



## RESEARCH ARTICLE

# Inferring past refugia and range dynamics through the integration of fossil, niche modelling and genomic data

Frederico Mestre<sup>1,2</sup> | Soraia Barbosa<sup>3,4</sup> | José A. Garrido-García<sup>5,6</sup> |  
Ricardo Pita<sup>2,7</sup> | António Mira<sup>2,7</sup> | Paulo C. Alves<sup>3,4,8,9</sup> | Joana Paupério<sup>3</sup> |  
Jeremy B. Searle<sup>3,10</sup> | Pedro Beja<sup>3,4,11</sup>

<sup>1</sup>Rui Nabeiro' Biodiversity Chair, Rua Dr. Joaquim Henrique da Fonseca, Évora, Portugal

<sup>2</sup>MED - Mediterranean Institute for Agriculture, Environment and Development & CHANGE - Global Change and Sustainability Institute, Institute for Advanced Studies and Research, Évora, Portugal

<sup>3</sup>CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, Vairão, Portugal

<sup>4</sup>BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Vairão, Portugal

<sup>5</sup>Estación Biológica de Doñana, Consejo Superior de Investigaciones Científicas (CSIC), Sevilla, Andalusia, Spain

<sup>6</sup>Estación Paleontológica Valle del Rio Fardes, Centro Nacional Instituto Geológico y Minero de España (CSIC-IGME), Granada, Spain

<sup>7</sup>Unidade de Biologia da Conservação, Departamento de Biologia, Escola de Ciências e Tecnologia, Universidade de Évora, Évora, Portugal

<sup>8</sup>Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal

<sup>9</sup>Estação Biológica de Mértola (EBM), Praça Luís de Camões, Mértola, Portugal

<sup>10</sup>Department of Ecology and Evolutionary Biology, Corson Hall, Cornell University, Ithaca, New York, USA

<sup>11</sup>CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Instituto Superior de Agronomia, Universidade de Lisboa, Lisbon, Portugal

## Correspondence

Soraia Barbosa, CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal.  
Email: [soraia Barbosa@cibio.up.pt](mailto:soraia Barbosa@cibio.up.pt)

Handling Editor: Anna Papadopoulou

## Abstract

**Aim:** Reconstructing species' glacial refugial history and demographic changes over time has greatly relied on comparing inferences from multiple methods while not sufficiently acknowledging their limitations. Here, we aim to integrate as fully as possible complementary methods in ecology, genomics and palaeobiology to improve the reconstruction of species biogeographical history.

**Location:** South-western Europe.

**Taxon:** Cabrera vole (*Microtus cabreræ*).

**Methods:** We compiled and mapped the raw fossil data of the Cabrera vole for the Last Glacial Maximum (LGM) and Mid-Holocene (MH). Alongside, we projected the calibrated ecological niche model (ENM) of the species' current distribution for the LGM, the Younger Dryas and the MH. Complementarily, we used previously obtained Genotyping-by-Sequencing data to evaluate the demographic history and range expansion patterns of all four Evolutionarily Significant Units of the species, in an integrative framework.

Frederico Mestre and Soraia Barbosa contributed equally to the manuscript.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Journal of Biogeography* published by John Wiley & Sons Ltd.



**Results:** ENM-inferred refugial areas and genomic modelling consistently supported northern Iberian glacial refugia for the Cabrera vole. This contrasted with the higher fossil abundance of the species in southern and eastern Iberia and southern France from the LGM to the MH. Our results suggest that populations in areas with high fossil abundance went extinct, and were later replaced by northern Iberian populations such that they did not contribute significantly to the current gene pool.

**Main conclusions:** Our integrative approach indicates how the range of the Cabrera vole fluctuated in response to environmental change during and following the LGM. Despite methodological limitations, the ENM and genomic approaches produced generally congruent results. Instead, the fossil record may misrepresent the ancestral distribution of this species and should be considered cautiously for ancestral distribution reconstruction, considering that it also reflects the fossilization conditions. Overall, our study supports the idea that integrative approaches are essential to provide an accurate and well-supported picture of historical refugial areas and range dynamics.

#### KEYWORDS

demographic inference, ecological niche modelling, fossilization, hindcasting, palaeoclimate, palaeo-distribution, range expansion

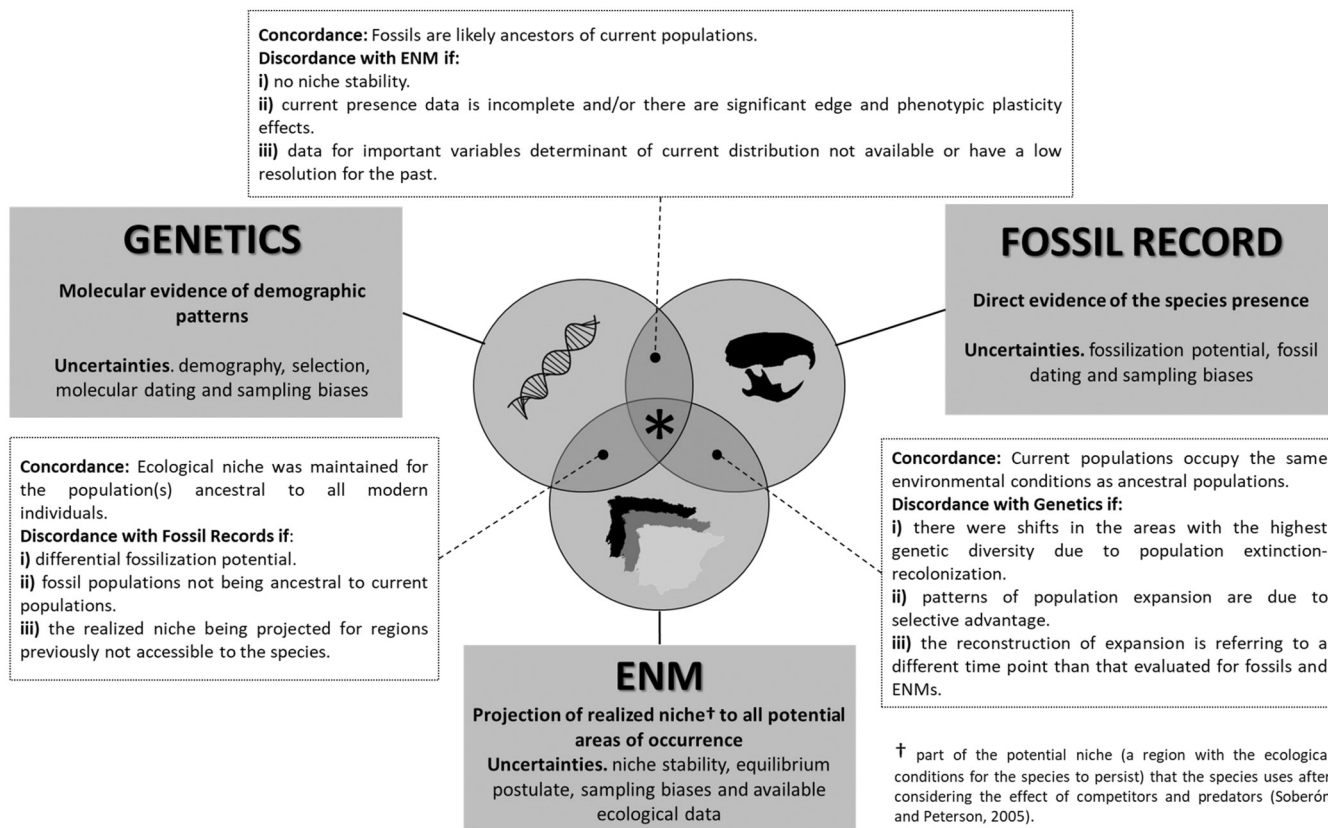
## 1 | INTRODUCTION

The reconstruction of glacial refugial history and distribution changes of species over time contributes to our understanding of the dynamics of species' responses to environmental change and can shed light on evolutionary capacity (Svenning et al., 2011). Species generally respond to changes in climate with contractions and expansions of their distribution ranges, where contraction periods typically involve population isolation, potentially resulting in differentiation or extinction, while expansion periods typically involve population homogenization or the formation of secondary contact zones (Hewitt, 2011a; Taberlet & Cheddadi, 2002). Range contractions and associated fragmentation processes in relation to the last glaciation are very well documented in temperate species in Europe (Hewitt, 2011b). Within refugial areas, populations can be further isolated in different refugia within the glacial refugium (Gómez & Lunt, 2007; Hewitt, 2000). While in isolation, populations can diverge and become locally adapted (Davis & Shaw, 2001; de Lafontaine et al., 2018). Therefore, the ecological niche of a species can also change over time and vary depending on different selective pressures across multiple within-refugial areas (Abellán & Svenning, 2014; Hewitt, 1996; Marske et al., 2013). After periods of contraction, generally experienced by temperate species during glacial maxima, surviving populations usually expand out of their refugia, as previously occupied areas become available with the improving conditions (Hewitt, 2000). The importance of local adaptation for species future distribution is clearly shown in forecasting studies, which have found significant differences in the predicted ranges of adaptively differentiated populations as they are expected to expand into new areas (Razgour et al., 2019). Thus, whether populations withstand the impacts of new environmental conditions will

be a function of genetic factors (e.g. genetic drift), biological factors (e.g. demographic stochasticity) and ecological factors (e.g. biotic interactions), which makes local extirpation a common phenomenon (Lasky et al., 2020).

Determining the geographical location of glacial/interglacial refugia and the range changes during and after glaciations is especially important in understanding a species' ability to respond to environmental change and to evaluate their adaptive potential (Anadón et al., 2015; Razgour et al., 2019). Inferences on species' refugial history and diversification have mostly been based on the existence of fossil records, on projecting the species' current niche to past climatic conditions, or on geographical tracing of genetic lineages including analysis of genetic diversity and demography (Gavin et al., 2014; Sommer & Nadachowski, 2006; Sommer & Zachos, 2009). However, no single method is likely to provide the complete history of the species distribution changes (e.g. Fernandez et al., 2021; Luna-Aranguré et al., 2020; Figure 1). More desirably, integrative approaches incorporate multiple lines of evidence in the same analysis to generate a result that best fits all input information (Fenderson et al., 2020). With multidisciplinary approaches integrating fossil record, ecological niche modelling and genetics, common patterns reflecting a closer approximation to the species' biogeographical history may emerge (Hoban et al., 2019).

The fossil record represents strong evidence of the species' presence in a given location in the past, and for that reason, it is frequently used to comparatively validate palaeo-distributions inferred from other approaches (Martínez-Meyer et al., 2004). However, the fossil record is constrained by several factors, such as (i) differential detectability due to variations in population density; (ii) differing propensity for fossilization, which can lead to biases in fossil abundance; (iii) dating accuracy or ability to identify fossil remains; (iv) sampling bias favouring particular regions, and temporal



**FIGURE 1** Examination of the relative merits and potential shortcomings of the fossil record, ecological niche modelling (ENM), and genetics in biogeographical interpretation. At each intersection of two lines of evidence, we describe the factors that might cause them to agree in their spatial predictions and how they may disagree with the predictions of the third line of evidence. The intersection of the three lines of evidence (\*) would result in the best prediction possible

or geographical scenarios or taxa; and (v) displacement from their point of origin (e.g. through water and air, which is likely important for plants or aquatic animals, or through predator-induced movements, for prey species, such as small mammals; Allison et al., 2011; Gavin et al., 2014; Lobo et al., 2010; Varela et al., 2011). Differences in the fossilization potential can affect the perception of abundances (Baker & Worley, 2014; Mitchell, 2015), which can be critical for the interpretation of refugial locations and their assumed contribution to the current populations, and also might explain possible discrepancies between the fossil record and other lines of evidence (Figure 1). Moreover, fossils can represent populations that have been replaced and, consequently, might not be ancestral to those found today (Svenning et al., 2011).

Ecological niche models (ENMs) relate the species distribution with the environmental conditions and can be used to infer the species past distribution range (Guisan & Zimmermann, 2000; Peterson, 2011) while relying on the following assumptions: (i) the species occurs wherever it has suitable conditions and it is absent from unsuitable regions (*equilibrium postulate*); (ii) the niche is maintained across space and time (*niche stability*); and (iii) all variables that significantly impact species distributions are considered in the models (e.g. Elith & Leathwick, 2009; Randin et al., 2006). While these assumptions might be acceptable for time periods up to nearly 22,000 years ago (ka BP; Svenning et al., 2011), different populations

of a species may adapt locally over time, or rapidly through selective sweeps, resulting in intraspecific niche variation as a consequence of the interaction between dispersal, selection and demographic asymmetries (Messer & Petrov, 2013; Pearman et al., 2010; Peterson & Holt, 2003; Voje, 2020). An additional source of uncertainty when projecting an ENM to a different period is the quality and variety of the Global Climate Models (GCMs; Beaumont et al., 2008). Thus, integration of the ENM with other sources of information is recommended to evaluate the accuracy of projections on geographical space (Hoban et al., 2019).

Patterns of current genetic/genomic diversity and differentiation add an additional line of evidence to infer demographic change and locations of refugia (Bertorelle et al., 2010; Csilléry et al., 2010; Pfenninger & Posada, 2002). To infer species demographic history, methods often use coalescence-based simulations to compare user-defined demographic models and determine patterns and timing of population divergence and merging (Excoffier et al., 2021). However, these methods are not geographically explicit, and thus do not allow glacial refugial areas to be determined. In contrast, range expansion analyses can unveil the genomic patterns resulting from serial founder events to trace back the geographical origin of population expansions, but are unable to infer their timing and demographic fluctuations (Peter & Slatkin, 2013, 2015). The reliability of the patterns inferred from demographic and range expansion methods

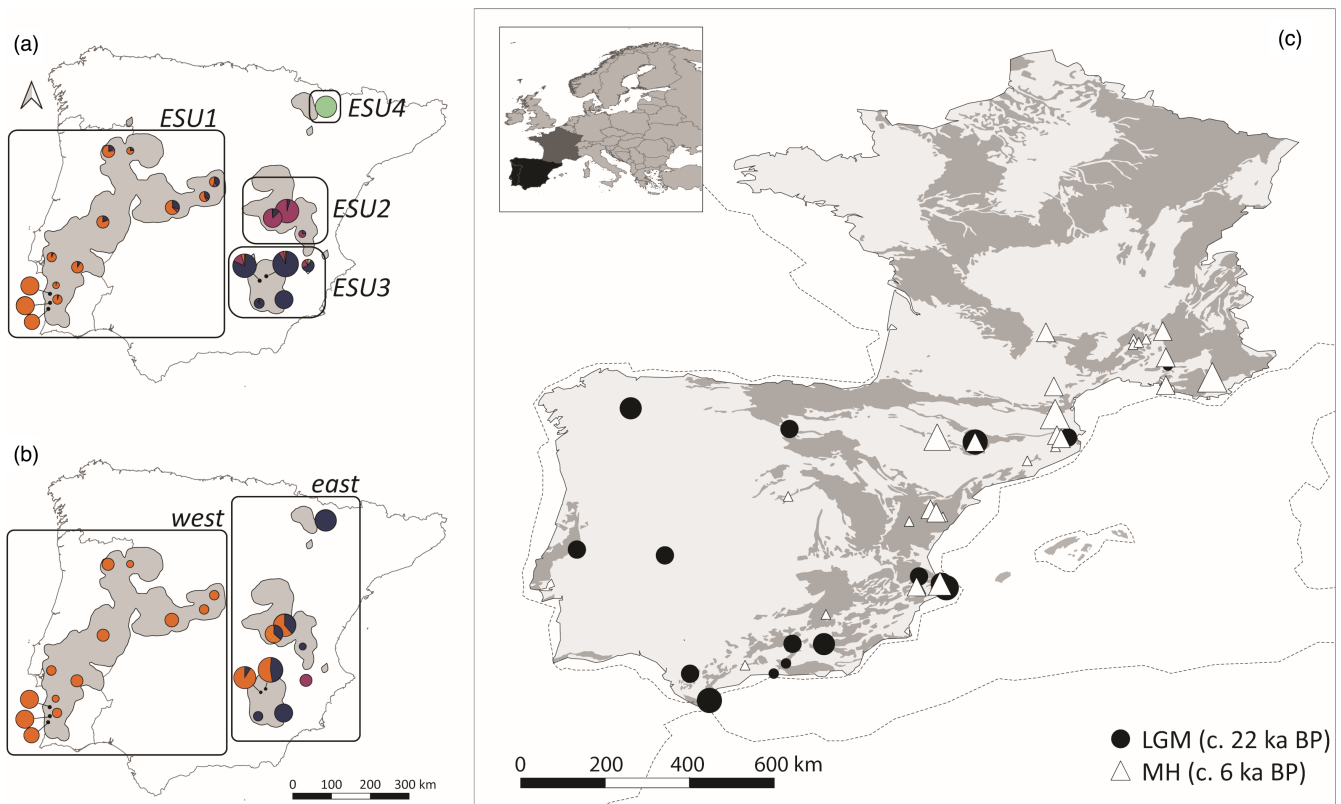


rests on several assumptions: (i) populations sampled are representative of the species genetic diversity; (ii) the genetic ancestral state of a genotype is known, to allow inferences on the directionality of expansion; and (iii) locations with higher genetic diversity are stable over time and have not changed due to, for example, drift or selection (Gavin et al., 2014; Hoban et al., 2019; Potter et al., 2016). Currently, there are methods that can integrate demographic summary statistics with range expansion, to consider aspects of genetic diversity and divergence in the identification of the origin of range expansion, while also accommodating inferred ENMs in a spatially explicit coalescent modelling framework (He et al., 2017).

Here, we evaluate the fossil record, ENMs and genomic data to advance our understanding of the biogeographical history of the near-threatened, Iberian-endemic Cabrera vole (*Microtus cabrerae*, Thomas 1906), which is considered highly vulnerable to climate change (Araújo et al., 2012; Mestre et al., 2015; Mestre et al., 2017). The Cabrera vole is the sole extant representative of the *Iberomys*

subgenus, which was widely distributed across the Iberian Peninsula and southern France in the Middle Pleistocene (Barbosa, Paupério, et al., 2018; Garrido-García & Soriguer-Escofet, 2012; Pita et al., 2014). Being a habitat specialist, the Cabrera vole is very dependent on habitat quality and is thought to respond strongly to environmental change (Pita et al., 2014; Rivas-Martinez, 2011; Santos et al., 2006). Since the Late Pleistocene, the distribution of the Cabrera vole has globally decreased (Laplana & Sevilla, 2013), and today the species is divided into four main Evolutionarily Significant Units (ESUs) within the Iberian Peninsula (Barbosa, Mestre, et al., 2018; Garrido-García et al., 2013; Figure 2a).

The fossil record suggests the first appearance of the Cabrera vole around 90 ka BP (Garrido-García & Soriguer-Escofet, 2012; Laplana & Sevilla, 2013; Pita et al., 2014), while previous phylogeographical studies using mitochondrial and single nuclear loci data estimated the time to the most recent common ancestor of extant populations to be around 22 ka BP (Barbosa et al., 2017). This



**FIGURE 2** Summary of previous genetic studies on Cabrera vole, using the same samples analysed in this study, together with the described fossil distribution of the species across the Iberian Peninsula and France (these regions shown, respectively, as black and grey in the map of Europe included as an inset in part [c]). For parts (a) and (b), the Cabrera vole present-day distribution is represented in dark grey (ETRS89), while sampled populations are represented by circles of size proportional to the number of individuals analysed from each locality (min 1 – max 15). (a) Population structure and identification of the four evolutionarily significant units (ESUs) defined from clustering analyses of nuclear single nucleotide polymorphism (SNP) data generated in Barbosa, Mestre, et al. (2018) and used in this study. (b) Phylogenetic assignment to west or east mitochondrial lineages, with boxes limiting the population assignment to the two groups; west haplotypes in the east group were found to represent recent secondary contact (Barbosa et al., 2017). (c) Presence data of the Cabrera vole fossil record for the Mid-Holocene (MH) and Last Glacial Maximum (LGM) in the Iberian Peninsula and France (light grey area). The size of the symbols is proportional to the number of studies that identified Cabrera vole fossils in separate excavation events for a given deposit (min 1 – max 5). Shaded dark grey represents limestone/karstic areas where bones are prone to fossilization, and dashed lines represent the LGM coastlines (WGS84)

suggests that the Cabrera vole has been contracting and expanding its range according to the climatic oscillations of the Quaternary (Garrido-García & Sorriquer-Escofet, 2012; Laplana & Sevilla, 2013).

In our 2017 study, we assumed that the most recent common ancestor of extant Cabrera vole was an LGM refugial population located in the south of the Iberian Peninsula—basing this on the highest fossil abundance for the species in that specific geographical area. Studies using ancient mitochondrial DNA data have shown that the species range contracted again during the Younger Dryas (YD, c. 13 ka BP; Laplana & Sevilla, 2013; Rodríguez-Varela, 2016). Specifically, during the YD, the widely expanded population is thought to have contracted into two sub-populations in east and west Iberia, and those are believed to represent the two main mitochondrial lineages that survived to the present (Barbosa et al., 2017; Figure 2b). The location of these LGM and YD refugia, however, has never been rigorously determined, given that it has so far been based only on the availability of the fossil record, and the genetic methods employed were not geographically explicit. In the present paper, we are rectifying that situation. By integrating an assessment of the fossil record, ENM and geographically explicit genomic analyses, we illustrate the importance of integrative multidisciplinary approaches to detect the glacial refugia of species and to better understand the change in the distribution and niche of species through time.

## 2 | MATERIALS AND METHODS

### 2.1 | Fossil record

We combined information from the fossil record of the Cabrera vole previously assembled by Garrido-García and Sorriquer-Escofet (2012), and Laplana and Sevilla (2013), comprising 18 locations from the LGM, and 27 locations from the mid-Holocene (MH; Figure 2c).

The distribution of the species during the LGM and the MH was based on (i) the LGM being within the Marine Isotopic Stage 2, the last and most severe glacial period (29–14.5 ka BP; Lisiecki & Raymo, 2005) and (ii) the MH being within the Atlantic Thermochron (7.9–5.3 ka BP; Zubakov & Borzenkova, 1990) characterized archaeologically for the Iberian Peninsula by the presence of Neolithic cultures (Almagro-Gorbea, 2014). To evaluate the impact of variation in fossilization potential for fossil detection, we considered a map depicting regions favourable to fossilization as the combination of karstic areas and soil with high pH generated by the abundance of carbonates in the geologic substrata (Baker & Worley, 2014; Schmidt et al., 2012; Figure 2c). Considering both the entire study area and the regions with fossilization potential, we evaluated the presence of the fossils relative to the ENM suitability values using a Student's *t*-test.

### 2.2 | Ecological niche modelling

For modelling the current distribution of the Cabrera vole, we used the most up to date records of species occurrence (Barbosa

et al., 2017; Garrido-García et al., 2013; Mestre et al., 2015; Vale-Gonçalves & Cabral, 2014), with 467 presences, and double that number as pseudo-absences (five sets of pseudo-absences were generated, totalling 4670, see e.g. Barbet-Massin et al., 2012). For model calibration and the projections to the present, climatic variables were downloaded from the WorldClim website (Hijmans et al., 2005). The model was calibrated considering 10 × 10 km UTM square grids: the presences were the centroids of UTM square grids where the species was recorded, and the bioclimatic variables were averaged for each UTM square grid. To estimate the ENM, we performed a preliminary screening of the bioclimatic variables to guarantee that only uncorrelated and relevant variables were considered in model building, using a two-step process: first, logistic regression models were fitted relating the presence of Cabrera voles and each candidate variable, considering both linear and unimodal (quadratic) effects, guaranteeing that only significant variables ( $p < 0.05$ ) were retained for the next step. Second, pairwise Spearman correlation coefficients ( $r_s$ ) were computed between all variables, and for each pair of highly correlated variables ( $|r_s| > 0.7$ ) the one considered to be potentially most influential for the species distribution, based on previous studies and knowledge about the species ecology, was retained (Dormann et al., 2013; Mestre et al., 2015, 2017; Mira et al., 2008; Pita et al., 2011).

Following the recommendations of Nogués-Bravo (2009), the inferred past distribution of the species, hindcasted using models calibrated with present distribution data, was projected to several climatic reconstructions, resorting to a multi-algorithm approach (Araújo & New, 2007; for modelling options, see Table S1). This was fundamental to account for the uncertainty in the modelling approach since different climatic reconstructions and modelling algorithms produce different results. To evaluate the models, the dataset was randomly split into calibration (70%) and evaluation subsets (30%) in a multiple cross-validation procedure, replicating the data splitting procedure five times (Guisan et al., 2017). The relative weight of each model in the ensemble was proportional to its evaluation scores, as given by true skill statistics (TSS; Allouche et al., 2006). TSS values range from -1 to +1, where +1 indicates perfect agreement, while TSS values  $\leq 0$  are indicative of a performance no better than random. Thus, only models with TSS values  $> 0.5$  were kept in the final ensemble model (Mestre et al., 2015). The contribution of each variable to the final model was evaluated through the variable importance metric, ranging from zero (no importance) to one (highest importance; Thuiller et al., 2013).

The area used for calibration of the ENM was the same used for model projection, the Iberian Peninsula and France. For the LGM and MH, the ENM was projected to three and nine GCMs, respectively (Table S2), using the variables of past climate reconstructions available in the WorldClim website (Hijmans et al., 2005). For the YD, the ENM was projected to one GCM, downloaded from the PalaeoClim website (Fordham et al., 2017). The individual geographical projections of alternative GCMs in the LGM and MH were compared using Schoener's *D* (Schoener, 1968), Warren's *I* (Warren et al., 2008) and Hellinger distance (van der Vaart, 1998) statistics as computed by the 'fuzzySim' R package, version 1.7.6 (Barbosa, 2015).



To conduct the GIS pre-analysis, we used the R packages 'raster' (Hijmans, 2020) and 'maptools' (Bivand et al., 2021). The ensemble modelling approach was run using the 'biomod2' package (Thuiller et al., 2013) of the R statistical software, version 3.3.1 (R Core Team, 2016). Details regarding the commands used can be found in a Figshare repository.

### 2.3 | Genomic data and range expansion analysis

We used Genotyping-by-Sequencing data from 107 Cabrera voles from 22 populations (Figure 2a), generated in Barbosa, Mestre, et al. (2018). For read mapping and single nucleotide polymorphism (SNP) calling, we used 'bwa' version 0.7.17 (Li, 2013) and 'samtools' version 1.9 (Li et al., 2009), using the prairie vole (*M. ochrogaster*, GCA\_000317375.1) genome as reference. The prairie vole genome also served as an outgroup to polarize SNP variation, that is, determine the ancestral state, which is essential for the demographic analyses. We considered only biallelic positions and loci with more than 1% minor allele frequency to ensure that SNPs were informative for demographic analyses, as well as a minimum of two reads per genotype per sample, no missing data, and kept only SNPs located more than 10,000 bp apart, to minimize Linkage Disequilibrium. These filters were employed using VCFtools version 0.1.16 (Danecek et al., 2011).

To re-evaluate the demographic history of the Cabrera vole using genomic data, previously analysed in Barbosa et al. (2017) using mitochondrial and single nuclear gene data, we first performed a demographic analysis to compare models potentially representing the demographic history of the current populations based on the current genetic diversity patterns (Barbosa et al., 2017; Barbosa, Mestre, et al., 2018; Figure 2a,b). These models represent different possible demographic scenarios (SC, Figure S1) considering a neutral hypothesis of a single splitting event within continuous migration between ESUs, likely representing an expansion from a central location in Iberia (SC1); two expansion models from the south of Iberia, consistent with the fossil record, with splitting events considering eastern and western lineages (SC2 and SC3, respectively); four expansion models from the north of Iberia, with splitting events from eastern (SC4 and SC5) and western lineages (SC6 and SC7), and also testing within each scenario for a model of isolation with migration leading to current incomplete lineage sorting (SC4 and SC6), vs. a model of population isolation with recent secondary contact (SC5 and SC7). For models representing secondary contact, migration was only allowed after TSEC (Figure S1).

We evaluated different splitting events and respective timings for current ESUs based on the site frequency spectrum (SFS). For that, we used the 'easySFS' tool implemented in 'dadi' (Gutenkunst et al., 2009), based on the unfolded SFS, given that the ancestral state of each variant is known. To obtain the SFS, we used all segregating sites, given that our dataset had no missing data. We then used these SFSs in the software 'fastsimcoal2' (Excoffier et al., 2021) to compare the seven demographic models (Figure S1), with different splitting patterns and admixture events, while allowing for divergence

time estimation. For each model we ran 100 replicates, each with 200,000 coalescent simulations to approximate the expected SFS in each cycle, and 50 optimization cycles to estimate the parameters. Parameter priors are found in Table S3. We then performed model comparison by calculating the Akaike information criterion (AIC) for the maximum estimated likelihood of the best run for each model. The best model was determined as the one with the lowest AIC and with more than 2 AIC units difference from the next best model.

To determine whether there was an expansion signal worth exploring in the genomic data, we first performed range expansion tests using the 'rangeExpansion' R package (Peter & Slatkin, 2013, 2015) using three different datasets: all populations combined (*all*), which refers to the origin of expansion of the current populations of the species; then dividing the populations based on the mitochondrial subdivision into *west* (ESU1) and *east* (ESU2, ESU3 and ESU4) lineages, and finally testing each ESU separately (Figure 2), resulting in a total of six tests.

We then used the 'X-ORIGIN' pipeline (He et al., 2017) to integrate the ENMs at the four different time points (LGM, YD, MH and present day) with summary statistics from the genomic data to perform a coalescent analysis of the origin of expansion of the current population. Briefly, this pipeline infers refugial or source populations accounting for the effects of temporal and spatial environmental heterogeneity, which may impact migration routes and, thus, present-day location of derived populations. For this analysis, the four ENMs corresponding to the four different timescales were statistically downscaled to 10×10km per cell. Considering a generation time of 1 year, we defined the following temporal correspondence for the Cabrera vole: LGM—22,000 generations ago, YD—13,000 generations ago; MH—6000 generations ago; and present—10 generations ago. As sampling locations, we used a median coordinate of each of the four ESUs identified for the Cabrera vole (Barbosa, Mestre, et al., 2018), and simulated the origin of expansion using the software SPLATCHE2 (Ray et al., 2010). Simulation parameters are shown in Table S4. We also determined a minimum and maximum potential origin in terms of latitude and longitude that encompassed the entire area of the Iberian Peninsula and southern France, including all known fossil locations for the LGM (Figure 2c). The empirical summary statistics used in X-ORIGIN were generated using ARLSUMSTAT (Excoffier & Lischer, 2010) including Tajima's D and Fu's FS for the entire dataset, heterozygosity (H) and number of segregating sites (S) for each ESU, genetic divergence (FST) and mean number of differences between pairs of populations (Pi) and between ESUs, and the directionality index from rangeExpansion,  $\psi$  (Peter & Slatkin, 2013). We used ABCSAMPLER (Wegmann et al., 2010) to explore the parameter space and iterate over the demographic simulations and corresponding calculation of summary statistics, using 20,000 simulations for each of the replicate runs. We excluded simulations that did not result in the final colonization of the present-day populations and then combined all remaining simulations from the 100 replicate runs for the post-processing analysis. Details regarding the commands used in all analyses can be found in the Figshare repository.

### 3 | RESULTS

#### 3.1 | Fossil record

The fossil record highlighted a concentration of Cabrera vole fossils on the eastern coast of the Iberian Peninsula, both during the LGM and MH, and in southern France during the MH, mostly associated with areas of higher fossilization potential (Figure 2c). To a lesser extent, fossils were also found across all the Iberian Peninsula for the LGM period and (minimally) in central Iberia for the MH. For some of the extreme northern regions of the Iberian Peninsula (Basque-Cantabrian mountains), there was strong support for the absence of Cabrera voles at both time periods, given the high fossilization potential and the numerous excavations performed in that area (Schmidt et al., 2012), which have not resulted in the identification of any *M. cabrerae* fossils (Garrido-García & Sorriquer-Escofet, 2012; Laplana & Sevilla, 2013).

Fossils occurring in regions of, apparently, lower fossilization potential, such as in north-western Iberia (Galicia) and central or western Iberia (Extremadura) are also located in karstic areas, which are not visible on the map as they are very small and coincident with the fossil locations (Figure 2c).

#### 3.2 | Ecological niche modelling

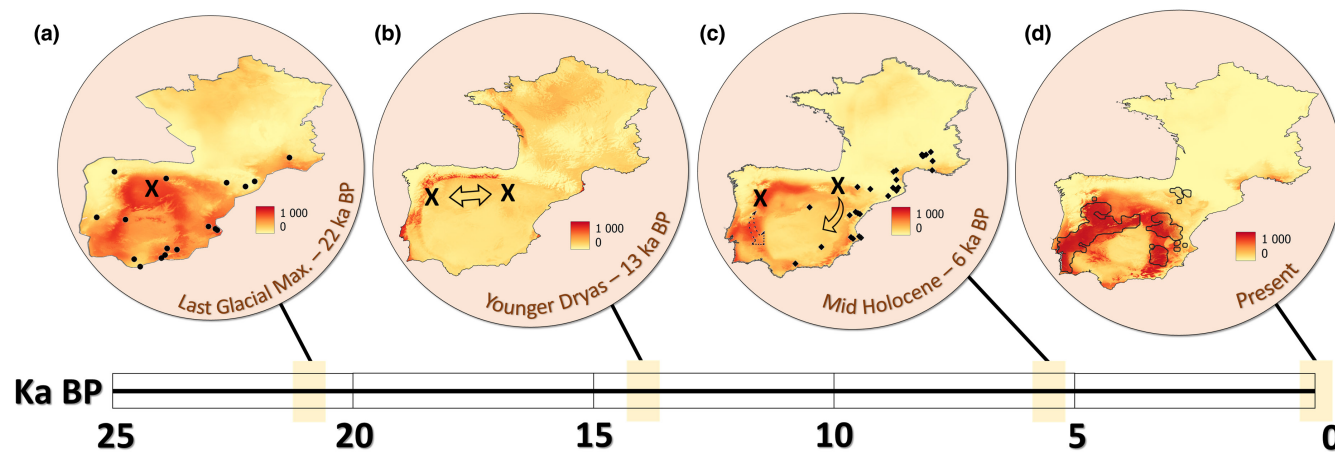
After the environmental variable pre-screening, the model retained bio4 (Temperature Seasonality), bio8 (Mean Temperature of Wettest Quarter), bio12 (Annual Precipitation) and bio18 (Precipitation of Warmest Quarter; Table S5). The most and least important variables explaining the current species distribution were bio18 and bio8, respectively (Table S6).

The projections of the ENM to the MH (Figure 3c) were reasonably similar across the nine GCMs, although some (e.g. BCC-CSM1-1) showed higher suitability to the north of the Iberian Peninsula, while others (e.g. MIROC-ESM) showed higher suitability in the south-western portion of the Peninsula (Figure S2; Table S7). Despite these differences, all models estimated that central Iberia had very low environmental suitability for the species during the MH. For the YD (c. 13ka BP), the regions of highest suitability were restricted to two narrow strips, one in the north-western portion of the Iberian Peninsula and another in western France (Figure 3b). The projections to the LGM (c. 22ka BP) also showed a reasonable similarity across the three GCMs considered (Figure 3a). The north-central and various southern portions of the Iberian Peninsula retained suitable environmental conditions for the Cabrera vole during the LGM, especially for two of the three GCMs (Figure S3; Table S8).

Comparing the ENMs and the fossil record for the MH, the suitability in the fossil locations was significantly higher only for one of the nine GCMs if we consider the entire study area as background; and in five out of nine, if we consider the area favourable to fossilization (Figure 3c; Table S8). For the YD, the lack of fossil records precluded the comparison between fossil location and ENM suitability. Finally, for the LGM, the ENM suitability values at the fossil locations were always higher than the background, whether considering the whole study region or only the areas favourable to fossilization (Figure 3a; Table S9).

#### 3.3 | Genomic data demography and range expansion analysis

We identified 4677 biallelic SNPs that were used for the demographic and range expansion analyses. For the demographic analyses, the estimates for the best run of each model varied greatly between models



**FIGURE 3** Schematic representation of the habitat suitability for the Cabrera vole during the Last Glacial Maximum (a), Younger Dryas (b), Mid-Holocene (c) and Present (d) in the Iberian Peninsula and France; low suitability (0, light yellow) to high suitability (1000, dark orange); Last Glacial Maximum: ENM and genetics suggest a potential refugium in the northern Iberian Peninsula (represented by an X as per the X-ORIGIN results), and the fossil record (represented by black circles) suggest a southern location; Younger Dryas—ENM suggests two narrow strips of higher suitability (north-western Iberia and western coastal France), no information concerning the fossil record, but genetics suggests a division into two populations (origin of expansion represented by X for each lineage as per *rangeExpansion* results); Mid-Holocene—Genetics suggests an expansion to the south, that might have occurred through the higher suitability regions identified by the ENM, while fossil record (represented by black circles) suggests the species presence in the south; and Present—Current range of the species (WGS84)



(Table S4). The best model was SC4, which was 20 AIC units lower than the second-best model (SC3), representing divergence of all ESUs through gradual isolation with migration (Figure S1; Table S4). Models representing secondary contact showed the highest AIC values. The estimates for time of divergence between lineages for SC4 were ~22ka BP for ESU1, ~13ka BP for ESU4 and ~1900years BP for ESU2 and ESU3. Effective population size was estimated as the largest in ESU1 (2651 voles), which is also the ESU with the largest geographical breadth, and for all models, migration rates were highest (0.002–0.004) between ESU2 and ESU3 (Table S4).

For the range expansion analyses, we did not detect an expansion signal for the entire dataset ('all', Table S10). However, the expansion model was strongly supported ( $p < 0.0001$ ) over natural isolation-by-distance considering the *west* (ESU1) and *east* (ESU2, ESU3 and ESU4) populations, for which we found a likely origin of both groups in western and north-eastern areas of the Iberian Peninsula, with an effective founder distance ('d1', representing the distance at which 1% of the founding population is lost) of c. 11 and 22 km, respectively (Figure S4a,b; Table S10). Of all ESUs analysed individually, we found a significant expansion signal in ESU3 alone ( $p < 0.01$ ), with  $d1 = 5.1$  km (Figure S4c; Table S10).

For the X-ORIGIN analysis, the summary statistics showed that both Tajima's  $D$  and Fu's  $FS$  across all populations were zero (Table S11). In terms of genetic diversity, ESU4 was the least diverse group across metrics (Table S12). Concerning genetic divergence, ESU4 was the most divergent population (Table S13). For the SPLATCHE2 simulations, we obtained a combined dataset of 236,429 replicates. Using this dataset, we estimated the origin of this dataset to be located at latitude 42 and longitude  $-6$ , representing a location in north-western Spain ('X' in Figure 3a; Figure S4).

In short, we estimated the time to the most recent common ancestor is around 22ka BP, consistent with a single population at the LGM (Figure 3a; Figures S1 and S4; Tables S3 and S10). Our demographic models and previous genetic and genomic results suggest that populations were restricted to at least two separate locations during the YD (Figure 3b; Figure S4), leading to the divergence between the two main mitochondrial lineages found today, named *west* and *east* (Figure 2b). These models support a divergence of ESU4 from other *east* lineage populations during this time (Figure S1; Table S3), consistent with a signal of southwards expansion of the *west* and *east* lineage from western and north-eastern areas, respectively (Figure S4a,b).

## 4 | DISCUSSION

Our study combining information on the fossil record, ecological niche modelling and genomics provides a novel integrative case study on the biogeographical history of a near-threatened small mammal, endemic to the Iberian Peninsula. When looking at and comparing all three lines of evidence, we found that the distribution of the Cabrera vole has been highly dynamic in response to environmental changes occurring since the LGM, with colder periods resulting in vicariance

of populations, likely due to population isolation, followed by range expansions during warmer periods, as expected for temperate species, particularly in southern European refugia such as the Iberian Peninsula (Hewitt, 2004, 2011b).

Comparatively, the patterns observed were not always consistent between fossils, the ENM and genomic data, mainly because the fossil record appears greatly biased towards areas with high fossilization potential. Overall, our results strongly support the need to integrate different lines of evidence to describe and understand the range dynamics of species in relation to past climate fluctuations while accounting for their strengths and limitations.

### 4.1 | An integrated approach for ENM and genomic data

Integration of ENM and genomic data is becoming the standard for refugial inference and biogeography due to limitations and biases in interpretation based on any single line of evidence (Bemmels et al., 2019; Fenderson et al., 2020; He et al., 2017; Hoban et al., 2019). In our previous molecular studies, we suggested that the Cabrera vole contracted to a single population as late as the Last Glacial Maximum (LGM, c. 22ka BP; Barbosa et al., 2017). At that time, we inferred the single LGM refugium to be in southern Iberia, based on the abundance of the fossil record (Figure 2).

The results of our demographic modelling and analyses of the geographical origin of expansion inferred from genomic data suggest that this single population or LGM refugium was probably located in northern Iberia, which is also an area of known fossil records of this and many other small and large vertebrates during that time period (Kotlík et al., 2022; Martínez-Freiría et al., 2020; Queirós et al., 2019; Querejeta et al., 2017). Through an integrative analysis of ENM and genomic data, which considers the full extent of the fossil record for localization, a northern LGM refugium was supported. This latter interpretation accounts for habitat suitability at the four different time points analysed in this study, and thus provides an integrated assessment of the origin of the post-LGM expansion considering both genetic and ecological data over time. It seems likely therefore that fossilization bias could mislead inferences of the location of the LGM refugium, especially when viewed in the context of extinction events with concomitant changes in the ecological niche of the species (Hinojosa et al., 2016). Contrarily, during the Early-MH, the increased fossil record throughout the Iberian Peninsula is in line with an increase in population genetic diversity, supporting the hypothesis that population expansion continued throughout the Late Holocene (Laplana & Sevilla, 2013; Rodríguez-Varela, 2016).

### 4.2 | Fossil data strengths and limitations

The fossil record of the Iberian Peninsula is fairly well documented and has been used to infer refugial history and population



bottlenecks using palaeodemography (Fernández-López de Pablo et al., 2019; Sommer & Zachos, 2009). In the present study focused on the Cabrera vole, there was a significant overlap between ENM suitability and the fossil record during the LGM, showing that the models identify regions with fossil presence as those with higher suitability, particularly when considering areas with high fossilization potential as background (Table S9). However, this overlap was not observed across the remaining time points, nor did it show concordance to the inferred refugial areas from genomic data.

Regions associated with a high fossilization potential might only represent a perceived high abundance, biasing the identification of potential refugia. Specifically, fossils located in south-eastern Iberia in regions with low habitat suitability (and high fossilization potential) might represent populations that became locally extinct. This was further supported by the lack of a genomic expansion signal from those southern populations. Although we argue that LGM populations from that region went locally extinct, the fact that it still has Cabrera vole populations could be explained by either a temporary decrease of habitat suitability at the LGM, allowing expanding populations to subsequently colonize the vacant habitat as conditions improved; or a permanent change in habitat condition may have allowed the establishment of different, and better adapted, populations (de Lafontaine et al., 2018).

It is, however, important to note that the fossil record, as it was used in our analyses, is a raw source of information. Thus, it may suffer from biases to a greater extent than model-derived results, such as the genomic and demographic analysis and ENM. Modelling of the potential areas of occurrence based on fossilization could in principle be considered by integrating fossil abundance with geological characteristics, producing a probabilistic fossil distribution landscape considering high absence certainty in excavation areas of high fossilization potential where the species was not found, and vice versa. This landscape would then need to be corrected for (i) sampling bias, as larger or more accessible fossil assemblages tend to concentrate most sampling efforts and (ii) taxonomic bias, as some species are more easily identified or preserved than others (Allison et al., 2011). Such layers could then be integrated with the ENM modelling to increase or decrease support of habitat suitability at each time point.

### 4.3 | ENM strengths and limitations

In this study focused on the Cabrera vole, ENM assigned high suitability to those regions where the genomic approach estimated the LGM refugia to be. However, the ENM also identified the regions with fossil presence as having higher suitability, in particular when compared with areas of high fossilization potential. While the fossil record indicates areas of known species presence in the past, caution should be taken when comparing it to ENMs developed from present-day populations, especially considering projections to increasingly early time periods. The presence of fossils in some sites with low ENM suitability may reveal the past presence of populations with particular niche requirements not captured by the ENM; while high ENM suitability in areas with no fossils might indicate

past occupied regions where no fossil trace of the species presence was preserved (Castellanos-Frías et al., 2018; Hinojosa et al., 2016; Hoban et al., 2019). Further integration of evidence can be achieved where it is possible to measure the genetic diversity of ancient DNA from fossils. This could allow a test of niche conservatism, by identifying which fossils represent ancestral individuals to the current populations, and thus are potentially representative of their niche, and the two lines of evidence can be integrated into a phylogenetic framework (Theys et al., 2019). In our system, fossils were present in southern France from the LGM until the MH, which suggests that the populations of Cabrera vole persisted in these locations until very recently. However, we found very little habitat suitability for that geographical area using ENM, suggesting that the ecological requirements from those populations are not comprehensively represented by the extant populations. In summary, those populations may represent distinct genetic lineages with slight variations in their ecological niche not fully captured by our models, and that recently went extinct. Studies of ancient DNA in Cabrera vole have shown continuity of mitochondrial DNA haplotypes from the Mesolithic until about a thousand years ago, not found in the extant Iberian populations, supporting the hypothesis of a replacement of the fossil populations by current populations (Rodríguez-Varela, 2016). In fact, Castellanos-Frías et al. (2018) suggested the possibility that the ecological niche of the Cabrera vole has changed with the climatic oscillations of the Quaternary, becoming broader during glacial times. Our results did not contradict this hypothesis, since the fossil record—namely in the MH—documents the species presence in regions outside the projected ecological niche calibrated with the current species distribution. However, as mentioned in the introduction, other explanations are also possible (Figure 1). Taken together, our data suggest that the surviving populations of a given species might not represent the complete ancestral ecological niche breadth, and that therefore the niche conservatism postulate should be considered on a lineage-by-lineage basis both in hindcasting and in forecasting (Napier et al., 2020; Razgour et al., 2019).

### 4.4 | Genomic data strengths and limitations

In this study, we found that glacial refugia for the ancestors of the current populations were located in northern regions of the species current distribution. The hypothesis of a north-to-south gradual expansion in an isolation-with-migration framework does not imply that other areas of the species' distribution had not been colonized previously. In fact, given the existence of other areas of the Iberian Peninsula with evidence of large populations at the time, inferred from the high fossil density in the south-eastern coast, there were likely refugia-within-refugia for this temperate species in Iberia during the LGM. This pattern has also been detected for many other small vertebrates during the same time period (García et al., 2020; Gómez & Lunt, 2007). The proposed north-to-south expansion simply means that local populations of the Cabrera vole located other than in the inferred origin of expansion did not contribute significantly to the



current gene pool. This is the greatest strength of the genetic data: to provide an unbiased snapshot of genetic diversity patterns (assuming that all major genetic units of the species have been sampled). There are, however, some potential limitations that need to be accounted for. Natural selection can change directionality of founder events by altering patterns of geographical distribution of genetic diversity (Peter & Slatkin, 2015). In our case, this limitation could be excluded as a potential bias, given that there are very few outlier loci that could be potentially under selection in our genomic dataset (as shown by Barbosa, Mestre, et al., 2018). Similarly, demographic changes (population expansions and contractions) can influence the patterns of genetic diversity through genetic drift, which would also bias the inference of origin of expansion to areas where larger and more diverse populations are found. In the present study system, given that smaller populations are located in the north (ESU4) and larger and more diverse populations in the south (ESU2 and ESU3), we would expect genetic drift to bias against our current predictions of a northern refugium (Barbosa, Mestre et al. 2018). There are also limitations associated with the demographic analysis itself, which, although not impacting the timing of splitting events, could bias our perception of how the patterns of genetic diversity were established. A recent study from Momigliano et al. (2021) showed that demographic inferences not accounting for changes in effective population size ( $N_e$ ) tend to favour secondary contact (SC) over isolation with migration (IM) models. As we did not infer changes in  $N_e$  over time, this could have caused a bias in model selection. However, in our results, we saw the opposite pattern, where IM models were favoured over those representing SC. We believe that, by allowing for changes in  $N_e$  with multiple events of population splitting, generally coinciding with times of  $N_e$  changes, we were able to avoid such bias across models.

#### 4.5 | Concluding remarks

By presenting the strengths and limitations of each line of evidence and integrating them, we have developed an alternative hypothesis on the glacial distribution of the Cabrera vole compared to those previously proposed (Barbosa et al., 2017; Barbosa, Mestre, et al., 2018; Laplana & Sevilla, 2013). Taken together, our results support the view that inferences based on a single line of evidence may provide an incomplete understanding of the palaeo-distribution and refugial locations of a species and highlight the importance of considering the merits and limitations of each data type. Moreover, we were able to identify specific sources of error that need to be considered by other researchers using similar approaches. Overall, we show that there is a need for this type of integrative frameworks, explicitly accounting for the uncertainties linked to each source of information for a robust assessment of biogeographical patterns and population dynamics over extended time periods and large spatial scales.

#### ACKNOWLEDGEMENTS

The authors would like to thank Dr. Qixun He for her help in troubleshooting and customizing the X-ORIGIN analyses and the reviewers

and editors for their helpful comments on the manuscript. No special permits were required to carry out the work.

#### CONFLICT OF INTEREST

The authors declare no conflicts of interest.

#### DATA AVAILABILITY STATEMENT

Data included in this manuscript are publicly available, including all genomic raw reads for 107 Cabrera vole (*Microtus cabrerarum*) samples, as well as the GBS VCF file (NCBI Sequence Read Archive Bioprojects PRJNA413751 and PRJNA419872), and the R code used for the ENM and genomic analyses (Figshare: <https://doi.org/10.6084/m9.figshare.14529138>).

#### ORCID

Frederico Mestre <https://orcid.org/0000-0002-7390-1120>

Soraia Barbosa <https://orcid.org/0000-0001-7806-4256>

José A. Garrido-García <https://orcid.org/0000-0003-4290-7885>

Ricardo Pita <https://orcid.org/0000-0001-8562-7301>

António Mira <https://orcid.org/0000-0001-6152-0454>

Paulo C. Alves <https://orcid.org/0000-0003-4797-0939>

Joana Paupério <https://orcid.org/0000-0003-2569-0768>

Jeremy B. Searle <https://orcid.org/0000-0001-7710-5204>

Pedro Beja <https://orcid.org/0000-0001-8164-0760>

#### REFERENCES

- Abellán, P., & Svenning, J.-C. (2014). Refugia within refugia – Patterns in endemism and genetic divergence are linked to Late Quaternary climate stability in the Iberian Peninsula. *Biological Journal of the Linnean Society*, 113, 13–28. <https://doi.org/10.1111/bj.12309>
- Allison, A., Bottjer, D., Allison, P., Allison, P. A., & Bottjer, D. J. (2011). Taphonomy: Bias and process through time. *Topics in Geobiology*, 32, 1–17. [https://doi.org/10.1007/978-90-481-8643-3\\_1](https://doi.org/10.1007/978-90-481-8643-3_1)
- Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43, 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Almagro-Gorbea, M. (2014). *Protohistory of the far west of Europe. From Neolithic to Roman conquest*. Burgos University - Atapuerca Foundation.
- Anadón, J. D., Graciá, E., Botella, F., Giménez, A., Fahd, S., & Fritz, U. (2015). Individualistic response to past climate changes: Niche differentiation promotes diverging Quaternary range dynamics in the subspecies of *Testudo graeca*. *Ecography*, 38, 956–966. <https://doi.org/10.1111/ecog.01163>
- Araújo, M. B., Guilhaumon, F., Neto, D. R., Pozo, I., & Gómez Calmaestra, R. (2012). In M. B. Araújo (Ed.), *Biodiversidade e alterações climáticas na península Ibérica/Biodiversidad y alteraciones climáticas en la península Ibérica*. Ministério da Agricultura, do Mar, do Ambiente e do Ordenamento do Território; Ministério de Agricultura, Alimentación y Medio Ambiente.
- Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, 22(1), 42–47. <https://doi.org/10.1016/j.tree.2006.09.010>
- Baker, P., & Worley, F. (2014). Animal bones and archaeology. Guidelines for best practice. In E. Fairnell (Ed.), *Cambridge manuals in archaeology*. English Heritage.
- Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution*, 3, 327–338. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>

- Barbosa, A. M. (2015). fuzzySim: Applying fuzzy logic to binary similarity indices in ecology. *Methods in Ecology and Evolution*, 6, 853–858. <https://doi.org/10.1111/2041-210X.12372>
- Barbosa, S., Mestre, F., White, T. A., Paupério, J., Alves, P. C., & Searle, J. B. (2018). Integrative approaches to guide conservation decisions: Using genomics to define conservation units and functional corridors. *Molecular Ecology*, 27, 3452–3465.
- Barbosa, S., Paupério, J., Herman, J. S., Ferreira, C., Pita, R., Vale-Gonçalves, H. M., Cabral, J. A., Garrido-García, J. A., Soriguer, R. C., Beja, P., Mira, A., Alves, P. C., & Searle, J. B. (2017). Endemic species may have complex histories: Within-refugium phylogeography of an endangered Iberian vole. *Molecular Ecology*, 26, 951–967. <https://doi.org/10.5061/dryad.9fc86>
- Barbosa, S., Paupério, J., Pavlova, S. V., Alves, P. C., & Searle, J. B. (2018). The *Microtus* voles: Resolving the phylogeny of one of the most speciose mammalian genera using genomics. *Molecular Phylogenetics and Evolution*, 125, 85–92. <https://doi.org/10.1016/j.ympev.2018.03.017>
- Beaumont, L. J., Hughes, L., & Pitman, A. J. (2008). Why is the choice of future climate scenarios for species distribution modelling important? *Ecology Letters*, 11, 1135–1146. <https://doi.org/10.1111/j.1461-0248.2008.01231.x>
- Bemmels, J. B., Lacey Knowles, L., & Dick, C. W. (2019). Genomic evidence of survival near ice sheet margins for some, but not all, North American trees. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 8431–8436. <https://doi.org/10.1073/PNAS.1901656116>
- Bertorelle, G., Benazzo, A., & Mona, S. (2010). ABC as a flexible framework to estimate demography over space and time: Some cons, many pros. *Molecular Ecology*, 19, 2609–2625. <https://doi.org/10.1111/j.1365-294X.2010.04690.x>
- Bivand, R., Lewin-Koh, N., Pebesma, E., Archer, E., Baddeley, A., Bearman, N., Bibiko, H., Brey, S., Callahan, J., Carrillo, G., & Dray, S. (2021). Package 'mapproj'. *R package version 1.1-2*. <https://CRAN.R-project.org/package=mapproj>
- Castellanos-Frías, E., García, N., & Virgós, E. (2018). Assessment of the effect of climate changes in the Late Pleistocene and Holocene on niche conservatism of an arvicolid specialist. *Scientific Reports*, 8(1), 9780. <https://doi.org/10.1038/s41598-018-28000-0>
- Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., & François, O. (2010). Approximate Bayesian computation (ABC) in practice. *Trends in Ecology & Evolution*, 25, 410–418. <https://doi.org/10.1016/j.tree.2010.04.001>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Davis, M. B., & Shaw, R. G. (2001). Range shifts and adaptive responses to Quaternary climate change. *Science*, 292, 673–679. <https://doi.org/10.1126/science.292.5517.673>
- de Lafontaine, G., Napier, J. D., Petit, R. J., & Hu, F. S. (2018). Invoking adaptation to decipher the genetic legacy of past climate change. *Ecology*, 99, 1530–1546. <https://doi.org/10.1002/ecy.2382>
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. <https://doi.org/10.1111/J.1600-0587.2012.07348.X>
- Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10, 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Excoffier, L., Marchi, N., Marques, D. A., Matthey-Doret, R., Gouy, A., & Sousa, V. C. (2021). FastSimcoal2: Demographic inference under complex evolutionary scenarios. *Bioinformatics*, 37, 4882–4885. <https://doi.org/10.1093/BIOINFORMATICS/BTAB468>
- Fenderson, L. E., Kovach, A. I., & Llamas, B. (2020). Spatiotemporal landscape genetics: Investigating ecology and evolution through space and time. *Molecular Ecology*, 29, 218–246. <https://doi.org/10.1111/MEC.15315>
- Fernandez, M. C., Hu, F. S., Gavin, D. G., de Lafontaine, G., & Heath, K. D. (2021). A tale of two conifers: Migration across a dispersal barrier outpaced regional expansion from refugia. *Journal of Biogeography*, 48, 2133–2143. <https://doi.org/10.1111/JBI.14209>
- Fernández-López de Pablo, J., Gutiérrez-Roig, M., Gómez-Puche, M., McLaughlin, R., Silva, F., & Lozano, S. (2019). Palaeodemographic modelling supports a population bottleneck during the Pleistocene-Holocene transition in Iberia. *Nature Communications*, 10(1), 1872. <https://doi.org/10.1038/s41467-019-09833-3>
- Fordham, D. A., Saltré, F., Haythorne, S., Wigley, T. M. L., Otto-Bliesner, B. L., Chan, K. C., & Brook, B. W. (2017). PaleoView: A tool for generating continuous climate projections spanning the last 21 000 years at regional and global scales. *Ecography*, 40, 1348–1358. <https://doi.org/10.1111/ECOG.03031>
- García, J. T., Domínguez-Villaseñor, J., Alda, F., Calero-Riestra, M., Pérez Olea, P., Fargallo, J. A., Martínez-Padilla, J., Herranz, J., Oñate, J. J., Santamaría, A., Motro, Y., Attie, C., Bretagnolle, V., Delibes, J., & Viñuela, J. (2020). A complex scenario of glacial survival in Mediterranean and continental refugia of a temperate continental vole species (*Microtus arvalis*) in Europe. *Journal of Zoological Systematics and Evolutionary Research*, 58(1), 459–474. <https://doi.org/10.1111/JZS.12323>
- Garrido-García, J. A., Rosário, I. T., Gisbert, J., García-Perea, R., Cordero, A. I., López-Alabau, A., Mathias, M. L., Mira, A., Pita, R., Santos, S., Sendra-Pérez, I., Vicente, V., & Soriguer, R. C. (2013). Revisión a nivel ibérico de la distribución del topillo de Cabrera o iberón, *Iberomys cabreræ* (Thomas, 1906). *Galemys*, 25, 35–49. <https://doi.org/10.7325/Galemys.2013.A4>
- Garrido-García, J. A., & Soriguer-Escofet, R. C. (2012). Cabrera's vole *Microtus cabreræ* Thomas, 1906 and the subgenus *Iberomys* during the Quaternary: Evolutionary implications and conservation. *Geobios*, 45, 437–444. <https://doi.org/10.1016/j.geobios.2011.10.014>
- Gavin, D. G., Fitzpatrick, M. C., Gugger, P. F., Heath, K. D., Rodríguez-Sánchez, F., Dobrowski, S. Z., Hampe, A., Hu, F. S., Ashcroft, M. B., Bartlein, P. J., Blois, J. L., Carstens, B. C., Davis, E. B., de Lafontaine, G., Edwards, M. E., Fernandez, M., Henne, P. D., Herring, E. M., Holden, Z. A., ... Williams, J. W. (2014). Climate refugia: Joint inference from fossil records, species distribution models and phylogeography. *New Phytologist*, 204, 37–54. <https://doi.org/10.1111/nph.12929>
- Gómez, A., & Lunt, D. H. (2007). Refugia within refugia: Patterns of phylogeographic concordance in the Iberian Peninsula. In S. Weiss & N. Ferrand (Eds.), *Phylogeography of southern European refugia* (pp. 155–188). Springer. [https://doi.org/10.1007/1-4020-4904-8\\_5](https://doi.org/10.1007/1-4020-4904-8_5)
- Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models: With applications in R*. Cambridge University Press.
- Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 2–3, 147–186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, 5, e1000695. <https://doi.org/10.1371/journal.pgen.1000695>
- He, Q., Prado, J. R., & Knowles, L. L. (2017). Inferring the geographic origin of a range expansion: Latitudinal and longitudinal coordinates

- inferred from genomic data in an ABC framework with the program X-ORIGIN. *Molecular Ecology*, 26, 6908–6920. <https://doi.org/10.1111/MEC.14380>
- 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. <https://doi.org/10.1111/j.1095-8312.1996.tb01434.x>
- Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907–913. <https://doi.org/10.1038/35016000>
- Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, 359, 183–195. <https://doi.org/10.1098/rstb.2003.1388>
- Hewitt, G. M. (2011a). Quaternary phylogeography: The roots of hybrid zones. *Genetica*, 139, 617–638. <https://doi.org/10.1007/s10709-011-9547-3>
- Hewitt, G. M. (2011b). Mediterranean peninsulas: The evolution of hotspots. In F. E. Zachos & J. C. Habel (Eds.), *Biodiversity hotspots* (pp. 123–147). Springer. <https://doi.org/10.1007/978-3-642-20992-5>
- Hijmans, R. J. (2020). *raster: Geographic data analysis and modeling. R package version 2.8-19*. <https://CRAN.R-project.org/package=raster>
- 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: A Journal of the Royal Meteorological Society*, 25, 1965–1978. <https://doi.org/10.1002/joc.1276>
- Hinojosa, L. F., Gaxiola, A., Pérez, M. F., Carvajal, F., Campano, M. F., Quattrocchio, M., Nishida, H., Uemura, K., Yabe, A., Bustamante, R., & Arroyo, M. T. K. (2016). Non-congruent fossil and phylogenetic evidence on the evolution of climatic niche in the Gondwana genus *Nothofagus*. *Journal of Biogeography*, 43, 555–567. <https://doi.org/10.1111/JBI.12650>
- Hoban, S., Dawson, A., Robinson, J. D., Smith, A. B., & Strand, A. E. (2019). Inference of biogeographic history by formally integrating distinct lines of evidence: Genetic, environmental niche and fossil. *Ecography*, 42, 1991–2011. <https://doi.org/10.1111/ecog.04327>
- Kotlík, P., Marková, S., Horníková, M., Escalante, M. A., & Searle, J. B. (2022). The bank vole (*Clethrionomys glareolus*) as a model system for adaptive phylogeography in the European theater. *Frontiers in Ecology and Evolution*, 10, 866605. <https://doi.org/10.3389/fevo.2022.866605>
- Laplana, C., & Sevilla, P. (2013). Documenting the biogeographic history of *Microtus cabrerai* through its fossil record. *Mammal Review*, 43, 309–322. <https://doi.org/10.1111/mam.12003>
- Lasky, J. R., Hooten, M. B., & Adler, P. B. (2020). What processes must we understand to forecast regional-scale population dynamics? *Proceedings of the Royal Society B: Biological Sciences*, 287, 20202219. <https://doi.org/10.1098/rspb.2020.2219>
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv, 1303.3997. <http://arxiv.org/abs/1303.3997>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Lisiecki, L. E., & Raymo, M. E. (2005). A Pliocene-Pleistocene stack of 57 globally distributed benthic  $\delta^{18}\text{O}$  records. *Paleoceanography*, 20(1), PA1003. <https://doi.org/10.1029/2004PA001071>
- Lobo, J. M., Jiménez-Valverde, A., & Hortal, J. (2010). The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, 33, 103–114. <https://doi.org/10.1111/j.1600-0587.2009.06039.x>
- Luna-Aranguré, C., Soberón, J., & Vázquez-Domínguez, E. (2020). A tale of four bears: Environmental signal on the phylogeographical patterns within the extant *Ursus* species. *Journal of Biogeography*, 47, 472–486. <https://doi.org/10.1111/JBI.13752>
- Marske, K. A., Rahbek, C., & Nogués-Bravo, D. (2013). Phylogeography: Spanning the ecology-evolution continuum. *Ecography*, 36, 1169–1181. <https://doi.org/10.1111/j.1600-0587.2013.00244.x>
- Martínez-Freiria, F., Freitas, I., Zuffi, M. A. L., Golay, P., Ursenbacher, S., & Velo-Antón, G. (2020). Climatic refugia boosted allopatric diversification in Western Mediterranean vipers. *Journal of Biogeography*, 47, 1698–1713. <https://doi.org/10.1111/JBI.13861>
- Martínez-Meyer, E., Peterson, A. T., & Hargrove, W. W. (2004). Ecological niches as stable distributional constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity. *Global Ecology and Biogeography*, 13, 305–314. <https://doi.org/10.1111/j.1466-822X.2004.00107.x>
- Messer, P. W., & Petrov, D. A. (2013). Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution*, 28, 659–669. <https://doi.org/10.1016/j.tree.2013.08.003>
- Mestre, F., Pita, R., Paupério, J., Martins, F. M. S., Alves, P. C., Mira, A. P. P., & Beja, P. (2015). Combining distribution modelling and non-invasive genetics to improve range shift forecasting. *Ecological Modelling*, 297, 171–179. <https://doi.org/10.1016/j.ecolmodel.2014.11.018>
- Mestre, F., Risk, B. B., Mira, A., Beja, P., & Pita, R. (2017). A metapopulation approach to predict species range shifts under different climate change and landscape connectivity scenarios. *Ecological Modelling*, 359, 406–414. <https://doi.org/10.1016/j.ecolmodel.2017.06.013>
- Mira, A. P. P., Marques, C. C., Santos, S. M., Rosário, I. T., & Mathias, M. L. (2008). Environmental determinants of the distribution of the Cabrera vole (*Microtus cabrerai*) in Portugal: Implications for conservation. *Mammalian Biology*, 73, 102–110. <https://doi.org/10.1016/j.mambio.2006.11.003>
- Mitchell, J. S. (2015). Preservation is predictable: Quantifying the effect of taphonomic biases on ecological disparity in birds. *Paleobiology*, 41, 353–367. <https://doi.org/10.1017/pab.2014.23>
- Momigliano, P., Florin, A. B., & Merilä, J. (2021). Biases in demographic modeling affect our understanding of recent divergence. *Molecular Biology and Evolution*, 38, 2967–2985. <https://doi.org/10.1093/MOLBEV/MSAB047>
- Napier, J. D., de Lafontaine, G., & Chipman, M. L. (2020). The evolution of paleoecology. *Trends in Ecology & Evolution*, 35, 293–295. <https://doi.org/10.1016/j.tree.2019.12.006>
- Nogués-Bravo, D. (2009). Predicting the past distribution of species climatic niches. *Global Ecology and Biogeography*, 18, 521–531. <https://doi.org/10.1111/j.1466-8238.2009.00476.x>
- Pearman, P. B., D'Amen, M., Graham, C. H., Thuiller, W., & Zimmermann, N. E. (2010). Within-taxon niche structure: Niche conservatism, divergence and predicted effects of climate change. *Ecography*, 33, 990–1003. <https://doi.org/10.1111/j.1600-0587.2010.06443.x>
- Peter, B. M., & Slatkin, M. (2013). Detecting range expansions from genetic data. *Evolution*, 67, 3274–3289. <https://doi.org/10.1111/evo.12202>
- Peter, B. M., & Slatkin, M. (2015). The effective founder effect in a spatially expanding population. *Evolution*, 69, 721–734. <https://doi.org/10.1111/evo.12609>
- Peterson, A. T. (2011). Ecological niche conservatism: A time-structured review of evidence. *Journal of Biogeography*, 38, 817–827. <https://doi.org/10.1111/j.1365-2699.2010.02456.x>
- Peterson, A. T., & Holt, R. D. (2003). Niche differentiation in Mexican birds: Using point occurrences to detect ecological innovation. *Ecology Letters*, 6, 774–782. <https://doi.org/10.1046/j.1461-0248.2003.00502.x>
- Pfenninger, M., & Posada, D. (2002). Phylogeographic history of the land snail *Candidula unifasciata* (Helicellinae, Stylommatophora): Fragmentation, corridor migration, and secondary contact. *Evolution*, 56, 1776–1788.
- Pita, R., Mira, A. P. P., & Beja, P. (2011). Assessing habitat differentiation between coexisting species: The role of spatial scale. *Acta Oecologica*, 37, 124–132. <https://doi.org/10.1016/j.actao.2011.01.006>
- Pita, R., Mira, A. P. P., & Beja, P. (2014). *Microtus cabrerai* (Rodentia: Cricetidae). *Mammalian Species*, 912, 48–70. <https://doi.org/10.1644/912.1>

- Potter, S., Bragg, J. G., Peter, B. M., Bi, K., & Moritz, C. (2016). Phylogenomics at the tips: Inferring lineages and their demographic history in a tropical lizard, *Carlia amax*. *Molecular Ecology*, 25, 1367–1380. <https://doi.org/10.1111/mec.13546>
- Queirós, J., Acevedo, P., Santos, J. P. V., Barasona, J., Beltran-Beck, B., González-Barrio, D., Armenteros, J. A., Díez-Delgado, I., Boadella, M., de Mera, I. F., Ruiz-Fons, J. F., Vicente, J., de la Fuente, J., Gortázar, C., Searle, J. B., & Alves, P. C. (2019). Red deer in Iberia: Molecular ecological studies in a southern refugium and inferences on European postglacial colonization history. *PLoS One*, 14(1), e0210282. <https://doi.org/10.1371/journal.pone.0210282>
- Querejeta, M., Fernández-González, A., Romero, R., & Castresana, J. (2017). Postglacial dispersal patterns and mitochondrial genetic structure of the Pyrenean desman (*Galemys pyrenaicus*) in the northwestern region of the Iberian Peninsula. *Ecology and Evolution*, 7, 4486–4495. <https://doi.org/10.1002/ece3.3034>
- R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Randin, C. F., Dirnböck, T., Dullinger, S., Zimmermann, N. E., Zappa, M., & Guisan, A. (2006). Are niche-based species distribution models transferable in space? *Journal of Biogeography*, 33, 1689–1703. <https://doi.org/10.1111/j.1365-2699.2006.01466.x>
- Ray, N., Currat, M., Foll, M., & Excoffier, L. (2010). SPLATCHE2: A spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics*, 26, 2993–2994. <https://doi.org/10.1093/BIOINFORMATICS/BTQ579>
- Razgour, O., Forester, B., Taggart, J. B., Bekaert, M., Juste, J., Ibáñez, C., Puechmaile, S. J., Novella-Fernandez, R., Alberdi, A., & Manel, S. (2019). Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 10418–10423. <https://doi.org/10.1073/pnas.1820663116>
- Rivas-Martínez, S. (2011). Mapa de series, geoseries y geopermaseries de vegetación de España. *Itineraria Geobotánica*, 18, 5–424.
- Rodríguez-Varela, R. (2016). *Variaciones en la distribución y la estructura genética de los linces paleárticos y el topillo de Cabrera desde el Pleistoceno Superior* [PhD thesis]. Universidad Complutense de Madrid, Facultad de Ciencias Geológicas.
- Santos, S. M., Simões, M. P., Mathias, M. L., & Mira, A. P. P. (2006). Vegetation analysis in colonies of an endangered rodent, the Cabrera vole (*Microtus cabrerae*), in southern Portugal. *Ecological Research*, 21, 197–207. <https://doi.org/10.1007/s11284-005-0104-3>
- Schmidt, I., Bradtmöller, M., Kehl, M., Pastoors, A., Tafelmaier, Y., Weninger, B., & Weniger, G. C. (2012). Rapid climate change and variability of settlement patterns in Iberia during the Late Pleistocene. *Quaternary International*, 274, 179–204. <https://doi.org/10.1016/j.quaint.2012.01.018>
- Schoener, T. W. (1968). The *Anolis* lizards of Bimini: Resource partitioning in a complex fauna. *Ecology*, 49, 704–726. <https://doi.org/10.2307/1935534>
- Sommer, R. S., & Nadachowski, A. (2006). Glacial refugia of mammals in Europe: Evidence from fossil records. *Mammal Review*, 36, 251–265.
- Sommer, R. S., & Zachos, F. E. (2009). Fossil evidence and phylogeography of temperate species: 'Glacial refugia' and post-glacial recolonization. *Journal of Biogeography*, 36(11), 2013–2020. <https://doi.org/10.1111/J.1365-2699.2009.02187.X>
- Svenning, J. C., Fløjgaard, C., Marske, K. A., Nógues-Bravo, D., & Normand, S. (2011). Applications of species distribution modeling to paleobiology. *Quaternary Science Reviews*, 21–22, 2930–2947. <https://doi.org/10.1016/j.quascirev.2011.06.012>
- Taberlet, P., & Cheddadi, R. (2002). Quaternary refugia and persistence of biodiversity. *Science*, 297, 2009–2010. <https://doi.org/10.1126/science.297.5589.2009>
- Theys, K., Lemey, P., Vandamme, A. M., & Baele, G. (2019). Advances in visualization tools for phylogenomic and phylogenetic studies of viral diseases. *Frontiers in Public Health*, 7, 208. <https://doi.org/10.3389/FPUHB.2019.00208/BIBTEX>
- Thuiller, W., Georges, D., & Engler, R. (2013). *biomod2: Ensemble platform for species distribution modeling*. R package version 3.1–25. <https://CRAN.R-project.org/package=biomod2>
- Vale-Gonçalves, H. M., & Cabral, J. A. (2014). New records on the distribution of three rodent species in NE Portugal from barn owl (*Tyto alba*) diet analysis. *Galemys*, 26, 100–104. <https://doi.org/10.7325/Galemys.2014.N3>
- van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511802256>
- Varela, S., Lobo, J. M., & Hortal, J. (2011). Using species distribution models in paleobiogeography: A matter of data, predictors and concepts. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 31, 451–463. <https://doi.org/10.1016/j.palaeo.2011.07.021>
- Voje, K. L. (2020). Testing eco-evolutionary predictions using fossil data: Phyletic evolution following ecological opportunity. *Evolution*, 74(1), 188–200. <https://doi.org/10.1111/evo.13838>
- Warren, D. L., Glor, R. E., & Turelli, M. (2008). Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution*, 62, 2868–2883. <https://doi.org/10.1111/j.1558-5646.2008.00482.x>
- Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010). ABCtoolbox: A versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, 11, 116. <https://doi.org/10.1186/1471-2105-11-116>
- Zubakov, V. A., & Borzenkova, I. I. (1990). *Global palaeoclimate of the Late Cenozoic*. Elsevier.

## BIOSKETCH

**Frederico Mestre** is a biologist whose main research interest is understanding the factors driving biodiversity patterns (whether species or communities) and their response to climate change and habitat modification.

**Soraia Barbosa** is a molecular biologist whose main research interests are understanding the processes associated with species endangerment using genetics and genomics and using molecular evidence to inform conservation planning.

**Author contributions:** Frederico Mestre, Soraia Barbosa, Ricardo Pita and Pedro Beja conceived the study. Frederico Mestre, Soraia Barbosa and José António Garrido-García provided the data. Frederico Mestre and Soraia Barbosa analysed the data. Frederico Mestre, Soraia Barbosa and Ricardo Pita wrote the manuscript, and all authors contributed to editing and revising the manuscript.

## SUPPORTING INFORMATION

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

**How to cite this article:** Mestre, F., Barbosa, S., Garrido-García, J. A., Pita, R., Mira, A., Alves, P. C., Paupério, J., Searle, J. B., & Beja, P. (2022). Inferring past refugia and range dynamics through the integration of fossil, niche modelling and genomic data. *Journal of Biogeography*, 49, 2064–2076. <https://doi.org/10.1111/jbi.14492>