were edited and aligned using Geneious Pro v.4.8.3 (Biomatters Ltd., 2009; http://www.geneious.com/). These analyses resulted in four final alignments: (1) 654 bp and 80 specimens for *P. sidae* COI, (2) 654 bp and 62 specimens for *P. cinarae* COI, (3) 611 bp and 23 specimens for *P. sidae* ITS2, and (4) 636 bp and 12 specimens for *P. cinarae* ITS2. As we were not estimating population sizes, a selection of unambiguous haplotypes using tcs v.1.21 (Clement *et al.*, 2000) was used for COI phylogenetic inference resulting in 15 haplotypes for *P. sidae* and 8 haplotypes for *P. cinarae*. *Pyrgus carthami* and *P. armoricanus* were used as outgroups for COI and ITS2, respectively. A coalescent approach was used to reconstruct Bayesian trees for all the alignments using beast v.1.5.3 (Drummond & Rambaut, 2007). For COI trees, HKY + I and GTR + I models of nucleotide substitution were used for *P. sidae* and *P. cinarae,* respectively, and HKY for both ITS2 alignments, according to jModelTest v.0.1 (Posada, 2008) suggestions for the Akaike information criterion (AIC). Parameters were estimated using two independent runs of 10 million generations each (with a pre-run burn-in of 100,000 generations) to ensure convergence.

Estimate of evolutionary entities

Genetic clusters representing evolutionary entities were established using the generalized mixed Yule-coalescent (GMYC) model (Pons *et al.*, 2006; Fontaneto *et al.*, 2007). This model tests for a change in branching rates at the species boundary to classify the observed intervals of genetic divergence to either inter-specific ('diversification') or intra-specific ('coalescent') processes to delimit 'independently evolving' mitochondrial DNA (mtDNA) clusters. All individual haplotypes of COI sequences for *P. sidae* and *P. cinarae*, as well as the outgroups *P. carthami* and *P. communis*, were used to perform the GMYC analysis to detect populations that are evolving independently within the study species. A maximum likelihood phylogeny was obtained with RAxML v.7.0.4 (Stamatakis, 2006) under a GTR + γ substitution model. The resulting topology was made ultrametric using a penalized likelihood as implemented in r8s v.1.7 (Sanderson, 2003). The GMYC analysis was conducted using 'Splits' from the R package (http://www.r-project.org/) with the 'single threshold' option.

Population genetics, demographic and genetic divergence analyses

To describe genetic diversity for each species, we calculated polymorphic sites and nucleotide diversity π_T (i.e. the average number of nucleotide differences per site between two sequences) and its variance (Nei, 1987). To visualize relationships among haplotypes, a statistical parsimony haplotype network was constructed with a 95% connection limit using tcs v.1.21 (Clement *et al.*, 2000), and these haplotypes were hierarchically nested in clades following Templeton's rules (Templeton & Sing, 1993). Closely related haplotypes were generally distributed in geographical proximity, and the tcs analysis resulted in three haplotype clades for *P. sidae* (West, Central and East) and two haplotype clades for *P. cinarae* (West and East) (Fig. 1, see Results). These haplotype clades included more than one population (defined in the sampling design) and both geographical levels (clades and populations) were used in the following analyses of genetic structure (Fig. 1).

To study genetic divergence and current gene flow among populations, we calculated pairwise Φ_{ST} values (an analogue to F_{ST} that incorporates genetic divergence between sequences) between them and their genetic divergence D_{xy} (i.e. the average number of nucleotide substitutions per site between populations; Nei, 1987). In order to assess the level of gene flow among populations we used S_{nn} statistics (nearest-neighbour statistics; Hudson, 2000), excluding the Iberian population of *P. cinarae*, where only one haplotype was observed. In addition, to assess at which spatial scale the genetic variability was structured, i.e. to detect if gene flow was more effective at the level of populations or haplotype clades, we used analysis of molecular variance (AMOVA) with hierarchical partitioning (Excoffier *et al.*, 2005). AMOVA was carried out by estimating Φ_{ST} at three hierarchical levels using 10,000 random permutations. Hierarchical level tests included the following: among haplotype clades; among populations in each haplotype clades; and within populations. A Mantel test was used to detect associations between genetic (i.e. pairwise Φ_{ST} between populations) and geographical distances, which were measured as the shortest distance between the centroids of two populations. We used the software DnaSP v.4.10 (Rozas *et al.*, 2003) to calculate the genetic diversity parameters and S_{nn} statistics. All other analyses were carried out using Arlequin v.3.1 (Excoffier *et al.*, 2005).

In order to detect evidence of population expansion and infer the demographic history of each species, we calculated the statistics Tajima's *D*, Fu's F_S, Fu and Li's *D* and *F* (Ramos-Onsins & Rozas, 2002) using DnaSP v.4.10 (Rozas *et al.*, 2003). Tajima's *D* tested neutrality (i.e. populations evolved under neutrality) against non-random processes (i.e. population expansion or natural selection). To distinguish between the non-random processes, we tested Fu and Li's *D* and *F* and Fu's F_S. The time of divergence between haplotype clades and time to most recent common ancestor (TMRCA) were inferred using the program mdiv (Nielsen & Wakeley, 2001) under the finite sites model. We reported the mode and 95% highest posterior densities (HPD) for each parameter (2,000,000 Markov chain interactions; burn-in = 500,000; $M = 0$; $T_{max} = 5$; θ = auto-initialize). In the absence of a specific mutation rate for *Pyrgus*, we assumed a generation time of 1 year based on the evidence that these species are univoltine (Hernández-Roldán *et al.*, 2009; Wagner, 2009), and a 1.5% Myr^{-1} divergence rate for arthropod COI (Quek *et al.*, 2004) in all historical demographic analyses.

Current and ancestral distribution modelling

The latitude and longitude of the centroid of each 10 × 10 km Universal Transverse Mercator (UTM) square, measured as *x* and *y* coordinates in metres, were selected as spatial variables, considering 50 different distribution points in *P. cinarae* and 154 in *P. sidae* based on the current known distribution of both species in Europe (Abadjiev, 2001; Kudrna, 2002; García-Barros *et al.*, 2004; J. G. Coutsis, Athens, Greece, pers. comm.).

The 19 WorldClim variables (http://www.worldclim.org/, described by Hijmans *et al.*, 2005) were considered. As WorldClim variables generally show a high collinearity that can distort the results obtained, a subselection of variables was employed. Taking into account the study area considered for the distribution modelling analyses (Europe), 10,000 randomly generated points were chosen to extract the values of the WorldClim variables. Using these values, the level of correlation between pairs of variables was analysed. When two variables shared a Pearson correlation coefficient of 0.8 or higher (Rissler & Apodaca, 2007), we selected the biologically most meaningful variable according to the physiological requirements of the *Pyrgus* species (usually that related to the activity of the more sensitive adult stage) or the variable that was easier to interpret (that encompassing a wider temporal range). In this way, 8 out of 19 variables were retained: Bio1 (annual mean temperature), Bio2 (mean diurnal range), Bio7 (temperature annual range), Bio8 (mean temperature of the wettest quarter), Bio12 (annual precipitation), Bio13 (precipitation of the wettest period), Bio15 (precipitation seasonality) and Bio18 (precipitation of the warmest quarter).

To predict the potential distribution models we employed Maxent v.3.3.2 (http://www.cs.princeton.edu/~schapire/maxent/), which uses a machine-learning algorithm to identify the areas in which the environmental conditions are suitable for the species considered in the model. For each species, starting from a uniform distribution, the program performs a number of iterations, each of which increases the probability of the sample locations for the species. The probability is displayed in terms of 'gain', and this gain increases iteration by iteration, until the change from one iteration to the next falls below a specified threshold, or a maximum number of iterations have been performed (Phillips *et al.*, 2006).

The default parameter settings were used (maximum number of background points 10,000; regularization multiplier 1; auto features; maximum iterations 500; convergence threshold 0.00001) as suggested by Phillips *et al.* (2006). Each model was run with 100 replicates and cross-validation, using 25% of the presence data to test the model and 75% to train the model (as suggested and used by other authors, e.g. Moffett *et al.*, 2007; Pawar *et al.*, 2007; Alba-Sánchez *et al.*, 2010). The logistic output was selected due to the easier interpretation of the results (interpreted as probability of presence of the species) compared to raw and cumulative output formats, and the results were presented on a linear scale (Phillips, 2008; Phillips & Dudík, 2008). Values of 0.5 indicate typical presence data points, and the most suitable sites are those whose logistic values are close to 1. The 95% confidence output files were chosen to represent the results, adjusting the threshold of maximum probability of presence to the closest one representing the real known distribution of the species.

To test the accuracy of the models, the area under the receiver operating characteristic curve (AUC) and the following set of 11 binomial tests were performed (Phillips *et al.*, 2006; Moffett *et al.*, 2007; Pawar *et al.*, 2007): (1) fixed

cumulative value 1, (2) fixed cumulative value 5, (3) fixed cumulative value 10, (4) minimum training presence, (5) tenth percentile training presence, (6) equal training sensitivity and specificity, (7) maximum training sensitivity plus specificity, (8) equal test sensitivity and specificity, (9) maximum test sensitivity plus specificity, (10) balance training omission, predicted area and threshold value, and (11) equate entropy of thresholded and original distributions. All 11 binomial tests were required to be significant at $P < 0.01$. If the predictions yielded by the model were not better than random, the AUC value would be equal to 0.5. Values of AUC higher than 0.7 (Pearce & Ferrier, 2000; Elith, 2002; Newbold *et al.*, 2009) or 0.85 (Newbold, 2009) are considered acceptable.

After calibrating the models for their current distributions in relation to the present climate, we modelled the distribution onto Last Glacial Maximum (LGM) WorldClim data, with two general atmospheric circulation models: CCSM (Community Climate System Model) and MIROC (Model for Interdisciplinary Research on Climate) models from the Paleoclimate Modelling Intercomparison Project Phase II (PMIP2) through WorldClim, using the same eight variables considered in the present models. The original variables (2.5 arcmin) were transformed according to the scale used with DIVA-GIS software v.7.1.7 (Hijmans *et al.*, 2005). Only those areas that were recovered as suitable according to both models were considered.

A comparison between the predicted areas in present and past times was made with STATISTICA v.7 (StatSoft, Inc., 2004), representing the elevation and latitude values for both species, in order to test their relationships.

Measurement of the ecological niche

To estimate the ecological niche of each species, we performed a principal components analysis (PCA) on the eight log-transformed WorldClim selected variables (see above). We plotted values on PCA axes 1 and 2 of 178 localities from Europe for *P. sidae*, 51 localities from Europe (except the Iberian Peninsula) for *P. cinarae* and 12 localities of *P. cinarae* from the Iberian Peninsula and assumed these limits as demarcating the ecological niches of each taxon. Additionally, we calculated Pearson correlations between species scores on PCA axes and the environmental data. We tested niche overlap between species conducting a multivariate analysis of variance (MANOVA) and *post hoc* Newman–Keuls tests using PCA species scores as dependent variables, and species as factors. Statistical analyses were computed using the 'ade4' of the R package (http://www.r-project.org/) and STATISTICA v. 7 software (StatSoft, Inc., 2004).

RESULTS

Phylogenetic analyses

The Bayesian phylogenetic tree for *P. sidae* COI haplotypes displayed three main clades with differing geographical origins: one clade included populations from the western part of the distribution, another from the central area, while the third included the eastern populations (Fig. 1a; see localities and haplotypes in Appendix S1). However, only the clade from the central region was statistically supported. For *P. cinarae*, the COI tree showed two strongly supported clades, one including the Iberian Peninsula populations and the other the rest (Fig. 1b). Interestingly, the divergence between these two clades (17 substitutions, 2.60%) was much deeper than those between clades of *P. sidae* (1 substitution, 0.15%), suggesting a much older origin for the disjunct distribution in *P. cinarae*.

The nuclear marker ITS2 resulted in a tree with low intraspecific divergences for *P. sidae* that failed to recover its eastern, central and western clades (Fig. 2a). However, it did recover the same two clades revealed by COI for *P. cinarae*, with monophyly of the western clade being statistically

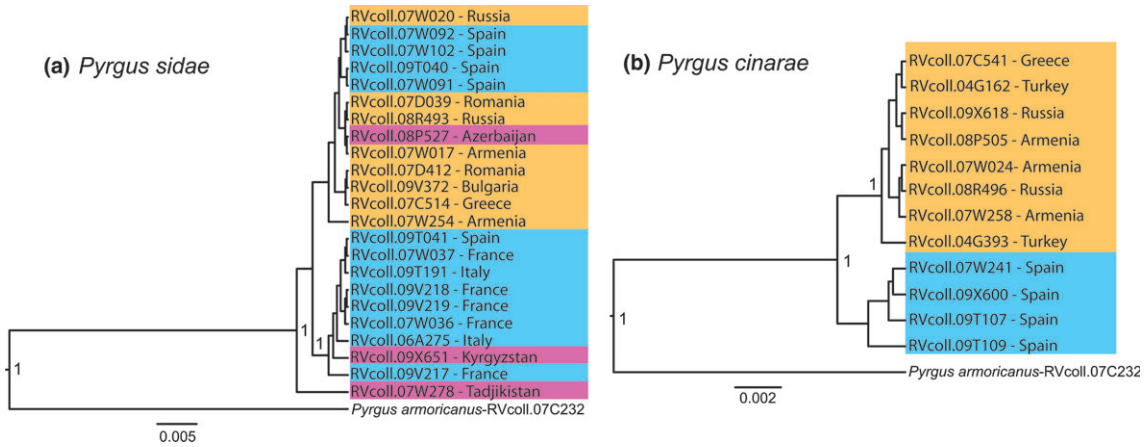

Figure 2 Internal transcribed spacer 2 (ITS2) gene trees for (a) *Pyrgus sidae* and (b) *P. cinarae*. BEAST v.1.5.8 under a coalescent model was used for Bayesian tree inference. Scale bars show divergence in substitutions per site. Only posterior probabilities > 0.9 are shown in the nodes. Colours correspond to the main regions determined by cytochrome *c* oxidase subunit 1 (COI) haplotype clades, as shown on the maps in Fig. 1.

significant (Fig. 2b). GMYC analysis grouped haplotypes into five evolutionarily independent entities (Appendix S2), namely the external groups *P. communis* and *P. carthami*, as well as *P. sidae* and two lineages of *P. cinarae,* one composed exclusively of the specimens from the Iberian Peninsula and the other of the remaining populations.

Population genetics and demographic analyses

From the COI alignments, 10 polymorphic sites (1.52%) with four singleton variable sites and nucleotide diversity (π) of 0.00408 ± 0.00048 were detected for *P. sidae*, whereas *P. cinarae* had 21 polymorphic sites (3.21%) with 15 singleton variable sites and a nucleotide diversity (π) of 0.01032 ± 0.00444. The parsimony haplotype network showed no haplotypes of any species present in all populations. On the contrary, all haplotypes were restricted to single populations, except haplotype 11 for *P. sidae* and haplotype 4 for *P. cinarae*, which were shared by two populations and are possibly ancestral haplotypes (Fig. 1, Table 1). Remarkably, TCS was unable to link the single Iberian haplotype (haplotype 8) of *P. cinarae* to the rest because of very high divergence (17 changes separate this haplotype from the closest). The ancestral haplotypes suggested by TCS for both *P. sidae* and non-Iberian *P. cinarae* were located in the Balkans–Caucasus and south Urals. In the haplotype network, haplotypes were grouped into clades strongly related to their geographical distribution. *Pyrgus sidae* showed three haplotype clades – Eastern (Kyrgyzstan and Tajikistan), Central (south Urals, Caucasus, Black Sea and Balkans) and Western (France–Italy and the Iberian Peninsula) clades. *Pyrgus cinarae* showed similar genetic groups, with two main independent haplotype clades – Iberia and the rest (south Urals, Caucasus, Black Sea and the Balkans), but this species is not present in Kyrgyzstan or Tajikistan. For *P. cinarae* there was one missing haplotype between the ancestral Balkans–Black Sea populations and south Urals or Caucasus, suggesting reduced historical gene flow among them despite the presence of the ancestral haplotype 4 in the south Urals. This is not the case for *P. sidae* because there were no missing haplotypes between populations.

All pairwise Φ_{ST} comparisons among populations were significant for both species, suggesting limited current gene flow among populations (Table 2). Both D_{xy} and Φ_{ST} values between populations were similar, with the notable exception of the *P. cinarae* Iberian population, which was shown to be genetically unique. S_{nn} tests of genetic structure showed differences among populations for *P. sidae* ($S_{nn} = 0.46250$, $P = 0.0010$), whereas *P. cinarae* (excluding the Iberian population) showed non-significant differences among populations ($S_{nn} = 0.625$, $P = 0.0620$). The absence of genetic structure for *P. cinarae* seems related to the low genetic intra-populational diversity detected for this species. This is most acute in the Iberian Peninsula where only one haplotype was detected among 12 individuals. The AMOVA analyses showed similar patterns for both species, with the highest genetic structure among haplotype clades (50.34% of genetic variation detected for *P. sidae*, 86.03% for *P. cinarae*) (Table 3). In both cases, the genetic differentiation at this spatial scale was non-significant ($P > 0.05$), probably due to the lack of statistical power (low replication on statistical inference based on permutations) because of the low number

Table 1 Cytochrome *c* oxidase subunit 1 (COI) haplotype composition of the sampled populations for *Pyrgus sidae* and *P. cinarae*: populations, nucleotide diversity (π), sampled sites (see Appendix S1 for site description), haplotypes present (n = number of individuals) and number of specimens sequenced (N). The ancestral haplotype for each species is highlighted.

Species	Haplotype clades	Populations	π	Sites	COI haplotype (n)	N
P. sidae	East	Kyrgyzstan	0.00306	Kyrgyzstan	2 (17), 3 (1)	18
				Tajikistan	1 (2),	2
	Central	South Urals	0.00204	Russia	6 (4), 8 (1), **11** (3)	8
		Caucasus–Black Sea–Balkans	0.00229	Azerbaijan	5 (2)	2
				Armenia	**11** (12)	12
				Turkey	**11** (2),	2
				Bulgaria	5 (2)	2
				Romania	5 (4), 7 (3)	7
				Greece	5 (2), 10 (1)	3
	West	France–Italy	0.00153	Italy	15 (2)	2
				France	4 (1), 15 (10)	11
		Iberian Peninsula	0.00306	Spain	9 (1), 12 (2), 13 (2), 14 (6)	11
P. cinarae	East	South Urals	0.00306	Russia	1 (19), 2 (1), **4** (1)	21
		Caucasus	0.00153	Armenia	6 (11), 7(2)	13
		Black Sea–Balkans	0.00204	Ukraine	3 (1), **4** (2)	3
				Turkey	**4** (5)	5
				Greece	**4** (7), 5 (1)	8
	West	Iberian Peninsula	0	Spain	8 (12)	12

Table 2 Average number of nucleotide substitutions per site (D_{xy}; Nei, 1987) between populations (above the diagonal), and values of Φ_{ST} from pairwise population comparisons (below the diagonal) for cytochrome *c* oxidase subunit 1 (COI) haplotypes of *Pyrgus sidae* and *P. cinarae*.

P. sidae	Kyrgyzstan	South Urals	Balkans	France	Spain
Kyrgyzstan		0.005	0.005	0.003	0.006
South Urals	0.796*		0.002	0.002	0.004
Balkans	0.770*	0.311*		0.002	0.004
France	0.839*	0.742*	0.647*		0.002
Spain	0.790*	0.665*	0.671*	0.603*	

P. cinarae	South Urals	Caucasus	Balkans	Spain
South Urals		0.005	0.003	0.0295
Caucasus	0.91*		0.004	0.0267
Balkans	0.869*	0.911*		0.029
Spain	0.99*	0.991*	0.992*	

*$P < 0.05$.

Table 3 Analysis of molecular variance (AMOVA) among *Pyrgus sidae* and *P. cinarae* populations grouped by main cytochrome *c* oxidase subunit 1 (COI) haplotype clades (see Fig. 1). Values for the variance components, the percentage of variation at each hierarchical level (%), *F*-statistics and *P*-values are shown.

	Variance components	Percentage of variation	Fixation indices*	*P*-value
P. sidae				
Among haplotype clades	0.615	50.34	$F_{CT} = 0.503$	0.074
Among populations within haplotype clades	0.293	23.99	$F_{SC} = 0.743$	0
Within populations	0.313	25.68	$F_{ST} = 0.743$	0
P. cinarae				
Among haplotype clades	8.086	86.03	$F_{CT} = 0.860$	0.245
Among populations within haplotype clades	1.202	12.79	$F_{SC} = 0.916$	0
Within populations	0.111	1.18	$F_{ST} = 0.988$	0

*Fixation indices for AMOVA measured the degree of genetic differentiation at different organizational levels: F_{CT}, differentiation among haplotype clades relative to total; F_{SC}, differentiation among populations relative to haplotypes clades; F_{ST}, differentiation within populations relative to total populations.

of populations per haplotype clade (Fitzpatrick, 2009). We calculated that the minimum expected *P*-value for the rejection of the null hypothesis for *P. sidae* (two groups of two populations each and one group of one population) was $\alpha = 0.066$, whereas for *P. cinarae* (one group of three populations and one group of one population) it was $\alpha = 0.25$. Thus, the genetic structure among haplotype clades was significant for *P. cinarae* ($P = 0.245$) (Table 3), and close to significant for *P. sidae* ($P = 0.074$). Genetic structure was significant among populations within haplotype clades and within populations for both species, especially for *P. sidae* (Table 3). Non-significant relationships were found between genetic and geographical distances in the Mantel test (*P. sidae* $r = 0.48$, $P = 0.06$; *P. cinarae* $r = 0.74$, $P = 0.25$), which discarded a pure isolation-by-distance model.

Non-random genetic divergence was detected for both species by the negative values of Tajima's *D* ($D = -0.816$ for *P. sidae* and $D = -0.869$ for *P. cinarae*, $P > 0.1$). The results obtained for Fu and Li's *D* ($D = -0.149$ for *P. sidae* and $D = -0.988$ for *P. cinarae*, $P > 0.1$) and Fu and Li's *F* ($F = -0.380$ for *P. sidae* and $F = -1.068$ for *P. cinarae*, $P > 0.1$) indicated that the populations have expanded or are under selection. Population expansion is confirmed by Fu's F_S ($F_S = -17.433$, $P = 0.001$ for *P. sidae* and $F_S = -3.088$, $P = 0.044$ for *P. cinarae*). Therefore, the demographic history of both species suggests the existence of a historical bottleneck and current population expansion. Times of divergence between clades calculated by MDIV models were different for the two species. For *P. sidae* the time of divergence between west and central clades was 0.25 Ma (95% HPD = 0.18–0.29 Ma), and between central and east clades was 0.27 Ma (95% HPD = 0.19–0.39 Ma). For *P. cinarae* the time of divergence between the Iberian clade and the rest was 1.07 Ma (95% HPD = 0.38–2.17 Ma). Estimates for the TMRCA were also different for the two species: TMRCA for *P. sidae* was 0.51 Ma (95% HPD = 0.37–0.72 Ma) for west and

central clades and 0.52 Ma (95% HPD = 0.31–0.71 Ma) for east and central clades, while TMRCA for *P. cinarae* was 1.90 Ma (95% HPD = 1.33–2.48 Ma) for the two main clades.

Current and ancestral distribution modelling

The models obtained showed high mean AUC scores (averaged across all 100 runs) in both species (*P. cinarae* 0.989, SD 0.006; and *P. sidae* 0.956, SD 0.011) according to the evaluation test provided by Maxent software. These high AUC values demonstrated a good model performance. Besides, predictions for *P. sidae* and *P. cinarae* were significantly different from random because all 11 binomial omission test thresholds proved significant (*P*-value < 0.01) across all 100 runs.

The predicted distribution for both species was quite similar to the actual one, although there were some differences. In the case of *P. sidae* (Fig. 3a) the areas with higher probability of presence are located in different mountain chains, such as the Iberian System and Pyrenees in the Iberian Peninsula, the Italian Alps, and the Balkans. The Hungarian populations of *P. sidae* seem to be located in a rather unfavourable area (i.e. with lower probability of presence). On the other hand, areas favourable for *P. cinarae* (Fig. 3b) were mainly in the Iberian System and the Balkans. There is a favourable area in the Iberian Peninsula from where *P. cinarae* has never been recorded that deserves deeper inspection in future faunistic surveys.

A heuristic estimate of relative contributions of the environmental variables to the Maxent model is shown in Appendix S3. Variables related to mean temperature and precipitation in the summer (such as precipitation of the warmest quarter, annual temperature range or mean diurnal range) had a prominent role in the model estimated by Maxent and thus seem to represent environmental factors affecting the distribution of both *Pyrgus* species. However, *P. sidae* was more influenced by the annual temperature, and *P. cinarae* by the availability of water in summer dry periods. Jackknife tests of variable importance indicated that annual range of temperature in the case of *P. sidae* and precipitation in the warmest quarter for *P. cinarae* had the highest gain when

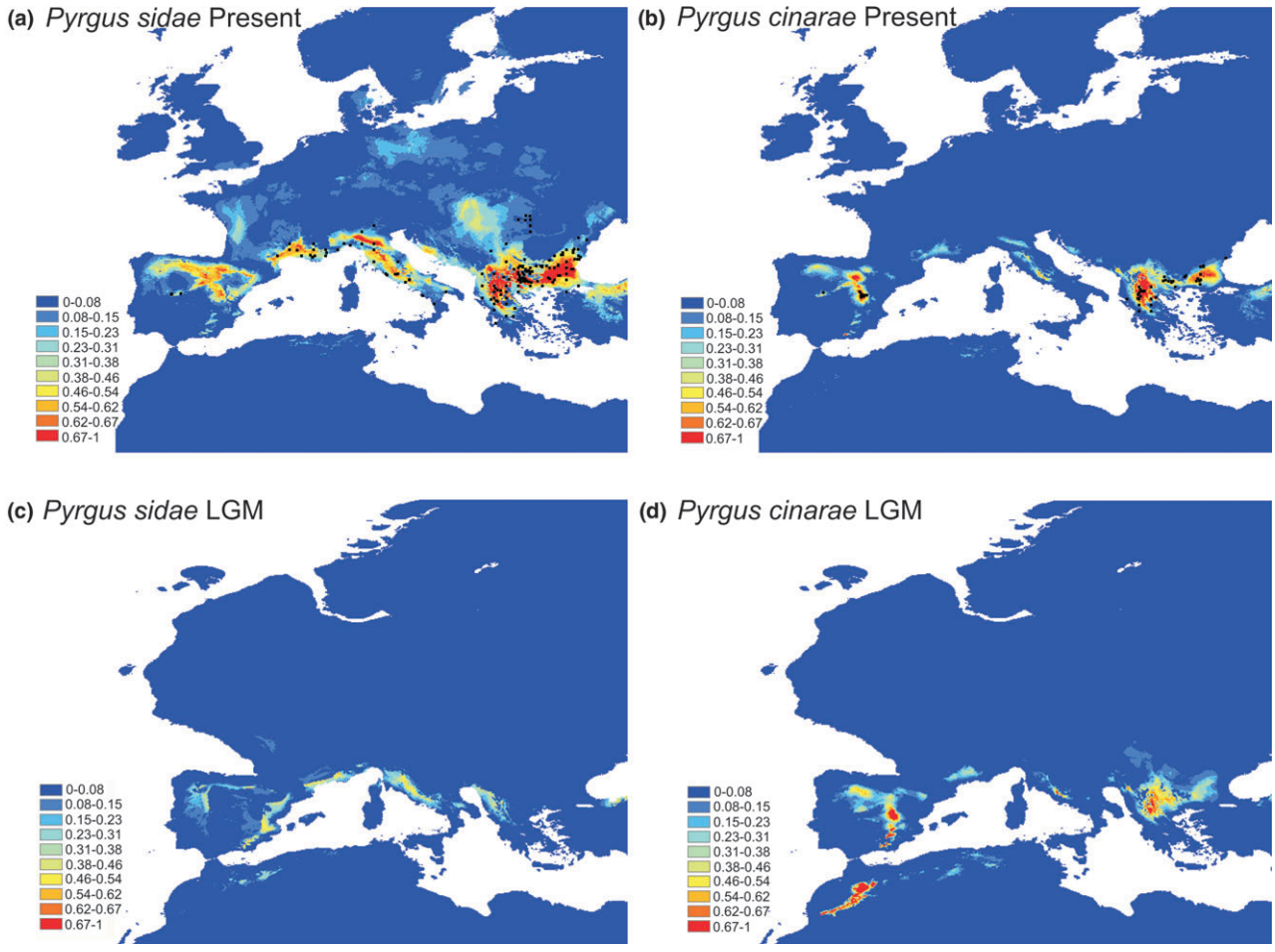

Figure 3 Potential distributions of *Pyrgus sidae* and *P. cinarae* in Europe obtained with Maxent based on current climatic data (a, b), and on the Last Glacial Maximum palaeoclimatic data (c, d). The most probable areas are shown in warm colours. The current known distribution of both species is represented in black.

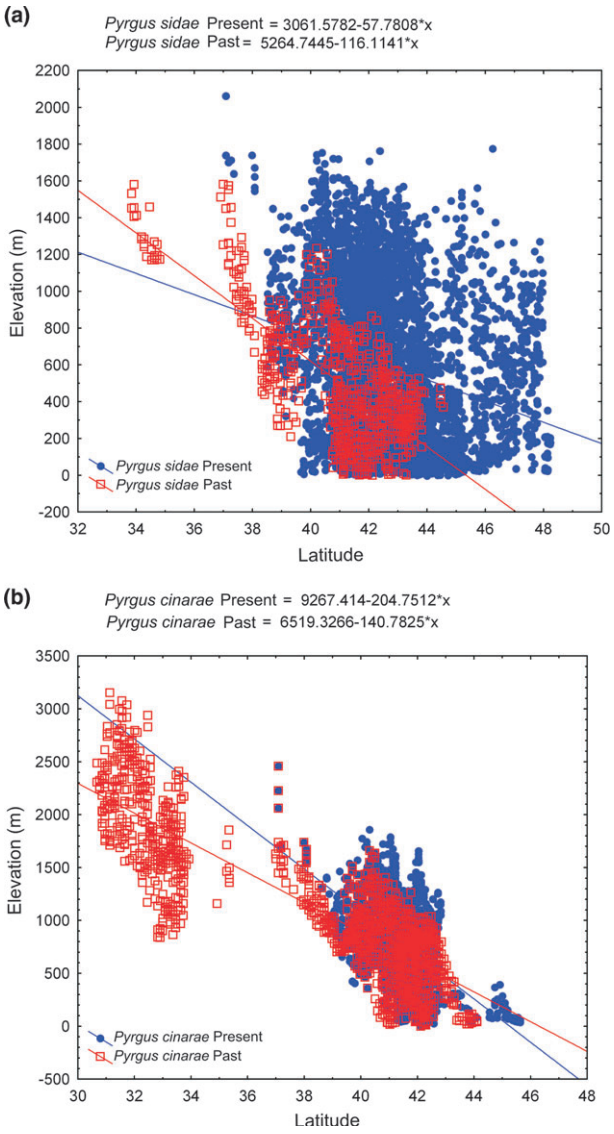

Figure 4 Elevation and latitude values of the predicted areas obtained with MAXENT at present (in blue) and during the Last Glacial Maximum (in red) for (a) *Pyrgus sidae* and (b) *P. cinarae*.

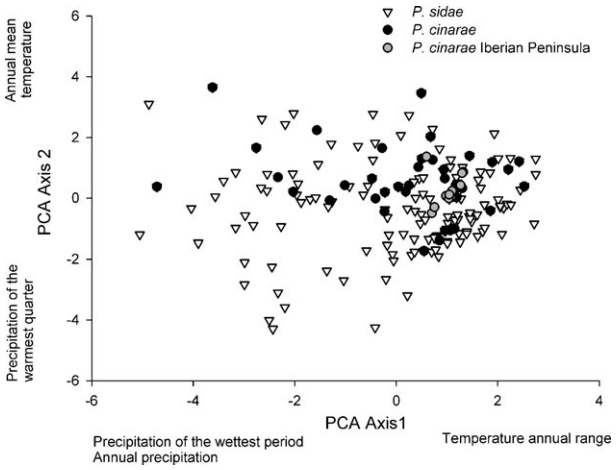

Figure 5 Results of the principal components analysis on environmental descriptors for *Pyrgus sidae* and the two main clades of *P. cinarae*. Environmental descriptors with the highest positive and negative correlations with species scores are indicated on the axes (see details of correlations in Appendix S4).

used in isolation, suggesting that these variables contained the most useful information, and also decreased the gain the most when omitted, suggesting that they contained the most information not present in the other variables.

A comparison of modern to LGM distributions (considering only minimum areas under both the CCSM and MIROC climatic models), revealed that climatically suitable areas have increased for *P. sidae* since the LGM. Its distribution now extends into the Balkan Peninsula (Fig. 3a,c), while suitable areas were previously much smaller, restricted to low elevations and latitudes (Fig. 4a). By contrast, *P. cinarae* encountered similarly favourable areas in both periods (Fig. 3b,d). Therefore, the distribution of *P. sidae* apparently increased in the interglacial periods while *P. cinarae* has been generally stable. Interestingly, connectivity between suitable areas remained quite stable for both species, at least along the coast in the case of *P. sidae*. In both cases there was no continuous suitable habitat connecting present-day populations, although connectivity was always much lower in *P. cinarae*. For both species, the existence of suitable habitat in North African mountains during the LGM was predicted, although these areas are not suitable for the species at present (Fig. 4). Thus, North Africa could have represented a glacial refugium for both species, although it is unlikely that they colonized this region because of their aversion to dispersal across water.

In the PCA, axes 1 and 2 explained 36.72% and 24.11%, respectively, of the total ecological niche variation. Axis 1 was correlated positively with annual temperature range and negatively with precipitation of the wettest period and annual precipitation (Appendix S4). Axis 2 was correlated with average annual temperature (positive values) and with rain in the warmest quarter (negative values). The two-dimensional niche defined by the PCA axes was significantly different among species (Wilks' $\lambda = 0.932$, d.f. = 474, $P < 0.05$). *Pyrgus sidae* occupied the largest ecological space and *P. cinarae*'s preferences were slightly more restricted, although the two species broadly overlapped. The Iberian *P. cinarae* populations displayed a rather small ecological space, clearly nested within that of the rest of *P. cinarae* populations (Fig. 5). Newman–Keuls post hoc tests only indicated niche differentiation for PCA axis 1 for Iberian specimens of *P. cinarae* with *P. sidae* ($P = 0.03$), but not between the remaining populations of *P. cinarae* with *P. sidae* ($P = 0.27$) or between the two main *P. cinarae* clades ($P = 0.14$). Any comparison showed non-significant niche differentiation on PCA axis 2 ($P > 0.05$).

DISCUSSION

Molecular data, both COI and ITS2 sequences, point to an old origin for the disjunct distribution of *P. cinarae*. This is

suggested by the topology of the phylogenetic trees and haplotype networks (Fig. 1b), which recover the Iberian population as sister to the rest, by the high divergences separating these two main clades (17 substitutions, 2.6% in COI; 2 substitutions, 0.03% in ITS2), and by the higher pairwise D_{xy} and Φ_{ST} values for *P. cinarae* than for *P. sidae* (Table 2). Age estimates based on MDIV analysis situate the origin of this disjunct distribution at 1.1 Ma (0.4–2.2 Ma), with a TMRCA of 1.9 Ma (1.3–2.5 Ma). Despite the potential error involved in coalescent age estimates, we can safely state that the disjunct distribution in *P. cinarae* is much older than that of *P. sidae*, and that it dates back to the initial Pleistocene glaciations. We can thus infer that during the latest several glacial and interglacial periods there has been no gene flow between the two disjunct groups of populations of *P. cinarae*, or at least it has left no signal in current genetic structure. Indeed, the GMYC analysis (Appendix S2) confirms that the *P. cinarae* Iberian isolate is an independent evolutionary lineage. This scenario is corroborated by distribution modelling results, which show that during the latest glacial to interglacial period the distribution of *P. cinarae* in Europe has been largely unaffected, without major elevational or latitudinal shifts (Fig. 4b). Such a distributional stasis (no important changes in distribution along time) at a general scale suggests that *P. cinarae*, unlike *P. sidae*, does not have the capacity to expand its distribution range to new habitats during interglacial periods. Interestingly, the potential present-day distribution shows that no substantial suitable habitat exists along the 1800 km separating Iberian and Balkan populations (Fig. 3b). The origin of the *P. cinarae* disjunct distribution is probably related to more general climatic trends that occurred during the last several million years, and has been maintained by distributional stasis during more recent glacial to interglacial fluctuations. Nevertheless, at a local scale, glacial cycles possibly had variable demographic and distributional effects on populations of this species. This seems true at least for the Iberian population, which is distributed in a very limited area and displays an extremely low genetic variability (a single COI haplotype and three haplotypes of ITS2 were detected). These particularities, coupled with evidence for demographic expansion in the species as a whole provided by the coalescence analysis, suggest that a recent bottleneck occurred in the Iberian Peninsula. Interestingly, the Iberian Peninsula is the only region where a substantial mismatch between potential and realized distribution is observed, suggesting that the capacity of these populations to expand may be hindered by current low genetic variability and population densities.

Genetic data indicate a much more recent origin for the observed distributional pattern of *P. sidae* compared with that of *P. cinarae*. All COI analyses recovered three main clades (west, central and east) that have low divergences between them (they only differ by single changes). The nuclear marker ITS2 was less informative and did not recover any clear pattern consistent with geographical distance for *P. sidae*, which could be interpreted as a lack of resolution of this marker or as the existence of recent gene flow between populations. The Iberian *P. sidae* population is embedded within the western clade, which also includes south France and central Italy. The MDIV age estimate for the divergence between the western and central clades is *c.* 0.27 Ma (0.18–0.29 Ma). Thus, we can safely assume that the 1000 km disjunct Iberian population originated earlier than this date, most probably during one of the latest glacial events. Given the COI divergences between Iberian and non-Iberian populations of *P. sidae* (one substitution, 0.15%) and of *P. cinarae* (17 substitutions, 2.60%), we believe that the difference in age estimates obtained by MDIV (less than 0.18–0.29 Ma and between 0.38 and 2.17 Ma, respectively) are reasonable despite uncertainty involved in coalescent age estimates.

Isolation since the LGM has been invoked as the cause for many cases of disjunct distributions in Europe and elsewhere (Avise, 2000; Hewitt, 2000). More generally, this has also been suggested to be the cause for patterns of population genetic structure, regardless of the existence of distributional discontinuity at present (Schmitt & Seitz, 2001, 2002; Schmitt & Krauss, 2004; Schmitt *et al.*, 2005; Schmitt, 2007). As the term 'disjunct distribution' can be applied at very different spatial scales, it has been used from rather local studies (Williams, 1980; Flinn *et al.*, 2010) to discontinuities of thousands of kilometres (Wahlberg & Saccheri, 2007; Garcia Collevatti *et al.*, 2009; Cagnon & Turgeon, 2010; Peña *et al.*, 2010). We could say that the case of *P. sidae* fits the conventional model of a species that suffers a marked distributional shift southwards and to lower elevations, a demographic reduction, and retains viable populations only in Mediterranean peninsulas that act as refugia during the glacial periods. During the interglacial periods, this species would expand to the European mainland, with the possibility of re-establishing gene flow between isolates, given enough time. In this case, the disjunct distribution is the product of the LGM because gene flow was not re-established. However, our results for *P. cinarae* show that the origin of long-distance disjunct distributions might not always be a direct consequence of one of the latest glacial periods, even in the case of a current distribution in known glacial refugia such as the Mediterranean peninsulas. This species, despite apparent similarity to *P. sidae*, displays unique particularities. One of them is an apparent stasis in the morphology and ecology after *c.* 1 million years of isolation of the Iberian population. *Pyrgus cinarae* also displays a surprising stasis in its potential distribution despite climatic oscillations. Being restricted to habitats not strongly affected by climatic oscillations could be the product of an unusually limited dispersal capability, which would account for the inability to have expanded during interglacial periods, producing a long-term disjunct distribution. If this hypothesis is correct, the predicted dispersal capability of *P. cinarae* should be substantially lower than that of *P. sidae*, which is difficult to imagine because a capture–mark–recapture study on Iberian *P. sidae* showed that this is already a typical sedentary species with low dispersal capabilities (Hernández-Roldán *et al.*, 2009).

In conclusion, we show that *P. cinarae* and *P. sidae* display different population genetic structures and ancestral potential distributions, which reveal that they have undergone different biogeographical histories. Indeed, the effect on these two

species of the last glacial to post-glacial period was radically different according to our results: while the distribution of *P. cinarae* was not substantially affected in Europe, that of *P. sidae* greatly changed in both latitude and elevation. Remarkably, the two dissimilar biogeographical histories resulted in very similar distributions across the Palaearctic at present, including a disjunct distribution with isolated populations on the Iberian Peninsula. This could thus be considered a case of convergence in biogeography, even more so when the two species involved belong to the same genus, frequently coexist in the same habitat, share host plants, are univoltine and are usually synchronic. Why then have they not followed parallel biogeographical histories? We demonstrate general ecological similarities, because their ecological niche space broadly overlaps in a PCA analysis. However, differences exist that are not readily obvious without conducting an ecological niche modelling exercise followed by PCA. We show that the rather subtle particularities of each species in their niche preferences result in a characteristic response to environmental change that subsequently determines population genetic structure. This case illustrates how minor ecological differences may lead to very different biogeographical histories and highlights the important role that ecology has played during the evolutionary history of species. By combining molecular data and ecological niche modelling it is possible to reconstruct and understand the history of a species to a fine level. Interpreting biogeographical patterns is nevertheless a complex exercise and the existence of convergence in biogeography should be a warning to avoid generalizations and not to extrapolate results for a taxon to other species, even if they are apparently equivalent from an ecological perspective. Finally, our results highlight the importance of integrating the spatial and temporal dimensions in biogeography.

ACKNOWLEDGEMENTS

We thank Ángel Blázquez, Ernst Brockmann, Vlad Dinca, Zdravko Kolev, Tristan Lafranchis, Vladimir Lukhtanov, Santi Viader, Juan Carlos Vicente and Wolfgang Wagner for help in collecting material. We are also grateful to John G. Coutsis for providing distribution records for both species in Greece, to Joan García-Porta for his very insightful comments regarding ecological niche modelling, and to Blanca Huertas for access to the collection of *Pyrgus* in the Natural History Museum, London. Support for this research was provided by the Spanish Ministerio de Ciencia e Innovación (projects CGL2007-60516/BOS and CGL2010-21226/BOS to J.L.H.-R., G.T. and R.V.; and fellowship BES-2008-002054 to G.T.) and by Genome Canada through the Ontario Genomics Institute to P.D.N.H.

REFERENCES

Abadjiev, S. (2001) *An atlas of the distribution of the butterflies in Bulgaria (Lepidoptera: Hesperioidea & Papilionoidea)*. Pensoft Publishers, Sofia-Moscow.

Alba-Sánchez, F., López-Sáez, J.A., Benito-de Pando, B., Linares, J.C., Nieto-Lugilde, D. & López-Merino, L. (2010) Past and present potential distribution of the Iberian *Abies* species: a phytogeographic approach using fossil pollen data and species distribution models. *Diversity and Distributions*, **16,** 214–228.

Alsos, I.G., Alm, T., Normand, S. & Brochmann, C. (2009) Past and future range shifts and loss of diversity in dwarf willow (*Salix herbacea* L.) inferred from genetics, fossils and modelling. *Global Ecology and Biogeography*, **18,** 223–239.

Avise, J.C. (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, MA.

Bocxlaer, I.V., Roelants, K., Biju, S.D., Nagaraju, J. & Bossuyt, F. (2006) Late Cretaceous vicariance in Gondwana amphibians. *PLoS ONE*, **1,** e74.

Cagnon, M.C. & Turgeon, J. (2010) Disjunct distribution in *Gerris* species (Insecta: Hemiptera: Gerridae): an analysis based on spatial and taxonomic patterns of genetic diversity. *Journal of Biogeography*, **37,** 170–178.

Clement, M., Posada, D. & Crandall, K. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9,** 1657–1660.

Cox, C.B. & Moore, P.D. (2005) *Biogeography: an ecological and evolutionary approach*, 7th edn. Blackwell Publishing, Oxford.

Davis, M.B. & Shaw, R.G. (2001) Range shifts and adaptive responses to Quaternary climate change. *Science*, **292,** 673–679.

Donoghue, M.J. & Moore, B.R. (2003) Toward an integrative historical biogeography. *Integrative and Comparative Biology*, **43,** 261–270.

Drummond, A.J. & Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7,** 214.

Elith, J. (2002) Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. *Quantitative methods for conservation biology* (ed. by S. Ferson and M. Burgman), pp. 39–58. Springer, New York.

Elith, J., Graham, C.H., Anderson, R. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29,** 129–151.

Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1,** 47–50.

Fitzpatrick, B.M. (2009) Power and sample size for nested analysis of molecular variance. *Molecular Ecology*, **18,** 3961–3966.

Flinn, K.M., Waterway, M.J. & Lechowicz, M.J. (2010) Disjunct performance and distribution in the sedge *Carex prasina*. *Oecologia*, **163,** 119–126.

Fontaneto, D., Herniou, E.A., Boschetti, C., Caprioli, M., Melone, G., Ricci, C. & Barraclough, T.G. (2007) Independently evolving species in asexual bdelloid rotifers. *PLoS Biology*, **5,** 914–921.

Garcia Collevatti, R., Gonçalves Rabelo, S. & Vieira, R.F. (2009) Phylogeography and disjunct distribution in *Lychnophora ericoides* (Asteraceae), an endangered cerrado shrub species. *Annals of Botany*, **104,** 655–664.

García-Barros, E., Munguira, M.L., Martín, J., Romo, H., García-Pereira, P. & Maravalhas, E.S. (2004) *Atlas de las mariposas diurnas de la Península Ibérica y Baleares (Lepidoptera: Papilionoidea & Hesperioidea)*. Monografías SEA, Vol. 11. Zaragoza.

Guisan, A. & Zimmerman, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135,** 147–186.

Habel, J.C., Schmitt, T. & Müller, P. (2005) The fourth paradigm pattern of postglacial range expansion of European terrestrial species: the phylogeography of the marbled white butterfly (Satyrinae, Lepidoptera). *Journal of Biogeography*, **32,** 1489–1497.

Hajibabaei, M., deWaard, J.R., Ivanova, N.V., Ratnasingham, S., Dooh, R.T., Kirk, S.L., Mackie, P.M. & Hebert, P.D.N. (2005) Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360,** 1959–1967.

Hernández-Roldán, J.L., Munguira, M.L. & Martín, J. (2009) Ecology of a relict population of the vulnerable butterfly *Pyrgus sidae* on the Iberian Peninsula (Lepidoptera: Hesperiidae). *European Journal of Entomology*, **106,** 611–618.

Hewitt, G.M. (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58,** 247–276.

Hewitt, G.M. (1999) Post-glacial re-colonization of European biota. *Molecular Genetics in Animal Ecology*, **68,** 87–112.

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

Hewitt, G.M. (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359,** 183–195.

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

Hudson, R. (2000) A new statistic for detecting genetic differentiation. *Genetics*, **155,** 2011–2014.

Hunn, C.A. & Upchurch, P. (2001) The importance of time/space in diagnosing the causality of phylogenetic events: towards a 'chronobiogeographical' paradigm? *Systematic Biology*, **50,** 1–17.

Ivanova, N.V., deWaard, J.R. & Hebert, P.D.N. (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes*, **6,** 998–1002.

Jakob, S.S., Ihlow, A. & Blattner, F.R. (2007) Combined ecological niche modelling and molecular phylogeography revealed the evolutionary history of *Hordeum marinum* (Poaceae) – niche differentiation, loss of genetic diversity, and speciation in Mediterranean Quaternary refugia. *Molecular Ecology*, **16,** 1713–1727.

Kudrna, O. (2002) The distribution atlas of European butterflies. *Oedippus*, **20,** 1–342.

Loader, S.P., Pisani, D., Cotton, J.A., Gower, D.J., Day, J.J. & Wilkinson, M. (2007) Relative timescales reveal multiple origins of parallel disjunct distributions of African caecilian amphibians. *Biology Letters*, **3,** 505–508.

Moffett, A., Shackelford, N. & Sarkar, S. (2007) Malaria in Africa: vector species' niche models and relative risk maps. *PLoS ONE*, **2,** e824.

Nei, M. (1987) *Molecular evolutionary genetics*. Columbia University Press, New York.

Newbold, T. (2009) *The value of species distribution models as a tool for conservation and ecology in Egypt and Britain*. PhD Thesis, University of Nottingham, UK.

Newbold, T., Gilbert, F., Zalat, S., El-Gabbas, A. & Reader, T. (2009) Climate-based models of spatial patterns of species richness in Egypt's butterfly and mammal fauna. *Journal of Biogeography*, **36,** 2085–2095.

Nielsen, R. & Wakeley, J. (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158,** 885–896.

Nix, H.A. (1986) A biogeographic analysis of Australian elapid snakes. *Atlas of elapid snakes of Australia* (ed. by R. Longmore), pp. 4–15. Australian Government Publishing Service, Canberra.

Noonan, B.P. & Chippindale, P.T. (2006) Vicariant origin of Malagasy reptiles supports Late Cretaceous Antarctic land bridge. *The American Naturalist*, **168,** 730–741.

Pawar, S., Koo, M.S., Kelley, C., Ahmed, F.M., Choudhury, S. & Sarkar, S. (2007) Conservation assessment and prioritization of areas in northeast India: priorities for amphibians and reptiles. *Biological Conservation*, **136,** 346–361.

Pearce, J. & Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, **133,** 225–245.

Peña, C., Nylin, S., Freitas, A.V.L. & Wahlberg, N. (2010) Biogeographic history of the butterfly subtribe *Euptychiina* (Lepidoptera, Nymphalidae, Satyrinae). *Zoologica Scripta*, **39,** 243–258.

Phillips, S.J. (2008) Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson *et al.* (2007). *Ecography*, **31,** 272–278.

Phillips, S.J. & Dudík, M. (2008) Modelling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31,** 161–175.

Phillips, S.J., Dudík, M. & Schapire, R.E. (2004) A maximum entropy approach to species distribution modelling. *Proceedings of the 21st International Conference on Machine Learning*, pp. 655–662. ACM Press, New York.

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

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D. & Vogler, A.P. (2006) Sequence-based species delimitation for the

DNA taxonomy of undescribed insects. *Systematic Biology*, **55,** 595–609.

Posada, D. (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25,** 1253–1256.

Quek, S.P., Davies, S.J., Itino, T. & Pierce, N.E. (2004) Codiversification in an ant-plant mutualism: stem texture and the evolution of host use in *Crematogaster* (Formicidae: Myrmicinae) inhabitants of macaranga (Euphorbiaceae). *Evolution*, **58,** 554–570.

Ramos-Onsins, S.E. & Rozas, J. (2002) Statistical properties of new neutrality test against population growth. *Molecular Biology and Evolution*, **19,** 2092–2100.

Rissler, L.J. & Apodaca, J.J. (2007) Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). *Systematic Biology*, **56,** 924–942.

Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X. & Rozas, R. (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19,** 2496–2497.

Sanderson, M.J. (2003) r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*, **19,** 301–302.

Schmitt, T. (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology*, **4,** 11.

Schmitt, T. & Hewitt, G.M. (2003) Molecular biogeography of the arctic-alpine disjunct burnet moth species *Zygaena exulans* (Zygaenidae, Lepidoptera) in the Pyrenees and Alps. *Journal of Biogeography*, **31,** 885–893.

Schmitt, T. & Krauss, J. (2004) Reconstruction of the colonization route from glacial refugium to the northern distribution range of the European butterfly *Polyommatus coridon* (Lepidoptera: Lycaenidae). *Diversity and Distributions*, **10,** 271–274.

Schmitt, T. & Seitz, A. (2001) Allozyme variation in *Polyommatus coridon* (Lepidoptera: Lycaenidae): identification of ice-age refugia and reconstruction of post-glacial expansion. *Journal of Biogeography*, **28,** 1129–1136.

Schmitt, T. & Seitz, A. (2002) Postglacial distribution area expansion of *Polyommatus coridon* (Lepidoptera: Lycaenidae) from its Ponto-Mediterranean glacial refugium. *Heredity*, **89,** 20–26.

Schmitt, T., Varga, Z. & Seitz, A. (2005) Are *Polyommatus hispana* and *Polyommatus slovacus* bivoltine *Polyommatus coridon* (Lepidoptera: Lycaenidae)? The discriminatory value of genetics in taxonomy. *Organisms Diversity & Evolution*, **5,** 297–307.

Schmitt, T., Hewitt, G.M. & Müller, P. (2006) Disjunct distributions during glacial and interglacial periods in mountain butterflies: *Erebia epiphron* as an example. *Journal of Evolutionary Biology*, **19,** 108–113.

Stamatakis, A. (2006) Raxml-vi-hpc: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22,** 2688–2690.

StatSoft, Inc. (2004) *STATISTICA (data analysis software system), version 7*. StatSoft, Inc., Tulsa, OK. Available at: http://www.statsoft.com.

Stehlik, I., Holderegger, R., Schneller, J.J., Abbott, R.J. & Bachmann, K. (2000) Molecular biogeography and population genetics of alpine plant species. *Bulletin of the Geobotanical Institute ETH*, **66,** 47–59.

Stockwell, D.R.B. & Peters, D. (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical information Science*, **13,** 143–158.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.G. & Cosson, J.E. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Evolution*, **7,** 453–464.

Templeton, A. & Sing, C. (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics*, **134,** 659–669.

deWaard, J.R., Ivanova, N.V., Hajibabaei, M. & Hebert, P.D.N. (2008) Assembling DNA barcodes: analytical protocols. *Methods in molecular biology: environmental genetics* (ed. by C. Martin), pp. 275–293. Humana Press, Totowa, NJ.

Wagner, W. (2009) Zur Ökologie von *Pyrgus cinarae* (Rambur 1839) nebst Beobachtungen zu *Spialia phlomidis* (Herrich-Schäffer 1845) – Larvalhabitat, Präimaginalstadien und Entwicklungszyklus (Lepidoptera: Hesperiidae). *Nachrichten des Entomologischen Vereins Apollo*, **29,** 199–204.

Wahlberg, N. & Saccheri, I. (2007) The effects of Pleistocene glaciations on the phylogeography of *Melitaea cinxia* (Lepidoptera: Nymphalidae). *European Journal of Entomology*, **104,** 675–684.

Weaver, K.F., Anderson, T. & Guralnick, R. (2006) Combining phylogenetic and ecological niche modeling approaches to determine distribution and historical biogeography of Black Hills mountain snails (Oreohelicidae). *Diversity and Distributions*, **12,** 756–766.

White, T.J., Bruns, T., Lee, S. & Taylor, J. (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications* (ed. by M.A. Innis, D.H. Gelfand, J.J. Sninsky and T.J. White), pp. 315–322. Academic Press, New York.

Williams, E.H. (1980) Disjunct distributions of two aquatic predators. *Limnology and Oceanography*, **25,** 999–1006.

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H. & Guisan, A. (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, **14,** 763–773.

SUPPORTING INFORMATION

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

Appendix S1 Details of *Pyrgus* spp. and sequences used in the present study.
Appendix S2 GMYC model applied to a maximum likelihood tree of cytochrome *c* oxidase subunit 1 (COI) haplotypes for *Pyrgus sidae* and *P. cinarae*.
Appendix S3 Heuristic estimate of relative contributions of the environmental variables to the MAXENT model in *Pyrgus sidae* and *P. cinarae*.
Appendix S4 Pearson correlations between species scores on principal components analysis (PCA) axes and environmental descriptors.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

BIOSKETCH

The Butterfly Diversity and Evolution Lab at the Institut de Biologia Evolutiva (CSIC-UPF) combines molecular, morphological and ecological data to study the biodiversity, biogeography and evolution of Lepidoptera. The Biodiversity Institute of Ontario (University of Guelph) is involved in the oversight of a large-scale programme on DNA barcoding, which employs this approach to probe biodiversity in varied groups of eukaryotes.

Author contributions: J.L.H.-R. and R.V. designed the research and obtained the specimens; J.L.H.-R., G.T. and E.Z. obtained the molecular data; G.T. performed phylogenetic analyses; C.M. performed population genetic, GMYC and principal components analyses; H.R. performed ecological niche modelling analyses based on distribution data compiled by J.L.H.-R. All authors contributed to discussing the results and to writing the paper.

Editor: Brett Riddle